Revised 3/04

Adjusting for Time-Varying Confounding in Survival Analysis[*]

Jennifer S. Barber

Susan A. Murphy

Natalya Verbitsky

University of Michigan

[*] This paper benefited substantially from conversations with William Axinn. This research was supported by National Institute of Child Health and Human Development grant HD32912, by National Science Foundation grants SBR 9811983 and DMS 9802885, and by grant P50 DA 10075 from National Institute on Drug Abuse to the Pennsylvania State University's Methodology Center. Address correspondence to the first author at Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, Michigan, 48106-1248; e-mail: jebarber@umich.edu.

Revised 3/04

Abstract

Adjusting for Time-Varying Confounding in Survival Analysis

In this paper we illustrate how directly including endogenous time-varying confounders in the model of the effect of an exposure on a response can lead to bias in discrete time survival analysis. An alternative to this method is Hernán, Brumback and Robins' (1999) use of sample weights to adjust for endogenous time-varying confounding. We discuss when this method can be used to provide unbiased estimators and we illustrate the method by addressing a substantive research question using existing survey data. We also critically examine the robustness of the weighting method to violations of the underlying assumptions via a simulation analysis.

Adjusting for Time-Varying Confounding in Survival Analysis

1. INTRODUCTION

In the social sciences it is frequently unethical or infeasible to conduct an experimental study. Thus, social scientists often must use observational data to address causal questions. A fundamental problem in assessing the effect of an exposure (key independent variable) on a response (dependent variable) in nonexperimental data is the presence of confounders. One statistical method used to control for confounding is to include measures of everything known to affect both the exposure and response in the model for the response, sometimes called "analysis of covariance" (Winship and Morgan 1999). Unfortunately, this approach leads to bias if the confounders are themselves affected by the exposure (i.e. they are *endogenous*) (Lieberson 1985; Robins 1987; Robins and Greenland 1994). This problem is particularly acute with time-varying confounders, because they are often affected by prior exposure. This paper describes, illustrates, and evaluates a recently-proposed class of models for survival analysis – called marginal structural models (MSMs) – which use sample weights (inverse-probability-of-exposure weights) to control for compositional differences due to endogenous time-varying confounders (Robins, Hernán and Brumback 2000).

2. A DRAWBACK OF COVARIANCE ANALYSIS

Our goal is to assess the effect of a time-varying exposure on the time to an event in a discrete-time survival analysis. If we use a covariance analysis, we would fit the model:

$$\text{logit}\,(p_{it}) = \beta_0 + \beta_1 X_{it} + \beta_2 U_i + \beta_3 V_{it} \qquad (1)$$

for individual i in year t, where $p_{it} = P[Y_{it} = 1 | X_{it}, U_i, V_{it}]$[1], β_0 is the intercept, β_1 is the coefficient for the exposure (X_{it}), β_2 is the coefficient for the exogenous confounders (U_i), and β_3 is the coefficient for the

[1] Y_{it} is 1 if individual i experiences the event at time t, and 0 otherwise.

potentially endogenous confounders (V_{it}). Note that the confounders may be time-varying (V_{it}) or time-invariant (U_i).

Only recently has the methodological community realized that covariance analysis may produce spurious correlations and thus produce further bias in the time-varying setting (Robins 1986, 1987, 1989; Robins and Greenland 1994; Robins et al. 1992, 2000). This bias is caused by endogenous confounders. Figure 1 illustrates one possible underlying reality for the relationship between exposure and response. In this example, the exposure, confounder, response, and unmeasured confounder are time-varying. (Exogenous time-invariant confounders are not pictured.) The arrows represent causal paths. (The arrows we discuss below are presented in black, all other arrows are in gray in Figure 1.) Note that no indirect or direct causal path can be traced in the direction of the arrows from exposure to response; that is, for illustrative purposes, the exposure is not causally related to the response. We use this to illustrate how covariance analysis (e.g., see the model in Equation 1), leads to the appearance of a relationship even when there is no relationship. To interpret this figure, it is necessary to think of time as progressing from left to right. For example, V_{i0} is pictured to the left of X_{i0}; this signifies that V_{i0} is not an outcome of X_{i0}.

(Figure 1, about here)

V_{i0} and V_{i1} are confounders because they are correlated with subsequent values of the exposure (X_{i0} and X_{i1}) and the response (Y_{i0} and Y_{i1}). To control for this confounding, covariance analysis includes the confounders as covariates in the regression of the response on the exposure. From the figure, however, we see that V_{i1} is endogenous (indicated by the arrow from X_{i0} to V_{i1}), causing the following problem: if we condition on V_{i1} (including it in our model), we create a spurious correlation between X_{i0} and Y_{i1} through V_{i1} and η_{i1} via the paths A, B and C. The spurious correlation occurs because V_{i1} is endogenous (a consequence of X_{i0}). (For a more detailed discussion, see Barber, Murphy, and Verbitsky (2004).)

3. A NEW METHOD: WEIGHTING

Hernán, Brumback, and Robins (1999) use a weighted survival analysis method, which they call a marginal structural model (MSM). This method uses sample weights (inverse-probability-of-exposure weights) to statistically control for time-varying confounders and thus produces an unbiased estimator of the effect of exposure on response. If the exposure were randomized, then a model such as

$$\text{logit}\,(p_{it}) = \beta_0 + \beta_1 X_{it} + \beta_2\, U_i \qquad (2)$$

could be used to assess the effect of exposure on response because randomized exposure results in an equal distribution of all variables across the exposure categories. β_1 represents an average effect of exposure (averaged across variables not included in the model). To use the weighting method, they fit the model in (2) with a weighted analysis. The goal of the weights is to equalize the distribution of variables across the exposure categories and thus mimic randomization.

3.1 The Weights

To implement the weighting method, we first estimate the weights using a model of the conditional probability of exposure in year t, among those still at risk of exposure for the first time. We use a logistic regression model (other models, such as probit, could be used instead). In each year j, the weight uses a ratio of two conditional probabilities. The denominator is the conditional probability of measured exposure status in year j, given that neither exposure nor response was experienced in a prior year, given variables indicating current and past endogenous confounder status, and given other exogenous variables. The numerator is the conditional probability of measured exposure status in year j, given that neither exposure nor response was experienced in a prior year, and given other exogenous variables. The weight at time j, w_{ij}, is the product of the ratios up to time j (t=1,…,j). Equation 3 shows the form of the weight.

$$w_{ij} = \prod_{t=1}^{j} \frac{P[X_{ij} \mid \overline{X}_{ij-1}, \overline{Y}_{ij-1}, U_i]}{P[X_{ij} \mid \overline{X}_{ij-1}, \overline{Y}_{ij-1}, \overline{V}_{ij-1}, U_i]} \qquad (3)$$

where the overbar indicates past history of the variable, e.g. $\overline{X}_{j-1}$ represents whether individual i was previously exposed, and if so, the time since the prior exposure. (See Appendix A in Barber, Murphy, and Verbitsky 2004 for detailed instructions for computing the weights.) After computing the weights, we fit the model in Equation 2 using a weighted logistic regression.

3.2 Intuition Underlying the Use of the Weights

Note that weighting the sample as described above *does not alter* the relationship between the exposure and the response. We do not use any information about the response in formulating the weights and on the individual-level, the x-y slope, or the relationship between exposure and response, is unaltered. This is intuitively similar to survey sample weighting – the weights are compensating for the oversampling (over-representation) of some confounder patterns among the different exposure categories. The sample weights equalize the confounder distribution across exposure categories, simulating the confounder pattern in an experiment that randomly assigned individuals to the exposure. (See Appendix B in Barber, Murphy, and Verbitsky (2004) for a detailed technical discussion.)

Because V_{it} is no longer a confounder in the weighted sample, we do not need to control or equivalently include it as a covariate in our (weighted) survival analysis model for the response; that is, we may fit the model in Equation 2. By not including V_{it} we avoid the spurious correlation problem, yet we control confounding by using the weights. That is, even though the correlations indicated by "A", "B", and "C" remain, we do not condition on V_{it} in the model, and thus a false correlation between X_{it} and Y_{it} is avoided.

As stated above, the weighting method works by equalizing the composition of individuals across confounder categories among the two groups of individuals – those with the exposure and those without. To see this, consider the following intuitive scenario. There are 110 individuals who have not

experienced the exposure prior to year t, and have not experienced the response prior to year t. Table 1 provides a hypothetical cross-tabulation of two variables in year t: the level of the confounder (V_{it}) and whether the individual experienced the exposure in year t (X_{it}). Note that of the 60 individuals where V_{it} = 1 (row 1), 30 individuals have X_{it} = 1 (column 1, row 1), whereas of the 50 individuals where V_{it} = 0 (row 2), only 10 individuals have X_{it} = 1 (column 1, row 2). This means that individuals with V_{it} = 0 are underrepresented among individuals with the exposure and individuals with V_{it} = 1 are overrepresented among those with the exposure.

(Table 1, about here)

In Table 1, among those individuals with V_{it} = 1 (row 1), one half have the exposure, and one half do not have the exposure. If this proportion also held true for those with V_{it} = 0 (row 2), then we would have Table 2. Here we have divided the individuals with V_{it} = 0 so that the column 1 and 2 proportions are the same as for row 1. This also results in a proportion with V_{it} = 0 among the total with X_{it} = 1(column 1) that equals the proportion with V_{it} = 0 among the total with X_{it} = 0 (column 2); both proportions are 25/55. The original sample will resemble Table 1, the weighted sample will resemble Table 2. This is accomplished by weighting each individual with the inverse of the conditional probability of exposure status given confounder status. Referring to Table 1, a weight of $(10/50)^{-1} = 5$ is assigned to the 10 individuals with X_{it} = 1, V_{it} = 0. The 40 individuals in the more common group with X_{it} = 0 and V_{it} = 0 are assigned a smaller weight of $(40/50)^{-1} = 5/4$. Because there is an equal number with X_{it} = 1 and X_{it} = 0 in the total group where V_{it} = 1, each of these 60 individuals is assigned an equal weight of $(30/60)^{-1} = 2$.

(Table 2, about here)

Note that after weighting the observations for each individual, the frequencies from Table 1 become double the weighted frequencies in Table 2. In practice, weights are assigned as the ratio of the probability of exposure status divided by the conditional probability of exposure status given

confounder status; this eliminates the elevation of the total sample size (see Equation 3). Additionally, Equation 3 shows that the probability of exposure is conditional on all available variables, not just confounder. Thus, the weights eliminate the correlation between confounder and exposure in the weighted sample.

4. A COMPARISON OF ASSUMPTIONS

Both analysis of covariance and the weighting method make a variety of assumptions to estimate causal effects. First, both analyses assume that there is no direct unmeasured confounding, or in other words, that all unmeasured confounders affect exposure only through measured confounders. This assumption is implied in Figure 1 by the absence of direct arrows from either η_{i0} or η_{i1} to either X_{i0} or X_{i1}. That is, there are no direct arrows from the unmeasured factors to the exposure. This assumption is sometimes called "sequential ignorability" (Robins, 1997).

Second, both methods make an additional assumption, although different in form. The weighting method assumes that no past confounder patterns exclude particular levels of exposure; that is, even if $V_{i0} = 0$, it is still possible to have the exposure and vice-versa (Robins, 1999). We call this weighting assumption #2. Covariance analysis does not require this assumption; however, if this situation does not hold, then covariance analysis extrapolates from the possible confounder × exposure patterns to the impossible confounder × exposure patterns. Thus covariance analysis replaces weighting assumption #2 by covariance analysis assumption #2: if some confounder patterns exclude particular levels of exposure, that is if the confounder × exposure patterns that do not exist in the current data *can occur* in a future setting, then covariance analysis assumes that the model holds for these confounder × exposure patterns. Thus, covariance analysis assumption #2 is extrapolation – extrapolating to nonexistent confounder × exposure patterns using the model for the existing confounder × exposure patterns. This is a potentially useful attribute that is not shared by the weighting analysis.

Covariance analysis makes one additional assumption: there are no unmeasured confounders

affecting measured endogenous confounders – if the arrows from η_{i1} to V_{i1} or from η_{i0} to V_{i1} were absent, this assumption would hold. The spurious correlation between X_{i0} and Y_{i1} through V_{i1} and η_{i1} discussed above results from a violation of this assumption. The weighting method does not require this assumption.

In the simulations, we examine the robustness of the weighting method to the first two assumptions (sequential ignorability and no past confounder patterns exclude particular levels of exposure), and compare the weighting method to covariance analysis when the first two assumptions do not hold.

5. EMPIRICAL EXAMPLE

We now illustrate the method described in section 3 using survey data. We address the sociological research question: (1) If more children in Nepal attend school, would more couples limit their total family size via sterilization? We use an existing data set and compare three types of methods: (a) the naive method, which ignores time-varying confounding (e.g., the model in Equation 2); (b) covariance analysis, which includes all known confounders as covariates in the model (e.g., the model in Equation 1); and (c) the weighting method (e.g., the model in Equation 2, fit with a weighted regression). We use path analysis rules to anticipate the direction of the bias from the three methods. Thus we first briefly explain the path analysis rules we use.

5.1 Path Analysis Rules

Following Davis (1985) we associate signs (+/-) with each path in a figure. Our goal is to ascertain the sign of the correlation between exposure and response, marginal over the variables that are uncontrolled (not included in the model) and conditional on the variables that are controlled (included in the model). Following Davis (1985) and Duncan (1966), we locate all paths connecting exposure to response. Paths without converging arrows, and for which no variable along the path is in the model, contribute to the correlation between exposure and response. For each such path, we multiply the signs

along the path and sum to find the sign of the correlation between exposure and response. (If the paths are of different signs the result is ambiguous). Paths with a variable at which there are converging arrows do not contribute to the above sum as long as the variable is not in the model. For example, using Figure 1, we see that if V_{i1} is not in the model, then the path A to B to C does not contribute to the sum. Furthermore if you control for (include in the model) a variable without converging arrows along a particular path between exposure and response then the path does not contribute to the above sum.

Suppose there is a path between exposure and response containing a variable with converging arrows that is included in the model. We do not know of a path analysis rule for including such a variable in the model. We derive a new rule by following Pearl's (1998) non-parametric d-separation rule. If a variable with converging arrows is included in the model, paths that include the variable contribute to the sum determining the sign of the correlation. To calculate the sign, multiply the signs of the coefficients along the path as before, but then multiply by -1 (see Appendix C in Barber, Murphy, and Verbitsky 2004 for a proof of the rule). Rules for deducing independencies in more complicated situations are provided by Pearl (2000).

5.2 Education and Fertility Limitation

To examine the research question, we use data from the Chitwan Valley Family Study. We focus on the relationship between sending a child to school and a couple's subsequent hazard[2] of sterilization. Briefly, our hypothesis is that couples who send a child to school become aware of the costs of doing so, and subsequently decide to terminate childbearing via permanent contraceptive use (Axinn and Barber 2001). For a more detailed description of the data and measures, see Axinn and Barber (2001).

There are at least two important measured variables that are likely confounders of this

[2] The hazard at time t is the conditional probability of sterilization at time t, given no prior sterilization and given all measured variables prior to time t included in the model.

relationship: the availability of schools near the couple's neighborhood, and the total number of children born to the couple. Both of these variables vary over the risk period. And, both are potentially endogenous to the relationship between children's education and sterilization. That is, the availability of a school near a couple's neighborhood is likely to increase the probability that a couple sends their child to school. In addition, the difficulties of sending a child to a distant school may increase the couple's (and neighborhood's) motivation to lobby the government for placement of a school near the neighborhood. In fact, many schools were placed in the study area as a result of such social action. Similarly, we expect that sending a child to school will influence how many children the couple has (similar to the effect on sterilization), and will also be influenced by the couple's total number of children. Note that this is the same example illustrated in Figure 1, with an additional confounder variable, and permitting a causal effect of exposure on response.

There are also other exogenous confounders that are likely to influence both the exposure (sending a child to school) and the response (sterilization) (see Axinn and Barber, 2001). These factors include whether the mother has ever attended school, the educational attainment of the father, whether the mother lived near a school from birth to age 12, the age (birth cohort) of the mother, religious/ethnic/racial group of the couple, miles to the nearest town, and years since the couple's first child turned 6. We have chosen to include these confounders in the regression model for the response (as well as in the models for the numerator and denominator of the ratio used in the weights) so that we can see the magnitude of their coefficients in our model.

Table 3 compares the naive, covariance analysis, and weighting methods using these data. In column 1, the use of the naive method yields an estimated causal effect of sending a child to school on the hazard of sterilization of .93. According to this method, the monthly log-odds of becoming sterilized is .93 higher among those couples who have sent a child to school compared to their peers who have not.

Of course, we suspect that this estimate is biased because we know that there are multiple confounders for which we have not accounted in this model. In other words, we suspect that *part* of the reason that sending a child to school is related to the hazard of sterilization is because some who sent a child to school probably did so because they live near a school or because they have relatively few children.

(Table 3, about here)

Thus, we next estimate parameters using covariance analysis, presented in column 2. This model includes all of the measured confounders as covariates in the model. As expected, including these confounders alters the estimate of the effect of sending a child to school on the hazard of sterilization. The magnitude is reduced to .74. In addition, column 2 suggests that living near a school and family size are strongly related to the hazard of sterilization.

Column 3 presents the results from the weighting method. Recall that this method adjusts for the time-varying confounders by including them in the weights in a weighted logistic regression. In this method, the estimated effect of sending a child to school on the hazard of sterilization is again slightly smaller compared to the naive and standard methods. Thus, the weighting method suggests that the naive and covariance analysis methods slightly overestimate the magnitude of the effect of sending a child to school on permanent contraceptive use. Overall, the differences in these estimates across the three types of methods are not substantial. In fact, the confidence intervals around these estimates substantially overlap, suggesting that the estimates do not statistically differ.

If the assumptions of (1) sequential ignorability and (2) no past confounder patterns exclude particular levels of exposure are true, then the similarity across columns in Table 3 suggests two possibilities. First, there may not be much unmeasured indirect confounding in this example. In other words, the correlations along the paths B and C illustrated in Figure 1 are very small or nonexistent. This is likely because the survey data were designed specifically to answer this research question, and

thus all known theoretically relevant confounders were measured in the survey. Second, it may be that previous values of the exposure have very little direct influence on subsequent values of the confounders (path A in Figure 1). This is possible if the link between whether a child has attended school and whether the couple lives nearby a school is weak, and if the link between whether a child has attended school and family size is weak (or mainly operates via an intermediate measured variable, such as sterilization).

Although the estimate from the weighted model is not dramatically different from the estimate given by covariance analysis, we provide some interpretation of this change for illustration. This is not, however, to suggest that the difference between these models is substantial. The decrease in the magnitude of the estimate comparing column 1 to column 3 in Table 3 suggests that there is a small path from X_{i0} to Y_{i1} via A, B, and C (shown in Figure 1). This means, because A and B arrows go in to the box V_{i1} (and thus we multiply our signs by a negative sign), either (1) all arrows are negative or (2) two of the A, B, C, arrows are positive and one is negative, to get an overall positive sign. From theory and our knowledge of the setting, we expect that sending a child to school may motivate couples to move closer to a school or to lobby the government for a school placed nearby. So, path A is likely to be positive. Thus, either path B or C must be negative, but not both. Fecundity is an unmeasured (and perhaps unmeasurable) potential confounder that demographers often consider. It is possible that sub-fecundity would be positively correlated with having a smaller family, but negatively related to sterilization, resulting in a positive path B and a negative path C.

6. SIMULATION

6.1 Description of the Simulated Data

Next, we describe the results of simulations performed to evaluate the weighting (MSM) method in assessing cause and effect. In these simulations we create and thus know the true underlying causal relationships. See Appendix D in Barber, Murphy, and Verbitsky (2004) for a detailed explanation of

how the simulated data were created.

We compare three methods that may be used to estimate the effect of a time-varying exposure on a time-varying outcome: (1) the naive method, which ignores time-varying confounding; (2) covariance analysis, which includes all known confounders as covariates in the model; and (3) the weighting method (MSM, or marginal structural model) proposed by Hernán, Brumback, and Robins (1999), which uses a weighted survival analysis method, using sample weights (inverse-probability-of-exposure weights) to statistically control for time-varying confounders.

The variables in our simulation are illustrated in Figure 2. We have 5 variables. There are two time periods in the simulated data: 0 and 1; they are indexed by t. Individuals are indexed by i.

- η_i is an unmeasured indirect confounder with three levels: 2=high, 1=medium, and 0=low. It is time-invariant.
- $V1_{it}$ = 1 or 0 (t = 0,1). This is a measured confounder.
- $V2_{it}$ = 1 or 0 (t = 0,1). This is a measured confounder.
- X_{it} = 1 or 0 (t = 0,1). This is the exposure.
- Y_{it} = 1 or 0 (t = 0,1). This is the response.

(Figure 2, about here)

We generated this data according to Figure 2, using the equations listed in Table 4. Note that in Figure 2, similar to our previous examples, there is no effect of exposure (X_{it}) on response (Y_{it}). Rather, the unmeasured confounder (η_i) is indirectly related to the exposure (X_{it}), and is also directly related to the response (Y_{it}). In fact, in our simulation, the unmeasured confounder (η_i) completely determines the response (Y_{it}). Note that there is no direct arrow in Figure 2 from η_i to X_{i0} or X_{i1}; recall that sequential ignorability must hold in order to produce an unbiased estimator of the causal effect – unmeasured confounders must *not* be directly related to exposure. In other words, we assume that we have good

surrogates for the unmeasured confounders in the form of the measured confounders. In Section 5.2.5 we analyze the simulated data acting as if one of the measured confounders is unavailable – in other words, we examine the robustness of the weighting method to the sequential ignorability assumption. Note that although we could add other arrows between variables, we have chosen to keep the problem as simple as possible for parsimony and clarity. Table 4 provides detailed information about the design of the simulation. All of the simulations are based on 1,000 simulated data sets of 1,000 observations each.

(Table 4, about here)

6.2 Results of the Simulation

6.2.1. Summary

In each of our simulations, we compare the weighting method to covariance analysis and use the naive method (regular logistic regression ignoring confounding) as the baseline.

As we illustrate below, when time-varying endogenous confounders are present, the naive method produces biased estimators of the causal effect of the exposure on response. Path analysis rules indicate that when the time-varying confounder is correlated with past and future exposure and is also correlated with the response, covariance analysis will produce biased estimators; our simulations demonstrate that this is true. Of course, the degree of bias depends on the strength of the correlations. The simulations also demonstrate that the weighting method provides unbiased estimators of the causal relationship between the exposure and the response. In our simulations, the weighting method provides estimators and standard errors no worse, and in many cases better, than covariance analysis even when time-varying covariates that are not confounders are included in addition to the confounders in the model.

In situations when weighting assumption #1 (sequential ignorability) is false but weighting assumption #2 (non-zero probability of all patterns of exposure and confounders) is true, our simulations indicate that the weighting method provides estimators that are usually less biased than the naive

method. The overall bias of estimators from covariance analysis lies in-between bias resulting from the use of the naive and weighting methods. However, when weighting assumption #2 is false but weighting assumption #1 (sequential ignorability) is true, the weighting method performs worse than both the naive method and covariance analysis. A detailed discussion of the simulations follows.

6.2.2 A Nonparametric Analysis

Table 5 presents an analysis of simulated data according to the model described by Equations (4a) and (4b):

$$logit\ (p_{i0}) = \beta_0 + \beta_1 X_{i0} + \delta_1 V1_{i0} + \delta_2 V2_{i0} \quad (4a)$$

$$logit\ (p_{i1}) = \beta_0 + \beta_4 + \beta_2 X_{i0} + \beta_3 X_{i1} + \delta_1 V1_{i1} + \delta_2 V2_{i1} + \delta_3 V1_{i0} + \delta_4 V2_{i0} \quad (4b)$$

where $p_{it} = P[Y_{it} = 1 | X_{it}, U_i, V_{it}]$[3], t=(0,1). In the naive method and the weighting method, the parameters, $\delta_1, \delta_2, \delta_3, \delta_4$ are set to zero. In the naive and weighting methods these are non-parametric models for Y_{it} because we have a parameter for all possible exposure-response combinations:

Proportion of $Y_{i0} = 1$:

$X_{i0} = 0$	$X_{i0} = 1$
expit(β_0)	expit($\beta_0 + \beta_1$)

[3] Y_{it} is 1 if individual i experiences the event at time t, and 0 otherwise.

Proportion of $Y_{i1} = 1$, given $Y_{i0} = 0$:

$X_{i0} = 0, X_{i1} = 0$	$X_{i0} = 0, X_{i1} = 1$	$X_{i0} = 1, X_{i1} = 1$
$\text{expit}(\beta_0 + \beta_4)$	$\text{expit}(\beta_0 + \beta_4 + \beta_3)$	$\text{expit}(\beta_0 + \beta_4 + \beta_2 + \beta_3)$

For parsimony and ease in presentation, we use the function expit(x) as a shorthand for $\frac{e^x}{1+e^x}$.

Note that p = expit(x) if and only if x = logit(p) where logit(p) = $\log[\frac{p}{1-p}]$.

Because the model is non-parametric, it fits the data perfectly – this means that if you substitute estimated betas for the above values you will get exactly the data proportions. For example substituting the estimated $\hat{\beta}_0, \hat{\beta}_4, \hat{\beta}_2, \hat{\beta}_3$ in $\text{expit}(\beta_0 + \beta_4 + \beta_2 + \beta_3)$ will give you the proportion of individuals exposed at both time 0 and time 1 who experience the response at time 1 among those exposed individuals who did not respond at time 0.

Covariance analysis includes all known confounders as covariates in the model[4], and the proportions can be computed as follows:

Proportion of $Y_{i0} = 1$:

$X_{i0} = 0$	$X_{i0} = 1$
$\text{expit}(\beta_0 + \delta_1 V1_{i0} + \delta_2 V2_{i0})$	$\text{expit}(\beta_0 + \beta_1 + \delta_1 V1_{i0} + \delta_2 V2_{i0})$

[4] Note that this model is no longer non-parametric when the confounders are added to the model. However, making a non-parametric model that includes the confounders would require the addition of many parameters to the model. Thus, for parsimony and ease in presentation, we estimate this simple parametric model.

Proportion of $Y_{i1} = 1$, given $Y_{i0} = 0$:

$X_{i0} = 0, X_{i1} = 0$	$X_{i0} = 0, X_{i1} = 1$	$X_{i0} = 1, X_{i1} = 1$
$\text{expit}(\beta_0 + \beta_4 + \delta_1 V1_{i1} + \delta_2 V2_{i1} + \delta_3 V1_{i0} + \delta_4 V2_{i0})$	$\text{expit}(\beta_0 + \beta_4 + \beta_3 + \delta_1 V1_{i1} + \delta_2 V2_{i1} + \delta_3 V1_{i0} + \delta_4 V2_{i0})$	$\text{expit}(\beta_0 + \beta_4 + \beta_2 + \beta_3 + \delta_1 V1_{i1} + \delta_2 V2_{i1} + \delta_3 V1_{i0} + \delta_4 V2_{i0})$

In each of these models, β_0 and β_4 are the intercept parameters, and δ_1, δ_2, δ_3, δ_4 are used to control for confounding. Our focus is on β_1, β_2, and β_3, which represent the observed relationship between exposure (X_{it}) and response (Y_{it}).

Naive method. Because the naive method ignores the time-varying confounding, it yields biased estimators of the effect of exposure on response when time-varying confounding exists. Our simulations confirm this. Table 5, columns 1 through 3 present the summaries of the estimates using the naive method.

(Table 5, about here)

For example, we expect that β_1 (estimator of the effect of exposure on response at time 0) will be a biased estimator of a zero effect because $V1_{i0}$ and $V2_{i0}$ are common correlates of X_{i0} as well as Y_{i0}. Using path analysis rules, $\hat{\beta}_1$ should be significantly positive. In a linear model with standardized covariates, $\hat{\beta}_1$ would be approximately $\gamma 1_0 * \alpha 1_0 + \gamma 2_0 * \alpha 2_0$. Because we are using a nonlinear model and we are not standardizing, we use this path analysis rule to ascertain sign, but not magnitude. Column 1 shows the average estimated value of $\hat{\beta}_1$ using the naive method. As predicted by the path analysis rule, instead of an average estimated $\hat{\beta}_1$ of 0, the average of the 1000 estimated $\hat{\beta}_1$'s is positive (=.31). The proportion of the 1000 data sets with |t-ratio| > 1.96 is .56. This means that if we ignored time-varying confounding, we would find false evidence that there is an effect of exposure on response (find an effect where none truly exists) for 56% of the data sets. If there were no bias, we would find false evidence of an effect for approximately 5% of the data sets (using a type 1 error rate of .05).

Covariance Analysis. Columns 4 through 6 of Table 5 present estimates from covariance analysis, which includes all known confounders as covariates in the model.

We expect from path analysis rules that β_1 (estimator of the effect of exposure on response at time 0) will be unbiased because we have included the measured confounders ($V1_{i0}$, $V2_{i0}$). Including the confounders in covariance analysis means that the path from X_{i0} to Y_{i0} via confounders at time 0 ($V1_{i0}$, $V2_{i0}$, and η_i) does not contribute to $\hat{\beta}_1$. Column 4 shows that the average of the 1000 estimated $\hat{\beta}_1$'s is .04 and the proportion of the 1000 data sets with |t-ratio| > 1.96 is only .06. This is not significantly different from the expected 5%. Thus, consistent with our prediction, covariance analysis produces a well performing estimator $\hat{\beta}_1$. Similarly, using path analysis rules, we expect $\hat{\beta}_3$ (estimator of the effect of exposure on response at time 1) to be unbiased because we have included the measured confounders ($V1_{i1}$, $V2_{i1}$). The average of the 1000 estimated $\hat{\beta}_3$'s is -.10 and the proportion of the 1000 data sets with |t-ratio| > 1.96 is only .05. Thus, consistent with our prediction, $\hat{\beta}_3$ is unbiased.

However, there is a problem in the estimation of β_2 (estimator of the effect of exposure at time 0 on response at time 1). Because we have included the confounders in the model, we might expect $\hat{\beta}_2$ to be unbiased. However, because covariance analysis includes $V1_{i1}$ and $V2_{i1}$ in the model and these are on the indirect path from X_{i0} to ($V1_{i1}$ and $V2_{i1}$) to η_i to Y_{i1}, the new path analysis rule (as explained in section 5.1) implies that the estimated effect of X_{i0} on Y_{i1} will be negatively biased ($\alpha 1_1$, $\alpha 2_1$, δ_1, δ_2 are all positive). Indeed in a linear model the bias would be $-\alpha 1_1 * \delta_1 / (\alpha 1_1^2 + 1) - \alpha 2_1 * \delta_2 / (\alpha 2_1^2 + 1)$. The simulations confirm that we do not get an unbiased estimator of β_2. Column 5 shows that the average of the 1000 estimated $\hat{\beta}_2$'s is *not* zero; rather, it is -.69. Moreover, in 80% of the data sets, we reach the false conclusion that there is an effect when none exists.

Weighting Method. Columns 7 through 9 in Table 5 present summaries of the estimates from the weighted logistic regression, as described in section 3 above. Note that we do not explicitly include

the confounders in the model because the weights account for the confounders.

We expect that $\hat{\beta}_1$ will be unbiased because we have adjusted for $V1_{i0}$ and $V2_{i0}$ in the weights; column 7 confirms this. We expect that $\hat{\beta}_2$ will also be unbiased for two reasons. First, we have adjusted for the confounders $V1_{i0}$ and $V2_{i0}$ in the weights. Second, we have not included outcomes of X_{i0} (i.e. $V1_{i1}$, $V2_{i1}$) in the model. Column 8 shows that the average of the 1000 estimated $\hat{\beta}_2$'s is -.003 and the proportion of the data sets with |t-ratio| > 1.96 is only .05. This average effect size is statistically indistinguishable from zero, and thus the weighted estimator of β_2 is unbiased. $\hat{\beta}_3$ is also unbiased (see column 9).

Summary. In sum, we find that the naive method leads to biased estimators of β_1, β_2, and β_3. Covariance analysis leads to biased estimators of β_2 only. And, the weighting method leads to <u>un</u>biased estimators of β_1, β_2, and β_3.

6.2.3 A Parsimonious Analysis

Next we fit a parsimonious model to the same data – excluding the term estimating the effect of X_{i0} on Y_{i1} (i.e., in the parsimonious model we set β_2=0). Social scientists rarely fit nonparametric models because there are usually more than two observation periods, which means the nonparametric model requires a prohibitive number of parameters. However, parsimonious models can include a variety of summary variables. Here, we include two parameters estimating the effect of current exposure on current response – β_1, the effect of X_{i0} on $Y_{i0;}$ and β_3, the effect of X_{i1} on Y_{i1}. If we had more than two observation periods, we would also include a variable summarizing past exposure, such as time since first exposure.

(Table 6, about here)

Similar to the nonparametric model estimated using the naive method (columns 1 through 3 in Table 5), columns 1 through 3 of Table 6 shows that the parsimonious model estimated using the naive method also leads to bias in $\hat{\beta}_1$ and $\hat{\beta}_3$.

Recall that in the *nonparametric* model estimated using covariance analysis (column 6 in Table 5), $\hat{\beta}_1$ and $\hat{\beta}_3$ were unbiased. However, following the path analysis rules, we expect $\hat{\beta}_3$ in the *parsimonious* model estimated using covariance analysis to be negatively biased. Indeed, column 4 of Table 6 shows that the average of 1000 estimated $\hat{\beta}_3$'s is -.59 and the proportion of the data sets with |t-ratio| > 1.96 is .43.

Because weights adjust for the confounding, path analysis rules suggest that $\hat{\beta}_1$ and $\hat{\beta}_3$ should be unbiased in the weighted analysis. Column 6 of Table 6 confirms that β_1 and β_3 are unbiased.

In sum, the naive method produces biased estimators of β_1 and β_3. Consistent with the intuitive description of the endogeniety problem, covariance analysis produces a biased estimator of β_3, the effect of X_1 on Y_1. In contrast, the weighting method produces unbiased estimators of both β_1 and β_3.

6.2.4 Varying the Magnitude of the Relationships

We expect that the strength of the correlation between the unmeasured confounder and exposure (through the α's and γ's) will be associated with the <u>degree</u> of bias in Table 7. This includes all estimators from the naive method, and β_2 from the standard method. Estimators from weighted logistic regression should not be affected by variations in correlations.

Table 7 presents simulations with smaller values for the α's and the δ's compared to Table 5 (refer to Figure 2 for the connection between these parameters and the covariates). In panel A, $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = 1.0$ and $\delta_1 = \delta_2 = 1.0$; these values are smaller than in Table 6. In panel B the values are even smaller relative to those in Table 6: $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = .5$ and $\delta_1 = \delta_2 = .5$. In other words, the correlation between the unmeasured confounder (η_i) and the confounders ($V1_{i1}$ and $V2_{i1}$) is smaller than in Table 5, and the correlation between exposure (X_{i0}) and subsequent levels of the confounders ($V1_{i1}$ and $V2_{i1}$) is smaller than in Table 5. Thus, we expect that the overall magnitude of the relationship between X_{i0} and Y_{i1} via these paths (similar to arrows A, B, and C in Figure 1) will be smaller. In other words, we expect the degree of bias to be smaller in Table 7 panel A than in Table 5, and to be still

smaller in Table 7 panel B.

(Table 7, about here)

The simulations confirm these expectations. Recall that the true value of *all* the β parameters in the simulated data is zero, so estimates closer to zero indicate less bias. Using both covariance analysis and the naive methods, *all* of the estimates decrease in magnitude (i.e., become closer to zero) from Table 5 to Table 7 panel A to Table 7 panel B. Additionally, the proportion of the data sets where we would reach an incorrect conclusion evaluating the significance of β_2 via covariance analysis decreases from 80% to 53% to 10%. We found this to be a general pattern, as well (not shown in tables): as the α and δ values decrease, the bias decreases. In contrast, in the weighting method we would reach an incorrect conclusion about 5% of the time for all parameters. Even when the confounders have a weak correlation to the exposure and the response, including them in the weights or in the weighted logistic regression does not lead to biased estimators.

A natural concern is that accidentally controlling for variables that are not confounders via the weights will lead to bias or instability of the estimators produced by the weighting method. However, additional simulations (not presented in tables) showed that this is not true; including in the weights a small number of covariates that are not confounders neither biased the results of the weighted analysis nor increased the standard errors. However, if the inclusion of the confounder(s) leads to a violation of weighting method assumption #2, it will cause problems, as described below in section 6.2.6.

6.2.5 Presence of Direct Unmeasured Confounders

In this section we test the robustness of the weighting method to weighting assumption #1 (sequential ignorability): *all* direct confounders are included in the weights. We perform this exercise because it is rarely true that *all* direct confounders are measured. We refer to this as "partially weighted" to emphasize that only some of the confounders are included in the weights.

In Table 8, the data were simulated so that $V1_{it}$ is more strongly related to Y_{it} than $V2_{it}$ ($\alpha 1_t >$

$\alpha 2_t$). In other words, $V1_{it}$ is a stronger predictor of the response than $V2_{it}$. As a shorthand, we refer to $V1_{it}$ as the "more important" confounder. When the weights adjust for only part of the confounding, we expect some bias in the estimators from the method using partial weights. Indeed we expect that when the weights adjust only for $V1_{it}$ (the more important confounder) but not $V2_{it}$, the bias will be smaller compared to when the weights adjust only for $V2_{it}$ but not $V1_{it}$. We also expect that when weights adjust only for $V2_{it}$ the bias will be smaller than the bias produced by the naive method that does not adjust for confounding at all.

(Table 8, about here)

Table 8 presents the estimates from these simulation models. The results confirm our expectations. Comparing columns 13 through 15, to columns 16 through 18, to columns 19 through 21, we see that as we adjust for more confounding in the weights, the estimators $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ become less biased and the error rates decline. Indeed, adjusting for even a small amount of confounding by using a model with partial weights (colums 13 through 15 and columns 16 through 18) leads to less bias than ignoring the confounding (columns 1 through 3). Again, we found this to be a general pattern (not shown in tables): as the proportion of confounders included in the weights increases, the bias decreases. In addition, even when $V2_{it}$ is not really a confounder (i.e., $\alpha 2_0 = 0$ or $\gamma 2_0 = 0$) and is included in the formulation of the weights along with $V1_{it}$, the weighting method provided estimators and standard errors no worse, and in many cases better, than the standard method. Also note that the same results hold for covariance analysis, but only for $\hat{\beta}_1$ and $\hat{\beta}_3$; as we adjust for more confounders in covariance analysis, the bias decreases for both $\hat{\beta}_1$ and $\hat{\beta}_3$. However, the bias of the estimator $\hat{\beta}_2$ increases as we include more confounders in covariance analysis (i.e., compare columns 5 and 11).

6.2.6 "Bad" Weights

Next, we examine the sensitivity of the weighting method to weighting method assumption #2: no past confounder patterns exclude particular levels of exposure. For this purpose, we constructed our

data so that a specific confounder pattern nearly determines the level of the exposure. For example, in the data for table 9, γ_0=-15.0; $\gamma 1_0$=$\gamma 1_1$=$\gamma 2_0$=$\gamma 2_1$=8.0; thus, the probability of exposure by time 0 (X_{i0} = 1) when $V1_{i0} = V2_{i0} = 0$ is 3.05 *10^{-7} (see Table 5 and Figure 3). This specific confounder pattern leads to a very low probability of exposure. As a result, we expect some very large weights, and we expect that our weighting method will produce biased estimators. This data has been generated so that covariance analysis assumption #2 is true.

(Table 9, about here)

The simulations confirm our expectations. As illustrated in Table 9, estimators from the weighting method are biased and results are worse than those from covariance analysis. The standard errors are poorly estimated (i.e., the estimated standard errors in row 1 in parentheses are different from the mean of standard errors in the second row) and the regression coefficients are poorly estimated (i.e., not close to zero). However, the probability of exposure when $V1_{i0} = V2_{i0} = 0$ has been set to a rather extreme value (3.05 *10^{-7}). When this probability is set to be less extreme, the weighting method did not show this level of bias in the simulations (not shown in tables). Thus, the "bad weights" problem may be unlikely – the correlations must be extremely high, and we found that in other simulations the weights cannot be estimated at all (which would provide a warning flag to the analyst).

7. CONCLUSIONS

In this paper, we have intuitively explained why including endogenous time-varying confounders in the analysis model – covariance analysis, a common method in social science – can produce biased and misleading results. We demonstrate a new method – weighting – to control for these endogenous time-varying confounders. We also evaluate this new method with simulated data. The simulations show the following: fitting nonparametric models with the naive method (ignoring time-varying confounding), all coefficients are biased. Fitting nonparametric models using covariance analysis, only the coefficient for the effect of exposure in the prior period on response in the current period is biased. Fitting

nonparametric models using the weighting method, none of the coefficients are biased. The analogous parametric models show a similar pattern. In addition, the greater the extent to which the exposure influences subsequent values of the confounder (i.e., the extent to which the confounders are endogenous to the exposure), the greater the bias. Our simulations suggest that when the relationship between exposure and subsequent values of the confounder is very small, the bias will also be very small. Furthermore, as more of the confounders are included in the weights, the bias in the weighted method decreases. We found that including *some* of the confounders in the weights leads to less bias than including *none* of the confounders in the weights.

Finally, we found that the weighting method performs quite well overall, except when weighting method assumption #2 (no past confounder patterns exclude particular levels of exposure) is violated severely – in other words, when a particular confounder pattern makes exposure *extremely* unlikely. Knowing when this will occur appears challenging. We expect that in many cases the data alone will not provide evidence about whether weighting assumption #2 is violated, and that detection of such a violation will require substantive input. Of course the data alone will not provide evidence about whether covariance analysis #2 holds, either.

An important and necessary generalization of this method would be to multilevel data. In fact, in the education and fertility example from the Chitwan Valley Family Study, women are grouped into neighborhoods. Future research should adapt these methods to multilevel data structures, and address whether and how our results would change given methods that accommodate multilevel data structures.

So, what is the overall worth of this new weighting method to sociologists? We have learned many lessons over the past few decades – embracing a "new" statistical method, only to be disappointed later that it does not perform well in a wide variety of sociological problems. We believe this new weighting method will be particularly useful in three situations. The first situation is when researchers must address their research question using data collected for a different purpose, and thus the data lack

some important (known) confounders. In this case their models cannot include some confounders, and to the extent these unmeasured confounders are important, estimators from covariance analysis will be biased. Second, sometimes even when researchers collect their own data, it may be too expensive or too difficult to get good measures of some confounders. Again, the models using these data cannot include these confounder(s), and to the extent the confounders are important, estimators from covariance analysis will be biased. A third situation for using the weighting method is when new information is discovered *after* data have been collected, indicating that there important confounders are missing from the data set. In all of these cases we recommend the use of the weighting method.

References

Axinn, William G. and Jennifer S. Barber. 2001. “Mass Education and Fertility Transition.” American Sociological Review 66(4): 481-505.

Barber, Jennifer S., Susan Murphy, and Natalya Verbitsky. 2004. "Adjusting for Time-Varying Confounding in Survival Analysis: A Technical Report." Population Studies Center Research Report 04-???. University of Michigan.

Davis, James A. 1985. The Logic of Causal Order. Beverly Hills: Sage.

Duncan, O. D. 1966. “Path Analysis: Sociological Examples.” American Journal of Sociology 17: 1-16.

Hernán, M. A., B. Brumback, and J. M. Robins. 1999. "Marginal Structural Models to Estimate the Causal Effect of Prophylaxis Therapy for Pneumocystis Carnii Pneumonia on the Survival of AIDS Patients." Epidemiology 98.

Lieberson, Stanley. 1985. Making It Count: The Improvement of Social Research and Theory. Berkeley and Los Angeles: University of California Press.

Pearl, Judea 1998. “Graphs, Causality, and Structural Equation Models.” Sociological Methods and Research 27:226-284.

Pearl, Judea 2000. Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press.

Robins, J. M. 1986. “A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods - Application to Control of the Healthy Worker Survivor Effect.” Mathematical Modeling 7:1393-1512.

Robins, J.M. 1987. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. Journal of Chronic Disease (40, Supplement), 2:139s-161s.

Robins, J. M. 1989. “The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies.” In Health Services Reserach Methodology: A Focus on AIDS, eds. L. Sechrest, H. Freedman, and A. Mulley. Rockville, MD: U.S. Department of Health and Human Services, pp.113-59.

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

Robins, J.M. (1999). "Marginal Structural Models versus Structural Nested Models as Tools for Causal

Inference." In Statistical Models in Epidemiology: The Environment and Clinical Trials. M.E. Halloran and D. Berry, Editors, IMA Volume 116, NY: Springer-Verlag, pp. 95-134.

Robins, J. M., Blevins, D., Ritter, G., and Wulfsohn, M. 1992. "G-estimation of the Effect of Prophylaxis Therapy for Pneumocystis Carinii Pneumonia on the Survival of AIDS Patients." Epidemiology 3:319-36.

Robins, J. M. and S. Greenland. 1994. "Adjusting for Differential Rates of Prophylaxis Therapy for PCP in High- Versus Low-Dose AZT Treatment Arms in an AIDS Randomized Trial." Journal of the American Statistical Association 89:737-749.

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

Winship, Christopher and Stephen L. Morgan. 1999. "The Estimation of Causal Effects From Observational Data." Annual Review of Sociology 25: 659-707.

TABLE 1
Crosstabulation of X_{it} by V_{it} for year t

		(1)	(2)	(3)
		$X_{it} = 1$	$X_{it} = 0$	Total
(1)	$V_{it} = 1$	30	30	60
(2)	$V_{it} = 0$	10	40	50
(3)	Total	40	70	110

TABLE 2
Hypothetical crosstabulation of X_{it} by V_{it} for year t,
if $p(X_{it} = 1 \mid V_{it} = 1) = p(X_{it} = 1 \mid V_{it} = 0)$

		(1)	(2)	(3)
		$X_{it} = 1$	$X_{it} = 0$	Total
(1)	$V_{it} = 1$	30	30	60
(2)	$V_{it} = 0$	25	25	50
(3)	Total	55	55	110

TABLE 3
Logistic regression estimates (with robust standard errors) of hazard of sterilization on children's education

	Naive	Standard	Weighted
	(1)	(2)	(3)
Any child has ever attended school	.93***	.74***	.68***
	(.10)	(.11)	(.11)
School is present within 5 minutes walk		.17*	
		(.08)	
Family size[a]			
Couple has 1 child		-.68***	
		(.11)	
Couple has 4 or more children		.49***	
		(.12)	
Mother ever attended school	-.11	-.12	-.07
	(.09)	(.10)	(.10)
Husband's years of education	.02*	.02*	.02*
	(.01)	(.01)	(.01)
Mother lived near school during childhood	.26**	.30**	.33**
	(.10)	(.10)	(.12)
Birth cohort[b]			
1952-1961 (age 35 - 44)	-.61***	-.65***	-.56***
	(.09)	(.09)	(.10)
1942-1951 (age 45 - 54)	-1.24***	-1.30***	-1.22***
	(.12)	(.12)	(.13)
Ethnic group[c]			
Low Caste Hindu	-.24	-.18	-.23
	(.13)	(.13)	(.15)
Newar	.13	.14	.13
	(.14)	(.14)	(.14)
Hill Tibeto-Burmese	-.11	-.06	-.02
	(.11)	(.11)	(.12)
Terai Tibeto-Burmese	-.84***	-.88***	-.84***
	(.13)	(.14)	(.14)
Miles to nearest town	-.001	-.001	-.001
	(.001)	(.001)	(.001)
Years since first birth	.0003	-.03**	.01
	(.01)	(.01)	(.01)
N (couples)	1,230	1,230	1,230
N (couple-years)	14,779	14,779	14,779

Notes: hazard starts 6 years after first birth.
[a] Reference group is couples with 2 or 3 children.
[b] Reference group is cohort born 1962-1971 (age 25 - 34).
[c] Reference group is Upper Caste Hindus; two tailed tests.
* $p < .05$, ** $p < .01$, *** $p < .001$, one tailed tests except where noted.

TABLE 4
Definitions of variables used to create simulated data

Variable	Formula used
η	$p(\eta_i = 0) = 1/3$ $p(\eta_i = 1) = 1/3$ $p(\eta_i = 2) = 1/3$
Y_{i0}	if $\eta_i = 0$, $Y_{i0} = 0$ if $\eta_i = 1$, $Y_{i0} = 0$ if $\eta_i = 2$, $Y_{i0} = 1$
Y_{i1}	if $\eta_i = 0$, $Y_{i1} = 0$ if $\eta_i = 1$, $Y_{i1} = 1$
X_{i0}	$p(X_{i0} = 1) = \frac{\exp(\gamma 0 + \gamma 1_0 * V1_{i0} + \gamma 2_0 * V2_{i0})}{1 + \exp(\gamma 0 + \gamma 1_0 * V1_{i0} + \gamma 2_0 * V2_{i0})}$
X_{i1}	$p(X_{i1} = 1 \mid X_{i0} = 0) = \frac{\exp(\gamma 0 + \gamma 1_1 * V1_1 + \gamma 2_1 * V2_1)}{1 + \exp(\gamma 0 + \gamma 1_1 * V1_1 + \gamma 2_1 * V2_1)}$ $p(X_{i1} = 1 \mid X_{i0} = 1) = 1$
$V1_{i0}$	$p(V1_{i0} = 1) = \frac{\exp(\alpha 0 + \alpha 1_0 * \eta_i)}{1 + \exp(\alpha 0 + \alpha 1_0 * \eta_i)}$
$V1_{i1}$	$p(V1_{i1} = 1) = \frac{\exp(\delta 0 + \delta 1_1 * X_{i0} + \alpha 1_1 * \eta_i)}{1 + \exp(\delta 0 + \delta 1_1 * X_{i0} + \alpha 1_1 * \eta_i)}$
$V2_{i0}$	$p(V2_{i0} = 1) = \frac{\exp(\alpha 0 + \alpha 1_0 * \eta_i)}{1 + \exp(\alpha 0 + \alpha 1_0 * \eta_i)}$
$V2_{i1}$	$p(V2_{i1} = 1) = \frac{\exp(\delta 0 + \delta 2_1 * X_{i0} + \alpha 2_1 * \eta_i)}{1 + \exp(\delta 0 + \delta 2_1 * X_{i0} + \alpha 2_1 * \eta_i)}$

TABLE 5

Mean of Estimates for Betas in Nonparametric Analysis Using Simulated Data (N=1,000 datasets of 1,000 cases each)

	Naive			Covariance Analysis			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Mean	.31	.22	.33	.04	-.69	-.10	-.002	-.003	.02
	(.15)	(.20)	(.30)	(.16)	(.25)	(.40)	(.14)	(.20)	(.30)
Mean of Estimated Standard Errors	.15	.19	.28	.16	.25	.39	.15	.20	.30
Proportion of (|T-Ratio| >1.96)	.56	.24	.22	.06	.80	.05	.04	.05	.05

Note: Standard deviation of the 1,000 estimates is in parentheses.

Note: The true values of β_1, β_2, β_3 are zero.

Values of parameters: α_0=0.0; $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1$ =1.5; γ_0 = 0.0; $\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1$ = 0.5; δ_0 = 0.0; $\delta_1 = \delta_2$ = 1.5

TABLE 6

Mean Estimates of Betas in Parsimonious Analysis Using Simulated Data (N=1,000 datasets of 1,000 cases each)

	Naive		Covariance Analysis		Weighted	
	(1)	(2)	(3)	(4)	(5)	(6)
	$\hat{\beta}_1$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_3$
Mean	.31	.49	.05	-.59	-.002	.02
	(.15)	(.25)	(.16)	(.34)	(.14)	(.25)
Mean of Estimated Standard Errors	.15	.25	.16	.34	.15	.26
Proportion of (|T-Ratio| >1.96)	.56	.51	.06	.43	.04	.04

Note: Standard deviation of the 1,000 estimates is in parentheses.

Note: Same data is analyzed as in Table 8.

Note: The true values of β_1, β_2, β_3 are zero.

Values of parameters: α_0=0.0; $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1$ =1.5; γ_0 = 0.0; $\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1$ = 0.5; δ_0 = 0.0; $\delta_1 = \delta_2$ = 1.5

TABLE 7

Mean of Estimates for Betas in Analysis Varying Magnitude of Relationships Using Simulated Data (N=1,000 datasets of 1,000 cases each)

	Naive			Covariance Analysis			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
PANEL A									
Values of parameters: α_0=0.0; $\boldsymbol{\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = 1.0}$; γ_0 = 0.0; $\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1$ = 0.5; δ_0 = 0.0; $\boldsymbol{\delta_1 = \delta_2 = 1.0}$									
Mean	.27	.16	.23	.02	-.47	-.07	-.003	.005	.001
	(.15)	(.19)	(.28)	(.15)	(.22)	(.32)	(.14)	(.19)	(.29)
Mean of Estimated Standard Errors	.15	.19	.28	.15	.23	.33	.15	.19	.29
Proportion of (|T-Ratio| >1.96)	.46	.12	.13	.05	.53	.06	.04	.04	.05
PANEL B									
Values of parameters: α_0=0.0; $\boldsymbol{\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = 0.5}$; γ_0 = 0.0; $\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1$ = 0.5; δ_0 = 0.0; $\boldsymbol{\delta_1 = \delta_2 = 0.5}$									
Mean	.17	.07	.10	.01	-.13	-.03	-.01	-.01	-.004
	(.14)	(.19)	(.27)	(.14)	(.20)	(.28)	(.14)	(.20)	(.28)
Mean of Estimated Standard Errors	.14	.19	.27	.15	.20	.28	.14	.19	.28
Proportion of (|T-Ratio| >1.96)	.21	.06	.07	.05	.10	.05	.05	.05	.04

Note: Standard deviation of the 1,000 estimates is in parentheses.

Note: The true values of β_1, β_2, β_3 are zero.

revised 3/04

TABLE 8

Mean Estimates of Betas in Analysis with Unmeasured Direct Confounders Using Simulated Data (N=1,000 datasets of 1,000 cases each)

	Naive			Covariance Analysis (model includes less important confounder, *V2*)			Covariance Analysis (model includes more important confounder, *V1*)			Covariance Analysis (model includes both measured confounders, *V1* and *V2*)			Partially Weighted (weights computed using less important confounder, *V2*)			Partially Weighted (weights computed using more important confounder, *V1*)			Weighted (weights computed using both measured confounders, *V1* and *V2*)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Mean	.27 (.15)	.20 (.19)	.31 (.28)	.16 (.16)	-.11 (.20)	.19 (.29)	.12 (.16)	-.42 (.25)	.07 (.39)	.02 (.16)	-.74 (.27)	-.04 (.40)	.14 (.15)	.14 (.19)	.20 (.28)	.10 (.14)	.05 (.19)	.07 (.28)	.004 (.14)	.01 (.19)	.01 (.28)
Mean of Estimated Standard Errors	.15	.19	.28	.15	.20	.30	.15	.25	.38	.16	.26	.39	.15	.19	.29	.15	.19	.29	.15	.19	.29
Proportion of (|T-Ratio| >1.96)	.45	.18	.17	.20	.09	.10	.13	.38	.06	.05	.82	.05	.16	.11	.10	.10	.05	.05	.04	.04	.05

Note: Standard deviation of the 1,000 estimates is in parentheses.

Note: The true values of β_1, β_2, β_3 are zero.

Values of parameters: α_0 =0.0; $\boldsymbol{\alpha 1_0 = \alpha 1_1 = 2.25}$; $\boldsymbol{\alpha 2_0 = \alpha 2_1 = .75}$; γ_0 = 0.0; $\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1 = 0.5$; $\delta_0 = 0.0$; $\delta_1 = \delta_2 = 1.5$

TABLE 9

Mean Estimates of Betas in Analysis with Bad Weights Using Simulated Data (N=1,000 datasets of 1,000 cases each)

	Naive			Covariance Analysis			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	b_1	b_2	b_3	b_1	b_2	b_3	b_1	$\hat{\beta}_2$	$\hat{\beta}_3$
Mean	.43	-.41	1.39	.08	-.59	-.25	.12	-.72	.90
	(.14	(.26)	(.29)	(.15)	(.31)	(.36)	(.15)	(.43)	(.47)
Mean of Estimated Standard Errors	.15	.26	.29	.15	.30	.37	.16	.34	.38
Proportion of (|T-Ratio| >1.96)	.86	.35	1.00	.08	.49	.09	.13	.70	.74

Note: Standard deviation of the 1,000 estimates is in parentheses.

Note: The true values of β_1, β_2, β_3 are zero.

Initial values of parameters: α_0**=2.0**; $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = 1.5$; γ_0 **= -15.0**; $\gamma \mathbf{1}_0 = \gamma \mathbf{1}_1 = \gamma \mathbf{2}_0 = \gamma \mathbf{2}_1 =$ **8.0**; $\delta_0 = 0.0$; $\delta_1 = \delta_2 = 1.5$

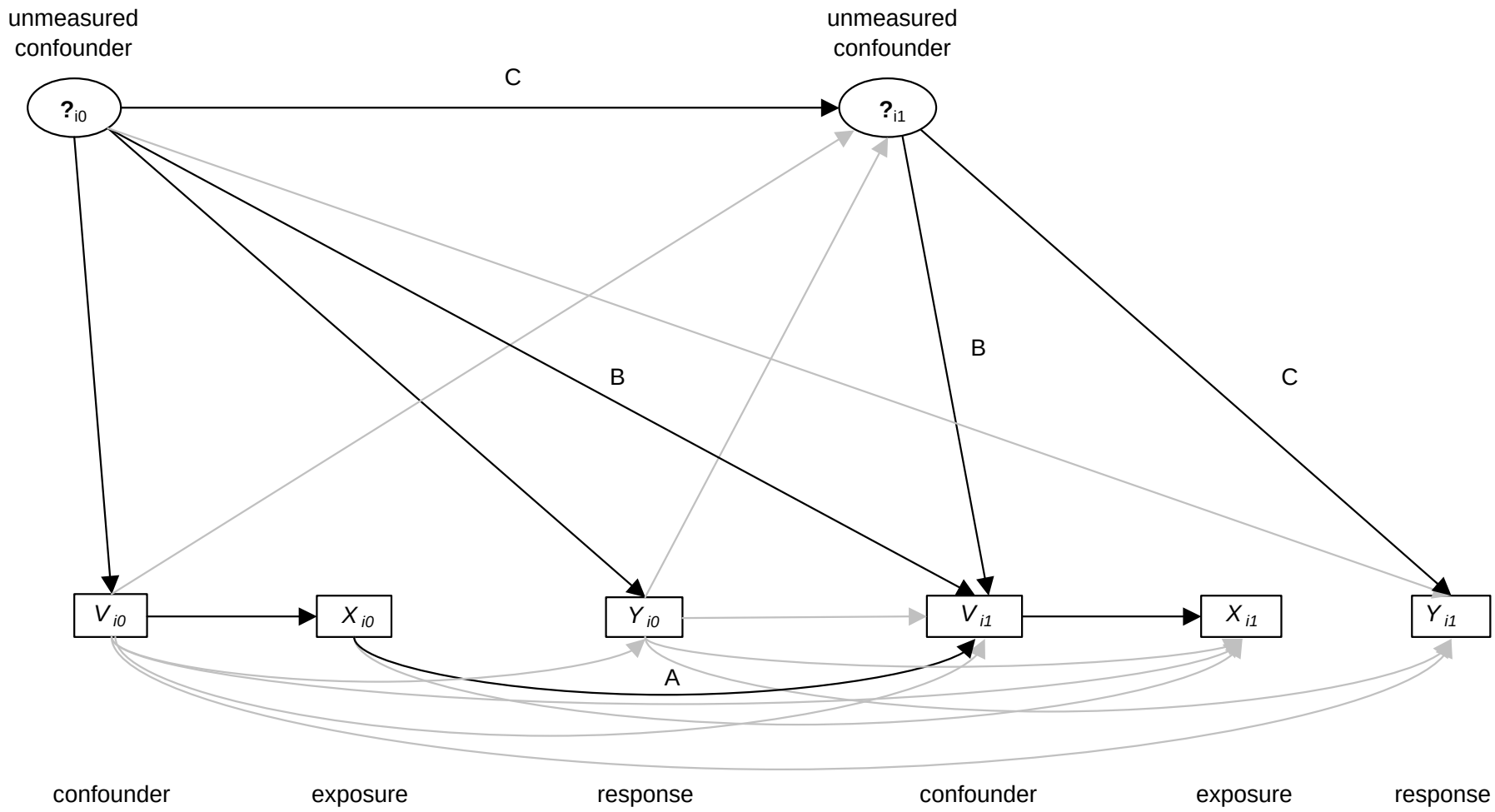

Figure 1
Graphical representation of the multivariate distribution of *unmeasured* , *confounder* , *exposure* , and *response*

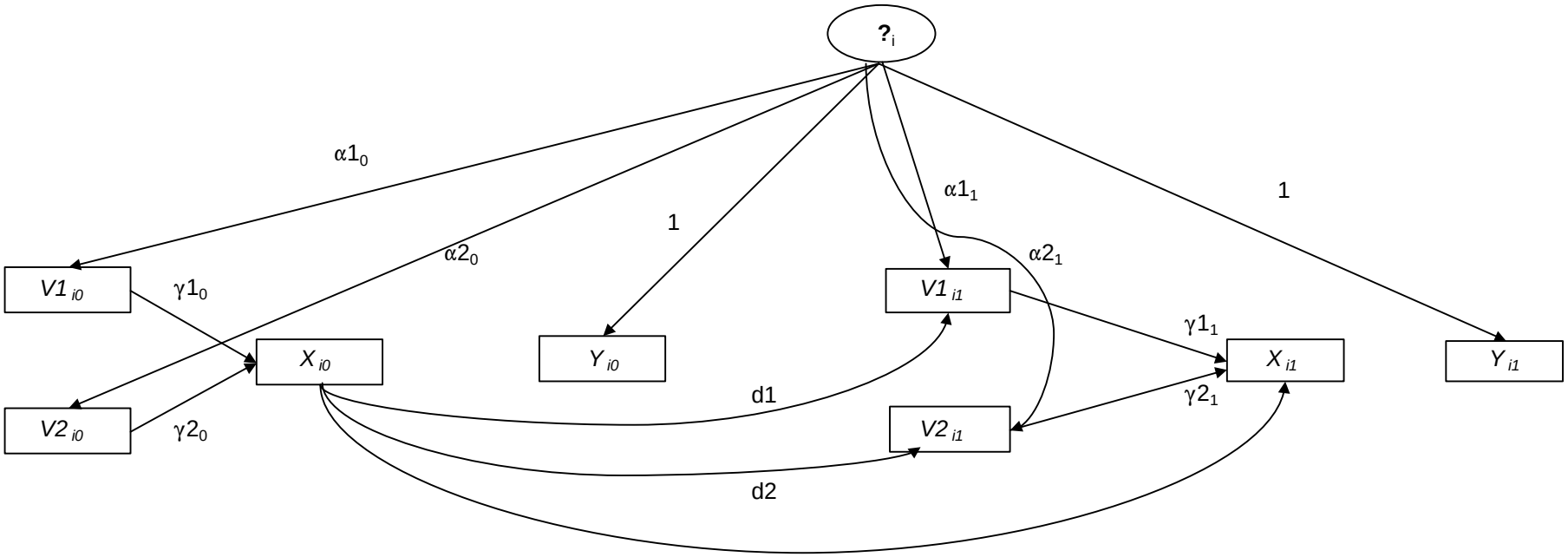

Figure 2
Multivariate distribution of all variables in simulation