Submitted to *International Statistical Reviews.*

12 November 2010

TECHNICAL REPORT
R-369
November 2010

Adjustments and their Consequences – Collapsibility Analysis using Graphical Models

Cognitive Systems Laboratory Technical Report R-369r

Sander Greenland
Departments of Epidemiology and Statistics
University of California, Los Angeles, CA 90095-1772, USA
lesdomes@ucla.edu

Judea Pearl
Cognitive Systems Laboratory
Departments of Computer Science and Statistics
University of California, Los Angeles
judea@cs.ucla.edu

SUMMARY

We consider probabilistic and graphical rules for detecting situations in which a dependence of one variable on another is altered by adjusting for a third variable (i.e., noncollapsibility), whether that dependence is causal or purely predictive. We focus on distinguishing situations in which adjustment will reduce, increase, or leave unchanged the degree of bias in an association of two variables when that association is taken to represent a causal effect of one variable on the other. We then consider situations in which adjustment may partially remove or introduce a potential source of bias in estimating causal effects, and some additional special cases useful for case-control studies, cohort studies with loss, and trials with noncompliance (nonadherence).

Keywords: Bias, causality, collapsibility, confounding, instrumental variables, mediation analysis, odds ratio

INTRODUCTION

A common analysis question is whether adjustment for a variable C will reduce, increase, or leave unchanged the degree of association between two other variables, say X and Y. The question comes into focus at two stages of the analysis. First, the investigator may have a simple qualitative model of the data-generating process and may wish to test whether predictions of that model match the observed changes in associations that are induced by various adjustments. Second, when the association of interest is taken to represent the causal effect of X on Y, the investigator may wish to minimize bias by adjusting for the proper set of variables. In both stages, predicting the effect of an adjustment on a given association becomes a question of central concern.

We will focus on graphical (and hence qualitative) tools for recognizing situations in which an adjustment for C can or cannot alter a measure of the dependence of Y on X. These tools apply whether that dependence is causal (i.e., a comparison of Y distributions under different interventions on X) or purely predictive (i.e., a comparison of Y distributions in subpopulations defined by X). When an adjustment alters the measure, we will say the measure is *noncollapsible* over C, although strictly speaking the measure is noncollapsible with respect to the adjustment for C. Thus our paper is about recognizing graphically causal or predictive structures in which we expect adjustment to alter our estimate, and whether the adjustment moves us toward or away from the target parameter.

Schistermann et al. (2009) and VanderWeele (2009) considered aspects of this problem in the context of overadjustment (adjustment that introduces bias) and unnecessary adjustment (adjustment with no impact on bias). We develop a more general framework to also consider when adjustment may only partially remove or introduce a source of bias, including selection

bias; to contrast conditions for odds ratios versus other measures such as risk differences and mean differences; and to show how adjustment may be used to discriminate among competing models. We also consider some additional special cases useful for estimating causal effects from case-control studies with differential selection, cohort studies with differential loss to follow-up, and trials with noncompliance (nonadherence).

The paper begins by reviewing necessary concepts and results from probability and graph theory in the form we will need. We then explain, in a series of examples, how these concepts can be used to determine the impact of various adjustments on bias in estimating causal effects. Following Greenland et al. (1999a), we will reserve the word "control" for those situations in which conditioning has a precise correspondence to experimental control (manipulation or intervention); this excludes most situations in observational studies. Unless stated otherwise, all subpopulations and distributions we discuss will be within the source population of the study, by which we mean the population serving as the source of study subjects (not person-time).

CONDITIONING, SUMMARIZATION, AND STANDARDIZATION

By *conditioning on* a variable (or set of variables) C we will mean examining relations within levels of C (i.e., within strata defined by single values of C). By *summarization* over C we will summarization of conditional (C-specific) dependencies across C. This definition includes pure conditioning, in which the summary is the list (vector) of C-conditional dependence measures, such as C-specific risk differences, risk ratios, log odds ratios, and so on; it also includes averaging these measures over C.

In practice, summarization is usually done using a regression coefficient under a highly fictional model in which the coefficient relating X to Y is assumed constant across C (known as homogeneity, uniformity, parallelism, "no interaction," or "no effect modification"). Any

average coefficient must then equal this constant, so the method of averaging does not matter. We will however focus on the general case, free of homogeneity assumptions, in which the averaging method can be important.

By *adjustment* for C we will then mean one of the many ways in which the dependence of Y on X might account for the relations of C to Y and X. This definition includes averaging of C-specific measures, but also includes standardization (comparisons of average outcomes), which can diverge from averaging of C-specific measures.

We will say a measure is *collapsible* or invariant with respect to an adjustment for C when the adjustment does not change the measure. We will draw primarily on results on collapsibility of risk, rate, and odds-based measures in contingency tables and binary regression (e.g., Whittemore, 1978; Samuels, 1981; Ducharme and LePage, 1986; Gail, 1986; Wermuth, 1987; Greenland and Mickey, 1988; Geng, 1992; Frydenberg, 1990; Clogg et al., 1995; Greenland, 1996; Geng and Li, 2002; Janes et al., 2010). Nonetheless, our discussion applies to continuous variables as well, due to the nonparametric nature of the formulas and graphical results we employ.

Of special focus will be comparisons across X of the distribution of Y given X and C, p(y|x,c), when this distribution is averaged over a specific distribution $p^*(c)$ for C. We will denote these averages by

$$p\{y|x;p^*(c)\} \equiv \sum_c p(y|x,c)p^*(c) \qquad (1)$$

(the sum is over all values of C). We will assume that $p^*(c) = 0$ whenever $p(x,c) = 0$ so that the average remains defined. Such averages are commonly known as the probability of Y=y given X=x, *standardized to* (averaged over) $p^*(c)$.

Important special cases of (1) include total-population averages over p(c),

$$p\{y|x;p(c)\} \equiv \sum_c p(y|x,c)p(c) \qquad (2)$$

and averages over $p^*(c) = p(c|x_r)$ where x_r is a specific reference value of X,

$$p\{y|x;p(c|x_r)\} \equiv \sum_c p(y|x,c)p(c|x_r)\ . \qquad (3)$$

Note when (3) is evaluated at $X = x_r$, C disappears from the expression:

$$p\{y|x_r;p(c|x_r)\} = \sum_c p(y|x_r,c)p(c|x_r) = p(y|x_r).$$

One may examine how averages such as (1), (2), or (3) vary with X. The resulting comparisons across X are called C-standardized measures of the dependence of Y on X. The C distribution $p^*(c)$ is held constant across these comparisons, thus removing this distribution as a factor contributing to variation in the Y distribution across X. When the standard (weighting) distribution $p^*(c)$ and the Y dependence p(y|x,c) are derived from the same population, as in (2) and (3), the resulting average is said to be *population standardized.* Examples include the standardized morbidity ratio (SMR), which divides (3) evaluated at $X=x_r$ by (3) evaluated at another value of X; it simplifies to $p(y|x_r)/p\{y|x;p(c|x_r)\}$.

Standardized distributions are equivalent to the distribution of Y given X obtained after inverse-probability reweighting of the joint distribution using the distribution of X given C (Robins et al., 2000; Sato and Matsuyama, 2002). For example, $p(y|x,c)p(c) = p(y,x,c)/p(x|c)$ and so $p\{y|x;p(c)\} = \sum_c p(y,x,c)/p(x|c)$, which is the joint distribution of Y,X, and C averaged over $1/p(x|c)$.

MEASURE AVERAGING AND COLLAPSIBILITY

The following basic collapsibility results have been noted in various forms at least since Yule (1934):

a) Any standardized probability (1) will simplify to p(y|x) if C is independent of Y given X, i.e., if $p(y|x,c) = p(y|x)$ then $p\{y|x;p^*(c)\} = p(y|x)$.

b) Population-standardized probabilities (2) and (3) will simplify to p(y|x) if C and X are marginally (unconditionally) independent, i.e., if p(c,x) = p(c)p(x) then

$$p\{y|x;p(c|x_r)\} = p\{y|x;p(c)\} = p(y|x).$$

Result (a) follows from the fact that if p(y|x,c) = p(y|x) then p(y|x) factorizes out of the summation over c and the latter summation becomes 1. Result (b) follows by noting that if p(c,x) = p(c)p(x) then expression (2) becomes

$$\sum_c p(y,x,c)p(c)/p(x)p(c) = \sum_c p(y,x,c)/p(x) = p(y,x)/p(x) = p(y|x),$$

and since $p(c|x_r) = p(c)$, expressions (2) and (3) are equal. It follows from these results that population-standardized measures (such as differences and ratios of population-standardized probabilities) are collapsible over C if either (a) C is independent of Y given X, or (b) C and X are marginally independent.

Standardized measures are constructed by taking averages over C *before* comparisons (e.g., ratios or differences) across X. Many other familiar adjusted measures are instead derivable by taking averages of comparisons within levels of C; that is, they average over conditional measures of association, *after* comparison across X. Examples include inverse-variance (information)-weighted averages. Recalling Jensen's inequality (an average of a nonlinear function does not equal the function applied to the averages), it should not be surprising to find divergences between collapsibility conditions depending on the step at which averaging is done (Samuels, 1981, sec. 3).

Standardized difference and ratio measures can be rewritten as averages of conditional measures. For example, in comparing two levels x_1 and x_0 of X using formula (2), the difference is

$$\sum_c p(y|x_1,c)p(c) - \sum_c p(y|x_0,c)p(c) = \sum_c \{p(y|x_1,c) - p(y|x_0,c)\}p(c)$$

which weights the C-specific differences by p(c). The standardized ratio is

$$\sum_c p(y|x_1,c)p(c)/\sum_c p(y|x_0,c)p(c) = \sum_c \{p(y|x_1,c)/p(y|x_0,c)\}p(y|x_0,c)p(c)/\sum_c p(y|x_0,c)p(c)$$

which weights the C-specific ratios by $p(y|x_0,c)p(c)$. This ratio must fall within the range of the C-specific ratios. The same is true of other averages such as Mantel-Haenszel risk ratios (Rothman et al., 2008, Ch. 15) and geometric mean ratios (such as those based on information weighting of log risk ratios).

If an unadjusted (unconditional) measure is outside the range of the C-conditional measures, then the measure cannot be collapsible with respect to any average of C-conditional measures (such as a standardized risk difference or risk ratio). Nonetheless, it may still be collapsible with respect to other adjustments. For example, standardized odds ratios constructed from (1)-(3) usually do not reduce to weighted averages of C-specific odds ratios. Thus an odds ratio may be collapsible with respect to a particular standardization, yet may be noncollapsible with respect to any average over the C-specific odds ratios. This conflict complicates their interpretation and has led to much confusion in the literature. For example, noncollapsibility over C with respect to averaging over odds ratios (or their logs) is sometimes called a "bias," but if C is sufficient for confounding control (see below), it does not correspond to a bias in estimating causal effects (Greenland et al., 1999b). We will return to this point later.

A measure is *simply collapsible* or strictly collapsible over C if the C-specific measures are constant and equal to the unconditional measure (Geng, 1992). Odds ratios are simply collapsible if X is independent of C given Y, as can be seen from the familiar XY "inversion" (symmetry) property of odds ratios: The C-specific odds ratios are

$$p(y_1|x_1,c)p(y_0|x_0,c)/p(y_1|x_0,c)p(y_0|x_1,c) = p(x_1|y_1,c)p(x_0|y_0,c)/p(x_1|y_0,c)p(x_0|y_1,c).$$

If C is independent of X given Y the latter term becomes

$$p(x_1|y_1)p(x_0|y_0)/p(x_1|y_0)p(x_0|y_1) = p(y_1|x_1)p(y_0|x_0)/p(y_1|x_0)p(y_0|x_1),$$

thus demonstrating simple collapsibility over C.

Now suppose instead that conditioning on C does not alter the dependence of Y on X, i.e., $p(y|x,c) = p(y|x)$ for all c and x (conditional independence of Y and C given X). Then conditioning on C cannot change *any* measure of dependence of Y on X, and any reasonable adjustment for C (whether standardization of probabilities or averaging of measures across levels of C) must produce a measure equal to the unconditional (unadjusted, marginal) measure. In other words, independence of C and Y given X implies simple collapsibility for *all* dependence measures. We will call this condition *complete collapsibility* over C: neither standardization nor conditioning nor averaging measures over C will change the dependence of Y on X. Complete collapsibility thus corresponds to a situation in which adjustments for C have no impact on bias.

All the above definitions and concepts can be applied if C represents a set of covariates, and all can be applied conditional on a set S of further covariates. For example, independence of Y and C given X and S implies *complete collapsibility given S* (after conditioning on S, further conditioning or adjustment for C does not change the dependence of Y on X given S). Similarly, an adjusted measure adjusted for C and S is *collapsible over C given S* if it equals its counterpart from adjusting for S only.

CONNECTIVITY AND ASSOCIATIONS IN DAGS

There are now many introductory reviews of causal analysis using directed acyclic graphs (DAGs) (e.g., Greenland et al., 1999a; Glymour and Greenland, 2008; Greenland and Pearl, 2010; Pearl, 2010a), as well as much more in-depth treatments (e.g., Pearl, 1995, 2009; Spirtes et al., 2001). Figure 1 gives three basic cases. We summarize the graphical concepts we will use to analyze them. Throughout, we will assume the graph represents relations in a specific population

under study. The results we will use apply even if the graph represents only conditional independencies rather than causal relations (as in Pearl, 1988 or Spiegelhalter et al., 1988), although their interest here derives from their causal interpretation.

Two *arrows* in a DAG are adjacent if each touches the same variable (whether by head or tail). A *path* between X and Y is a sequence of adjacent arrows going through the DAG from X to Y. A variable on a path between X and Y is an *interceptor* on the path. An interceptor C where two arrowheads meet (two arrows collide, as in the path from X to Y in Fig. 1c) is a *collider* on the path, and the path is said to be *blocked* at C. If instead C is where an arrowhead meets a tail (as in the path from X to Y in Fig. 1b) it is a *mediator* on the path. Finally, if C is where two arrowtails meet (as in the path from X to Y in Fig. 1a) it is a *fork* on the path. Note that all three of these conditions are only relative to a path; for example, in Fig. 1a, C is a fork on the path X←C→Y, a mediator on the path A→C→Y, and a collider on the path A→C←X; thus it is not meaningful to speak of a variable as a mediator or collider without reference to the path on which it is so.

A path is said to be *unconditionally closed* or blocked at every collider and *unconditionally open* at every mediator or fork. Thus, a path is *unconditionally open* if it contains no collider; conversely, if the path contains a collider it is *unconditionally closed* or blocked. Two variables in a DAG are said to be *d-connected* if there is an open path between them, and are *d-separated* (Pearl, 1988, 1995, 2009) if there is no such path. The "d-" in these definitions stands for "directionally" and distinguishes these conditions from other concepts of separation. Nonetheless, because the popular DAG literature uses only directional concepts, in what follows we will shorten "d-connected" to "connected" and "d-separated" to "separated" (as

in Greenland et al., 1999a). We may then say that two variables are connected by all the open paths between them.

A path is *open given a set of variables* S if (i) S contains no mediator or fork on the path and (ii) any collider on the path is either in S or has a descendant in S; otherwise it is *closed given* S or *blocked by* S. Two variables in a DAG are *connected given* S if there is a path between them that is open given S; otherwise they are *separated given* S. Two *variables* are adjacent if they have an arrow between them. The variable at the tail is called the *parent* of the variable at the head, which is called the *child* of the tail variable. The set of parents of a variable X in a given DAG is denoted pa(X). If X has no parent in the DAG, as in Fig. 1b, pa(X) is empty and X is said to be *exogenous* in the DAG; otherwise X is *endogenous*, as in Fig. 1a where pa(X) = {C}.

A path is *directed* if it contains only mediators (so that one moves from arrowhead to arrowtail at each variable in the path). If there is a directed path from one variable to another, the tail-end variable (the start) is called an *ancestor* of the variable at the ending arrowhead; and the variable at the final arrowhead (the end) is called a *descendant* of the starting variable. In a DAG, no variable is its own ancestor (i.e., there are no feedback loops). If the DAG is taken as a causal model, a variable is said to causally affect its descendants and causally affected by its ancestors.

A distribution p and a DAG over a set of variables are said to be *compatible* if p factorizes into $\prod$ p(x|pa(X)), where the product is over all the variables (this product is called the *Markov factorization* implied by the DAG). It can be shown (Pearl, 1988, 2009; Spirtes et al., 2001) that for any compatible p, two variables X and Y in a DAG will be independent given another set of variables S in the DAG if X and Y are separated by S. The converse is not true in

general, but exceptions in which connected variables are nonetheless independent in p correspond to very special cases involving perfect cancelations among associations (and hence are sometimes referred to as "unstable" or "unfaithful" properties of the distribution; Pearl, 2009; Spirtes et al., 2001).

The remainder of this paper is concerned primarily with describing properties of distributions compatible with a given graph. As simple examples, adjacent variables will always be connected and hence cannot be assumed independent, no matter what information we obtain about the remaining variables in the DAG. In other words, adjacent variables may remain dependent at any level of conditioning on the remaining variables in the DAG. Conversely, two nonadjacent variables X and Y in a DAG are separated and hence will be unconditionally independent if neither is a descendant of the other; if instead Y is a descendant of X, then X and Y will be separated by pa(Y) and hence independent given pa(Y).

Considering the DAG as a probabilistic influence network or *Bayes net* (Lauritzen and Spiegelhalter, 1988; Pearl, 1986, 1988), information can flow from one point to another along open paths. In particular, if two variables are connected, then information can flow between them. This means we should not assume that connected variables are independent; in particular, new information obtained about a variable C may (upon conditioning on that information) alter our probabilities regarding any variable connected to it. Furthermore, if a variable C is connected to both X and Y, we should not be surprised if obtaining and conditioning on information about C alters the connection between X and Y.

SEPARATION AND COLLAPSIBILITY

Because separation implies independence in compatible distributions, we obtain the following two criteria for detecting collapsibility in a distribution given a compatible graph (i.e.,

for recognizing graphically when conditioning and standardization will *not* change a dependency):

a) If C is separated from Y given X, then the dependence of Y on X will be completely collapsible over C (i.e., unaltered by adjustment for C).

b) If C is separated from X unconditionally, then population-standardized measures of dependence of Y on X will be collapsible over C.

Both these criteria also apply conditional on a set S of covariates, and with C replaced by a set of covariates.

When comparing two levels x_1 and x_0 of X, criterion (a) applies to standardized differences and ratios of probabilities, such as $p\{y|x_1;p(c)\} - p\{y|x_0;p(c)\}$ and $p\{y|x_1;p(c)\}/p\{y|x_0;p(c)\}$ derived from expression (2). Under criterion (b) (unconditional separation of C and X), both these measures equal the analogous unconditional (unadjusted) measures $p(y|x_1) - p(y|x_0)$ and $p(y|x_1)/p(y|x_0)$ obtained by dropping p(c) from the expressions. (These measures usually take Y to be a binary disease indicator with y denoting disease; our results apply to any Y and y.) Both criteria also apply when comparing two levels y_1 and y_0 of the outcome Y via standardized odds such as $p\{y_1|x;p(c)\}/p\{y_0|x;p(c)\}$, as well as to differences and ratios of these odds: Under either criterion, the resulting measures will be unchanged by standardization.

Each of criteria (a) and (b) is sufficient alone, but neither is necessary and so the converse of each is not quite correct. Nonetheless, if C is connected to Y conditional on X, then without more restrictions we will not have complete collapsibility for the dependence of Y on X; in particular, we would expect that at least one of the risk differences and one of the risk ratios conditional on C and their summaries will differ from the corresponding unconditional risk

differences and risk ratios. Furthermore, if C is also connected to X unconditionally, we would expect noncollapsibility for averages across C of the risk differences and risk ratios. We say "expect" because if C is not binary, there are special cases in which conditioning on C has no impact on certain summaries of these measures (due to cancellations that occur upon averaging); again, these correspond to unstable (unfaithful) properties of the distribution.

Whether the changes upon conditioning on C or adjustment for C represent increased or decreased bias depends upon further details, especially on the effect targeted for estimation (Glymour and Greenland, 2008; VanderWeele, 2009). Intuitively, we might expect conditioning on C to remove bias for estimating any effect of X on Y in Fig. 1a and direct effects in Fig. 1b, whereas we might expect it to create bias for estimating net effects in Fig. 1b and any effect in Fig. 1c. As discussed below, these intuitions are correct when targeting total-population effects.

When we consider odds $p(y_1|x,c)/p(y_0|x,c)$ and their comparisons conditional on C, instead of those computed from standardized probabilities, criterion (b) is no longer relevant. In its place we have

c) If C is separated from X conditional on Y, then the odds ratio will be simply collapsible over C.

As a partial converse, if C is connected to X conditional on Y and is connected to Y conditional on X, then we expect noncollapsibility over C-conditional odds ratios. Again, we say "expect" because of special exceptions when C is not binary (Whittemore, 1978) and caution that, even in Fig. 1a, odds-ratio noncollapsibility partly represents a mathematical peculiarity of odds ratios rather than pure confounding (Greenland et al., 1999b). Parallel remarks apply to differences and ratios of rates (hazards) (Greenland, 1996), with Y now understood to contain both time at risk and the outcome indicator.

SEPARATION AND COLLAPSIBILITY TESTING

A causal model can only be tested through its statistical implications, which are the conditional independencies implied by separation criteria. Consequently, the bulk of causal assumptions embedded in such a model will remain untested (Pearl, 2009; Greenland, 2010). Nonetheless, Pearl (2009, p.345-48) suggests how separation and collapsibility tests can be combined with substantive knowledge to screen candidate graphical models.

To test a separation criterion, higher statistical power can be attained by testing the independency implied by the criterion rather than by testing the implied collapsibility. On the other hand, collapsibility is often more relevant to causal inference (as may be seen from the examples below). Thus, Pearl and Paz (2010) suggest using collapsibility tests and related procedures as diagnostics for graphical models, analogous to collapsibility-based tests of regression models (e.g., Clogg et al., 1995). This is because collapsibility holds under either or both of two conditional independencies (as well as under other conditions); therefore, if a test rejects collapsibility, it rejects all graphical models having an independency that implies collapsibility.

When C and S are vectors of covariates, such tests can be performed using familiar modeling strategies. Suppose a graph predicts that a measure of the dependence of the outcome Y on the exposure X is collapsible over C given S. One approach starts with a model for p(y|x,c,s), such as a logistic regression model, and then tests whether the X coefficient is equal to that obtained when C is dropped to produce a model for p(y|x,s), which is collapsibility of the X coefficient over C (Clogg et al., 1995).

Asymptotic tests can however falter with very high-dimensional S, especially when S strongly predicts X and Y. An alternative for these cases replaces the pair of vectors (C,S) and

the vector S with fitted values from models for the exposure-propensity scores p(x|c,s) and p(x|s), respectively; that is, (C,S) and S are replaced by their fitted X-propensity scores. We may then test equality of the adjusted measures derived from the two scores, which is equivalent to testing collapsibility of the measures over C (Pearl, 2009, p. 349). The two approaches can be combined by using the propensity scores to fit the Y (outcome) model, as in doubly robust estimation (Kang and Shafer, 2007). Again, rejection of equality (collapsibility) of a measure after deleting C from both the Y and X models implies rejection of all graphical models that entail collapsibility of the X coefficient over C.

RELATIONS TO CAUSAL EFFECTS

The collapsibility results we have described do not assume the quantities at issue are related to causal effects. To make that connection, define p(y|do[x]) as the distribution Y would have upon setting X to the value x for everyone in the population, when that is possible. This do[x] formalism is closely related to the potential-outcome (counterfactual) model of causation, in which each individual is presumed to have a well-defined potential-outcome variable Y_x when administered level x of X, whether or not x is the level actually administered; in that case p(y|do[x]) = p(y_x) (see Pearl, 2009, Ch. 7 for further details of the relation). In either formalism, care is needed in choice of X in order for the setting of X to a level x represented by do[X=x] or Y_x to make sense (Greenland, 2005; Hernán, 2005). This would be so if X were a treatment such as a vaccination indicator, but not if X were a defining property of an individual such as a gender indicator (but see Pearl 2009, p. 361 for a more liberal view of do[X=x]).

When do[X=x] is well defined, we say a set of covariates S is *sufficient* or *admissible* for estimating total-population effects of X on Y if p{y|x;p(s)} = p(y|do[x]); that is, S is sufficient precisely in the case that standardization by p(s) yields the effect of setting X=x. S is *minimal*

sufficient if it is sufficient but no subset of S is. If S is sufficient for the total-population effects and we assume no contagion, standardization by $p(s|x_r)$ is sufficient for estimating effects in the subpopulation with $X=x_r$; that is, $p\{y|x;p(s|x_r)\}$ equals the effect of having set this subpopulation to $X=x$ instead of its actual setting of $X=x_r$ (Shpitser and Pearl, 2009). The converse is not correct: A set may be sufficient for some choices for x_r (some subpopulations) but insufficient for other choices of x_r; more generally, a set may be sufficient for some subpopulation effects but insufficient for others or for total-population effects (Joffe et al., 2010). Contagion further complicates analysis of subgroups because then the distribution in one subgroup may depend on the distribution and hence alteration of other subgroups (Halloran and Struchiner, 1995).

Turning to graphical criteria for recognizing sufficiency, a path from X to Y is said to be *back-door* (relative to X) if it starts with an arrow into X. A set S then satisfies the *back-door criterion* for estimating the effect of X on Y if it (i) contains no descendant of X and (ii) blocks every back-door path from X to Y. Such a set is sufficient for effect estimation (Pearl, 1995, 2009; Greenland et al., 1999ab). It is often said that measures of relations of X to Y conditioned on a sufficient set S are "unconfounded," because the exposure X will be connected to outcome Y given S only via directed paths from X to Y, which in a causal graph represent the effects of X on Y.

EFFECTS OF CONDITIONING ON AN INTERCEPTOR

In each graph in Fig. 1, C is connected to X unconditionally and conditional on Y, and to Y conditional on X. Thus we expect noncollapsibility (change) over C for all measures; that is, we expect conditioning on C will change one or more of the risk differences, risk ratios, and odds ratios relating X to Y. Nonetheless, the meaning of this change is quite different across the graphs.

In Fig. 1a, the path between X and Y via C (X←C→Y) has a fork at C, and so is an open path; hence X and Y may be associated via this path. The association transmitted along this path is a source of bias for estimating the effect of X on Y, sometimes called "classical confounding" because the path contains a shared cause of X and Y (Greenland, 2003). Note that the key source of this confounding is that C is *uncontrolled*, not that it is unmeasured. For example, C may have been measured but left uncontrolled because it failed to have a "statistically significant" association with X or with Y. Conversely, C may have been controlled without being measured by virtue of design features (e.g., for practical purposes, a population-based study in Finland will have controlled for conventional American "race" categories of white, black, etc.).

Conditioning on C will block (close) the confounding path in Fig. 1a; hence if X has no effect on Y, then X and Y will be independent given C (independent within every stratum or level of C), reflecting correctly this absence of effect. Put more generally, C alone satisfies the back-door criterion and thus is sufficient for estimating effects; furthermore, it is minimal sufficient. Hence, to estimate an effect of X on Y, we should condition on C. If we modified Fig. 1a by inserting a mediator M or fork F between C and X or between C and Y, C would remain sufficient (as would M alone, F alone, or any combination of C, M or F) and the value of expressions (1) and (2) would not change.

In Fig. 1b, the path between X and Y via C (X→C→Y) is direct through C, and so is an open path; hence X and Y may be associated via this path. In both Fig. 1a and 1b, the open path will be blocked by conditioning fully on C; hence if X had no direct effect on Y, X and Y would be separated given C, reflecting correctly this absence of effect. Thus, to estimate a net effect of X on Y, we should not condition on C because C is a mediator (intermediate) between X and Y,

carrying part of the net effect; but if we want to estimate a C-specific direct effect of X on Y, we would condition fully on C.

Behavior opposite of the confounding case in Fig. 1a arises in Fig. 1c, where the path the path between X and Y via C (X→C←Y) is blocked at C, and hence X and Y cannot be associated via this path. This closed path will however be unblocked (opened) by conditioning on C; this means that X and Y may be dependent given C (dependent within at least one level of C) even if there is no effect of X on Y. This would continue to be so if we inserted a mediator M or fork F between X and C or between Y and C. Thus, to estimate an effect of X on Y, we should *not* condition on C, the opposite situation from Fig. 1a. Fig. 1c arises when C is an indicator of selection in case-control studies or a censoring indicator in cohort studies and trials with losses.

IMPACTS OF CONDITIONING ON A DESCENDANT OF AN INTERCEPTOR

Considering first a child Z of C, suppose that the connection C→Z is not perfect, so that at most conditioning on Z corresponds to only partial adjustment for C. The situations in Fig. 1 would arise if (for example) C was unmeasured and Z was an imperfect but nondifferential measurement of C or proxy for C (i.e., Z is independent of X and Y given C). Again, if we do not condition, in Fig. 1a and 1b, X and Y remain connected through C, which is a source of bias for estimating any effect of X on Y in Fig. 1a or a direct effect in Fig. 1b; in Fig. 1c, X and Y remain separated and there is no bias for estimating any effect of X on Y.

What if we condition on Z only? For each case in Figure 1 we see that Z is connected to X both unconditionally and conditional on Y, and is connected to Y given X. Thus we expect noncollapsibility over Z for all measures. In Fig. 1a and 1b, one way to interpret the changes in risk differences and risk ratios is that conditioning on Z partially closed the open path connecting X and Y through C. In Fig. 1c, however, these changes correspond to a partial opening of the

unconditionally closed path X→C←Y. Another interpretation, based on virtual colliders, is given in Pearl (2009, p.338).

Conditioning on Z can be viewed as adjustment for C using a nondifferentially misclassified proxy. In Fig. 1a with binary C and Z, this has long been known to induce partial control of confounding (Greenland, 1980), and so conditioning on Z moves us part way from the confounded unconditional (unadjusted) association of X and Y toward the total effect of X on Y. (For nonbinary covariates this reasoning is only correct in an average sense, as some but not all strata of Z may end up more confounded than the original unadjusted association; see Brenner, 1993.) Analogously, in Figure 1b we are partially adjusting for the effect of X on Y mediated through C; for risk differences and risk ratios, we expect this to move us partway from the net effect of X on Y toward a direct effect of X and Y. Again, Fig. 1c brings an opposite phenomenon from Fig. 1a: Conditioning on Z produces an open noncausal path from X to Y, which may introduce bias. In each figure, however, Z is separated from Y by C and X, implying that we will have complete collapsibility over Z given C.

The results just described extend to any descendant Z of C that does include X or Y among its ancestors or descendants.

IMPACTS OF CONDITIONING ON AN ANCESTOR OF AN INTERCEPTOR

Turning now to a parent A of C, suppose that the connection A→C is not perfect, so that at best conditioning on A corresponds to only partial control of C. We may now say "control" rather than adjustment because A actually does control C in a causal sense. Hence, under the diagrams we present, the consequences of conditioning on A parallel the consequences that would follow if A were an intervention to set the level of C. Because of this parallel, we will see some telling divergences from what happens when conditioning on the child Z of C: This is

because, unlike with A or C, intervention to change the level of Z could have no effect on any other variable.

What if we condition on A only? In Fig. 1a we see that A (like Z) is connected to X both unconditionally and conditional on Y, and is connected to Y conditional on X, so we expect noncollapsibility for all measures. For risk differences and risk ratios we can interpret these changes as reflecting partial closure of the open path connecting X and Y through C. Thus we expect partial control of confounding if we condition on A in place of C, moving us from the confounded unconditional (unadjusted) association of X and Y toward the total effect of X on Y (again, for nonbinary covariates this reasoning is only correct in an average sense). As before, the interpretation for average odds ratios is more complex.

In Figure 1b however A is separated from X unconditionally. Thus, from (b), population-standardized measures will be collapsible over A. Nonetheless, A is connected to Y conditional on X, so we expect some some A-specific measures to differ from their corresponding unconditional measure, which we might interpret as partial control of the effect of X on Y mediated through C. Furthermore, A is connected to X conditional on Y, so we expect average odds ratios to differ from unconditional odds ratios. It might be tempting to think that these odds-ratio changes represent partial control of the effect of X on Y mediated through C, but again the reality is more complex.

In Fig. 1c, A is not connected to Y conditional on X, and so, unlike with C or Z, we have complete collapsibility over A. In graphical terms, conditioning on A does not even partially open the path from X to Y through C, and thus induces no bias; this is so even if A determines C completely (C=A), for in that case X and Y will no longer affect C. Nonetheless, A is connected to X given C, to X given Y and C, and to Y given X and C. Thus, unlike with Z, we expect

noncollapsibility over A given C for all measures. If C represents selection, this means that A will appear to be a confounder among the selected, even though it is not. In contrast, in Fig. 1a and 1b, A is separated from Y by C given X, so there will be complete collapsibility over A given C.

The results just described extend to any ancestor A of C that does include X or Y among its ancestors or descendants.

SOME SPECIAL CASES

Figure 2 displays some special cases of Figure 1 that are often discussed. Fig. 2a and 2b drop the arrows between C and Y in Fig. 1a and 1b. Furthermore, all are separated from Y given X and so we have complete collapsibility over them (whether considering them singly, in pairs, or all together).

In Fig. 2c, A, C, and Z are all unconditionally separated from X; hence population-standardized measures of dependence of Y on X are collapsible over C, A, or Z, since $p\{y|x;p(c)\} = p\{y|x;p(a)\} = p\{y|x;p(z)\} = p(y|x)$. Moreover, those measures will be unconfounded, since $p(y|x) = p(y|do[x])$. Nonetheless, C, A, and Z are all connected to X given Y and to Y given X; hence we should expect noncollapsibility over conditional odds ratios.

In Fig. 2d, C and Z are connected to X unconditionally and to Y given X; hence we expect population-standardized measures of dependence of Y on X to be noncollapsible over C and Z, and this remains so if we condition on A as well. On the other hand, A, C and Z are separated from X given Y, implying that conditional odds ratios will be collapsible over all of them. Furthermore, if we do not condition on C or Z, A will be separated from Y conditional on X and so we have complete collapsibility over A.

Fig. 2d can be taken as representing a case-control study with exposure X, disease Y, C indicating selection, and Z indicating consent. In such a study, Y by definition affects selection C very strongly, resulting in severe noncollapsibility over C of all measures except odd ratios. Because the unconditional measures are not confounded, this noncollapsibility over C represents a strong bias from conditioning on C. This bias afflicts all familiar measures that depend on absolute frequencies of Y values in some fashion, such as risk differences, odds differences, and risk ratios (See Pearl 2009, p. 338 for graphical explanation). Risk ratios, for example, cannot exceed 2 if the absolute frequency of Y=1 is never below ½. In contrast, odds ratios relating X to Y depend only on relative frequencies of Y values and hence are collapsible; this collapsibility can be viewed as a graphical generalization of the famous result by Cornfield (1951), and justifies use of the odds ratios from participants (C=Z=1) to estimate the unconditional odds ratios.

Nonetheless, an effect of X (or an ancestor of X) on selection or consent will connect X to Y via C or both, and thus introduce bias in the odds ratio; this bias is the familiar Berksonian form of selection bias (Greenland et al., 1999a; Glymour and Greenland, 2008; Pearl, 2009). Similar concerns arise in cohort studies in which C represents loss to follow-up or other forms of censoring, and in trials in which C is a compliance indicator and the analysis discards noncompliers ("per-protocol" analysis).

EXTENSIONS

There are many ways to extend the previous graphical results. We present some examples to illustrate how the rules we have described may guide us in selecting adjustment variables that are not confounders in the classical sense seen in Fig. 1a. In each example in Fig. 3a-3d, C

exhibits some form of noncollapsibility, but in the first example this noncollapsibility amplifies a bias, in the second and third it reduces a bias, and in the fourth it does both.

Bias Amplification. Fig. 3a adds an uncontrolled confounder U to Fig. 2a. A, C, and Z satisfy graphical conditions for instrumentality, i.e., unconditional connection to X and connection to Y only through X as a mediator (Pearl, 2009); they are also connected to X given Y. Unlike Fig. 2a, A, C, and Z are connected to Y conditional on X via the path C→X←U→Y because X is a collider on that path. Thus we expect noncollapsibility over A, C, and Z for all measures. If we condition on U, however, we are back in a completely collapsible situation like that in Fig. 2a.

Considering cases in which effects can be given a sign (positive or negative), Bhattacharya and Vogt (2007) and Pearl (2010b) show how the unconditional noncollapsibility over A, C, Z in Fig. 3a corresponds to increased bias from the confounding back-door path X←U→Y. To see this, suppose effects represented by the arrows in X←U→Y have the same sign. Then the XY association transmitted along the X←U→Y path will be positive, and hence the bias from failing to condition on U (the confounding by U) will be upward (VanderWeele and Robins, 2010); the change in the XY association from adjusting for A, C, or Z will also be upward, resulting in more bias after the adjustment than before (i.e., the biases will not cancel). Parallel reasoning shows that if the arrows in X←U→Y have opposite signs, the XY association transmitted through this path will be negative, so the bias from failing to condition on U will be downward (VanderWeele and Robins, 2010); and the change from adjusting for A, C, or Z will also be downward, resulting in more bias.

In either case, adjusting for A, C, or Z can only result in more bias in the same direction as the confounding by U (hence is bias amplifying). Intuitively, when we consider the

unconditional (crude) association between X and Y, systematic variation in X is partly explained by variation in C and partly by variation in U. The U component is transmitted to Y via the confounding path X←U→Y and so counts toward bias. If we condition on C, however, the C component vanishes; hence all systematic variation in X comes from variation in U and is transmitted to Y via the confounding path, with larger bias as a result.

The bias-amplification problem does not arise in traditional instrumental-variable adjustment methods (e.g., Sommer and Zeger, 1991; Hernán and Robins, 2006) because the instruments are used to correct the unconditional association, instead of being conditioned on as in outcome-regression and propensity-score adjustment. Nonetheless, some authors recommend selecting all variables that influence exposure X for the latter adjustments, without regard to their relation to the outcome Y (e.g., Hirano and Imbens, 2001; Rubin, 2002, 2009). Unfortunately adjusting for variables related only to exposure may not only amplify bias, but may also unnecessarily inflate variances (e.g., see Brookhart et al., 2006; Austin et al. 2007).

Conditioning on apparent instrumental variables can also amplify certain types of selection bias (Pearl, 2010b). Consider Fig. 3b, which modifies Fig. 3a by replacing U→Y with U→S←Y. Before conditioning on S, the association of X and Y will be collapsible over A, C, or Z because X separates those variables from Y. Nonetheless, conditioning on S opens a path from X to Y via U and S, introducing selection bias. Furthermore, X no longer separates A, C, or Z from Y, so we should expect the association of X and Y to be noncollapsible over A, C, or Z. This noncollapsibility again reflects bias amplification.

On the other hand, if the selection bias is transmitted only through an effect of X, conditioning on A, C, or Z will not amplify that bias. This is because A, C, and Z will remain separated from Y by X after selection, and thus remain independent of Y given X. As an

example, consider Fig. 3c, which modifies Fig. 3a by replacing U→X with U→S←X. Again, conditioning on S produces bias because it opens a path from X to Y via U and S, but the association remains unchanged and hence is not further biased by conditioning on A, C, or Z. This example illustrates bias equivalence (bias from conditioning on S is equivalent to bias from conditioning on S and any combination from A,C, or Z), which is discussed further below.

Bias Removal. Fig. 3b modifies Fig. 1c by replacing X→C←Y with X→C→S←Y, as could arise when S is an indicator of analysis inclusion and C is an adherence indicator. The unconditional XY association unbiased for the effect of X on Y. S is connected to X both unconditionally and given Y, and is connected to Y given X, so we expect conditioning on S alone to change and thus introduce bias in estimating that effect.

As in Fig. 1c, in Fig. 3d both C and Z are connected to X unconditionally and given Y, while A is not connected to X. Nonetheless, A, C and Z are independent of Y given X, so we have complete collapsibility over them all. But A, C, and Z are connected to Y given X and S, and are connected to X given S and given Y and S. Thus we expect noncollapsibility over A, C, and Z given S.

In Fig. 3d, S is separated from X given C and Y, so we have collapsibility of the XY odds ratio over S given C as well as over C; hence this odds ratio is collapsible over the compound variable {C,S} even though we cannot assume that it is collapsible over S or over C given S. Put another way, conditioning on C removes the selection bias in the odds ratio produced by conditioning on S, making C a "bias-breaking" variable for the odds ratio (Geneletti et al., 2009). The same situation holds if the X→C relation is reversed to C→X so that C is a fork rather than a mediator between X and S, or if C is mediator or fork between Y and S (Geneletti et al., 2009).

Note however that in these cases, removal of bias by conditioning on C is limited to the odds ratio; that is, we expect only odds-ratio collapsibility over {C,S}. Because {C,S} is unconditionally connected to X as well as connected to Y given X, population-standardized measures are not collapsible over {C,S}. This noncollapsibility corresponds to the well-known fact that conditioning on a variable affected by the outcome variable Y (as in case-control sampling) will alter the observed proportions with a specific outcome (such as disease) and so alter risk differences, risk ratios, and odds differences.

In Fig. 3d, we expect noncollapsibility of the XY odds ratio over A and over Z given S, but (in contrast to {C,S}) we also expect noncollapsibility over {A,S}, {Z,S}, and {A,Z,S}. This means that, after conditioning on S, we might ordinarily expect bias reduction from the change induced by conditioning on A, Z, or both, but we would not expect complete bias removal.

Bias Equivalence. Fig. 3e adds an uncontrolled unconditional confounder U of XY to Fig. 3d. Now the unconditional XY odds ratio is biased, being a mix of the study effect X→Y and the association over the confounding back-door path X←U→Y. Because this confounding path has no overlap with the selection-bias path X→C→S←Y, the previous observations about the latter path continue to apply: We have noncollapsibility over S but collapsibility over both C and {C,S} relative to the U-confounded (unconditional) XY odds ratio. This collapsibility is a basic example of bias equivalence (Pearl and Paz, 2010): We are left with the same degree of confounding (from U) whether we condition on nothing, on C, or on both C and S.

Conditional on U we also have noncollapsibility over S but collapsibility over both C and {C,S} relative to the unconfounded (U-conditional) XY odds ratio, so we also have bias equivalence given U (which in this case is no bias whether in addition to U we condition on nothing, on C, or on both C and S). If instead we condition on A, Z or both after conditioning on

S, we no longer have such equivalencies, since we have unconditional collapsibility over A, Z, or both, but we expect noncollapsibility over {A,S}, {Z,S}, and {A,Z,S}. Thus as in Fig. 3d, we would not expect conditioning on A or Z to be sufficient for removal of the bias from conditioning on S, even for odds ratios.

Overlapping Bias Paths. Fig. 3f adds an uncontrolled variable U to Fig. 3d, one which does not unconditionally confound the XY relation but does confound other relations. Hence there is no bias unconditionally. Nonetheless, conditioning on C, S, or Z now opens a new path from X to Y, X→C←U→Y. As a consequence, we no longer have collapsibility over C or Z, and conditioning on S opens two paths from X to Y (the new path, as well as X→C→S←Y).

As in Fig. 3e, we still have odds-ratio collapsibility over S given C, so C and {C,S} remain bias equivalent for odds ratios, as do {C,U} and {U,C,S}; and, once we condition on S (as we are forced to do when S is selection), we would have to condition on U as well as C to remove all odds-ratio bias. Unlike Fig. 3e, however, in Fig. 3f U could be ignored for odds-ratio estimation if there were no conditioning on C, S or Z. Furthermore, we would ordinarily expect the bias from the X→C←U→Y path to be larger if C were conditioned than if only S were conditioned. In this sense, after conditioning on S, we would expect further conditioning on C to amplify the bias from the X→C←U→Y path even though it would remove the bias from the X→C→S←Y path; the net impact of conditioning on C given S is thus hard to predict.

If instead of C we condition on A, Z or both after conditioning on S, we no longer have bias equivalencies. We have unconditional collapsibility over A, but after conditioning on S we expect noncollapsibility over any combination from A, U, or Z. Thus, as in Figs. 3d and 3e, we would not expect conditioning on A or Z to be sufficient for complete removal of the bias from conditioning on S, even after conditioning on U.

DRAWBACKS AND ADVANTAGES OF ODDS AND HAZARD MEASURES

Even if C and X are marginally independent, not all averages of C-specific measures will be collapsible. Suppose however, the C-specific measures are constant across C. If the unconditional measure equals this constant value, it is said to be *strictly collapsible* or *simply collapsible* over C (Whittemore, 1978; Ducharme and LePage, 1986; Geng, 1992). Since all averages of the C-specific measures must equal this constant, simple collapsibility implies collapsibility of these averages. Because standardized risk differences and risk ratios are averages of C-specific values, simple collapsibility of risk differences and risk ratios implies collapsibility of standardized risk differences and risk ratios. Simple collapsibility of odds ratios and differences does not however imply collapsibility of the standardized odds ratios and differences. Instead, rather paradoxically, if C and Y are dependent given X but C is marginally independent of X, all population-standardized odds ratios will be collapsible but simple collapsibility cannot hold (Miettinen and Cook, 1981; Greenland et al., 1999b). Again, parallel results hold for odds differences, as well as for hazard ratios when Y comprises time at risk and the outcome indicator.

On the other hand, as illustrated above with Fig. 3e and 3f, odds ratios have the potential to remain unbiased when conditioning on variables affected by the outcome Y, provided that conditioning does not open a path from X to Y. The application of these results extends from odds-ratio to rate-ratio analysis when sampling or conditioning is done in a manner that forces sample odds ratios to estimate hazard (rate) ratios, as is typical in case-control studies with risk-set (density) sampling and in survival analysis (Rothman et al., 2008, p. 113-114 and 294-295).

DISCUSSION

We have reviewed algebraic results and introduced graphical criteria to answer questions about when adjustment for particular variables will increase or reduce bias from a particular source in a given graphical model. There is considerably more that could be researched and discussed regarding implications of adjustment for statistical efficiency and mean-squared error (or more generally, net loss), and quantification of the bias added or removed by a given adjustment. Basic results on these topics are available, especially for ratio measures (e.g., Yanagawa, 1984; Flanders and Khoury, 1990; Greenland, 1991, 2003; De Stavola and Cox, 2008; Janes et al., 2010), but many details and extensions remain to be worked out (which is unsurprising given the many parameters that must be modeled to quantify efficiency and bias). We also caution that the use of preliminary tests for model and covariate selection (whether independence or collapsibility testing) can distort the final *P*-values and confidence intervals for the effect of interest (Leamer, 1978); see Greenland (2008) for a review and suggested alternatives to preliminary testing.

One semi-quantitative guideline that has been noted before is that associations and hence noncollapsibilities tend to attenuate when they arise from more extended paths (Greenland, 2003). This attenuation arises when adjacencies are of similar magnitude or when comparing paths to their subpaths. For example, in all panels of Fig. 1 and in ordinary settings in health and social sciences, we would expect the strength of associations of Z with X and Y to be less than the strength of associations of C with X and Y. That is because each path connecting Z to X or Z to Y properly contains the corresponding path connecting C to X or C to Y. As a result, we expect a smaller degree of XY noncollapsibility over Z than over C, which means that adjusting for Z will move us less from the unconditional association than will adjusting for C. In Fig. 1a, this means Z-adjustment does not remove as much bias as C-adjustment; in Fig. 1c, it means that

Z adjustment will not produce as much bias as C-adjustment; and in Fig. 1b the bias implication depends on whether we are interesting in a direct or total effect.

Another avenue for extending qualitative results is in terms of direction of bias, which as mentioned above can be derived by adding signs to path arrows (VanderWeele et al., 2008; VanderWeele and Robins, 2010). Quantitative considerations will have to enter when one considers multiple bias sources, as occur in Figs. 3e and 3f after conditioning on S. We expect that the net bias in most such situations will be not be simple in form and will be heavily dependent on contextual details; thus general results that can simplify context-specific analyses would be valuable. We hope that the results provided here provide a reasonable starting or reference point for further extensions.

Acknowledgments: We wish to thank Charles Poole for prompting this investigation with penetrating questions, and Tyler VanderWeele and Thomas Richardson for helpful comments and correspondence on the topic.

REFERENCES

Austin, P.C., Grootendorst, P., and Anderson, G.M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine*, 26, 734-753.

Bhattacharya, J. and Vogt, W. Do instrumental variables belong in propensity scores? Technical Working Paper 343, National Bureau of Economic Research, Cambridge, MA, 2007, available at http://www.nber.org/papers/t0343.

Brenner, H. (1993). Bias due to non-differential misclassification of polytomous confounders. *Journal of Clinical Epidemiology*, 46, 57–63.

Brookhart, M., Schneeweiss, S., Rothman, K.J., Glynn, R., Avorn, J. and Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology* 163, 1149–1156.

Clogg, C.C., Petkova, E., and Haritou A. (1995). Statistical methods for comparing regression coefficients between models (with discussion). *American Journal of Sociology*, 100,1261–312.

Cornfield, J. (1951). A method of estimating comparative rates from clinical data: application to cancer of the lung, breast and cervix. *J Natl Cancer Inst*, 11, 1269–1275.

Ducharme, G.R. and Lepage, Y. (1986). Testing collapsibility in contingency tables. *Journal of the Royal Statistical Society*, Series B, 48, 197-205.

De Stavola, B.L. and Cox, D.R. (2008). On the consequences of overstratification. *Biometrika*, 95, 992–996.

Flanders. W.D. and Khoury, M.J. (1990). Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology*, 1, 199–246.

Frydenberg, M. (1990). Marginalization and collapsibility in graphical statistical models. *Annals of Statistics*,18, 790-805.

Gail, M. H. (1986). Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In: *Modern Statistical Methods in Chronic Disease Epidemiology* (S. H. Moolgavkar and R. L. Prentice, eds.), 3-18. Wiley, New York.

Geneletti, S., Ricequalityson, S. and Best, N. (2009). Adjusting for selection bias in retrospective case-control studies. *Biostatistics*, 10, 17-31.

Geng, Z. (1992). Collapsibility of relative risk in contingency tables with a response variable. *Journal of the Royal Statistical Society*, Series B, 54, 585-593

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 S. Greenland (2008). Causal diagrams. Ch. 12 in: Rothman, K.J., S. Greenland and T.L. Lash, eds. *Modern Epidemiology*, 3rd ed. Philadelphia: Lippincott.

Greenland, S. (1980). The effect of misclassification in the presence of covariates. *American Journal of Epidemiology*, 112, 564-569.

Greenland, S. (1991). Reducing mean squared error in the analysis of stratified epidemiologic studies. *Biometrics*, 47, 773-775.

Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*, 7, 498-501.

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

Greenland, S. (2005a). Epidemiologic measures and policy formulation: Lessons from potential outcomes (with discussion). *Emerging Themes in Epidemiology* (online journal) 2:1–4. (Originally published as "Causality theory for policy uses of epidemiologic measures," Chapter 6.2 in: Murray, C.J.L., J.A. Salomon, C.D. Mathers and A.D. Lopez, eds. (2002) *Summary Measures of Population Health*. Cambridge, MA: Harvard University Press/WHO, 291-302.)

Greenland, S. and Mickey, R.M. (1988). Closed-form and dually consistent methods for inference on collapsibility in 2xJxK tables. *Applied Statistics*, 37, 335-343.

Greenland, S. and Pearl, J. (2010). Causal diagrams. In: Lovric, M. (ed.). *International Encyclopedia of Statistical Sciences*. New York: Springer.

Greenland, S., Pearl, J. and Robins, J. M. (1999a). Causal diagrams for epidemiologic research. Epidemiology, 10, 37-48.

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

Greenland, S. (2010). Overthrowing the tyranny of null hypotheses hidden in causal diagrams. Ch. 22 in: Dechter, R., Geffner, H., and Halpern, J.Y. (eds.). *Heuristics, Probabilities, and Causality: A Tribute to Judea Pearl*. London: College Press, 365-382.

Halloran, M.A. and Struchiner, C.J. (1995). Causal inference for infectious diseases. *Epidemiology*, 6, 142-151.

Hernán, M.A. (2005). Hypothetical interventions to define causal effects—afterthought or prerequisite? *American Journal of Epidemiology* 162, 618–620.

Hernán, M.A. and Robins, J.M. (2006). Instruments for causal inference. *Epidemiology* 17, 360-372.

Hirano, K. and Imbens, G.W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services & Outcomes Research Methodology* 2, 259–278.

Janes, H., Dominici, F. and Zeger, S. (2010). On quantifying the magnitude of confounding. *Biostatistics*, 11, 572–582

Joffe, M.M., Yang, W.P., and Feldman, H. I. (2010). Selective ignorability assumptions in causal inference. *International Journal of Biostatistics*, 6 (2), article 11, available at http://www.bepress.com/ijb/vol6/iss2/11

Kang, J.D., Shafer, J.L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). *Statistical Science*, 22, 523-580.

Lauritzen, S. L. and D. J. Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of Royal Statistical Society*, Series B, 50, 157–224.

Leamer, E.E. (1978). *Specification Searches*. New York: Wiley.

Miettinen, O.S. and Cook E.F. (1981). Confounding: essence and detection. *American Journal of Epidemiology* 114, 593–603.

Pearl, J. (1986). Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 9, 241–288.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika*, 82, 669–710.

Pearl, J. (2009). *Causality*, 2nd ed. New York: Cambridge University Press.

Pearl, J. (2010a). An introduction to causal inference. *The International Journal of Biostatistics* (online journal), 6(2), Article 7, available at http://www.bepress.com/ijb/vol6/iss2/7.

Pearl, J. (2010b). On a class of bias-amplifying variables that endanger effect estimates. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. AUAI, Corvallis, OR, 2010, 417-424, available at http://ftp.cs.ucla.edu/pub/stat_ser/r356.pdf.

Pearl, J. and Paz, A. (2010). Confounding equivalence in observational studies. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. AUAI, Corvallis, OR, 2010, 433-441, available at http://ftp.cs.ucla.edu/pub/stat ser/r343.pdf.

Robins JM, Hernán MA and Brumback B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*,11, 550–560.

Rothman, K.J., S. Greenland and T.L. Lash, eds. *Modern Epidemiology*, 3rd ed. Philadelphia: Lippincott.

Rubin, D.B. (2002). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.

Rubin, D.B. (2009). Should observational studies be designed to allow lack of balance in covariate distributions across treatment group? *Statistics in Medicine*, 28, 1420-1423.

Samuels, M.L. (1981). Matching and design efficiency in epidemiological studies. Biometrika, 68, 577-588.

Sato, T. and Matsuyama, Y. (2003). Marginal structural models as a tool for standardization. *Epidemiology*, 14, 680–686.

Schisterman, E., Cole, S. and Platt, R. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*, 20, 488–495.

Sommer, A.S. and Zeger, S. (1991). On estimating efficacy from clinical trials. *Stat Med*, 10, 45–52.

Shpitser, I. and Pearl, J. (2009). Effects of treatment on the treated: Identification and generalization. In J. Bilmes and A. Ng (eds.). *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, Montreal, Quebec, 2009.

Spirtes, P., Glymour, C. and Scheines, R. (2001). *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA: MIT Press.

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

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.

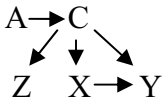
VanderWeele, T.J., Hernán, M.A., Robins, J.M. (2008). Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology*, 19, 720-728.

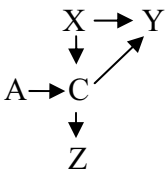
Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *Journal of the Royal Statistical Society*, Series B, 49, 353-364.

Yanagawa, T. (1984). Case-control studies: assessing the effect of a confounding factor. *Biometrika*, 71, 191-194.

Yule, G.U. (1934). On some points related to vital statistics, more especially statistics of occupational mortality. *Journal of the Royal Statistical Society*, 97, 1–84.

Figure 1. Graphs with C connected to X and Y under all conditions.

a) 

b) 

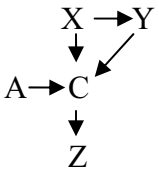
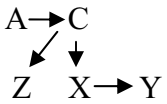
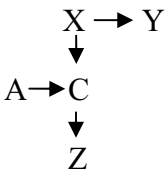
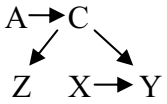
c) 

Figure 2. Graphs with C separated from X or Y under some condition.

a) 

b) 

c) 

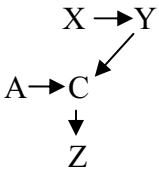
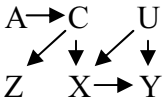
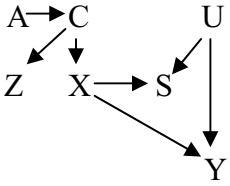
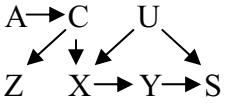
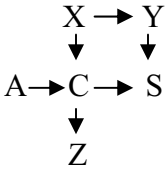
d) 

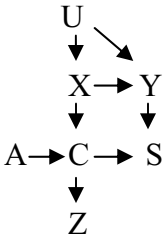
Figure 3. Graphs with an additional ancestor U or descendant S of X or Y.

a)


b)


c)


d)


e)


f)