

On the Distinction Between Interaction and Effect Modification

Tyler J. VanderWeele



Abstract: This paper contrasts the concepts of interaction and effect modification using a series of examples. Interaction and effect modification are formally defined within the counterfactual framework. Interaction is defined in terms of the effects of 2 interventions whereas effect modification is defined in terms of the effect of one intervention varying across strata of a second variable. Effect modification can be present with no interaction; interaction can be present with no effect modification. There are settings in which it is possible to assess effect modification but not interaction, or to assess interaction but not effect modification. The analytic procedures for obtaining estimates of effect modification parameters and interaction parameters using marginal structural models are compared and contrasted. A characterization is given of the settings in which interaction and effect modification coincide.

(*Epidemiology* 2009;20: 863–871)

This paper revisits certain issues concerning differences between various concepts of interaction^{1–7} and shows how ideas from causal inference can be used to draw a distinction between the concepts of interaction and effect modification. As formally defined below using counterfactuals, “interaction” may be used to describe instances in which potential interventions on a secondary exposure are in view, while “effect modification” may be used to describe instances in which merely conditioning on a secondary exposure is in view. Often the terms interaction and effect modification are used interchangeably. However, when epidemiologists run regression analyses with product terms, sometimes only one of 2 possible interpretations is warranted. In some cases, the coefficient for the product term in the regression can be interpreted as a measure of what will be formally defined below as effect modification; in other cases, the coefficient for the product term can be interpreted as a measure

of what will be formally defined below as an interaction of effects. Sometimes the coefficient for the product term can be interpreted both as a measure of effect modification and as a measure of interaction; sometimes only one of the 2 interpretations (or neither) is warranted.

The paper is structured as follows. First, I provide and contrast formal counterfactual definitions for interaction and effect modification. Second, examples are given showing that it is possible to have effect modification without interaction or interaction without effect modification. Third, further examples are given showing that in some cases it is possible to identify effect modification but not interaction and that in other cases it is possible to identify interaction but not effect modification. Fourth, analytic procedures to estimate interaction and effect modification parameters in marginal structural models are contrasted. Finally, I characterize those settings in which the concepts of interaction and effect modification do in fact coincide. The goal is to help clarify the distinction between the concepts of interaction and effect modification.

Throughout this paper causal directed acyclic graphs (DAGs) will be employed to illustrate the ideas and examples.^{8–10} The focus, however, is not on causal DAGs but rather on the concepts of interaction and effect modification. Most examples will be accessible to readers who are not familiar with such graphs; accessible introductions to causal DAGs can be found elsewhere.^{10–13}

DEFINITIONS AND NOTATION

I will use the following notation throughout the paper. I consider either interaction between the effects of 2 exposures E and Q on some outcome D or alternatively, effect modification by Q of the effect of exposure E on outcome D . Let D_e denote the counterfactual outcome D under an intervention to set the exposure variable E (possibly contrary to fact) to level e . Let D_{eq} denote the counterfactual outcome D under interventions to set the exposure variables E and Q to levels e and q , respectively.

The presence of interaction or effect modification may depend on the scale being used. Common scales in epidemiologic research include the risk difference scale, the risk ratio scale, and the odds ratio scale. Interaction and effect modification can be present on one scale and absent on another.^{14,15} The focus in this paper will be on the risk difference scale because this scale is arguably of greatest

Submitted 11 November 2008; accepted 31 July 2009.

From the Department of Epidemiology and Biostatistics, Harvard University, Boston, MA.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Tyler J. VanderWeele, Department of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115. E-mail: tvanderw@hsph.harvard.edu.

Copyright © 2009 by Lippincott Williams & Wilkins

ISSN: 1044-3983/09/2006-0863

DOI: 10.1097/EDE.0b013e3181ba333c

public health importance^{1–3} and also has certain relations to biologically based notions of interaction.⁷ However, the remarks apply equally to other scales.

The formal counterfactual definitions given below for effect modification and interaction are arguably present in the literature, sometimes in words, sometimes implicitly and sometimes formally; however, the distinction between the 2 concepts is usually not made explicit. The concept of effect modification concerns whether the effect of one variable on another differs across strata of a third.⁷

Definition 1

A variable Q is said to be an effect modifier on the causal risk difference scale for the effect of E on D , conditional on X , if Q is not an effect of E and if there are 2 levels of E , e_0 and e_1 , and 2 levels of Q , q_0 and q_1 , and some x , such that $E[D_{e_1}|Q = q_1, X = x] - E[D_{e_0}|Q = q_1, X = x] \neq E[D_{e_1}|Q = q_0, X = x] - E[D_{e_0}|Q = q_0, X = x]$.

Until the penultimate section, the discussion will focus on unconditional effect modification (ie, without conditioning on additional covariates X). Unconditionally, we will thus say that a variable Q is an effect modifier on the causal risk–difference scale for the effect of E on D if Q is not an effect of E and if there are 2 levels of E , e_0 , and e_1 , and 2 levels of Q , q_0 , and q_1 , such that $E[D_{e_1}|Q = q_1] - E[D_{e_0}|Q = q_1] \neq E[D_{e_1}|Q = q_0] - E[D_{e_0}|Q = q_0]$. This definition of effect modification requires that for some e_0 and e_1 , $E[D_{e_1}|Q = q] - E[D_{e_0}|Q = q]$ is not constant in q . Although the restriction that Q is not an effect of E is perhaps not strictly necessary in the counterfactual framework,¹⁶ methods that condition on effects of the exposure tend to bias effect measures.^{16–21} Effect modification may be of public health interest because the effect of E (the primary exposure or treatment under study) may vary across subpopulations defined by some variable or set of variables Q ; different interventions may be beneficial for different populations.

Note there is an asymmetry between E and Q in the definition of effect modification: only the effect of one of the 2 exposures is principally in view, namely the effect of E on D . The role of the other exposure Q in the definition simply concerns whether the effect of primary interest varies across strata of this other exposure Q . In contrast to this definition of effect modification, the roles of E and Q in the definition of interaction,^{2,6,22} given below, are symmetric.

Definition 2

There is said to be an interaction on the causal risk difference scale between the effects of E and Q on D , conditional on X , if there are 2 levels of E , e_0 , and e_1 , and 2 levels of Q , q_0 , and q_1 , such that for some x , $E[D_{e_1q_1}|X = x] - E[D_{e_0q_1}|X = x] \neq E[D_{e_1q_0}|X = x] - E[D_{e_0q_0}|X = x]$.

Most of the following discussion will be on unconditional interaction of effects (ie, without conditioning on additional covariates X). Unconditionally, interaction on the

causal risk–difference scale occurs between the effects of E and Q on D if there are 2 levels of E , e_0 , and e_1 , and 2 levels of Q , q_0 , and q_1 , such that $E[D_{e_1q_1}] - E[D_{e_0q_1}] \neq E[D_{e_1q_0}] - E[D_{e_0q_0}]$. In essence, this requires the effect of 2 exposures together to be different from the combination of the 2 effects considered separately. From a public health perspective, knowing about interaction is important in determining whether or not 2 exposures or causes of interest can be studied separately without losing some of the important features of the way the 2 exposures affect the outcome together.

Note that the condition in Definition 2 can be rewritten as $E[D_{e_1q_1}] - E[D_{e_0q_0}] \neq \{E[D_{e_1q_0}] - E[D_{e_0q_0}]\} + \{E[D_{e_1q_1}] - E[D_{e_0q_1}]\}$. If the direction of the inequality is that $E[D_{e_1q_1}] - E[D_{e_0q_0}] > \{E[D_{e_1q_0}] - E[D_{e_0q_0}]\} + \{E[D_{e_1q_1}] - E[D_{e_0q_1}]\}$, then a superadditive interaction on the causal risk difference scale would be said to be present; if $E[D_{e_1q_1}] - E[D_{e_0q_0}] < \{E[D_{e_1q_0}] - E[D_{e_0q_0}]\} + \{E[D_{e_1q_1}] - E[D_{e_0q_1}]\}$, then a subadditive interaction on the causal risk difference scale would be said to be present. With the condition for the interaction of effects written as $E[D_{e_1q_1}] - E[D_{e_0q_0}] \neq \{E[D_{e_1q_0}] - E[D_{e_0q_0}]\} + \{E[D_{e_1q_1}] - E[D_{e_0q_1}]\}$, it is clear that the role of E and Q in Definition 2 are symmetric. Note further that in Definition 2, interventions on both E and Q are being considered. In the definition given for effect modification, only interventions on exposure E were being considered (setting E to e_1 in contrast to setting E to e_0) whereas interventions were not being considered for Q ; the role of Q in Definition 1 for effect modification was that of a conditioning variable. Thus, Q is a conditioning variable in the definition of effect modification, and an intervention variable in the definition of interaction. This distinction between effect modification and interaction also underlies the examples given in the following sections.

One further point merits attention. Definition 2 allows for the assessment of the interaction between the 2 effects with regard to their magnitude. Whether quantitative interaction (as defined in Definition 2) corresponds to an interaction in a mechanistic or biologic sense (sometimes referred to as synergism or antagonism) is another matter. It has been shown that interactions (even causal interactions as given in Definition 2) on the risk difference or risk ratio or odds ratio scale need not correspond to interactions in any biologic or mechanistic sense.^{7,22–24} The relation between mechanistic interaction and statistical interaction for particular measures of effect is considered elsewhere^{7,22,24,25} and is beyond the scope of this paper.

EFFECT MODIFICATION WITH NO INTERACTION AND INTERACTION WITH NO EFFECT MODIFICATION

VanderWeele and Robins²⁶ showed that a variable Q could serve as an effect modifier for the effect of E on D even

if Q had no causal effect on D . In such cases Q effectively served as a proxy for some other variable that did have a causal effect on D . Such examples will also serve to illustrate how there can be effect modification without interaction. For the first example, I return to an example of VanderWeele and Robins,²⁶ although the emphasis here is different. Consider the causal DAG given in Figure 1.

Suppose that E denotes some drug in a randomized trial, D denotes hypertension, X denotes a person's genotype, and Q denotes the person's hair color. In this example, Q might serve as an effect modifier for the effect of E on D since Q is a proxy for X . We will thus likely have a situation in which $E[D_{e=1}|Q=q] - E[D_{e=0}|Q=q]$ is not constant over q . However, because Q is not itself a cause of D , there will be no interaction between the effects of E and Q on D . Because Q has no effect whatsoever on D , we will have $E[D_{e=1,q}] - E[D_{e=0,q}] = E[D_{e=1}] - E[D_{e=0}]$ for all values of q . Interventions on Q do nothing to D , so there can be no interaction between E and Q . A variable Q can thus be an effect modifier for the effect of E on D without there being an interaction between the effects of E and Q on D .

It is also clear from the discussion above that this example also demonstrates a case in which there is effect modification of the effect of E on D by Q , but not of the effect of Q on D by E . Once again, effect modification is not symmetric in the 2 exposures.

I now turn to an example that demonstrates the converse: interaction between the effects of E and Q on D without

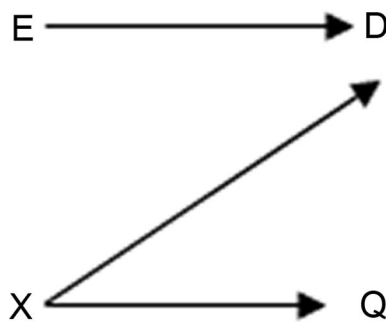


FIGURE 1. Effect modification by Q of the effect of E on D without interaction between the effects of E and Q on D .

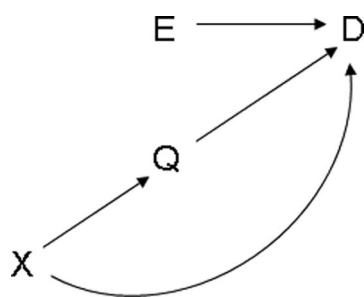


FIGURE 2. Potential interaction between the effects of E and Q on D without effect modification by Q of the effect of E on D .

effect modification. Consider the causal DAG given in Figure 2, and suppose that E denotes a drug in a randomized trial for weight loss for obese children, D denotes the final weight of a child 6 months after the trial completion, X denotes the sugar intake in a child's diet, and Q denotes some measure of exercise. Sugar intake might, for example, affect exercise by making children more hyperactive. Suppose further that the effects of the diet drug E and exercise Q interact so that $E[D_{e_1,q_1}] - E[D_{e_0,q_1}] \neq E[D_{e_1,q_0}] - E[D_{e_0,q_0}]$. In other words, if we were to intervene to give children the drug and also to force exercise, then the effect of these 2 interventions together, compared with the baseline of no drug and no exercise, would be greater than the sum of the effects of each intervention considered separately. We would then say that there is an interaction between the effects of E and Q on D . However, it does not follow from this that Q is an effect modifier for the effect of E on D . Suppose that sugar intake had 2 effects on weight; first, high levels of sugar intake within the diet will likely increase weight gain; second, sugar intake might give children more energy, making them hyperactive and thus more likely to exercise, and thereby lowering their weight through exercise. It is furthermore possible that the indirect effect of sugar intake that lowers weight through exercise essentially cancels out the direct effect of sugar intake which increases weight. In such cases, exercise Q may not be an effect modifier for the effect of drug E on weight D .

A numerical example is given for this illustration in Appendix 1 but the intuition can be given more simply. If we let $Q = 1$ denote a high level of exercise and $Q = 0$ denote a low level of exercise, then $E[D_{e=1}|Q=1] - E[D_{e=0}|Q=1]$ denotes the effect of the drug for children with high levels of exercise and $E[D_{e=1}|Q=0] - E[D_{e=0}|Q=0]$ denotes the effect of the drug for children with low levels of exercise. Although high levels of exercise will tend to decrease weight, children with high levels of exercise might also be those with high levels of sugar intake, which itself increases weight. Likewise, although low levels of exercise will tend to increase weight, children with low levels of exercise might also be those with low levels of sugar intake which itself decreases weight. These cancellations may occur irrespective of whether a child receives drug E . It is therefore possible that we would observe $E[D_{e=1}|Q=1] - E[D_{e=0}|Q=1] = E[D_{e=1}|Q=0] - E[D_{e=0}|Q=0]$ and thus conclude that Q is not an effect modifier for the effect of E on D . In this case we would have interaction of effects, without effect modification. Although in practice it is unlikely that the effects related to sugar intake and exercise will cancel out exactly, it is entirely possible, in this example and in others, that we observe substantial interaction between effects measured by $(E[D_{e_1,q_1}] - E[D_{e_0,q_1}]) - (E[D_{e_1,q_0}] - E[D_{e_0,q_0}])$ but very slight effect modification as measured by $(E[D_{e_1}|Q=q_1] - E[D_{e_0}|Q=q_1]) - (E[D_{e_1}|Q=q_0] - E[D_{e_0}|Q=q_0])$ or vice versa. The conditioning on Q in effect

modification will not in general yield the same results as intervening on Q in the assessment of the interaction of 2 effects.

Note that in the first example, the absence of interaction was structural; interaction between the effects of E and Q is absent irrespective of the distribution of the variables. In contrast, in the second example, the absence of effect modification depended on the cancellation of effects. In general, if interaction between the effects of E and Q is present, then the absence of effect modification will require the cancellation of effects. If there is an interaction between the effects of E and Q , then if effect modification is absent it will not be for structural reasons. This is because, in this case, Q itself must have a nonadditive effect on D . Conditioning on Q will capture this effect, and thus effect modification will be absent only if there is cancellation by the effect of some other variable related to Q .

THE IDENTIFICATION OF INTERACTION AND EFFECT MODIFICATION

The previous section considered only whether effect modification and interaction were present or absent, not whether effect modification and interaction could be assessed given data on a particular set of variables. This section focuses on the identification of effect modification and of interaction. The point of this section is somewhat more subtle: it may be possible to identify effect modification from data but not interaction, or to identify interaction from data but not effect modification. Consider first the causal DAG given in Figure 3.

In this example, we can identify $E[D_{e_1}|Q = q] - E[D_{e_0}|Q = q]$ from data on E, Q and D alone; we do not need data on U .⁸ We can thus assess whether Q is an effect modifier for the effect of E on D by using data only on E, Q , and D . This is because conditioning on Q suffices to control for confounding of the effects of E on D ; Q blocks all “back-door paths” from E to D .^{8–10} Suppose, however, we were interested in assessing not just effect modification but interaction between the effects of E and Q on D . In this case we would need to identify joint effects of the form $E[D_{eq}]$ so that we could test whether $E[D_{e_1q_1}] - E[D_{e_0q_1}] \neq E[D_{e_1q_0}] - E[D_{e_0q_0}]$. It can be shown, however, that the joint effects of E and Q on D cannot be identified with data on E, Q , and D alone (see corollary 4.5.4 of Pearl⁹). This is because the unmeasured variable U is a common cause of Q and D , and thus confounds the effect of Q on D . If we had data on U as well as E, Q , and D , we could identify joint effects of the form $E[D_{eq}]$ because we could block all back-door paths from E and Q to D . Because data are not available on U , joint effects of the form $E[D_{eq}]$ are not identified. Thus, we cannot assess whether there is an interaction between the effects of E and Q on D . Consequently, Figure 3 constitutes an example in which we can identify whether or not effect modification is present, but not whether there is an interaction between effects.

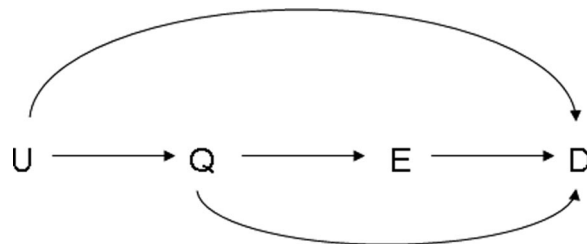


FIGURE 3. Identification of effect modification of the effect of E on D by Q without identification of the joint effects of E and D .

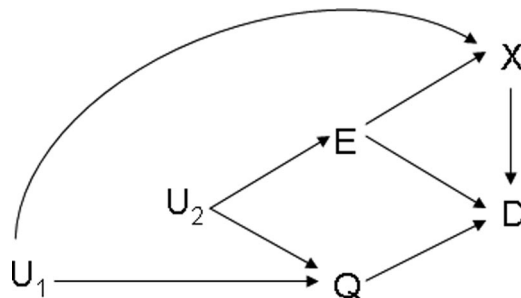


FIGURE 4. Identification of the joint effects of E and D without identification of effect modification of the effect of E on D by Q .

We now consider the converse situation, in which we can identify whether an interaction between effects is present but not whether there is effect modification. Consider the causal DAG given in Figure 4. Technical details for this example are given in the eAppendix (<http://links.lww.com/EDE/A340>) of this paper and require various theoretical results concerning causal DAGs.^{8,9,27–31} In Figure 4, suppose that data are available on E, Q, X , and D , but that U_1 and U_2 are unmeasured. It is shown in the eAppendix that joint effects of the form $E[D_{eq}]$ are identified with data on just E, Q, X , and D and thus the condition $E[D_{e_1q_1}] - E[D_{e_0q_1}] \neq E[D_{e_1q_0}] - E[D_{e_0q_0}]$ can be tested and interaction between the effects of E and Q on D can be assessed. However, it is furthermore shown in the eAppendix (using Theorem 6 of Shpitser and Pearl²⁸ and Theorem 3 of Tian and Pearl³¹) that effects of the form $E[D_{e_1}|Q = q] - E[D_{e_0}|Q = q]$ are not in general identified, so that effect modification cannot be assessed with data on E, Q, X , and D alone. In Figure 4 we thus have an example in which we can identify whether an interaction between effects is present but not whether effect modification is present. In Appendix 2 we discuss the identification of interaction and effect modification in greater detail, and give 2 results that relate the identification of interaction and effect modification to causal DAGs.^{27–31}

INTERACTION AND EFFECT MODIFICATION IN MARGINAL STRUCTURAL MODELS

The concepts of interaction and effect modification can also be distinguished in the different analytic procedures used to assess interaction or effect modification with marginal structural models. For simplicity, I assume that E and Q are binary and that D is continuous. A marginal structural model is a model for expected counterfactual outcomes. Ordinary regression models are models for expected outcomes conditional on covariates. Thus a regression model might be a model for the conditional expected outcome $E[D|E = e, Q = q, X = x]$ for some measured covariates X , whereas a marginal structural model is a model for the counterfactual outcomes $E[D_{eq}]$. A marginal structural model for the effect of E on D conditional on Q that allows for potential effect modification of the effect of E on D by Q takes the form

$$E[D_{eq}|Q = q] = \alpha_0 + \alpha_1 e + \alpha_2 q + \alpha_3 eq \quad (1)$$

A marginal structural model for the joint effects of E and Q on D and that allows for an interaction between the effects of E and Q on D takes the form

$$E[D_{eq}] = \beta_0 + \beta_1 e + \beta_2 q + \beta_3 eq \quad (2)$$

Although the 2 marginal structural models look very similar in form, the models themselves and the interpretation of the model coefficients ($\alpha_0, \alpha_1, \alpha_2, \alpha_3$) and ($\beta_0, \beta_1, \beta_2, \beta_3$) are subtly different. Model 1 is a model for the counterfactual outcome intervening on E only, but conditional on Q . Model 2 is a model for counterfactual outcomes intervening on both E and Q . In model 1, α_3 represents the parameter for assessing whether there is effect modification. In model 2, β_3 represents the parameter for assessing whether there is an interaction between the effects of E and Q on D .

The distinction between interaction and effect modification in marginal structural models can best be seen by contrasting the analytic techniques used to fit models 1 and 2. In a marginal structural model, control for confounding variables is made not by conditioning but by weighting.^{32,33} This weighting technique is generally referred to as inverse probability of treatment weighting.^{32,33} We will consider the procedures for fitting models 1 and 2 in turn. We assume that the unmeasured confounding assumptions required to fit these models are satisfied. In particular we assume that we have some measured covariates X such that the effect of E on D is unconfounded given X and Q ; this will allow us to estimate the parameters in model 1. If, in addition to this assumption, the effect of Q on D is unconfounded given X , then we can also estimate the parameters in model 2. (See Appendix 2 for further details on evaluating these no-unmeasured-confounding assumptions.) Assuming that these assumptions hold, we can estimate model 1 using inverse probability of treatment weighting by calculating the following weights:

$$w_i^E = \frac{P(E = e_i|Q = q_i)}{P(E = e_i|Q = q_i, X = x_i)}$$

where e_i , q_i , and x_i denote individual i 's values of E , Q , and X , respectively. The probabilities in the weights can be estimated using regular logistic regression. The inclusion of the probability in the numerator is optional, but tends to lead to more efficient estimates.³³ To fit the marginal structural model in equation 1, one can use a weighted regression of D on E and Q including an $E \times Q$ product term where the weights are given by w_i^E . The estimates of the parameters from the weighted regression will correspond to those of the marginal structural model in equation 1. For the estimates of the standard errors to be valid, robust estimation of standard errors must be used. (See Robins et al for further details on fitting marginal structural models.³³)

To fit model 2, an additional set of weights is needed, namely weights for Q that can be constructed as follows:

$$w_i^Q = \frac{P(Q = q_i)}{P(Q = q_i|X = x_i)}$$

These numerator and denominator probabilities can once again be estimated using regular logistic regression. Once these weights are calculated, the parameters of the marginal structural model in equation 2 can be estimated by again using a weighted regression. One uses a weighted regression of D on E and Q including an $E \times Q$ product term, but this time the observations are then weighted by the product $w_i^E * w_i^Q$ and not just w_i^E . The estimates of the parameters from the weighted regression will correspond to those of the marginal structural model in equation 2. If Q is not a cause of E (nor E of Q), then it may not be necessary to condition on Q in the numerator and denominator probabilities in the weights w_i^E for E ; if one is in doubt, or if E and Q are correlated even conditional on X , then conditioning on Q is advisable. If, on the other hand, there are variables V that are effects of Q confounding the relationship between E and D , then these need to be included in the denominator probabilities for the weight for E (but are not included in the weights for Q). The weight for E is then

$$w_i^E = \frac{P(E = e_i|Q = q_i)}{P(E = e_i|Q = q_i, X = x_i, V = v_i)}$$

Positivity and consistency assumptions must hold for the estimation of the parameters of marginal structural models through inverse-probability-of-treatment weighting; these assumptions are discussed elsewhere.³²⁻³⁶ The estimation procedures described above are a special case of an entire class of estimators for interaction parameters that include doubly robust and multiply robust estimators.³⁷

Robins et al³³ showed that the analytic procedures required to fit marginal structural models clarify the distinction between confounding and effect modification, a distinc-

tion that regression analysis often obscures. To fit a marginal structural model using inverse-probability-of-treatment weighting, confounding is taken care of by including the confounding variables X as covariates in the denominator probabilities of the weights; effect modification on the other hand is evaluated by including the potential effect modifier Q in the weighted regression model. Similarly, in the discussion above, the analytic procedures required to fit marginal structural models make clear the distinction between effect modification and interaction. The assessment of interaction and effect modification both use a weighted regression model of the outcome D on exposures E and Q ; however, to estimate the effect modification parameter α_3 , the regression is weighted by only one set of weights, w_i^E , whereas to estimate the interaction parameter β_3 the regression is weighted by the product of 2 sets of weights, $w_i^E * w_i^Q$.

WHEN INTERACTION AND EFFECT MODIFICATION COINCIDE

Throughout this paper, I have emphasized the distinction between effect modification and interaction. In many settings, of course, interaction and effect modification will coincide. The question thus arises whether we can characterize settings in which interaction and effect modification do indeed coincide. A characterization can be given in terms of causal DAGs. It follows from the rules of causal DAGs (see Rule 2 of Pearl's do-calculus⁸) that we will have that interaction and effect modification will coincide (because $E[D_{eq}] = E[D_e|Q = q]$), whenever all paths between D and Q are blocked by E on a graph obtained by modifying the original graph to remove the arrows into E and the arrows emerging from Q . The characterization can be made conditional on X if conditional effect modification or conditional interaction is under consideration. In this case, (E, X) must block all of the aforementioned paths⁸; if this condition holds then $E[D_{eq}|X = x] = E[D_e|Q = q, X = x]$. In any case, it can be verified that the condition needed for interaction and effect modification to coincide fails (in its unconditional form) for all of the examples shown in Figures 1–4. In each case there is a common cause of Q and D , sometimes measured (X), sometimes unmeasured (U), that leads to the failure of the condition. In these cases, conditioning on Q gives some information about a common cause of Q and D , and because this common cause affects D , conditioning on Q will not coincide with intervening on Q . Effect modification and interaction are thus not unconditionally equivalent in Figures 1–4.

In other settings, however, the equivalence will hold. For example, a common setting in the literature on interaction is to assume that the effects of both E and Q on D are unconfounded given X . If this holds, and the exposure Q precedes the exposure E then the condition above will be satisfied and we will have $E[D_{eq}|X = x] = E[D_e|Q = q, X = x]$. Thus, interaction between E and Q conditional on X and

effect modification of E by Q conditional on X will coincide in this setting.

In practice, the distinction between effect modification and interaction is often ignored. The researcher might simply regress D on E , Q , $E \times Q$ and possibly some covariates X . However, while effect modification and interaction may often coincide, they do not always. For example, a regression of D on E , Q and $E \times Q$ for the variables given in Figure 3 could be used to assess effect modification, but not interaction. In Figure 4, a regression of D on E , Q , $E \times Q$ (with or without X) cannot be used to assess either effect modification or interaction; data on E , Q , D , and X can be used to assess the interaction between the effects of E and Q on D but methods other than regression are needed (see Appendix 2 for further details).

In Figure 2, the interpretation of the coefficient for the product term in a regression of D on E , Q , $E \times Q$ and X is somewhat more straightforward. In Figure 2, X blocks all paths from Q to D on the graph with the arrows into E and the arrows emerging from Q removed. Thus in Figure 2, effect modification conditional on X and interaction conditional on X will coincide. The estimate of the coefficient for the product term in a regression of D on E , Q , $E \times Q$, and X can, in Figure 2, be interpreted as a measure of effect modification conditional on X , and as a measure of interaction conditional on X . To obtain unconditional measures of interaction, one could standardize the conditional measures over the distribution of X ; to obtain unconditional measures of effect modification, one could standardize the conditional measures over the distribution of X given Q . As noted above, these unconditional measures of interaction and effect modification (unlike the conditional measures) will not in general coincide for Figure 2.

Finally, in Figure 1, it can be seen that conditional on X there is neither interaction between the effects of E and Q , nor effect modification of E by Q ; clearly Q has no effect on D , and once we condition on X , Q is independent of D . Thus interaction conditional on X and effect modification conditional on X coincide trivially.

DISCUSSION

This paper discusses the distinction between interaction and effect modification. The distinction is manifest in a variety of ways: in the definition of these concepts, in situations where one is present but the other is absent, in the conditions required for their identification, and in the analytic procedures required to estimate parameters in marginal structural models corresponding to interaction or effect modification.

Although the definitions above and the series of contrasting examples make clear the formal distinction between these concepts, there is a certain ambiguity or duality in the language that is used. As noted above, the 2 terms are often used interchangeably in practice. Moreover, what has been

formally defined above as effect modification is considered by many to fall under a very general category of interaction, conceived as broadly encompassing several distinct concepts.⁷ If one conceives of interaction broadly in this manner, then Definition 1 concerning effect modification and Definition 2 concerning the interaction of the causal effects of 2 exposures would both be considered as 2 distinct concepts within the broader category of interaction.

Conversely, what has formally been defined above as an “interaction of effects” one could also conceive of as “effect modification by intervention.” Although the phrase effect modification is customarily used when the effect of one variable on another differs conditionally within strata of a third, it seems not unreasonable to use the term effect modification (by intervention) to refer to cases in which the effect of one variable on another varies under different interventions on a third variable.

We see then a certain duality or ambiguity in the language we use; the formal definitions given above, however, allow for a clear distinction between concepts to be drawn. Although it is understandable that the language is often used interchangeably, it is nevertheless important to be aware that there are 2 distinct concepts, irrespective of what we might choose to call them. When epidemiologists run regression analyses with product terms, sometimes only one of 2 possible interpretations is warranted. In some cases the regression coefficient for the product term can be interpreted as a measure of (what has been defined above as) effect modification; in other cases the coefficient can be interpreted as a measure of (what has been defined above) as interaction. Sometimes both interpretations are warranted; sometimes only one, or the other, and sometimes neither. In general, some knowledge of the causal structure among variables is necessary to determine which interpretation (or interpretations) is warranted.

The distinction between interaction and effect modification has implications for the analysis of randomized trials. Often subgroup analyses are conducted in such trials. Provided the variable Q defining the subgroups is a pretreatment baseline covariate, the subgroup analyses will warrant an interpretation as measures of effect modification because the treatment itself is randomized (and thus its effect on the outcome will in general be unconfounded). However, these subgroup analyses will not in general allow for an interpretation as measures of interaction (as given in Definition 2 above) because the effect of the secondary variable Q , defining the subgroups, will not in general be unconfounded. To interpret contrasts between subgroup analyses as measures of interaction, one would also need to control for confounders of the relationship between the secondary variable, Q , and the outcome. For each distinct secondary exposure variable Q for which an interaction interpretation of the subgroup analyses is desired, it will likely be necessary to consider a different

set of confounding variables. Again, randomization of the primary treatment does not ensure unconfoundedness of the effects of a secondary exposure variable used to define subgroups.

In future research, more complex examples could be considered that include time-varying effect modifiers^{38,39} and interactions, or that explicitly consider time itself as potential effect modifier.

ACKNOWLEDGMENTS

I thank Sander Greenland and James Robins for helpful comments on an earlier draft of this paper.

REFERENCES

1. Blot WJ, Day NE. Synergism and interaction: are they equivalent? *Am J Epidemiol.* 1979;110:99–100.
2. Saracci R. Interaction and synergism. *Am J Epidemiol.* 1980;112:465–466.
3. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol.* 1980;112:467–470.
4. Koopman JS. Interaction between discrete causes. *Am J Epidemiol.* 1981;113:716–724.
5. Miettinen OS. Causal and preventive interdependence: Elementary principles. *Scand J Work Environ Health.* 1982;8:159–168.
6. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scand J Work Environ Health.* 1988;14:125–129.
7. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
8. Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995;82:669–688.
9. Pearl J. *Causality: Models, Reasoning, and Inference.* Cambridge, United Kingdom: Cambridge University Press; 2000.
10. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10:37–48.
11. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15:615–625.
12. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology.* 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2008:183–209.
13. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes and the properties of conditioning on a common effect. *Am J Epidemiol.* 2007;166:1096–1104.
14. Mantel N, Brown C, Byar DP. Tests for homogeneity of effect in an epidemiologic investigation. *Am J Epidemiol.* 1977;106:125–129.
15. Kupper LL, Hogan MD. Interaction in epidemiologic studies. *Am J Epidemiol.* 1978;108:447–453.
16. Joffe M, Small D, Hsu C-Y. Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Stat Sci.* 2007;22:74–97.
17. Greenland S, Neutra RR. Control of confounding in the assessment of medical technology. *Int J Epidemiol.* 1980;9:361–367.
18. Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J Roy Stat Soc Ser A.* 1984;147:656–666.
19. Robins JM, Morgenstern H. The foundations of confounding in epidemiology. *Math Modeling.* 1987;14:869–916.
20. Robins JM. The control of confounding by intermediate variables. *Stat Med.* 1989;8:679–701.
21. Weinberg CR. Toward a clearer definition of confounding. *Am J Epidemiol.* 1993;127:1–8.
22. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component cause framework. *Epidemiology.* 2007;18:329–339.
23. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol.* 1981;10:383–387.
24. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology.* 2009;20:6–13.
25. VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika.* 2008;95:49–61.

26. VanderWeele TJ, Robins JM. Four types of effect modification: a classification based on directed acyclic graphs. *Epidemiology*. 2007;18: 561–568.
27. Pearl J, Robins JM. Probabilistic evaluation of sequential plans from causal models with hidden variables. In: *Proceedings of 11th Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann; 1995:444–453.
28. Shpitser I, Pearl J. Identification of conditional interventional distributions. In: *Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence*. Corvallis, OR: AUAI Press; 2006:437–444.
29. Shpitser I, Pearl J. Identification of joint interventional distributions in recursive semi-Markovian causal models. In: *Proceedings of 21st Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press; 2006: 1219–1226.
30. Huang Y, Valorta M. Pearl’s calculus of interventions is complete. In: *Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence*. Corvallis, OR: AUAI Press; 2006:217–224.
31. Tian J, Pearl J. A general identification condition for causal effects. In: *Proceedings of 18th National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press/MIT Press; 2002:567–573.
32. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. New York: Springer-Verlag; 1999:95–134.
33. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
34. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168:656–664.
35. Cole SR, Frangakis CE. The consistency assumption in causal inference: a definition or an assumption? *Epidemiology*. 2009;20:3–5.
36. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20:880–883.
37. Vansteelandt S, VanderWeele TJ, Tchetgen EJ, Robins JM. Multiply robust inference for statistical interactions. *J Am Stat Assoc*. 2008;103: 1693–1704.
38. Petersen ML, Deeks SG, Martin JN, van der Laan MJ. History-adjusted marginal structural models for estimating time-varying effect modification. *Am J Epidemiol*. 2007;166:985–993.
39. Robins JM, Hernán MA, Rotnitzky A. Effect modification by time-varying covariates. *Am J Epidemiol*. 2007;166:994–1002.

APPENDIX 1. NUMERICAL EXAMPLE OF INTERACTION WITHOUT EFFECT MODIFICATION

This appendix provides a numerical example (related to Fig. 2) to show interaction between effects without effect modification. Suppose that E denotes a drug in a randomized trial for weight loss for obese children, D denotes the final weight of a child 6 months after the trial completion, X denotes the sugar intake in a child’s diet (standardized on a scale from 0–1), and Q denotes some measure of exercise (standardized on a scale from 0–1). For simplicity we will consider an example in which the link between exercise and sugar intake is deterministic so that $Q = X$. More complicated examples with stochastic links could also be constructed. Suppose further now that $E[X] = 0.6$ and $E[D|E, Q, X] = (150 - 10E - 5EQ + 30X + 5EX)$. Then on the causal DAG given in Figure 2 we have that $E[D_{eq}] = \sum_x E[D|E = e, Q = q, X = x]P(X = x) = \sum_x (150 - 10e - 5eq + 30x + 5ex)P(X = x) = (150 - 10e - 5eq + 30E[X] + 5eE[X]) = 150 - 10e - 5eq + (30)(0.6) + 5e(0.6) = 168 - 7e - 5eq$. Thus $E[D_{e=1,q=1}] - E[D_{e=0,q=1}] = (168 - 7 - 5) - (168) = -12$ and $E[D_{e=1,q=0}] - E[D_{e=0,q=0}] = (168 - 7) - (168) = -7$ and thus there is interaction on the causal risk

difference scale. However, $E[D_e|Q = q] = \sum_x E[D|E = e, Q = q, X = x]P(X = x|Q = q) = \sum_x (150 - 10e - 5eq + 30x + 5ex)P(X = x|Q = q) = (150 - 10e - 5eq + 30E[X|Q = q] + 5eE[X|Q = q]) = (150 - 10e - 5eq + 30q + 5eq) = (150 - 10e + 30q)$. Thus $E[D_{e=1}|Q = 1] - E[D_{e=0}|Q = 1] = (150 - 10 + 30) - (150 + 30) = -10$ and $E[D_{e=1}|Q = 0] - E[D_{e=0}|Q = 0] = (150 - 10) - (150) = -10$ and thus there is no effect modification by Q on the causal risk difference scale of the effect of E on D .

APPENDIX 2. IDENTIFICATION RESULTS FOR EFFECT MODIFICATION AND INTERACTION

This appendix provides results concerning the identification of effect modification and of the interaction of effects. The notation $X \perp\!\!\!\perp Y|Z$ denotes that X is conditionally independent of Y given Z . We will assume data is available on some set of covariates X . In general, to identify effect modification by Q of the effect of E on D the effect of E on D must be unconfounded given X and Q . More formally, in counterfactual notation, this condition is that $D_e \perp\!\!\!\perp E|(X, Q)$. Conditions required to identify joint effects $E[D_{eq}]$ (and thereby interactions between effects) are more subtle. On the crudest level, it will suffice to identify joint effects $E[D_{eq}]$ if the effects of E and Q on D are unconfounded given X or in counterfactual notation, $D_{eq} \perp\!\!\!\perp (E, Q)|X$. Somewhat more generally, the joint effect $E[D_{eq}]$ will be identified if there exist 2 sets of measured covariates X and V such that either $D_{eq} \perp\!\!\!\perp Q|X$ and $D_{eq} \perp\!\!\!\perp E|(X, V, Q)$, or $D_{eq} \perp\!\!\!\perp E|X$ and $D_{eq} \perp\!\!\!\perp Q|(X, V, E)$. In words, it will suffice either if the effect of Q on D is unconfounded given X and the effect of E on D is unconfounded given (X, V, Q) or if the effect of E on D is unconfounded given X and the effect of Q on D is unconfounded given (X, V, E) . Below, however, we give an even more general set of identification conditions. The results for the identification of effect modification and the interaction of effects concern causal DAGs and follow immediately from previous results in the literature. The first result concerns the identification of effect modification and is an immediate consequence of the rules of Pearl’s do-calculus.⁸

Result 1

If on a causal directed acyclic graph, Q is not a descendent of E and there exists a set X of non – descendents of E such that all the backdoor paths from E to D are blocked by (X, Q) then $E[D_e|Q = q]$ is identified and is given by $E[D_e|Q = q] = \sum_x E[D|E = e, X = x, Q = q]P(X = x|Q = q)$.

If the conditions of Result 1 hold, then effect modification can be identified since each term in $(E[D_{e_1}|Q = q_1] - E[D_{e_0}|Q = q_1]) - (E[D_{e_1}|Q = q_0] - E[D_{e_0}|Q = q_0])$ is identified. The second result concerns the identification of the interaction of effects. The result is a special case of Theorem 1 in the work of Pearl and Robins.²⁷

Result 2

Consider a causal directed acyclic with 2 exposures A_1 and A_2 and some outcome D . Suppose there exists a set W of nondescendants of A_1 such that all backdoor paths from A_1 to D are blocked by W in the graph with the arrows going into A_2 removed. Suppose further that there exists a set V of nondescendants of A_2 such that all backdoor paths from A_2 to D on the original graph are blocked by (V, W, A_1) then $E[D_{a_1a_2}]$ is identified and is given by $E[D_{a_1a_2}] = \sum_{v,w} E[D|A_2 = a_2, V = v, A_1 = a_1, W = w]P(V = v|A_1 = a_1, W = w)P(W = w)$.

If the conditions of Result 2 hold with A_1 and A_2 corresponding to Q and E respectively, or vice versa, then the

interaction of effects can be identified since each term in $(E[D_{e_1q1}] - E[D_{e_1q0}]) - (E[D_{e_0q1}] - E[D_{e_0q0}])$ is identified.

Figure 3 above constitutes an example in which we can apply Result 1 but not Result 2; it is shown in the eAppendix (<http://links.lww.com/EDE/A340>) that Figure 4 above constitutes an example in which we can apply Result 2 but not Result 1. Results 1 and 2 represent reasonably general sufficient conditions for the identification of effect modification and the interaction of effects but they are not necessary conditions. Conditions for conditional effects²⁸ and for joint effects^{29,30} that are both necessary and sufficient have been derived, but it is not easy to provide a simple characterization of these results.