Epidemiologic Methods

Confounding and Effect Modification: Distribution and Measure

Tyler J. VanderWeele, *Harvard University*

Confounding and Effect Modification:
Distribution and Measure

Tyler J. VanderWeele

Abstract

The paper considers the properties of and relations between confounding and effect modification from the perspective of causal inference and with a distinction drawn as to how each of these two epidemiologic concepts can be defined both with respect to a distribution of potential outcomes or with respect to a specific effect measure. Both concepts are conditional on other covariates but the form this conditionality takes differs. Both concepts are also properties of the population, and both are relative to the specific exposure and to the specific outcome. For a particular population, the presence of confounding depends on how the exposure was assigned; the presence of effect modification does not. The possibility of confounding without effect modification and vice versa is discussed both with respect to distribution and measure. Discussion is given as to how confounding and effect modification relate to statistical models and to the relevance of the points made in the paper to data analysis and interpretation.

KEYWORDS: Causal inference, confounding, effect heterogeneity, effect modification

Introduction

This paper revisits the properties of and relationships between confounding and effect modification. The topic has of course received attention in the past (Miettinen, 1974; Greenland and Morgenstern, 1989; Geng and Li, 2002; Stürmer et al., 2006; Rothman et al., 2008). Here, I would like to revisit it in light of insights that can be drawn from the causal inference literature and also with an eye towards a further distinction that can be drawn concerning how these two epidemiologic concepts relate both to overall distributions and to specific measures.

The paper describes how both confounding and effect modification may be defined so as to make reference to an entire distribution of potential outcomes or so as to reference a specific measure. The paper then also considers (i) the conditionality of both concepts, (ii) the relation of both concepts to study design, (iii) that both concepts are properties of the population, (iv) that both concepts are relative with respect to exposure and the outcome, (v) implications that hold between confounding and effect modification and (vi) the relation of both concepts to statistical models. The paper concludes by discussing a few points concerning how relations between confounding and effect modification, as they relate to both distribution and measure, are relevant for data analysis and interpretation.

Notation

We will let A denote an exposure of interest, Y an outcome of interest, C a set of covariates, and Q a specific covariate of interest occurring prior to the exposure. We will use the notation Y_a to denote the potential outcome (or "counterfactual outcome", Rubin, 1990; Hernán, 2004) for an individual if exposure A had been set, possibly contrary to fact, to value a. We assume throughout the consistency assumption that if actual exposure $A = a$ then $Y = Y_a$. We use the notation $X \perp\!\!\!\perp Y | Z$ to denote that X is independent of Y conditional on Z. For simplicity we will generally assume binary treatment with $A \in \{0, 1\}$ but the remarks here are applicable more generally. The average causal effect for a population is then denoted by $E(Y_1 - Y_0)$. Some of the early literature in epidemiology placed emphasis on the effect of exposure on the exposed, i.e. $E(Y_1 - Y_0 | A = 1)$. There has been more recent emphasis on average causal effect, $E(Y_1 - Y_0)$, but which of the two is of interest will vary by context.

Confounding: Distribution and Measure

The basic notion of exchangeability or "no confounding" is that the outcomes observed amongst the unexposed (or exposed) are representative of what would have been observed had the exposed been unexposed (or had the unexposed been exposed). If the outcomes observed amongst the unexposed are representative of what would have been observed if the exposed had been unexposed, then the group are effectively "exchangeable"; there is no confounding. If the exposed and unexposed are not comparable in this way then confounding is said to be present. Some of the epidemiologic literature uses the terms "no confounding" and "exchangeability" interchangeably but sometimes a distinction is drawn between the two with only the latter denoting also the absence of selection bias (and possibly measurement error). In the absence of selection bias and measurement error, the two could be taken as equivalent.

This general notion of confounding or exchangeability can be defined both with respect to the distribution of potential outcomes and with respect to a specific measure. The distinction has been drawn before (Greenland et al., 1999).

Definition 1. We say that there is no confounding *in distribution* of the effect of A on Y conditional on C if $P(Y_a|C = c) = P(Y|A = a, C = c)$ for all a, c.

Definition 1 states that within strata of C, the group that actually had exposure status $A = a$ is representative of what would have occurred had the entire population with $C = c$ been given exposure $A = a$. If this holds, we could use the observed data to reason about the effect of intervening to set $A = a$ for the entire population. If the conditions of Definition 1 are not satisfied, we will say that there is confounding in distribution (conditional on C). In this paper, we will use the expressions, "no confounding in distribution" and "unconfounded in distribution" interchangeably. The condition in Definition 1 is sometimes also referred to as "weak ignorability" or "ignorable treatment assignment" (Rosenbaum and Rubin, 1983), "exchangeability" (Greenland and Robins, 1986), "no unmeasured confounding" (Robins, 1992), "selection on observables" (Barnow et al., 1980; Imbens, 2004), or "exogeneity" (Imbens, 2004). The condition in Definition 1 is often written in terms of a conditional independence assumption, namely, $Y_a \perp\!\!\!\perp A|C$. This can also be written as $P(Y_a|A = 1, C = c) = P(Y_a|A = 0, C = c) = P(Y_a|C = c)$ indicating that, conditional on C, the exposed and unexposed groups are comparable in their potential outcomes.

A further distinction can be drawn between confounding "in expectation" and "realized" confounding (Fisher, 1935; Rothman, 1977; Greenland, 1990; Greenland et al., 1999). In a randomized trial the groups receiving the placebo and the treatment will be comparable in their potential outcomes on average over repeated experiments. However, for any given experiment, the particular randomization may result in chance imbalances due to the particular allocation. Such a scenario would be one in which there is no confounding "in expectation" but there is realized confounding for the particular experiment (conditional on the allocation). Some authors (Greenland et al., 1999; Greenland and Robins, 2009) prefer to restrict the use of "no confounding" to that that is realized; a number of authors (e.g. Rubin, 1991; Robins, 1992; Stone, 1993) use terms like "no confounding" to refer to that in expectation; here we will adopt the latter practice. See Greenland and Robins (2009) for further discussion.

Note that Definition 1 makes reference to the whole distribution of potential outcomes, $P(Y_a|C = c) = P(Y|A = a, C = c)$. In some of the earlier causal inference literature (Greenland and Robins, 1986), confounding and exchangeability were discussed not in terms of distributions of potential outcomes but principally in terms of mean differences of potential outcomes. This gives rise to the notion of confounding in measure (Greenland et al., 1999). We will denote measures of interest by μ where $\mu(p_1, p_2)$ is a function of two population parameters. Common measures in epidemiologic research include the risk difference $\mu(p_1, p_2) = p_1 - p_2$, the risk ratio, $\mu(p_1, p_2) = p_1/p_2$ and the odds ratio, $\mu(p_1, p_2) = p_1(1 - p_1)/\{p_2(1 - p_2)\}$.

Definition 2. We say that there is no confounding *in measure* μ of the effect of A on Y conditional on C if $\mu(E(Y_1|C = c), E(Y_0|C = c)) = \mu(E(Y|A = 1, C = c), E(Y|A = 0, C = c))$.

Whereas Definition 1 requires that an entire distribution of potential outcomes be comparable, Definition 2 makes reference to a specific measure. For example, with the risk difference measure $\mu_D(p_1, p_2) = p_1 - p_2$, Definition 2 requires only that $E(Y_1|C = c) - E(Y_0|C = c) = E(Y|A = 1, C = c) - E(Y|A = 0, C = c)$. This may hold without Definition 1 holding for either of two reasons. First, if the outcome is not binary, then it is possible that the mean potential outcomes in the exposed and unexposed are equal even though their distributions are not. For example, it may be the case that the mean potential outcome Y_0 is comparable in the exposed and unexposed, i.e. $E(Y_0|A = 1, C = c) = E(Y_0|A = 0, C = c)$, even though the distributions $P(Y_0|A = 1, C = c)$ and $P(Y_0|A = 0, C = c)$ are different (as might occur if

the distribution of Y_0 was more disperse for the exposed than the unexposed). Second, it is possible that even if it is not the case that $E(Y_1|C = c) = E(Y|A = 1, C = c)$ and $E(Y_0|C = c) = E(Y|A = 0, C = c)$, it may be that the bias for $E(Y_1|C = c)$ and for $E(Y_0|C = c)$ effectively cancel one another out so that the associational risk difference, $E(Y|A = 1, C = c) - E(Y|A = 0, C = c)$, is in fact equal to the causal risk difference, $E(Y_1|C = c) - E(Y_0|C = c)$. In other words, there is no confounding for the risk difference measure. Such cases would probably be quite rare in practice.

This second possibility also indicates that there can be confounding for one measure but not another. If $E(Y_1) = 0.4$, $E(Y_0) = 0.2$, $E(Y|A = 1) = 0.3$, $E(Y|A = 0) = 0.1$ then $E(Y_1) - E(Y_0) = 0.4 - 0.2 = 0.2$ and $E(Y|A = 1) - E(Y|A = 0) = 0.3 - 0.1 = 0.2$ so we would have no confounding for the risk difference measure but $E(Y_1)/E(Y_0) = 0.4/0.2 = 2$ and $E(Y|A = 1)/E(Y|A = 0) = 0.3/0.1 = 3$ so we would have confounding for the risk ratio measure.

In some of the earlier literature, often the assumption was only made that the mean outcome of unexposed was representative of what would have been observed if the exposed had been unexposed i.e. only $E(Y_0|A = 1, C = c) = E(Y_0|A = 0, C = c)$. This would allow one to estimate the effect of the exposure on the exposure, $E(Y_1 - Y_0|A = 1)$ but not the average causal effect, $E(Y_1 - Y_0)$. Requiring that $E(Y_a|A = 1, C = c) = E(Y_a|A = 0, C = c)$ for only one value of a is sometimes referred to as an assumption of "partial exchangeability" (Greenland and Robins, 1986).

The definition above defines confounding in measure as the equality of an associational and causal measure. It is important to note that, in data analysis, to get a valid estimate, one must also use an estimator of $E(Y|A = a, C = c)$ that is consistent. If C is multivariate this can be a difficult modeling task.

Definition 2 concerns conditional measures; one might also consider marginal or standardized measure. If μ is a standardized measure then there is no confounding *in measure* μ of the marginal effect of A on Y adjusting for C if $\mu(E(Y_1), E(Y_0)) = \mu(\sum_c E(Y|A = 1, C = c)P(C = c), \sum_c E(Y|A = 0, C = c)P(C = c))$. Similar definitions could be given for the marginal effect on the exposed or unexposed i.e. $\mu(E(Y_1|A = a_1), E(Y_0|A = a_1))$ or $\mu(E(Y_1|A = a_0), E(Y_0|A = a_0))$. If the exposure groups are exchangeable in that $E(Y_a|A = 1, C = c) = E(Y_a|A = 0, C = c)$ for all a and c then the effect of A on Y will be unconfounded for both the conditional and marginal measures. With conditional measures, there may be no confounding for the measure for some strata of C but not for others. Likewise, for standardized measures,

there may be no confounding for one standardized measure but not another; for example, if the effect of A on Y is unconfounded in distribution conditional on C and C has an effect on the outcome only in the presence of exposure, then there will be no confounding for $\mu(E(Y_1|A = a_1), E(Y_0|A = a_1))$ without controlling for C but there will in general be confounding for $\mu(E(Y_1), E(Y_0))$ if control is not made for C.

In the examples above, we have seen that the effect of A on Y may be unconfounded for a specific measure but not in distribution. Definitions 1 and 2 are nevertheless related by the following proposition.

Proposition 1. If Y is binary, then there is confounding in distribution if and only if there is some measure μ for which there is confounding in measure μ.

Proposition 1 need not hold true if Y is not binary because of the possibility that the mean, but not the distribution, of potential outcomes in the exposed and unexposed are equal. An immediate corollary of Proposition 1 is that if the effect of A on Y is unconfounded in distribution then it will be unconfounded for all measures μ. This implication holds true also if Y is not binary. Definition 1 is the more stringent and more general definition.

Confounding in distribution is sometimes assessed using causal diagrams and in some of the examples in subsequent sections, we will make use of such causal diagrams (Pearl, 1995, 2009). An introduction to causal diagrams can be found elsewhere (Pearl, 1995, 2009; Glymour and Greenland, 2008). An important result that we will draw upon here is that if the causal diagram is such that all common causes of any two variables on the graph are also on the graph then Pearl's backdoor path theorem applies (Pearl, 1995). A backdoor path from a variable A to another variable Y is a sequence of consecutive edges that begins with an edge pointing into A. Pearl's backdoor path theorem can be stated as follows: if a set of variables C that are not effects of A blocks all backdoor paths from A to Y then $Y_a \perp\!\!\!\perp A|C$ for all a. See Pearl (1995, 2009) or Glymour and Greenland (2008) for formal definitions of blocked paths and other related concepts. Essentially, if C satisfies this backdoor path criterion then the effect of A on Y is unconfounded in distribution conditional on C. When unmeasured confounding is present its influence can sometimes be assessed through sensitivity analysis (Schlesselman, 1978; Rothman et al. 2008; Lash et al., 2009; VanderWeele and Arah, 2011) or reasoning about the sign of the bias (VanderWeele, 2008; VanderWeele and Robins, 2010).

Effect Modification: Distribution and Measure

We have seen that a distinction can be drawn between confounding in distribution and confounding in measure. A similar distinction can in fact also be drawn with regard to effect modification as the following two definitions make clear.

Definition 3. We say that there is effect modification *in distribution* across strata of Q for the effect of A on Y if $P(Y_a|Q = q)$ varies with q.

Definition 4. We say that there is effect modification *in measure* μ across strata of Q for the effect of A on Y if $\mu(E(Y_1|Q = q), E(Y_0|Q = q))$ varies with q.

Definition 3 considers different subpopulations defined by their level of some variable Q and considers what would happen if treatment for all individuals were set to level a. It effectively asks whether the distribution of the outcome Y_a would be comparable across different strata of Q. If not then effect modification in distribution is said to be present. Definition 4, concerning effect modification "in measure," is the definition more commonly employed in the epidemiologic literature. With definition 4, effect modification is said to be present *in measure* μ if the effect of the exposure using measure μ (e.g. the risk difference, risk ratio, or odds ratio scale) varies across strata of Q. Definition 4 is the definition that is generally considered when effect modification is in view. Definition 3 concerning effect modification *in distribution* helps to see the parallel with confounding *in distribution*, but, a little reflection makes clear that Definition 3 concerning effect modification in distribution is a fairly trivial concept insofar as if Q has any effect whatsoever on Y then there will in general be effect modification in distribution (even if the exposure A has no effect on the outcome). We do not advocate the use of Definition 3 in practice but employ it in this paper simply so as to draw the appropriate parallels and distinctions with confounding.

More recently, the expression "effect-measure modification" (Rothman, 2002; Brumback and Berg, 2008) has been used in place of the expression, "effect modification." This has arguably occurred for two reasons. First, as has often been pointed out (Miettinen, 1974; Rothman, 2002; Brumback and Berg, 2008; Rothman et al., 2008), there may be effect modification for one measure (e.g. the risk difference) but not for another (e.g. the risk ratio). Effect modification in measure is thus scale-dependent and the expression "effect-measure modification" makes this more explicit. Second, with observational data, control for confounding is often inadequate; the quantities we estimate

from data may not reflect true causal effects. The expression "effect-measure modification" suggests only that our measures (which may not reflect causal effects) vary across strata of Q, rather than the effects themselves (which we may not be able to consistently estimate).

If the effect of A on Y is unconfounded in distribution conditional on C with Q a subset of C then $P(Y_a|Q=q) = \sum_c P(Y|A=a, C=c)P(Q=q|C=c)$ and one can then assess effect modification in distribution by assessing whether this quantity varies with q. If effect of A on Y is unconfounded in distribution conditional on C so that $E(Y_a|A=1, C=c) = E(Y_a|A=0, C=c)$ for all a and c, then one can assess effect modification in measure by assessing whether $\mu(\sum_c P(Y|A=1, C=c)P(Q=q|C=c), \sum_c P(Y|A=0, C=c)P(Q=q|C=c))$ varies with q.

In the definition of effect modification given above, we considered whether the effect of an exposure on an outcome varies across strata defined by another variable. This is arguably how the term "effect modification" has traditionally been used within epidemiology. However, a distinction should be drawn between this setting and one in which interventions on both factor, e.g. on both A and Q, are considered. When interventions on two factors are considered (rather than interventions on one factor assessed across strata defined by another factor), the resulting measures might be better referred to as measures of "causal interaction." The distinction has important implications for confounding control - whether confounding for one or two factors needs to be controlled for - and is considered in greater detail elsewhere (VanderWeele 2009a; VanderWeele and Knol, 2011). Here our focus will be on "effect modification"/"effect heterogeneity" rather than "causal interaction." For effect modification/effect heterogeneity, the secondary factor (i.e. the "effect modifier") may or may not itself have a causal effect on the outcome; it may serve as a proxy for a variable that does. For causal interaction the secondary factor must have an effect on the outcome. It should be noted that some authors suggest refraining from the use of "effect modifier" for a variable that does not itself have a causal effect on the outcome (Shahar and Shahar, 2010; cf. VanderWeele, 2010). The term "effect heterogeneity" may better capture the notion that an effect varies across strata defined by another variable. Here, however, we will retain the traditional use of the term "effect modification" or the more recent variant "effect-measure modification" (Rothman, 2002; Brumback and Berg, 2008) for the phenomenon of an effect varying across strata of another variable.

As with Definitions 1 and 2, Definitions 3 and 4 are also related as indicated by the following proposition.

Proposition 2. If Y is binary then there is effect modification in distribution across strata of Q if and only if there exists a measure μ such that there is effect modification in measure μ across strata of Q.

That effect modification in measure implies effect modification in distribution holds for arbitrary Y but the reverse implication requires binary Y.

Note that with observational data, neither statements about confounding nor about effect modification can be definitively verified. The definitions for confounding and effect modification in either distribution or measure are statements about counterfactual outcomes; because we do not observe the potential outcomes for each individual under the two different exposure states, we cannot check these conditions. At best we can attempt to collect data on a sufficiently rich set of covariates C such that the assumption of no confounding (in distribution or measure) is thought reasonable, conditional on C. Under this assumption we can then assess effect modification.

In a study in which A is randomized the effect of A on Y will be unconfounded (in distribution and measure) both unconditionally and conditional on any set of pre-randomization covariates C. In a randomized study, we can thus also consistently estimate measures of effect modification. Note that even though the secondary factor Q is not randomized, effect modification, as defined in Definitions 3 and 4, concerns whether the effect of A on Y varies across strata of Q. This need not indicate a causal effect of Q itself on Y; Q may be serving as a proxy for another variable that has a causal effect on Y (VanderWeele, 2009a, 2010).

In many definitions of confounding and effect modification, effect modification is taken as scale-dependent and confounding as being scale-independent. We have seen in this section, however, that there are analogues between confounding and effect modification for both distribution and measure. Although confounding is often taken as confounding in distribution (Definition 1) and is thus scale-independent, Definition 2 for confounding in measure makes reference to a particular measure and is thus scale-dependent. Likewise although effect modification is often taken as effect modification in measure (Definition 4) and is thus scale-dependent, Definition 3 for effect modification in distribution does not make reference to a particular measure; it is scale independent. In subsequent sections we will consider other similarities and differences in properties of confounding and effect modification. We will discuss how while confounding depends on how the exposure was assigned, effect modification does not. Likewise, we will see that although confounding and effect modification are both relative to other variables being conditioned on, and to the

population, and to the exposure and outcome of interest, the ways in which confounding and effect modification are relative to these different factors varies.

Conditionality of Confounding and Effect Modification

Both confounding (in distribution and measure) and effect modification (in distribution and measure) are dependent on what other variables are being conditioned upon (Miettinen, 1974; Rothman et al., 2008; VanderWeele, 2009b). We will consider confounding and effect modification in turn.

Conditional on X, an additional variable C clearly assists in confounding control if $Y_a \perp\!\!\!\perp A|(C, X)$ but it is not the case that $Y_a \perp\!\!\!\perp A|X$ i.e. if the effect of A on Y is unconfounded in distribution conditional on (C, X) but not conditional on X. If this is the case we will say that C is a confounder in distribution (conditional on X). Whether a variable assists in confounding control will depend on the other variables for which control is made. In the causal diagram in Figure 1 for instance, C_1 blocks all backdoor paths from A to Y and thus $Y_a \perp\!\!\!\perp A|C_1$.
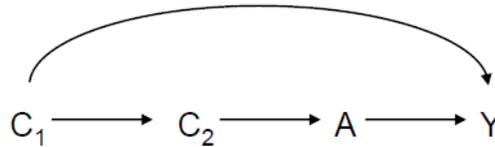

Fig. 1. Diagram illustrating that confounding is relative to the other variables for which control is made.

The variable C_1 thus unconditionally assists in confounding control. However, conditional on C_2, control for C_1 is irrelevant. The effect of A on Y is unconfounded in distribution conditional on C_2 or (C_1, C_2). Conditional on C_2, the variable C_1 does not assist in confounding control. Conversely, in Figure 2, C_1 does not unconditionally assist in confounding control.
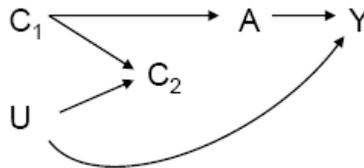

Fig. 2. Diagram illustrating collider bias.

The effect of A on Y is unconfounded when conditioning on nothing or when conditioning on C_1. However, conditional on C_2 (e.g. if we wanted to compute

effect measures within strata of C_2), C_1 does assist in confounding control. The effect of A on Y is unconfounded in distribution conditional on (C_1, C_2) but not conditional on C_2 alone because of what is sometimes called "collider stratification" or "M-bias" (Greenland, 2003; Cole et al., 2010).

Likewise effect modification is dependent on the other variables being conditioned upon. We could say that there is effect modification in distribution across strata of Q for the effect of A on Y conditional on C if $P(Y_a | Q = q, C = c)$ varies with q and that there is effect modification in measure μ across strata of Q for the effect of A on Y conditional on C if $\mu(E(Y_1 | Q = q, C = c), E(Y_0 | Q = q, C = c))$ varies with q. On a causal diagram, a necessary condition for Q to serve as an effect modifier (in distribution or measure) for the effect of A on Y conditional on C is that Q be associated with the parents of Y (other than A) conditional on C (VanderWeele and Robins, 2007).

We give two further diagrams to illustrate the conditionality of effect modification. In Figure 3, Q may be an effect modifier (in distribution or measure) for the effect of A on Y unconditionally because it serves as a proxy for C_1 which might interact with A in its effects on Y.
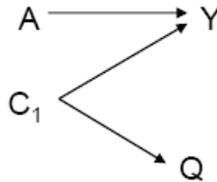

Fig. 3. Diagram illustrating that effect modification is relative to the other variables for which control is made.

However, conditional on C_1, Q will no longer be an effect modifier (in distribution or measure) for the effect of A on Y. Conversely in Figure 4, Q is not an effect modifier (in distribution or measure) unconditionally but may be an effect modifier for the effect of A on Y conditional on C_1 (VanderWeele and Robins, 2007).
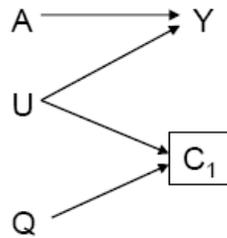
Fig. 4. Diagram illustrating that effect modification may be present conditionally but not unconditionally.

Note that although confounding and effect modification are both clearly relative to what other variables are being conditioned upon, they are relative in different ways. In Figure 1, C_1 no longer assists with confounding control when conditioning on C_2. However, C_1 may still be an effect modifier conditional on C_2. In Figure 3, whether Q is an effect modifier for the effect of A on Y depends on whether or not we are conditioning on C; however, irrespective of whether or not we are conditioning on C, Q will not assist in control of confounding for the effect of A on Y. The effect of A on Y is unconfounded irrespective or whether or not we control for Q (or for C).

In the appendix, to illustrate the conditionality of confounding and effect modification further, we show that it is always possible to hypothetically construct a single variable E such that there is no further confounding conditional on that variable and that it is also possible to hypothetically construct a variable S such that no other variable serves as an effect modifier conditional on S.

Relation to Study Design

For a fixed population and exposure and outcome, whether another variable is a confounder or assists in confounding control depends on how the treatment or exposure was administered. A variable may be a confounder for an exposure-outcome relationship in an observational study, but would not be a confounder if a randomized trial for the effect of the exposure had been conducted in the same population. For example, an observational study by Charig et al. (1986) compared open surgery with percutaneous nephrolithotomy in the treatment of kidney stones. Individuals with open surgery had larger stones on average. The difference in cure rates adjusted for kidney stone size was in fact in the opposite direction of the crude difference in cure rates. In the study, kidney stone size confounded the effect of treatment on the rate of cure. Had the same population been used but if the study design had actually been one in

which the treatments were randomized (rather than an observational study) then kidney stone size would no longer have been a confounder in the study. Whether kidney stone size is a confounder for the study thus depends on the study design and is not an intrinsic property of a variable. As will be seen below confounding is also relative to the population, the exposure and the outcome.

For a fixed population and exposure and outcome, effect modification does not depend on how the exposure or treatment was administered. This difference between confounding and effect modification can also be seen by the definitions given above for confounding and effect modification. The definitions for confounding depend on the distribution of the exposure A in the population (and the distribution of A will depend on whether or not exposure was randomized in a balanced manner). The definitions for effect modification do not make reference to the distribution of the exposure A, only to the distribution of potential outcomes Y_a, which is essentially viewed as a fixed feature of the population in question. The presence of effect modification does not depend on whether the exposure is randomized. In the Charig et al. (1986) study, the investigators also found a larger effect on the risk difference scale comparing open surgery with percutaneous nephrolithotomy for those with smaller kidney stones than those with larger kidney stones. Let us assume that the estimates within strata of kidney stone size obtained by Charig et al. (1986) from observational data accurately reflect the true causal effects, then kidney stone size is an effect modifier for the risk difference measure comparing the two treatments. Suppose now that with the same population treatment had been randomized, the effects comparing open surgery with percutaneous nephrolithotomy by kidney stone size would remain the same (assuming there was no confounding in the observational study conditional on kidney stone size) and once again kidney stone size would be an effect modifier for the risk difference measure. Although effect modification does not depend on how treatment was assigned, it is, as noted above, relative to the other variables being conditioned upon and is also, as noted below, relative to the population, the exposure and the outcome.

Relativity of Confounding and Effect Modification to Population

Confounding depends on how treatment was assigned; effect modification does not. Both are however relative to a population. A variable might serve as a confounder for a cohort design of one population but not serve as a confounder for a cohort design of another population. This may be because the potential confounder is related to the exposure in one population but

not in another; or it may be because the potential confounder is related to the outcome in one population but not in another. For example, Kwok et al. (2010) noted that in observational studies of breast feeding in western countries, higher socioeconomic status both increased the likelihood of breast-feeding and decreased the likelihood of having an obese child; however, in a study in Hong Kong, lower socioeconomic status increased the likelihood of breast-feeding but was not as clearly related to obesity in children. In examining the effects of breastfeeding on obesity in children, socioeconomic status would thus likely be a confounder in the western studies but perhaps not in the study in Hong Kong. Interestingly, a randomized trial on breastfeeding promotion (Kramer et al., 2007) found an effect of breastfeeding on IQ but not on obesity; the studies of breastfeeding in western countries (perhaps subject to confounding by SES) suggested an effect of breastfeeding on both childhood obesity and IQ whereas the study in Hong Kong (where confounding by SES was less likely an issue) indicated an effect only for IQ.

Effect modification is, like confounding, relative to a population. For example, suppose Q modifies the risk difference for the effect of A but that it is also the case that A only has an effect in the presence of some genetic factor $G = 1$ i.e. there is no effect of A if $G = 0$. Suppose that in population 1, some individuals have the genetic factor ($G = 1$) but that the genetic factor is entirely absent in population 2. Then Q might serve as an effect modifier for the risk difference for A in population 1 but it would not in population 2 since the effect of A in population 2 would be 0 for all levels of Q. Said another way, the prevalence of factors other than the exposure A and the potential effect modifier Q, may differ across populations. This point was also illustrated in a very succinct manner by Rothman using sufficient-component cause diagrams (Rothman, 1976).

That effect modification is relative to the population also points to the fact that population itself can also serve as an effect modifier. Population is clearly an effect modifier for the exposure A in the hypothetical example with the gene just given. Also in the breast-feeding example, it is possible that the effects of breast-feeding in Hong Kong and in western countries are in fact different. Kwok et al. (2010) notes that breastfed infants are more likely to be given glucose drinks in Hong Kong; this might interact with the effects of breastfeeding on obesity, effectively cancelling them out.

Relativity of Confounding and Effect Modification to Both Exposure and Outcome

From Definitions 1-4, we see that both confounding and effect modification make reference to a specific exposure and a specific outcome. A variable is not simply a confounder (or an effect modifier) for a treatment nor is it simply a confounder (or effect modifier) for an outcome. Rather it will be or not be a confounder (or effect modifier) for a specific exposure-outcome relationship. Whether a variable is a confounder for a specific exposure is relative to the particular outcome. In the causal diagram in Figure 5, C_1 is a confounder for the effect of A on Y; however it is not a confounder for the effect of A on V. Again, a variable is a confounder for a specific exposure-outcome relationship not simply for a specific exposure.
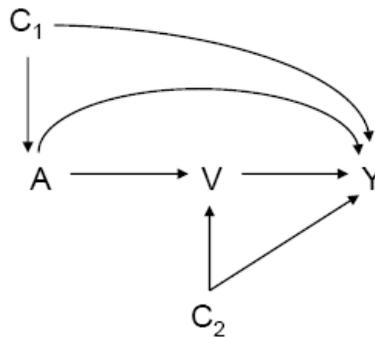

Fig. 5. Diagram illustrating that confounding is relative to the exposure and to the outcome: C_1 is a confounder for the effect of A on Y; however it is not a confounder for the effect of A on V; C_2 is a confounder for the effect of V on Y; however it is not a confounder for the effect of A on Y.

Likewise, whether a variable is a confounder for a specific outcome is relative to the particular exposure. In Figure 5, C_2 is a confounder for the effect of V on Y; however it is not a confounder for the effect of A on Y; control for C_2 would not be necessary if the effect of A on Y were of interest. We see a variable is a confounder not simply for a specific outcome but for a specific exposure-outcome relationship.

Similarly, a variable is an effect modifier for a specific exposure-outcome relationship not simply for a specific exposure. In the causal diagram in Figure 6, Q may serve as an effect modifier for the effect of A on Y; however, in Figure 6, Q cannot serve as an effect modifier of the effect of A on V (VanderWeele and Robins, 2007). We see then from Figure 6 that whether a variable is an effect modifier is relative to the particular outcome.
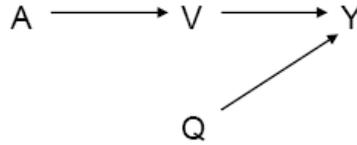
Fig. 6. Diagram illustrating effect modification is relative to the outcome: Q may serve as an effect modifier for the effect of A on Y but Q cannot serve as an effect modifier of the effect of A on V.

Likewise, a variable is an effect modifier for a specific exposure-outcome relationship not simply for a specific outcome. In the causal diagram in Figure 7, Q may serve as an effect modifier for the effect of A on Y if A and Q interact in their effects on Y; however, in Figure 7, Q cannot serve as an effect modifier of the effect of V on Y (VanderWeele and Robins, 2007). We see then from Figure 7 that whether a variable is an effect modifier is relative to the particular exposure.
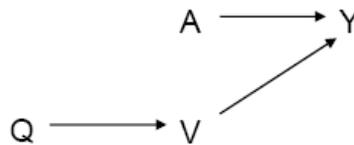

Fig. 7. Diagram illustrating effect modification is relative to the exposure: Q may serve as an effect modifier for the effect of A on Y but Q cannot serve as an effect modifier of the effect of V on Y.

Relations Between Confounding and Effect Modification: Distribution and Measure

A variable may be an effect modifier without being a confounder as occurs with subgroup analyses in randomized trials. We saw this also in Figure 3. In fact Figure 3 suffices to demonstrate that a variable can be an effect modifier in distribution for the effect of A on Y without it being a confounder in distribution and also that a variable can be an effect modifier for measure μ without it being a confounder for measure μ. Likewise, it has been noted previously (Miettinen, 1974; Fisher and Patil, 1974; Greenland and Morgenstern, 1989; Rothman et al., 2008) that a variable can be a confounder for measure μ (and thus also a confounder in distribution) without it being an effect modifier for measure μ. To see this suppose that $Y_a \perp\!\!\!\perp A|C$ and $E(Y|A = 1, C = 1) = 0.6, E(Y|A = 0, C = 1) = 0.5, E(Y|A = 1, C =$

$0) = 0.3, E(Y|A = 0, C = 0) = 0.2$ so that $E(Y_1|C = 1) - E(Y_0|C = 1) = 0.1 = E(Y_1|C = 0) - E(Y_0|C = 0)$ and thus $E(Y_1 - Y_0) = 0.1$ so that C is not an effect modifier for the risk difference measure. Suppose, however, $P(C = 1|A = 1) = 0.8$ and $P(C = 1|A = 0) = 0.4$ then $E(Y|A = 1) = (0.6)(0.8) + (0.3)(0.2) = 0.54$ and $E(Y|A = 0) = (0.5)(0.4) + (0.1)(0.6) = 0.26$. Thus $E(Y|A = 1) - E(Y|A = 0) = 0.28 \neq 0.1 = E(Y_1 - Y_0)$ and so C serves as a confounder for the risk difference measure for the effect of A on Y.

The question remains, however, whether a variable may be a confounder in distribution without it being an effect modifier in distribution. The following proposition answers this question negatively.

Proposition 3. If $Y_a \perp\!\!\!\perp A|(C, X)$ but it is not the case that $Y_a \perp\!\!\!\perp A|X$ then C must be an effect modifier in distribution for the effect of A on Y conditional on X.

Proposition 3 essentially states that if C is a confounder conditional on X (so that the effect of A on Y is unconfounded in distribution conditional on (C, X) but not conditional on X alone) then C must be an effect modifier in distribution for the effect of A on Y conditional on X. Thus while a variable can be a confounder in measure but not an effect modifier in measure, or can be an effect modifier in measure but not a confounder in measure, or an effect modifier in distribution but not a confounder in distribution, a variable that is a confounder in distribution must also be an effect modifier in distribution.

Confounding and Effect Modification in Models

Confounding and effect modification, as conceived in this paper, and in much of modern epidemiology are causal concepts: they relate to the distribution of counterfactual variables. In practice, however, statistical models are often used to reason about the presence or absence of confounding and effect modification.

To assess whether an additional variable C is a confounder for the effect of an exposure A on an outcome Y when already controlling for covariates X, an investigator will often fit the two models:

$$g\{E[Y|A = a, X = x, C = c]\} = \beta_0 + \beta_1 a + \beta_2' x + \beta_3 c$$
$$g\{E[Y|A = a, X = x]\} = \beta_0^* + \beta_1^* a + \beta_2^{*\prime} x$$

where g is a link function, and will examine whether β_1 is equal to β_1^*. If they are equal, then often C is discarded as a confounder. Although this approach will in some settings give valid results (Greenland et al. 1999; VanderWeele and Shpitser, 2011), several caveats are important. First, the procedure assumes

that the set of variables (X, C) with which one begins suffices to control for confounding for the effect of A on Y, at least for the measure corresponding to the link function g e.g. for a difference measure if g is the identity link. If the original set (X, C) does not suffice to control for confounding it may be the case that β_1 and β_1^* are equal but that if a sufficient set of confounders were included in the model, the coefficients in models with and without C would differ. Backwards selection techniques, as are often used in practice by iteratively applying the procedure above, will in general only be valid if the original set of covariates considered itself suffices to control for confounding (VanderWeele and Shpitser, 2011).

A second and perhaps even more important and neglected caveat is that although the change-in-coefficient procedure above will, provided (X, C) suffices to control for confounding, be valid for difference and risk ratio measures, it fails for logistic regression and odds ratio measures (Greenland et al., 1999). This is because the odds ratio is not a collapsible measure: even if C is not a confounder (or if the exposure A is randomized so there is no confounding), controlling for an additional covariate C will in general change the odds ratio (Greenland et al., 1999). If the exposure is randomized, controlling for more and more covariates will in general increase the odds ratio measure (Robinson and Jewell, 1991). The change-in-coefficient procedure should not be used for logistic regression unless the outcome is rare in which case odds ratios approximate risk ratios and the procedure may thus apply approximately (Greenland et al., 1999). Third, if the set (X, C) does not suffice to control for confounding, a change in coefficients may occur in settings in which absence of control for a pre-exposure covariate yields an unbiased estimate but control for the covariate does not. This occurs in settings with collider stratification such as the variable C_2 in Figure 2. Controlling for C_2 versus no covariates would in general change the regression coefficient for A, but it would be the estimate without controlling for C_2 that would be unbiased. A change in coefficient would not in such settings indicate that control should be made for the covariate. Fourth, even when the change-in-coefficient procedure is valid, when it is actually applied, what is being compared in practice is coefficient estimates rather than the true coefficients and thus the approach is subject to error due to sampling variability. Decisions about confounder control are in general best made on substantive rather than statistical grounds. Finally, even in settings in which the change-in-coefficient procedure yields valid conclusions, the conclusions concern confounding in the measure for the scale corresponding to link function g. As noted in previous sections, a change in scale may alter whether a variable is a confounder in measure.

To assess whether a variable Q is an effect modifier for the effect of an

exposure A on an outcome Y conditional on covariates X, an investigator will often fit a model

$$g\{E[Y|A = a, X = x, Q = q]\} = \beta_0 + \beta_1 a + \beta_2' x + \beta_3 q + \beta_4 aq$$

and assess whether the coefficient, β_4, for the product term is non-zero. Provided that the model is correctly specified and that (X, Q) suffices to control for confounding of the effect of A on Y, β_4 will provided a measure of effect modification of the effect of A on Y conditional on X for the measure corresponding to the link function g. However, as noted in previous sections, the presence or absence of effect modification on one scale does not imply the presence or absence of effect modification for another. This approach of examining the coefficient for the product term relates to effect modification in measure. For effect modification in distribution, if either β_3 or β_4 are non-zero then there will be effect modification in distribution in so far as the distribution of counterfactual outcomes Y_a, conditional on (X, Q), will vary across strata of Q. As noted above, however, effect modification in distribution is a very weak notion of effect modification; it may be present even if A has no effect on Y at all.

It is sometimes commented that if a variable is an effect modifier, we no longer are concerned whether it is a confounder. Such comments arguably arise from treating confounding and effect modification as statistical concepts rather than causal concepts (i.e. not making reference to counterfactuals). If confounding is defined, rather than merely assessed, by the change-in-coefficient method, then once we fit the model above for effect modification, including the interaction term $\beta_4 aq$, we no longer have a single coefficient for A. Rather we have two, namely, β_1 and β_4, and so it seems that the change-in-coefficient method to assess confounding breaks down. However, in observational studies, whether a variable is a confounder is always a concern, regardless of whether it is an effect modifier, or of whether we are interested in assessing effect modification. Depending on the context, effect modification may or may not be of intrinsic interest. Often we will be interested in effect modification in order to target populations in which some intervention will be most effective. However, in other contexts concerning policy decisions which may result in either the entire population being exposed or unexposed, the overall treatment effect, rather than effect modification measures, may be what is most important. Even when assessing the overall treatment effect is the primary study goal, product terms may have to be included in statistical models to yield accurate estimates of the overall effect; one may have to average over the distribution of effect modifiers to estimate the overall treatment effect; assessing effect modifi-

cation itself may then not be what is of central interest. Confounding is always a concern in observational research; we should be concerned if a variable is a confounder even when it is an effect modifier.

It should finally be noted that more recently, instead of using regression models to estimate effects, marginal structural model, fit using an inverse probability of treatment weighting (IPTW) are now often being employed. One of the conceptual advantages of the marginal structural model approach is that it more clearly distinguishes the analytic procedures for handling confounding and effect modification (Robins et al., 2000). Suppose we have data on exposure A, outcome Y and covariates (X, Q). A marginal structural model for the overall effect of A on Y takes the form:

$$g\{E[Y_a]\} = \alpha_0 + \alpha_1 a.$$

The parameters of the marginal structural model can be estimated by fitting a conditional regression model:

$$g\{E[Y|A = a]\} = \alpha_0 + \alpha_1 a$$

where each subject i is weighted by the inverse probability of treatment weight

$$w^i = \frac{P(A = a^i)}{P(A = a^i|X = x^i, Q = q^i)}$$

where a^i, x^i, and q^i are the actual values of A, X, and Q for subject i. Provided (X, Q) suffices to control for confounding for the effect of A on Y, this inverse probability of treatment weight procedure for the conditional model

$$g\{E[Y|A = a]\} = \alpha_0 + \alpha_1 a$$

will give consistent estimators of the parameters of the marginal structural model. Control for confounding is made, not by covariate adjustment as in regression, but by weighting. The weights themselves, w^i, may be estimated by modeling both the numerator and denominator probabilities using logistic regression. Provided the models for these probabilities are correctly specified, the procedure will still yield consistent estimators of the marginal structural model even if the estimated weights, rather than the true weights, are used (Robins et al., 2000).

On the other hand, if effect modification is of interest, one may use a

marginal structural model of the form:

$$g\{E[Y_a|Q=q]\} = \alpha_0 + \alpha_1 a + \alpha_2 q + \alpha_3 aq.$$

Note that α_3 in this marginal structural model gives a measure of effect modification that is marginalized over X rather than conditional on X as in the regression based approach above. The parameters of the marginal structural model for effect modification can be estimated by fitting a conditional regression model:

$$g\{E[Y|A=a, Q=q]\} = \alpha_0 + \alpha_1 a + \alpha_2 q + \alpha_3 aq$$

where each subject i is weighted by the inverse probability of treatment weight

$$v^i = \frac{P(A=a^i|Q=q^i)}{P(A=a^i|X=x^i, Q=q^i)}$$

where a^i, x^i, and q^i are again the actual values of A, X, and Q for subject i. Provided (X, Q) suffices to control for confounding for the effect of A on Y, this inverse probability of treatment weighting procedure for the conditional model

$$g\{E[Y|A=a, Q=q]\} = \alpha_0 + \alpha_1 a + \alpha_2 q + \alpha_3 aq$$

will give consistent estimators of the parameters of the marginal structural model for effect modification. Using the marginal structural model / IPTW approach, the distinction between confounding and effect modification is made clear in the analytic procedure itself insofar as the model that is fit and the weights that are used are both different in the procedures used for confounding control versus effect modification assessment. See Robins et al. (2000) for further details on fitting marginal structural models. Finally, it should be noted that the marginal structural models used for assessing effect modification are different from those assessing causal interaction i.e. when assessing the effects of interventions on two exposures, rather than interventions on one exposure within strata of another factor; for causal interaction there are two sets of confounding variables and two sets of weights are used (VanderWeele, 2009).

Discussion

This paper has considered the properties of and relationships between confounding and effect modification. We have seen that both confounding and effect modification can be defined with respect to distributions of potential

outcomes or with respect to specific measures. We can summarize the properties of confounding and effect modification as follows. When defined with respect to distribution, neither confounding nor effect modification is scale-dependent. When defined with respect to measure, both confounding and effect modification are scale-dependent. Both confounding and effect modification are relative to what other variables are being conditioned upon; however the ways in which confounding and effect modification are relative differ. The presence of confounding depends on the manner in which the exposure was assigned; the presence of effect modification does not. Both confounding and effect modification are relative to the population in question. Both confounding and effect modification are relative to the specific exposure and outcome under study; a variable is not a confounder or an effect modifier for a particular exposure, nor simply for a particular outcome, but for the relation between a specific exposure and a specific outcome. A variable may be an effect modifier for a specific measure without it being a confounder; likewise a variable may be a confounder for a specific measure without it being an effect modifier. A variable can be an effect modifier in distribution without it being a confounder in distribution. However, a variable cannot be a confounder in distribution without it being effect modifier in distribution.

The purpose of this paper has been primarily conceptual. However, the properties considered and the distinctions drawn have important implications for data analysis. Several points merit attention with regard to effect modification. First, it has been noted repeatedly that effect modification is relative to the effect measure (Miettinen, 1974; Rothman, 2002; Brumback and Berg, 2008; Rothman et al., 2008); one may have effect modification on one scale but not on another. However, this is not the only factor to which effect modification is relative and which must be considered in interpretation. In addition to scale, the conditionality of effect modification on other covariates is important whenever one is interpreting effect modification analyses. Two studies of the same population may report different conclusions concerning effect modification because different variables are controlled for in the analysis. A variable may be an effect modifier because it serves as a proxy for another variable that actually interacts with the exposure of interest; a different analysis that controlled for this variable that truly interacted with the exposure might then have that the effect of exposure no longer varies across strata of the original effect modifier (VanderWeele, 2009a; VanderWeele and Robins, 2007; VanderWeele and Knol, 2011). That effect modification is relative to a population should also be taken into account in the interpretation of effect modification analyses. Two analyses of different populations that study the same exposure, outcome, effect modifier and condition on the same covariates may yield dif-

ferent conclusions about effect modification. This should not necessarily be taken as indicating that one of the analyses must be wrong; it is possible for there to be effect modification in one population but not in another. The relative nature of effect modification to the scale, to the other covariates in the analysis, and to the population are important to consider when interpreting effect modification analyses.

Our discussion of the properties of confounding is also relevant for data analysis. First, that confounding is relative to the population may be helpful in reducing biases in causal effects. A covariate that is strongly related to the outcome in one population, may be unassociated with the outcome in a second population. Analyses of causal effects may be subject to much less confounding bias in one population than in another. When possible, it may thus be advantageous to undertake observational studies in populations where confounding is thought to be less problematic. Second, as has been previously noted (Miettinen, 1974; Greenland and Robins, 1986; Rothman et al., 2008; Pearl, 2009), the conditionality of confounding implies that one cannot simply check whether each covariate is unconditionally associated with the exposure and with the outcome (conditional on exposure) in determining whether or not a variable is a confounder. The associations conditional on all other covariates must be considered. Due to the need to consider all associations conditionally, backward selection techniques may be more relevant if reduction in the number of covariates is thought desirable (Robins, 1997; VanderWeele and Shpitser, 2011); even then we must consider whether the initial set of covariates suffice to control for confounding on substantive grounds. Third, the distinction between confounding in distribution versus measure becomes important when considering "collapsibility" approaches to confounding assessment i.e. in settings in which an investigator evaluates confounding by comparing an adjusted and unadjusted estimate. Greenland et al. (1999) showed that for the risk difference and the risk ratio scales, collapsibility follows from no-confounding and vice versa. However, this implication holds for confounding in measure, not confounding in distribution. One may have collapsibility on the risk difference scale and therefore conclude that a particular variable is not a confounder of the risk difference (conditional on the other covariates); however, this does not imply that the variable is not a confounder for the risk ratio; it might be necessary to make control for that variable in evaluating the risk ratio. Collapsibility of the risk difference implies no confounding in measure for the risk difference; collapsibility of the risk ratio implies no confounding in measure for the risk ratio; however, neither implies no confounding in distribution. One must be careful when changing scales - not only in assessing effect modification - but also when thinking about confounding.

Notions of counterfactuals or potential outcomes from the causal inference provide a formal framework in which to conceptualize causation. The phenomena of confounding and effect modification are concerned respectively with how such causal effects relate to the observed data and how they may vary across strata of other variables. These two phenomenon are distinct but, as has been seen, also intimately related. The concepts and formalizations that have developed from within causal inference literature more clearly shed light on the properties of and the relationships and distinctions between these two important epidemiologic concepts.

References

Barnow, B.S., Cain, G.G., and Goldberger, A.S. (1980). Issues in the analysis of selectivity bias. In: E. Stromsdorfer and G. Farkas (Eds.), *Evaluation Studies* vol. 5. San Francisco: Sage.

Brumback, B. and Berg, A. (2008). On effect-measure modification: relationships among changes in the relative risk, odds ratio, and risk difference. *Statistics in Medicine*, 27:3453-3465.

Charig, C.R., Webb, D.R., Payne, S.R., and Wickham, O.E. (1986). Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. *British Medical Journal*, 292:879-82.

Cole, S.R., Platt, R.W., Schisterman, E.F., Chu, H., Westreich, D., Richardson, D. and Poole, C. (2010). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, 39:417-420.

Fisher, R.A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Fisher, L. and Patil, K. (1974). Matching and unrelatedness. *American Journal of Epidemiology*, 100:347-349.

Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics*, 58:21-29.

Geng, Z., and Li, G. (2002). Conditions for non-confounding and collapsibility without knowledge of completely constructed causal diagrams. *Scandinavian Journal of Statistics*, 29:169-181.

Glymour, M.M. and Greenland, S. (2008). Causal diagrams In: Rothman, K.J., Greenland, S., Lash, T.L. (eds.) *Modern Epidemiology*, 3rd edition, Chapter 12. Philadelphia: Lippincott Williams and Wilkins.

Greenland, S. (1990). Randomization, statistics and causal inference, *Epidemiology*, 1:421-429.

Greenland, S. (2003). Quantifying biases in causal models: classical confounding versus collider-stratification bias. *Epidemiology*, 14:300-306.

Greenland, S. and Morgenstern, H. (1989). Ecological bias, confounding, and effect modification. *American Journal of Epidemiology*, 18:269-274.

Greenland, S. and Robins, J.M. (1986). Identifiability, exchangeability, and epidemiologic confounding. *International Journal of Epidemiology*, 15:413-419.

Greenland, S. and Robins, J.M. (2009). Identifiability, exchangeability and confounding revisited. *Epidemiologic Perspectives and Innovations*, 6, Article 4.

Greenland, S., Robins, J.M., and Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, 14:29-46.

Hernán, M.A. (2004). A definition of causal effect for epidemiological studies. *Journal of Epidemiology and Community Health*, 58:265–271.

Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics*, 86:4-29.

Kramer, M.S., Matush, L., Vanilovich, I. et al. (2007). Effects of prolonged and exclusive breastfeeding on child height, weight, aidposity, and blood pressure at age 6.5 y: evidence from a large randomized trial. *American Journal of Clinical Nutrition*, 86:1717-1721.

Kwok, M.K., Scholling, C.M., Lam, T.H., Lueng, G.M. (2010). Does breastfeeding protect against childhood overweight? Hong Kong's 'Children of 1997' birth cohort. *International Journal of Epidemiology*, 39:297-305.

Lash, T.L., Fox, M.P. and Fink, A.K. (2009). Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer.

Miettinen, O. (1974). Confounding and effect modification. *American Journal of Epidemiology*, 100:350-353.

Pearl, J. (1995). Casual diagrams for empirical research (with discussion). *Biometrika*, 82:669-710.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference.* 2nd Edition Cambridge: Cambridge University Press.

Robins, J.M. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79:321-334.

Robins, J.M. (1997). Causal Inference from Complex Longitudinal Data. In: *Latent Variable Modeling and Applications to Causality.* Lecture Notes in Statistics (120), M. Berkane, Editor. NY: Springer Verlag, pp. 69-117.

Robins, J.M., Hernán, M.A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550-560.

Robinson, L. and Jewell, N.P. (1991). Some surprising results about covariate adjustment in logistic regression models. *International Statistical Rev*iew, 59, 227–240.

Rosenbaum, P.R. and Rubin, D.B. (1983).The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983; 70:41-55.

Rothman, K.J. (1976). Causes. *American Journal of Epidemiology*, 104:587-592.

Rothman, K.J. (1977). Epidemiologic methods in clinical trials. *Cancer*, 39:1771-1775.

Rothman, K.J. (2002). Measuring interactions. *Epidemiology: An introduction.* Oxford: University Press, pp. 168-180.

Rothman, K.J., Greenland, S., and Lash, T.L. (2008). *Modern Epidemiology.* 3rd ed., Philadelphia, Lippincott Williams & Wilkins.

Rubin, D.B. (1990). Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25:279-292.

Rubin, D.B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 47:1213-1234.

Schlesselman, J.J. (1978). Assessing effects of confounding variables. *American Journal of Epidemiology*, 1978; 108, 3-8.

Shahar, E. and Shahar, D.J. (2010). On the definition of effect modification. *Epidemiology*, 21:586-587.

Stone, R. (1993). The assumptions on which causal inference rest. *Journal of the Royal Statistical Society Series B*, 455-466.

Stürmer, T., Rothman, K.J., and Glynn, R.J. (2006). Insights into different results from different causal contrasts in the presence of effect-measure modification. *Pharmacoepidemiology and Drug Safety*, 15:698–709.

VanderWeele, T.J. (2008). The sign of the bias of unmeasured confounding. *Biometrics*, 64:702-706.

VanderWeele T.J. (2009a). On the distinction between interaction and effect modification. *Epidemiology*, 20:863-871.

VanderWeele TJ. (2009b). On the relative nature of over-adjustment and unnecessary adjustment. *Epidemiology*, 20:496-499.

VanderWeele, T.J. (2010). Response to "On the definition of effect modification," by E. Shahar and D.J. Shahar. *Epidemiology*, 21:587-588.

VanderWeele, T.J. and Arah, O.A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments and confounders. *Epidemiology*, 22:42-52.

VanderWeele, T.J. and Knol, M.J. (2011). The interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Annals of Internal Medicine*, 154:680-683.

VanderWeele, T.J. and Robins, J.M. (2007). Four types of effect modification – a classification based on directed acyclic graphs. *Epidemiology*, 18:561-568.

VanderWeele, T.J. and Robins, J.M. (2010). Signed directed acyclic graphs for causal inference. *Journal of the Royal Statistical Society, Series B*, 72:111-127.

VanderWeele, T.J. and Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics*, 67:1406-1413.

Appendix 1. Proofs

Proof of Proposition 1. Suppose there is confounding in distribution then there must be some value a and c such that $P(Y_a|C = c) \neq P(Y|A = a, C = c)$; we thus have $E(Y_a|C = c) \neq E(Y|A = a, C = c)$. Let a' be some other value of A. Either $E(Y_{a'}|C = c) = E(Y|A = a', C = c)$ or $E(Y_{a'}|C = c) \neq E(Y|A = a', C = c)$. If $E(Y_{a'}|C = c) = E(Y|A = a', C = c)$ then since $E(Y_a|C = c) \neq E(Y|A = a, C = c)$ we have that $E(Y_a|C = c) - E(Y_{a'}|C = c) \neq E(Y|A = a, C = c) - E(Y|A = a', C = c)$ and thus the effect of A on Y is not unconfounded in the risk difference measure conditional on C. If, on the other hand, $E(Y_{a'}|C = c) \neq E(Y|A = a', C = c)$, then either $\frac{E(Y_a|C=c)}{E(Y_{a'}|C=c)} = \frac{E(Y|A=a,C=c)}{E(Y|A=a',C=c)}$ or $\frac{E(Y_a|C=c)}{E(Y_{a'}|C=c)} \neq \frac{E(Y|A=a,C=c)}{E(Y|A=a',C=c)}$. If the latter, then the effect of A on Y is not unconfounded in the risk ratio measure conditional on C. If the former then $E(Y_a|C = c) - E(Y_{a'}|C = c) = E(Y_a|C = c) - E(Y_a|C = c)\frac{E(Y|A=a',C=c)}{E(Y|A=a,C=c)} = E(Y_a|C = c)\{1 - \frac{E(Y|A=a',C=c)}{E(Y|A=a,C=c)}\} \neq E(Y|A = a, C = c)\{1 - \frac{E(Y|A=a',C=c)}{E(Y|A=a,C=c)}\} = E(Y|A = a, C = c) - E(Y|A = a', C = c)$ and thus the effect of A on Y is not unconfounded in the risk difference measure conditional on C. From this it follows that if there is confounding in distribution then there must be confounding either for the risk difference or risk ratio measure. The reverse implication follows essentially immediately: if $P(Y_a|C = c) = P(Y|A = a, C = c)$ then $\mu(E(Y_1|C = c), E(Y_0|C = c)) = \mu(E(Y|A = 1, C = c), E(Y|A = 0, C = c))$.

Proof of Proposition 2. Suppose there is effect modification in distribution then there exist a, q_1 and q_2 such that $P(Y_a|q_1) \neq P(Y_a|q_2)$ so $E(Y_a|q_1) \neq E(Y_a|q_2)$. Let a' be some other value of A. Either $E(Y_{a'}|q_1) = E(Y_{a'}|q_2)$ or $E(Y_{a'}|q_1) \neq E(Y_{a'}|q_2)$. If $E(Y_{a'}|q_1) = E(Y_{a'}|q_2)$ then since $E(Y_a|q_1) \neq E(Y_a|q_2)$ we have that $E(Y_a|q_1) - E(Y_{a'}|q_1) \neq E(Y_a|q_2) - E(Y_{a'}|q_2)$ and thus there is effect modification in the risk difference measure. If, on the other hand, $E(Y_a|q_1) = E(Y_a|q_2)$, then either $E(Y_a|q_1)/E(Y_{a'}|q_1) = E(Y_a|q_2)/E(Y_{a'}|q_2)$ or $E(Y_a|q_1)/E(Y_{a'}|q_1) \neq E(Y_a|q_2)/E(Y_{a'}|q_2)$. If the latter, then there is effect modification in the risk ratio measure. If the former then $E(Y_a|q_1) - E(Y_{a'}|q_1) = E(Y_a|q_1) - E(Y_a|q_1)\frac{E(Y_{a'}|q_2)}{E(Y_a|q_2)} = E(Y_a|q_1)\{1 - \frac{E(Y_{a'}|q_2)}{E(Y_a|q_2)}\} \neq E(Y_a|q_2)\{1 - \frac{E(Y_{a'}|q_2)}{E(Y_a|q_2)}\} = E(Y_a|q_2) - E(Y_{a'}|q_2)$ and thus there is effect modification in the risk

difference measure. From this it follows that if there is effect modification in distribution then there must be effect modification either for the risk difference or risk ratio measure. The reverse implication follows essentially immediately: if $P(Y_a|Q = q)$ does not vary with q then $\mu(E(Y_1|Q = q), E(Y_0|Q = q))$ will not vary with q.

Proof of Proposition 3. Since $Y_a \perp\!\!\!\perp A|(X, C)$ we have that $P(Y_a|C = c, X = x) = P(Y|A = a, C = c, X = x)$. If C is not an effect modifier in distribution for the effect of A on Y conditional on X then $P(Y_a|C = c, X = x)$ does not vary with c. Let c' be some specific value of C so that $P(Y_a|C = c, X = x) = P(Y_a|C = c', X = x)$ for all c. We would then have that

$$
\begin{aligned}
P(Y_a|X = x) &= \sum_c P(Y_a|C = c, X = x)P(C = c|X = x) \\
&= \sum_c P(Y_a|C = c', X = x)P(C = c|X = x) \\
&= P(Y_a|C = c', X = x) \\
&= \sum_c P(Y_a|C = c', X = x)P(C = c|A = a, X = x) \\
&= \sum_c P(Y_a|C = c, X = x)P(C = c|A = a, X = x) \\
&= \sum_c P(Y|A = a, C = c, X = x)P(C = c|A = a, X = x) \\
&= P(Y|A = a, X = x)
\end{aligned}
$$

where the first equality holds by the law of iterated expectations, the second through fifth because $P(Y_a|C = c, X = x) = P(Y_a|C = c', X = x)$ for all c, the sixth because $Y_a \perp\!\!\!\perp A|(X, C)$ and the seventh again by the law of iterated expectations. We have thus shown that if $Y_a \perp\!\!\!\perp A|(X, C)$ and if C is not an effect modifier in distribution for the effect of A on Y conditional on X then $Y_a \perp\!\!\!\perp A|X$. Consequently if $Y_a \perp\!\!\!\perp A|(X, C)$ but it is not the case that $Y_a \perp\!\!\!\perp A|X$ then C must be an effect modifier in distribution for the effect of A on Y conditional on X.

Appendix 2. Conditionality of Confounding and Effect Modification.

To illustrate the conditionality of confounding and effect modification further, we show that it is always possible to hypothetically construct a single variable E such that there is no further confounding conditional on that variable and that it is also possible to hypothetically construct a variable S such that no other variable serves as an effect modifier conditional on S. With regard to confounding, if we define $E = P(A = 1|Y_0, Y_1)$ then by the theory of propensity

scores (14) we will have that $Y_a \perp\!\!\!\perp A|E$ since for all a, $Y_a \perp\!\!\!\perp A|(Y_0, Y_1)$. Thus conditional on E, no further variable is needed to control for confounding. We can construct this variable hypothetically but we of course cannot construct it in practice as we do not observe both (Y_1, Y_0) for any individual.

For effect modification we can define $S = (Y_0, Y_1)$ i.e. S indicates the values of the outcome under each possible exposure condition. If Y were binary there would be four possible values of S: $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. The subgroups defined by the variable S are sometimes referred to as principal strata (39). Conditional on the principal stratum $S = (y_0, y_1)$, no other variable can be an effect modifier. To see this note that $P(Y_a|Q = q, S = (y_0, y_1))$ and $\mu(E(Y_1|Q = q, S = (y_0, y_1)), E(Y_0|Q = q, S = (y_0, y_1)))$ cannot vary with q since conditioning on $S = (y_0, y_1)$ suffices to fix both Y_0 and Y_1. For example, if Y is binary then in the principal stratum with $(0,1)$ all individuals will have $Y_0 = 0$ and $Y_1 = 1$ and thus there can be no further effect modification within this principal stratum.

The remarks here were made with respect to dichotomous exposure A but apply to more general exposures as well. If A takes values in some set $\mathcal{A}$ then we can let $E = P(A = 1|(Y_a)_{a \in \mathcal{A}})$ and we will have $Y_a \perp\!\!\!\perp A|E$ and if we let $S = (Y_a)_{a \in \mathcal{A}}$ then no other variable can be an effect modifier conditional on S.