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.

Abstract

Adjusting for Time-Varying Confounding in Survival Analysis

In this paper we illustrate how the standard social science method of controlling for time-varying confounding (i.e., directly including time-varying common correlates in the model of the effect of the exposure on the response) can lead to bias in discrete time survival analysis. An alternative to this standard method is Hernán, Brumback and Robins' (1999) method of using sample weights to adjust for endogenous time-varying confounding. We discuss when this method can be used to provide unbiased estimates of effects and we illustrate the method by addressing two substantive research questions 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

Many pressing social science questions concern cause and effect.[1] For example, consider the question, "If more children in poor countries attend school, would more couples limit their total family size via sterilization?,"[2] or "Does lower educational attainment in the United States lead to earlier first births?" In general, causal questions are best answered in an experimental setting – a setting in which subjects are randomly assigned to levels of the exposure (Freedman 1991; Mason 1991; Sobel 2000; Winship and Morgan 1999; Winship and Sobel 2000; see Smith 1990 and Blalock 1991 for insightful critiques of experimental methods). However in the social sciences it is frequently unethical, impractical, 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 (our key independent variable of interest) on a response (dependent variable) is the presence of confounders. Confounders are common correlates of both the exposure and the response. Or, put another way, confounders are alternate explanations for the observed relationship between the exposure and the response. In the first research question above, for example, parents' education is a common correlate of both sending a child to school and the timing of sterilization. Educated couples are more likely to send their children to school, and

---

[1] Whether assessing causation is or should be an important goal in social science has been addressed at length elsewhere (e.g. see Abbott 1998; Blalock 1991; Smith 1990; Winship and Sobel 2000). This paper does not comment on *whether* causal analysis is appropriate or desirable; rather, our focus is on *when* social scientists seek to answer a causal question. In addition, we do not focus on the definition of *cause* in this article. For excellent reviews of the philosophical literature on causation, see Abbott 1998; Marini and Singer 1988; Winship and Sobel 2000.

[2] In this paper, we focus on this research question in a specific setting, rural Nepal. Although the question is considered generally in these introductory comments, we later use data on Nepal to examine the question empirically. Thus, for consistency, we focus our theoretical reasoning on the same setting, rural Nepal.

educated couples also have higher rates of sterilization. Thus parents' education is a confounder.

As is well known, confounding due to common correlates of sending a child to school and the timing of sterilization will lead to bias in the estimator of the effect of sending a child to school on the timing of sterilization. In order to reduce this bias, scientists control for all measured confounders (Bohrnstedt and Knoke 1982; Bollen 1989; Heckman and Robb 1985; Manski 1995; Manski and Nagin 1998; Winship and Morgan 1999; Winship and Sobel 2000). This bias is due to an unequal distribution of the confounders between the types of couples who send their children to school and those who do not; that is, the confounders such as parents' education are distributed differently in the two exposure groups (exposed and not exposed). The unequal distribution is called a *compositional difference*. When compositional differences exist in the exposure groups, the estimated coefficient of the effect of the exposure (sending a child to school) on the response (timing of sterilization) will reflect the differences between the types of couples who send their children to school and those who do not *in addition to* the causal effect that sending a child to school has on sterilization. In experimental settings compositional differences in the exposure groups are minimized by randomization of the subjects to different exposure levels; in observational studies we use statistical methods along with scientific assumptions to adjust for the compositional differences.

A standard 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). This provides an estimate of the exposure-response relationship within levels of the confounders. For example, George W. Bohrnstedt and David Knoke refer to "…clear-cut cases of spurious relationships which disappear when the appropriate common cause of both variables is held constant (1982, p. 73)."[3] For example, if we are interested in the relationship between sending children to school and the timing of sterilization in Nepal, we will

[3]For overviews of the methods typically used in social sciences to statistically control for confounding, see Winship and Morgan 1999; Sobel 2000; Winship and Sobel 2000.

include measures of all known factors that are related to both children's education and sterilization in Nepal in addition to the timing of sending children to school in the appropriate regression model for the timing of sterilization. Thus, we would control for parents' education, access to schools, and family size, in addition to variables indicating subpopulations of interest, such as mother's birth cohort, and racial/ethnic/religious group (Axinn and Barber 2001). Then, we compare the timing of sterilization among those who sent their children to school with the timing of sterilization among those who did *not* send their children to school, but only within similar groups of people in terms of these control variables.

Unfortunately, this approach leads to bias and, thus, mistaken conclusions if the confounders are themselves affected by the exposure (i.e. *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. For example, in Nepal, access to nearby schools is a confounder of the relationship between sending a child to school (exposure) and the timing of sterilization (response) because easy access to schools near the community increases the probability of sending a child to school and also encourages sterilization (Axinn and Barber 2001). In addition, the confounder is endogenous because the exposure influences subsequent values of the confounder – sending a child to a distant school (exposure) increases the probability that local residents will subsequently lobby the government for placement of a school near their community (confounder).

In Section 2, we discuss and illustrate how controlling for endogenous time-varying confounders by including them in the model of the response results in biased estimators. Consequently, this method does not address the desired question, "If more children in poor countries attend school, would more couples limit their total family size?" Robins, Hernán and Brumback (2000) propose a class of models called marginal structural models (MSMs), which use sample weights (inverse-probability-of-exposure weights) to control for the compositional differences due to endogenous time-varying confounders. In

Section 3, we describe this method for discrete-time survival analysis[4], as proposed by Hernán, Brumback and Robins (1999) and we provide an intuitive explanation of why this method eliminates bias. In Section 4, we discuss and contrast the assumptions underlying the weighting (MSM) method and the standard approach of controlling for endogenous confounding. Section 5 illustrates the suggested method using data from the Chitwan Valley Family Study in Nepal and the Intergenerational Panel Study of Parents and Children in the United States. Finally, in section 6, we critically examine the feasibility and usefulness of the weighting method with a simulation analysis. The simulations further illustrate the bias that results when we use the standard approach of controlling for endogenous time-varying confounders by including them in the model. Furthermore we find that the weighting method is quite robust to unmeasured confounding. In order to anticipate the biases resulting from inappropriately controlling for confounders, we derive and use a new path analysis sign rule.

2. A DRAWBACK OF THE STANDARD METHOD

Our goal is to assess the effect of sending a child to school (exposure, or *expos*) on the timing of sterilization (response, or *resp*). There are multiple confounders of the relationship between sending a child to school and the timing of sterilization – e.g., parents' education, access to nearby schools, and family size. Thus, following the standard practice of including all confounders as covariates in the analysis (which we call the "standard method"), we might fit the model:

$$\text{logit}\,(p_{ij}) = \beta_0 + \beta_1\, expos_{ij} + \beta_2\, subpop_j + \beta_3\, confounders_{ij} \qquad (1)$$

where p_{ij} is the hazard of sterilization by couple j during year i (given no prior sterilization), $\text{logit}(p_{ij}) = p_{ij}/(1-p_{ij})$, and $expos_{ij}$ is 1 if a child of couple j has ever attended school prior to year i and 0 otherwise. $Subpop_j$ is a vector of variables indicating subpopulations of interest to which couple j belongs, such as birth cohort or racial/ethnic/religious group and $confounders_{ij}$ is a vector of the confounder variables measured for couple j in year i. One possible confounder can be defined as 1 if a couple lives near a

[4] There is nothing intrinsic about the relevance of the weighting method to survival analysis; rather the response in our example, time until sterilization, requires a survival analysis method.

school in year i and 0 otherwise. Note that some of the confounders vary over time (i.e., whether the couple currently lives near a school), and some do not (i.e., whether the woman (or man) lived near a school during childhood). Finally, there are unmeasured factors that may be confounders of the relationship of interest. For example, political power of the leaders in couple j's neighborhood may influence the exposure (sending a child to school) as well as the response (timing of sterilization).

Only recently has the methodological community come to realize that the standard method of including confounders in the 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. To see this, consider the following figures and examples.

Figure 1 illustrates one possible underlying reality for the relationship between exposure and response.[5] In this example, *expos* represents a time-varying measure of whether a child has attended school; *resp* is a time-varying measure of whether the couple has been sterilized; *conf* represents a time-varying measure of whether the couple lives near a school, and *unmeas* is a vector of unmeasured, time-varying confounders; in this illustration, we discuss political power of the neighborhood leaders. The arrows represent causal paths. (Additional arrows could also be drawn in Figure 1 – 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 along with the direction of the arrows from exposure to response; that is, for illustrative purposes, we are assuming that the exposure is not causally related to the response (i.e., the total effect of exposure on response is 0). We use this to illustrate how the standard method of controlling for confounding (e.g., see the model in Equation 1), will lead to the appearance of a

[5]This figure represents the causal relationships between the variables, $unmeas_0$, $conf_0$, $expos_0$, $resp_0$, $unmeas_1$, $conf_1$, $expos_1$ and $resp_1$ (Pearl 1998). The absence of arrows implies conditional independence; for example, the absence of an arrow from $unmeas_1$ to $expos_1$ implies that $unmeas_1$ and $expos_1$ are independent, conditional on $conf_1$, $unmeas_0$, $conf_0$, $expos_0$, and $resp_0$. Subscripts refer to time period.

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, $conf_0$ is pictured to the left of $expos_0$ in Figure 1; this signifies that we are assuming $conf_0$ is not an outcome of $expos_0$.

(Figure 1, about here)

Notice, from Figure 1, that $conf_0$ and $conf_1$, are confounders because they are correlated with subsequent values of the exposure ($expos_0$ and $expos_1$) and the response ($resp_0$ and $resp_1$). The measured confounder *conf* is related to *expos* because living near a school makes it easier for a couple to send their child to school. Also, $conf_0$ is associated with sterilization because political power of the couple's neighborhood leaders is likely to influence their hazard of sterilization (e.g., via lobbying for access to sterilization services in clinics) as well as their access to schools. In order to control for this confounding, the standard method is to include *conf* as a covariate in the regression of *resp* on *expos*, in order to ultimately assess the effect of sending a child to school (*expos*) on the timing of sterilization (*resp*). From the figure, however, we see that $conf_1$ is endogenous (indicated by the arrow from $expos_0$ to $conf_1$), causing the following problem: if we condition on $conf_1$ (including it in our model), we create a spurious correlation between $expos_0$ and $resp_1$ through $conf_1$ and $unmeas_1$ via the paths A, B and C[6]. The spurious correlation occurs because $conf_1$ is endogenous (a consequence of $expos_0$).

To illustrate how including endogenous confounders in a regression model can cause bias, consider the following non-sociological example. We use a simple example in order to facilitate a clearer illustration of the problem than could be provided with a realistic, yet more complicated sociological example. In this example the bias in estimated causal effect is caused by gross spurious correlation, and is modeled after Pearl (1998). See Figure 2. The causal question is, "Does having the front yard sprinkler on cause the back yard grass to be wet?" Of course the answer should be no, but as we shall see, conditioning on an endogenous confounder leads to the answer yes. In this example, the

[6]This path, A to B to C, is not along the direction of the arrows; Pearl (1998) calls this a "backdoor" path.

exposure (analagous to sending a child to school) is whether the front yard sprinkler is on or off; the response (analagous to the timing of sterilization) is whether the back yard grass is wet or dry. The measured confounder (analagous to the presence of a school nearby) is whether the front yard grass is wet or dry. Assume for the purpose of this illustration that we do not know whether it is raining; thus, this is an unmeasured confounder. We use this example to intuitively explain the problem that can arise in the model depicted in Figure 1; thus, Figure 2 looks like Figure 1 except that it refers to the sprinkler example. Similar to Figure 1, time progresses from left to right, and variables measured within the same time period (e.g., 0 or 1) cannot affect another variable to the left (temporally preceding) in the figure.

(Figure 2, about here)

Logically, we know that whether the front yard sprinkler is on has no effect on whether the backyard grass is wet; in other words, the exposure has no effect on the response. This is illustrated in Figure 2 – there are no paths along the direction of the arrows from $expos_0$ to $resp_0$ or $resp_1$, or from $expos_1$ to $resp_1$. We also know that whether the front yard sprinkler is on ($expos_0$) subsequently affects whether the front yard grass is wet ($conf_1$) (arrow A); $conf_1$ is endogenous. In addition, we know that rain causes the front yard grass to be wet ($unmeas_0$ affects $conf_0$ and $unmeas_1$ affects $conf_1$), and if the front yard grass is wet, the front yard sprinkler will not turn on ($conf_0$ affects $expos_0$ and $conf_1$ affects $expos_1$). Finally, whether it is raining (*unmeas*) has direct effects on whether the front (*conf*) and back (*resp*) yard grass are wet (arrows B and C), but only affects whether the front yard sprinkler is on via whether the front yard grass is wet.

The spurious correlation arises as follows: Suppose that we know the front yard grass is wet at the beginning of time 1 ($conf_1 = 1$). (Knowing or conditioning on the value of $conf_1$ is analogous to including it as a covariate along with the exposure in a regression analysis.) If it is also the case that front yard sprinkler was not on during time 0 ($expos_0 = 0$) then we know that the back yard grass at the end of time 1 must be wet ($resp_1 = 1$). Thus conditional on wet front yard grass at time 1, it appears that

not turning on the front yard sprinker at time 0 causes the backyard grass at time 1 to be wet! This occurs because the combination of knowing the front yard sprinker is off and that the front yard grass is wet implies that it must be raining and thus the back yard grass must be wet at the end of time 1. In a regression model, this makes it appear that $expos_0$ causes $resp_1$. Thus, by including the endogenous confounder as a covariate in a regression analysis of the effect of the front yard sprinkler on the back yard grass, it can appear that the front yard sprinkler causes the back yard grass to be wet.

Thus, conditional on the value of the confounder, a spurious correlation from previous values of the exposure to present values of the response is created. In Figure 2, the spurious correlation is depicted by the paths from $expos_0$ to $resp_1$ via A, B, and C created when conditioning on the confounder[7]. The spurious correlations result in biased estimates of the causal effect, which in turn result in in false conclusions regarding the consequences of the exposure for the response, which in turn lead to inaccurate conclusions and knowledge.

3. A NEW METHOD: WEIGHTING

Hernán, Brumback, and Robins (1999) use a weighted survival analysis method, which they call marginal structural models (MSMs), that uses sample weights (inverse-probability-of-exposure weights) to statistically control for time-varying confounders. This method can be used to produce an unbiased estimator of the effect of exposure[8] on response. We refer again to our example of sending children to school and timing of sterilization in Nepal (see Figure 1). If the exposure had been randomized, then a model such as

---

[7]This path, A to B to C is not along the direction of the arrows; Pearl (1998) calls this a "backdoor" path.

[8] Note that the MSM/weighting method can also handle multiple exposures of interest. For example, the two exposures could be combined into a single variable; two binary exposures could be combined into one multivariate variable, or two continuous exposures could be combined into a variable with a bivariate normal distribution, or some other bivariate distribution.

$$\text{logit}\,(p_{ij}) = \beta_0 + \beta_1\, expos_{ij} + \beta_2\, subpop_j \qquad (2)$$

can be used to assess the effect of sending a child to school on the timing of sterilization. Note that randomization of the exposure results in all variables being distributed equally across exposure categories; β_1 represents an average effect of exposure (averaged across variables not included in the model). To use the weighting method, 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 couple's hazard of exposure (sending a child to school) in year i, among couples who are still at risk of exposure for the first time. In our example we use a logistic regression model because our exposure is dichotomous (other models such as a probit model could be used instead). In each year i, the weight uses a ratio of two conditional probabilities. The denominator is the conditional probability of a couple's measured *expos* status in year i, given that they did not previously send a child to school or get sterilized, variables indicating their current and past *conf* status (i.e., number of years lived near a school), and variables indicating subpopulation status (i.e., mother's birth cohort, racial/ethnic/religious group). The numerator is the conditional probability of a couple's measured *expos* status in year i, given that they did not previously send a child to school or get sterilized, and variables indicating subpopulation status. The weight at time t, w_t, is the product of the ratios up to time t (i=1,…,t). Equation 3 shows the form of the weight.

$$w_t = \prod_{i=1}^{t} \frac{P\left[expos_i \mid \overline{expos}_{i-1}, \overline{resp}_{i-1}, subpop\right]}{P\left[expos_i \mid \overline{expos}_{i-1}, \overline{resp}_{i-1}, \overline{conf}_{i-1}, subpop\right]} \qquad (3)$$

where subpop indicates status in multiple subpopulations, and the overbar indicates past history of the

variable, e.g. $\overline{expos}_{i-1}$ represents whether the couple has been previously exposed, and if so, the time since the prior exposure.

After computing the weights, we fit the model in Equation 2 using a weighted logistic regression. *Resp* is the response, *expos* is the predictor, and subpopulation indicators are additional covariates. If desired, some or all of the time-independent confounders can be included in the set of variables designating subpopulations. These time-independent confounders must be included in both the numerator and denominator models for the weights. We detail the estimation and computation of the weights with examples of SAS code in Appendix A.

3.2 Intuition Underlying the Use of the Weights

In this section, we provide an intuitive justification of the weights. (See Appendix B for a detailed technical discussion.) 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 groups. In sum, the sample weights equalize the confounder distribution across exposure categories, simulating the confounder pattern if an experiment was conducted that randomly assigned individuals to the exposure.

Intuitively, the weights eliminate the correlation of *conf* with *expos*, as represented pictorially in Figure 3. Now there is no correlation between *conf* and *expos*; that is, the path marked with an "X" is not present in the weighted sample.

(Figure 3, about here)

Because *conf* is no longer a confounder in the weighted sample, we do not need to control or equivalently include *conf* as a covariate in our (weighted) survival analysis model for *resp*; that is, we

may fit the model in Equation 2. By not including *conf* 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 *conf* in the model, and thus a false correlation between *expos* and *resp* is avoided. This can be seen in Figure 4.

(Figure 4, about here)

As stated above, the weighting method works by equalizing the composition of couples with varying school locations (and other confounders) among the two groups of couples – those with the exposure (a child has attended school) and those with no exposure. To see this, consider the following intuitive scenario. There are 110 couples with the following two characteristics prior to year i: (a) they have not yet sent a child to school and (b) neither partner has been sterilized. Thus, they are still at risk of the exposure and the response. Table 1 provides a hypothetical cross-tabulation of two variables in year i: whether the couple lives near a school (*conf*) and whether a child of the couple has attended school (*expos*). Note that of the total of 60 couples where *conf* = 1 (row 1), 30 couples have *expos* = 1 (column 1, row 1), whereas of the total of 50 couples where *conf* = 0 (row 2), only 10 couples have *expos* = 1 (column 1, row 2). This means that couples who do not live near a school are underrepresented among couples with the exposure (a child who has attended school) and couples who *do* live near a school are overrepresented among couples with the exposure.

Table 1 Crosstabulation of *expos* by *conf* for year i

		(1)	(2)	(3)
		***expos* = 1** Child has attended school	***expos* = 0** Child has *not* attended school	Total
(1)	***conf* = 1** Couple lives near a school	30	30	60
(2)	***conf* = 0** Couple does *not* live near a school	10	40	50
(3)	Total	40	70	110

In Table 1, among those couples who live near a school (row 1), ½ of the couples have the exposure (a child who has attended school), and ½ of the couples do *not* have the exposure. If this proportion also held true for those couples who do *not* live near a school (row 2), then the table would be as in Table 2.

Table 2. Hypothetical crosstabulation of *expos* by *conf* for year i, if p(*expos* = 1 | *conf* = 1) = p(*expos* = 1 | *conf* = 0)

		(1)	(2)	(3)
		***expos* = 1** Child has attended school	***expos* = 0** Child has *not* attended school	Total
(1)	***conf* = 1** Couple lives near a school	30	30	60
(2)	***conf* = 0** Couple does *not* live near a school	25	25	50
(3)	Total	55	55	110

Here we have divided the couples who do *not* live near a school so that the same proportions are in columns 1 and 2 as the row 1 proportions. This also results in a proportion of couples with *conf* = 0 among the total couples with the exposure (column 1) that is equal to the proportion of couples with

conf = 0 among the total couples without the exposure (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 couple with the inverse of the conditional probability of *expos* status given *conf* status. Referring to Table 1, a weight of $(10/50)^{-1} = 5$ is assigned to the 10 couples with *expos* = 1, *conf* = 0. The 40 couples in the more common group with *expos* = 0 and *conf* = 0 are assigned a smaller weight of $(40/50)^{-1} = 5/4$. Because there is an equal number of couples with *expos* = 1 and *expos* = 0 in the total group where *conf* = 1, each of these 60 couples is assigned an equal weight of $(30/60)^{-1} = 2$. After weighting the observations for each couple, the frequencies from Table 1 become the weighted frequencies in Table 3.

Table 3. Weighted crosstabulation of *expos* by *conf* for year i

		(1)	(2)	(3)
		***expos* = 1** Child has attended school	***expos* = 0** Child has *not* attended school	Total
(1)	***conf* = 1** Couple lives near a school	60	60	120
(2)	***conf* = 0** Couple does *not* live near a school	50	50	100
(3)	Total	110	110	220

Note that in Table 3, the representation of couples with *conf* = 0 among the couples with the exposure is equal to the representation of couples with *conf* = 0 among the couples without the exposure; both proportions are 50/110=25/55. The only difference between Table 2 and 3 is that all cell sizes in Table 3 are double those in Table 2. In practice, weights are assigned as the ratio of the probability of *expos* status divided by the conditional probability of *expos* status given *conf* status; this eliminates the

elevation of the total sample size (see Equation 3). Additionally, Equation 3 shows that the probability of *expos* status is conditional on all available confounders and all subpopulation indicators, not just *conf* (whether the couple lives near a school, in our example). Thus the weights eliminate the correlation between *conf* and *expos* in the weighted sample; that is, the path marked with an "X" in Figure 3 is not present in the weighted sample.

4. A COMPARISON OF ASSUMPTIONS

Both the standard method and the weighting method make a variety of assumptions in order to estimate causal effects. First, both methods 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 $unmeas_0$ or $unmeas_1$ to either $expos_0$ or $expos_1$. That is, there are no direct arrows from the unmeasured factors (here, political power of the neighborhood leaders) to the exposure (a child has attended school). This means that political power of the neighborhood leaders may have brought a school to the area, but it does not directly affect whether any specific couple sends their child to school beyond making it easier to do so because of the new school. In Figure 2, this means that whether the front yard sprinkler is "on" is only affected by whether the grass is wet; it does not respond to the rain directly. This assumption is sometimes called sequential ignorability (Robins, 1997)[9] and can be stated precisely using the potential outcomes model of Rubin (1978) (see Robins 1999a for an nice discussion) or by causal diagrams (see Greenland, Pearl, and Robins 1999; Pearl and Robins 1995).

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 the couple does not live near a school ($conf_0 = 0$), it is still possible that a child of the couple

[9]Robins also uses the term “no unmeasured confounders” instead of sequential ignorability. We use "sequential ignorability" to avoid confusion. The phrase “no unmeasured confounders” might be better stated as “no unmeasured direct confounders.”

has attended school ($expos_0 = 1$) (in this case a distant school) and vice-versa (Robins, 1999b). We call this weighting assumption #2. The standard method does not require this assumption; however, if this situation does not hold, then the standard method extrapolates from the possible confounder × exposure patterns to the impossible confounder × exposure patterns. Thus the standard method replaces weighting assumption #2 by standard 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 the standard method assumes that the model we are using for the standard method holds for these confounder × exposure patterns. Thus, standard method assumption #2 is extrapolation – the standard method extrapolates 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 method.

The standard method makes one additional assumption: there are no unmeasured confounders affecting measured endogenous confounders – if the arrows from $unmeas_1$ to $conf_1$ OR from $unmeas_0$ to $conf_1$ were absent, this assumption would hold. The spurious correlation between $expos_0$ and $resp_1$ through $conf_1$ and $unmeas_1$ discussed above in section 2 results from a violation of the second assumption. Note that 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 weighting assumption #2), and compare the weighting method to the standard method when the first two assumptions do not hold.

5. EMPIRICAL EXAMPLES

We now illustrate the method described in section 3 using survey data. We address two sociological research questions: (1) If more children in Nepal attend school, would more couples limit their total family size via sterilization? and (2) Does low educational attainment in the United States lead to early first births? For each research question, 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) the standard method, 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 arising in the three methods and in the simulations. Thus we first briefly explain the path analysis rules we use.

5.1 Path Analysis Rules

Using the ideas in Davis (1985) we associate signs (+/-) with each path in a figure. We wish to ascertain the sign of the correlation between *expos* and *resp*, 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 *expos* to *resp*. Paths without converging arrows and for which no variable along the path is in the model contribute to the correlation between *expos* and *resp*. For each such path, we multiply the signs along the path. We then sum over all such paths to find the sign of the correlation between *expos* and *resp*. (If the paths are of different signs the result is ambiguous). Paths with a variable at which there are converging arrows (Pearl {1998} calls these variables "colliders"), do not contribute to the above sum as long as the variable is not in the model. For example in Figure 2, $conf_1$ is a collider on the path A to B to C. Furthermore if you control for (include in the model) a noncollider along a particular path between *expos* and *resp* then the path is "blocked" and this path does not contribute to the above sum.

Suppose there is a path between *expos* and *resp* containing a collider that is included in the model. We do not know of a path analysis rule for incorporating the consequences of including a collider in the model; we derive a new rule by following Pearl's (1998) non-parametric d-separation rule. This rule notes that if a collider is included in the model, paths from *expos* to *resp* that include the collider contribute to the sum determining the sign of the correlation between *expos* and *resp*. To calculate the sign for a path with a collider, multiply the signs of the coefficients along the path as

before (ignoring that the variable is a collider), but then multiply by -1 (see Appendix C for the derivation in a Gaussian version). The rationale for multiplying by -1 can be seen in Figure 2. First, the correlation represented by paths A and B must be positive (i.e., turning the front yard sprinkler on causes the front yard grass to be wet and rain causes the grass to be wet). Yet the partial correlation between whether the front yard sprinkler is on at time 0 and whether it is raining at time 1, given that the front yard grass is wet at time 1, must be negative. That is, given that the front yard grass is wet, it is less likely to be raining if the front yard sprinkler was on than if the front yard sprinkler was off. These rules are sufficient for the diagrams in Figures 1 through 6; rules for deducing independencies in more complicated situations are provided by Pearl (2000).

5.2 Education and Fertility Limitation

To examine the first research question, we use data from the Chitwan Valley Family Study. The purpose of the study was to collect detailed information about historical social changes in the neighborhoods in the Chitwan Valley, and to analyze how those social changes are related to individual level behavioral changes in marriage and childbearing behavior. Thus, the data were collected expressly to analyze the research question on which we focus here. Data were collected from a representative sample of 171 neighborhoods in the valley, located in central Nepal. Each adult residing in the 171 neighborhoods was interviewed (n = 5,271), with a 97% response rate. The study collected retrospective histories of change in each neighborhood using the Neighborhood History Calendar method (Axinn, Barber, and Ghimire 1997), and retrospective histories of each individual's behavior using a life history calendar adapted specifically to the setting (Axinn, Pearce, and Ghimire 1999). We focus on the relationship between sending a child to school and a couple's subsequent hazard of sterilization.[10] Briefly, our hypothesis is that couples who send a child to school become aware of the

[10] We use the word *sterilization* as a shorthand for all permanent contraceptive use to stop childbearing. Similar to previous research, we consider a woman's own sterilization, her husband's sterilization, IUD, Norplant, and Depo-provera to be permanent methods in this setting (Axinn and

costs of doing so, and subsequently decide to terminate childbearing via permanent contraceptive use (Axinn and Barber 2001). Because sterilization among couples who have no children is extremely rare in this setting, we consider couples to be at risk of sterilization only after they have given birth to their first child. In addition, we consider couples where the wife is age 55 or older to no longer be at risk of sterilization. Finally, because we are focusing on the experience of sending a child to school, we examine only those couples who are at risk of having sent a child to school – in other words, only those couples with at least one child who is old enough to attend school. Thus, our analysis includes couples where the wife is age 54 or younger, and whose oldest child is at least age 6. Our hypothesis is that sending a child to school will motivate couples to become sterilized sooner than their peers who have never sent a child to school.[11]

There are at least two important measured variables that are likely confounders of this 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.[12]

Barber 2001). Our data show that among married women age 25-54 with at least one child, 98.77% who had ever used any of these contraceptive methods said that they wanted no more children.

[11] Much has been written about the effects of education on contraceptive use, and we refer readers to this literature rather than elaborating on this hypothesis here (e.g., see Axinn 1993; Axinn and Barber 2001).

[12] Using a combination of ethnographic, survey, and archival methods, each of the 147 schools in the study area was categorized as resulting from one (or a combination) of the following: (1) community-based lobbying and fund-raising; (2) government allocation; or (3) privately funded and built. In all, 76% of the schools were categorized as resulting from community-based action, 25% as

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 Figures 1 and 3, with an additional confounder variable, and permitting a causal effect of exposure on response.

In addition, there are multiple other time-invariant confounders that are likely to influence both the exposure (sending a child to school) and the response (sterilization), as well as multiple indicators of subpopulations of interest (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 first child turned 6. These measures are consistent with analyses by Axinn and Barber (2001). We present descriptive statistics for these measures in Table 4. We have chosen to include these time-invariant confounders and subpopulation indicators 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 4, about here)

Table 5 compares the naive, standard, 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 of those

---

resulting from a government allocation, and 21% as resulting from private individuals' efforts. (Many schools were included in more than one category.)

who sent a child to school probably did so because they live near a school or because they have relatively few children.

(Table 5, about here)

Thus, we next estimate parameters using the standard method, 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 of the estimate is reduced to .74, a slightly smaller effect. 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 standard 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 sequential ignorability and weighting assumption #2 hold, the similarity across columns in Table 5 suggests two possibilities. First, it may be that there is not a lot of unmeasured indirect confounding in this example. In other words, the correlations along the paths B and C illustrated in Figure 3 are very small or nonexistent. This may be 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 3). 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 the standard model, we provide some interpretation of this change in order to illustrate the difference in greater detail. This is not, however, to suggest that the difference between these models is substantial. The slight decrease in the magnitude of the estimate across the columns in Table 5 suggests that there is a small path from $expos_0$ to $resp_1$ via A, B, and C (shown in Figures 2 and 3). This means, because A and B arrows go in to the box $conf_1$ (and thus we multiply our signs by a negative sign), two of the A, B, C, arrows must be positive and one must be 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, the path A is likely to be positive. Fecundity is an unmeasured (and perhaps unmeasurable) potential confounder that demographers often consider. It is possible that fecundity would be positively correlated with having a smaller family, but negatively related to sterilization.

5.3 Educational Attainment and First Birth Timing

To examine the second research question, we use data from the Intergenerational Panel Study of Parents and Children (IPS), an eight-wave, 31-year panel study. The original sample, which was drawn from1961 birth records in the Detroit metropolitan area, included married white mothers who had recently given birth to a first, second, or fourth child. The mothers were interviewed shortly after the focal child's birth in 1961, and at multiple points afterwards. The focal children were interviewed at ages 18 (1980), 23 (1985), and 31 (1993). The response rate for the 1962 survey was 92 percent (n = 1,113 families); approximately 85 percent of those families responded to the 1993 survey (n = 882).

Education, work, cohabitation, marriage, and childbearing histories from age 15 to age 31 were obtained from the focal children at ages 23 and 31 via the life history calendar (Freedman et al. 1988). We use these histories to construct the dependent variable for the analyses, the timing of the child's first birth[13], and measures of early adult experiences with school, work, and marriage. Although the unit of analysis is the focal children, we use questions asked of the mothers in the early waves of the study to construct other variables described below.

We focus on the relationship between dropping out of school for the first time (for six months or more) and the hazard of first birth. The period of risk begins at age 15, when all of the respondents are still enrolled in school. Our hypothesis is that dropping out of school will cause individuals to have a first birth sooner than their peers who have not dropped out of school, or who delay dropping out of school.[14]

There are multiple important factors measured in these data that are likely confounders of this relationship: educational attainment (separate from enrollment), cohabitation, and marriage. Each of these variables varies over the risk period. And, each is potentially endogenous to the relationship between dropping out of school and first birth timing. For example, dropping out of school increases the probability of entering a cohabiting or marital relationship; the converse is also true: entering a cohabiting or marital relationship while still in school increases the probability of dropping out of school (Thornton, Axinn, and Teachman 1995).

In addition, there are multiple other time-invariant confounders that are likely to influence both the exposure (dropping out of school) and the response (first birth timing), as well as indicators of

---

[13] We code the timing of the first birth as the month the birth occurred minus nine months. This times the hazard to conception rather than birth, and limits our analysis to those pregnancies that resulted in a live birth. For further discussion, see Barber (2000).

[14] Much has been written about the reciprocal relationship between education and first birth timing; we refer readers to this vast literature rather than elaborating on this hypothesis here. For a good review of this literature, see Hoffman (1998) and Hofferth, Reid, and Mott (2001).

subpopulations that may be correlated with the exposure and the response (e.g., see Barber 2000, 2001). These factors include whether the respondent ever "went steady", the mother's completed family size, the family's financial assets, whether the mother worked outside the home, whether the mother was Catholic, the mother's experiences with divorce and remarriage, and respondent's age and gender. We present descriptive statistics for these measures in Table 6. We have chosen to include these time-invariant confounders and subpopulation indicators 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).

(Table 6, about here)

Table 7 compares the naive, standard, and weighting methods using these data. In column 1, using the naive method yields an estimate of the effect of dropping out of school on first birth timing of 1.46. This means that the monthly log-odds of having a first birth is 1.46 higher among those who have dropped out of school (for at least six months) compared to their peers who are still enrolled in school. 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.

(Table 7, about here)

Thus, we next estimate parameters using the standard method, 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 dropping out of school on the hazard of first birth. The magnitude of the estimate is reduced to 1.20, still a substantial effect.

Column 3 presents the estimates 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 model, the effect of dropping out of school on the hazard of first birth is slightly larger (1.51), although

not statistically different from the estimates in columns 1 and 2.[15]

Although the estimate from the weighted model is again not dramatically different from the estimate given by the standard model, we provide some interpretation for illustrative purposes, but do not mean to suggest that the difference is substantial. Because the standard method slightly underestimates the relationship between dropping out of school and first timing relative to the weighting method, we know that the path composed of arrow A, B, and C must exert an overall negative influence on the response. Because our confounders are "colliders" (two arrows go in to the box), we multiply the product of the signs in the path by a negative sign. Thus, either (1) all of the A, B, C paths are positive, or (2) two of the separate paths must be negative and one must be positive to get an overall positive sign. From theory and previous research, we know that dropping out of school dramatically increases both marriage and cohabitation rates (Thornton, Axinn, and Teachman 1995). Thus, A represents a positive correlation. This means that both B and C must be either positive or negative. This is reasonable because most factors that are strongly related to marriage have a similar effect on first birth timing; for example, a desire for many children speeds both the entry into marriage and the timing of first birth.

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

[15] We originally estimated the models in column 2 including work status (i.e., full-time, part-time, or none). Including this variable caused problems in our weighted analysis because of the high correlation between working full-time and dropping out of school, which violates weighting method assumption #2 (no past confounder patterns exclude particular levels of exposure). Fortunately, work status is not a true confounder of the relationship between dropping out of school and first birth timing, because it is not significantly related to first birth timing, net of the other variables included in our model. Thus, we subsequently dropped it from the model in column 2 and from the model of dropping out of school used to create the weights. This experience illustrates the importance of investigating the relationship between potential confounders and the exposure.

relationships. This strategy is similar to that employed by Stolzenberg and Relles (1990) in their statistical evaluation of Heckman's sampling bias correction.

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) the standard method, 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, utilizing sample weights (inverse-probability-of-exposure weights) to statistically control for time-varying confounders.

For ease in interpretation, and to make the simulation more concrete and easily understandable, we assign a substantive meaning to each of the variables (modeled after the Chitwan Valley Family Study example described above). These variables are illustrated in Figure 5. The exposure is whether a child has attended school; the response is the timing of sterilization. The two time-varying confounders are whether the couple lives near a school and family size (operationalized as having a small vs. a large family). Again, there are typically many unmeasured confounders (some of them time-varying) that may influence both the exposure and the outcome – for example, aspirations for children, genetic propensity for childbearing, fecundity, cognitive ability, or political power of the neighborhood leaders to bring a school to the neighborhood. We use political power in this simulation to denote the unmeasured confounder.

We have 5 variables. Subscripts refer to the time period under observation. There are two time periods in the simulated data: 0 and 1.

- *unmeas* = 2 if the couple lives in a neighborhood where the leaders have a high level of political power, 1 if the neighborhood leaders have medium-level political power, and 0 if the neighborhood leaders have low political power. This is an unmeasured indirect confounder.
- $conf1_i$ = 1 if the couple lives near a school, 0 otherwise (i = 0,1). This is a measured confounder.

- $conf2_i$ = 1 if family size is small (1 or 2 children), 0 if family size is not small (3 or more children) (i = 0,1). This is a measured confounder.
- $expos_i$ = 1 if any child of the couple has attended school by time i, 0 otherwise (i = 0,1). This is the exposure.
- $resp_i$ = 1 if sterilization is initiated at time i, 0 otherwise (i = 0,1). This is the response.

(Figure 5, about here)

We generated this data according to Figure 5, using the equations listed in Table 8. Note that in Figure 5, similar to our previously described examples, there is no effect of sending a child to school ($expos_i$) on the timing of sterilization ($resp_i$). In other words, the data is generated so that the exposure is *not* causing the response. Rather, *unmeas* (the unmeasured confounder – political power) is indirectly related to $expos_i$ (the exposure – a child attending school), and is also directly related to $resp_i$ (the response – timing of sterilization). In fact, in our simulation, *unmeas* (the unmeasured confounder – political power) completely determines $resp_i$ (the response – timing of sterilization). Note that there is no direct arrow in Figure 5 from *unmeas* to $expos_0$ or $expos_1$; 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; thus in this section 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 8 provides detailed information concerning the design of the simulation. All of the simulations are based on 1,000 simulated data sets of 1,000 observations each.

(Table 8, about here)

6.2 Results of the Simulation

6.2.1. Summary

In each of our simulations, we compare the weighting method to the standard method (regular logistic regression with confounders as covariates) 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. Our use of path analysis rules indicates that when the time-varying confounder is correlated with past and future exposure and is also correlated with the response, the standard 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 the standard method 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 the standard method 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 the standard method. A detailed discussion of the simulations follows.

6.2.2 A Nonparametric Analysis

Table 9 presents an analysis of simulated data according to the model described by Equations

(4a) and (4b):

$$logit\ (resp_0) = \beta_0 + \beta_1(expos_0) + \delta_1\,(conf1_0) + \delta_2\,(conf2_0) \quad (4a)$$

$$logit\ (resp_1) = \beta_0 + \beta_4 + \beta_2\,(expos_0) + \beta_3\,(expos_1) + \delta_1\,(conf1_1) + \delta_2\,(conf2_1) + \delta_3\,(conf1_0) + \delta_4\,(conf2_0) \quad (4b)$$

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 *resp* because we have a parameter for all possible exposure-response combinations:

Proportion of $resp_0 = 1$:

$expos_0 = 0$	$expos_0 = 1$
$expit(\beta_0)$	$expit(\beta_0 + \beta_1)$

Proportion of $resp_1 = 1$:

$expos_0 = 0, expos_1 = 0$	$expos_0 = 0, expos_1 = 1$	$expos_0 = 1, expos_1 = 1$
$expit(\beta_0 + \beta_4)$	$expit(\beta_0 + \beta_4 + \beta_3)$	$expit(\beta_0 + \beta_4 + \beta_2 + \beta_3)$

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 $expit(\beta_0 + \beta_4 + \beta_2 + \beta_3)$ will give you the proportion of people who were

exposed at both time 0 and time 1 who are sterilized at time 1.

The standard method includes all known confounders as covariates in the model[16], and the proportions can be computed as follows:

Proportion of $resp_0 = 1$:

$expos_0 = 0$	$expos_0 = 1$
$\text{expit}(\beta_0 + \delta_1 conf1_0 + \delta_2 conf2_0)$	$\text{expit}(\beta_0 + \beta_1 + \delta_1 conf1_0 + \delta_2 conf2_0)$

Proportion of $resp_1 = 1$:

$expos_0 = 0$, $expos_1 = 0$	$expos_0 = 0$, $expos_1 = 1$	$expos_0 = 1$, $expos_1 = 1$
$\text{expit}(\beta_0 + \beta_4 + \delta_1 conf1_1 + \delta_2 conf2_1 + \delta_3 (conf1_0) + \delta_4 (conf2_0))$	$\text{expit}(\beta_0 + \beta_4 + \beta_3 + \delta_1 conf1_1 + \delta_2 conf2_1 + \delta_3 (conf1_0) + \delta_4 (conf2_0))$	$\text{expit}(\beta_0 + \beta_4 + \beta_2 + \beta_3 + \delta_1 conf1_1 + \delta_2 conf2_1 + \delta_3 (conf1_0) + \delta_4 (conf2_0))$

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 $expos_i$ (sending a child to school) and $resp_i$ (timing of sterilization).

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 9, columns 1 through 3 present the summaries of the estimates using the naive method.

(Table 9, about here)

[16] 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.

For example, we expect that β_1 (estimator of the effect of $expos_0$ on $resp_0$) will be a biased estimator of a zero effect because $conf1_0$, $conf2_0$ (measured confounders at time 0) are common correlates of $expos_0$ (whether any child has attended school at time 0) as well as $resp_0$ (ever sterilized by time 0). Using path analysis rules, $\hat{\beta}_1$ should be significantly positive. In a linear model with standardized covariates, $\hat{\beta}_1$ would be approximately $\gamma 1_0 * \alpha 1_0 + \gamma 2_0 * \alpha 2_0$, which is positive. 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 sending a child to school on sterilization (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).

Standard Method. Columns 4 through 6 of Table 9 present estimates from the standard method, which includes all known confounders as covariates in the model.

In the standard method, we expect from path analysis rules that β_1 (estimator of the effect of $expos_0$ on $resp_0$) will be unbiased because we have included the measured confounders ($conf1_0$, $conf2_0$). (Including the confounders in model using the standard method blocks the path from $expos_0$ to $resp_0$ via measured confounders at time 0 ($conf1_0$, $conf2_0$, and *unmeas*).) 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, the standard method produces a well performing estimator $\hat{\beta}_1$. Similarly, using path analysis rules, we expect β_3 (estimator of the effect of $expos_1$ on $resp_1$) to be unbiased because we have included the measured confounders ($conf1_1$, $conf2_1$). 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, β_3 is unbiased.

However, there is a problem in the estimation of β_2 (estimator of the effect of $expos_0$ on $respos_1$). Because we have included the confounders in the model for the standard method, we might expect β_2 to be unbiased. In other words, we might expect $\hat{\beta}_2$ to be approximately 0 because we created these simulation data so that $expos_0$ does not cause $resp_1$. However, because the standard method includes $conf1_1$ and $conf2_1$ in the model and these are colliders on the path from $expos_0$ to ($conf1_1$ and $conf2_1$) to *unmeas* to $respos_1$, the new path analysis rule (as explained in section 4) implies that the estimated effect of $expos_0$ (ever sent a child to school by time 0) on $resp_1$ (timing of sterilization) will be negatively biased ($\alpha 1_1$, $\alpha 2_1$, η_1, η_2 are all positive). Indeed in a linear model the bias would be $-\alpha 1_1 * \eta_1 / (\alpha 1_1^2 + 1) - \alpha 2_1 * \eta_2 / (\alpha 2_1^2 + 1)$, which is negative (see Appendix C for this derivation). 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 9 present summaries of the estimates from use of 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 β_1 (estimator of the effect of $expos_0$ on $resp_0$) will be unbiased because we have adjusted for the confounders ($conf1_0$, $conf2_0$) in the weights; column 7 confirms this. We expect that β_2 (estimator of the effect of $expos_0$ on $resp_1$) will also be unbiased for two reasons. First, we have adjusted for the confounders ($conf1_0$, $conf2_0$) in the weights. Second, we have not included outcomes of $expos_0$ (child attended school by time 0) (i.e. $conf1_1$, $conf2_1$) 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. β_3 is also unbiased (see column 9).

Note that if we included confounders as covariates in the model to be used in the weighting method, this would lead to bias in the estimator of β_2 (estimator of the effect of $expos_0$ on $resp_1$). This is because including the confounders at time 1 opens up two new paths via the confounders, $conf1_1$, $conf2_1$ (the A, B, C paths) between $expos_0$ and $resp_1$.

Summary. In sum, we find that the naive method leads to biased estimators of β_1, β_2, and β_3. The standard method leads to biased estimators of β_2 only. And, the weighting method leads to unbiased 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 $expos_0$ on $resp_1$ (i.e., in the parsimonious model we set β_2=0). This model is analogous to those typically estimated by many scientists. We rarely fit nonparametric models because there are usually more than two observation periods in the data set, which means the nonparametric model requires a prohibitive number of parameters. 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 $expos_0$ on $resp_{0;}$ and β_3, the effect of $expos_1$ on $resp_1$. If we had more than two observation periods, we would also include a variable summarizing past exposure, such as time since first exposure.

(Table 10, about here)

Similar to the nonparametric model estimated using the naive method (columns 1 through 3 in Table 9), columns 1 and 2 of Table 10 show that the parsimonious model estimated using the naive method also leads to bias in both β_1 and β_3.

Recall that in the *nonparametric* model estimated using the standard method (column 6 in Table 9), β_3 was unbiased. However, following the path analysis rules, we expect β_3 in the *parsimonious* model estimated using the standard method to be negatively biased. Indeed, column 4 of Table 10 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 β_3 (the estimator of the effect of $expos_1$ on $resp_1$) should be unbiased in the weighted analysis. Column 6 of Table 10 confirms that β_3 is unbiased.

In sum, the naive method produces biased estimators of β_1 and β_3. Consistent with the intuitive description of the endogeniety problem with the standard method (discussed in section 2 and illustrated in Figure 2), the standard method produces a biased estimator of β_3, the effect of $expos_1$ on $resp_1$. 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 degree of bias for the biased estimators in Table 9. 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 11 presents simulations with smaller values for the α's and the η's compared to Table 9 (refer to Figure 6 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 $\eta_1 = \eta_2 = 1.0$; these values are smaller than in Table 9. In panel B the values are even smaller relative to those in Table 9: $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = .5$ and $\eta_1 = \eta_2 = .5$. Specifically, the correlation between the unmeasured confounder (*unmeas*) and the confounders ($conf1_1$ and $conf2_1$) is smaller than in Table 9, and the correlation between exposure ($expos_0$) and subsequent levels of the confounders ($conf1_1$ and $conf2_1$) is smaller than in Table 9. Thus, we expect that the overall magnitude of the relationship between $expos_0$ and $resp_1$ via these paths (similar to arrows A, B, and C in Figure 3) will be smaller. In other words, we expect the degree of bias to be smaller in Table 11 panel A than in Table 9, and to be still smaller in Table 11 panel B.

(Table 11, 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 biased estimators. Using both the standard and the naive methods, *all* of the estimates decrease in magnitude (i.e., become closer to zero) from Table 9 to Table 11 panel A to Table 11 panel B. Additionally, the proportion of the data sets where we would reach an incorrect conclusion evaluating the significance of β_2 via the standard method decreases from 80% to 53% to 10%.[17] In contrast, in the weighting method we would reach an incorrect conclusion about 5% of the time for all parameters. This suggests that 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) violates weighting method assumption #2, then we will have 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; thus, usually some direct confounders are excluded from the model. We refer to this as "partially weighted" to emphasize that only some of the confounders are included in the weights.

In Table 12, the data were simulated so that $conf1_i$ is more strongly related to $resp_i$ than $conf2_i$

[17] We found this to be a general pattern, as well (not shown in tables): as the α and η values decrease, the bias decreases.

($\alpha 1_i > \alpha 2_i$). In other words, $conf1_i$ is a stronger predictor of the response than $conf2_i$. As a shorthand, we refer to $conf1_i$ 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 $conf1_i$ (the more important confounder) but not $conf2_i$, the bias will be smaller compared to when the weights adjust only for $conf2_i$ but not $conf1_i$. We also expect that when weights adjust only for $conf2_i$ (the less important confounder), the bias will be smaller than the bias produced by the naive method that does not adjust for confounding at all.

(Table 12, about here)

Table 12 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).[18] Also note that the same results hold for the standard method, but only for $\hat{\beta}_1$ and $\hat{\beta}_3$; as we adjust for more confounders with the standard method, 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 the standard method (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

[18] 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 *conf2* is not really a confounder, and is included in the formulation of the weights along with *conf1*, the weighting method provided estimators and standard errors no worse, and in many cases better, than the standard method.

the data for table 13, γ_0=-15.0; $\gamma 1_0$=$\gamma 1_1$=$\gamma 2_0$=$\gamma 2_1$=8.0; thus, the probability of sending a child to school by time 0 ($expos_0$ = 1) when there is no school nearby and one has a large family ($conf1_0$ = $conf2_0$ = 0) is 3.05 *10^{-7} (see Table 8 and Figure 6). 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.

(Table 13, about here)

The simulations confirm our expectations. As illustrated in Table 13, estimators from the weighting method are biased and results are worse than those from the standard method. 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 sending a child to school when (a) there is no nearby school and (b) the family is large, 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).[19]

7. CONCLUSIONS

In this paper, we have intuitively explained why including endogenous time-varying confounders in the analysis model – this is the standard 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

[19] Recall that we encountered a similar situation in our second substantive example in section 4. The probability of dropping out of school (exposure) while working full-time (potential confounder) was very high. However, this situation did not lead to strange weights, nor did we have trouble estimating the weights or the weighted logistic regression. This highlights the need for further research on how we might detect violations of weighting method assumption #2.

following: fitting nonparametric models with the naive method (ignoring time-varying confounding), all coefficients are biased. Fitting nonparametric models using the standard method, 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 model 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 intend to explore this possibility in the future. 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.

An important and necessary generalization of this method would be to multilevel data. In fact, in the education and fertility examples, 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 "new" statistical methods, 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 the standard method 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 the standard method 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

Abbott, Andrew. 1998. “The Causal Devolution.” Sociological Methods and Research 27(2):148-81.

Axinn, William G. 1993. “The Effects of Children's Schooling on Fertility Limitation.” Population Studies 47:481-493.

Axinn, William G., Jennifer S. Barber and Dirgha Ghimire. 1997. “The Neighborhood History Calendar: A Data Collection Method Designed for Dynamic Multilevel Modeling.” Sociological Methodology 27: 355-92.

Axinn, William G., Lisa D. Pearce, and Dirgha Ghimire. 1999. “Innovations in Life History Calendar Applications.” Social Science Research 28: 243-64.

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

Barber, Jennifer S. 2000. "Intergenerational Influences on the Entry into Parenthood: Mothers’ Preferences for Family and Nonfamily Behavior." Social Forces 79(1):319-348.

Barber, Jennifer S. 2001. "Ideational Influences on the Transition to Parenthood: Attitudes Toward Childbearing and Competing Alternatives." Social Psychology Quarterly 64(2):101-127.

Blalock, Herbert. 1991. "Are There Really any Constructive Alternatives to Causal Modeling?" Sociological Methodology 21:325-336.

Bohrnstedt, George W. and David Knoke (1982). Statistics for Social Data Analysis. F. E. Peacock Publishers, Inc. Itasca, IL.

Bollen, Kenneth A. (1989). Structural Equations with Latent Variables. New York: John Wiley & Sons.

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.

Freedman, David A. 1991. "Statistical Models and Shoe Leather." Sociological Methodology 21:291-314.

Freedman, Deborah, Arland Thornton, Donald Camburn, Duane Alwin, and Linda Young-DeMarco. 1988. "The Life History Calendar: A Technique for Collecting Retrospective Data." Sociological Methodology 18:37-68.

Greenland, S., Pearl, J., Robins, J.M. (1999). "Causal Diagrams for Epidemiologic Research." Epidemiology 10(1):37-48.

Heckman, J. J. and R. Robb. 1985. “Alternative Methods for Evaluating the Impact of Interventions.” In Longitudinal Analysis of Labor Market Data, eds. J. J. Heckman and B. Singer. Cambridge, U.K.:

Cambridge University Press, pp. 156-245.

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.

Hofferth, Sandra L., L. Reid, and Frank L. Mott. 2001. "The Effects of Early Childbearing on Schooling over Time." Family Planning Perspectives 33:259-267.

Hoffman, Saul D. 1998. "Teenage Childbearing Is Not So Bad After All... or Is It? A Review of the New Literature." Family Planning Perspectives 30:236-40.

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

Manski, C. F. 1995. Identification Problems in the Social Sciences. Cambridge, MA: Harvard University Press.

Manski C. F. and D. S. Nagin. 1998. "Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism." Sociological Methodology 28: 99-137.

Marini, Margaret M. and B. Singer. 1988. "Causality in the Social Sciences." Sociological Methodology 18:347-409.

Mason, William M. 1991. "Freedman is Right as Far as He Goes, But There is More, And It's Worse. Statisticians Could Help." Sociological Methodology 21:337-352.

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.

Pearl, J. and J.M. Robins. 1995. "Probabilistic Evaluation of Sequential Plans from Causal Models with Hidden Variables." Uncertainty in Artificial Intelligence, Proceedings of the 11th Conference, pg. 444-453.

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. (1999a). "Association, causation, and marginal structural models." Synthese 121: 151-179.

Robins, J.M. (1999b). "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.

Rubin, Donald B. 1978. "Bayesian inference for causal effects: The role of randomization." The Annals of Statistics 6:34-58.

Smith, Herbert. 1990. "Specification Problems in Experimental and Nonexperimental Social Research." Sociological Methodology 20:59-92.

Sobel, Michael E. 2000. "Causal Inference in the Social Sciences." Journal of the American Statistical Association 95(450): 647-651.

Stolzenberg, Ross M. and Daniel A. Relles. 1990. "Theory Testing in a World of Constrained Research Design: The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research." Sociological Methods and Research 18:395-415.

Thornton, Arland, William G. Axinn, and Jay D. Teachman. 1995. "The Influence of School Enrollment and Accumulation on Cohabitation and Marriage in Early Adulthood." American Sociological Review 60:762-774.

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

Winship, Christopher and Michael Sobel. 2000. "Causal Inference in Sociological Studies." Unpublished manuscript, Harvard University.

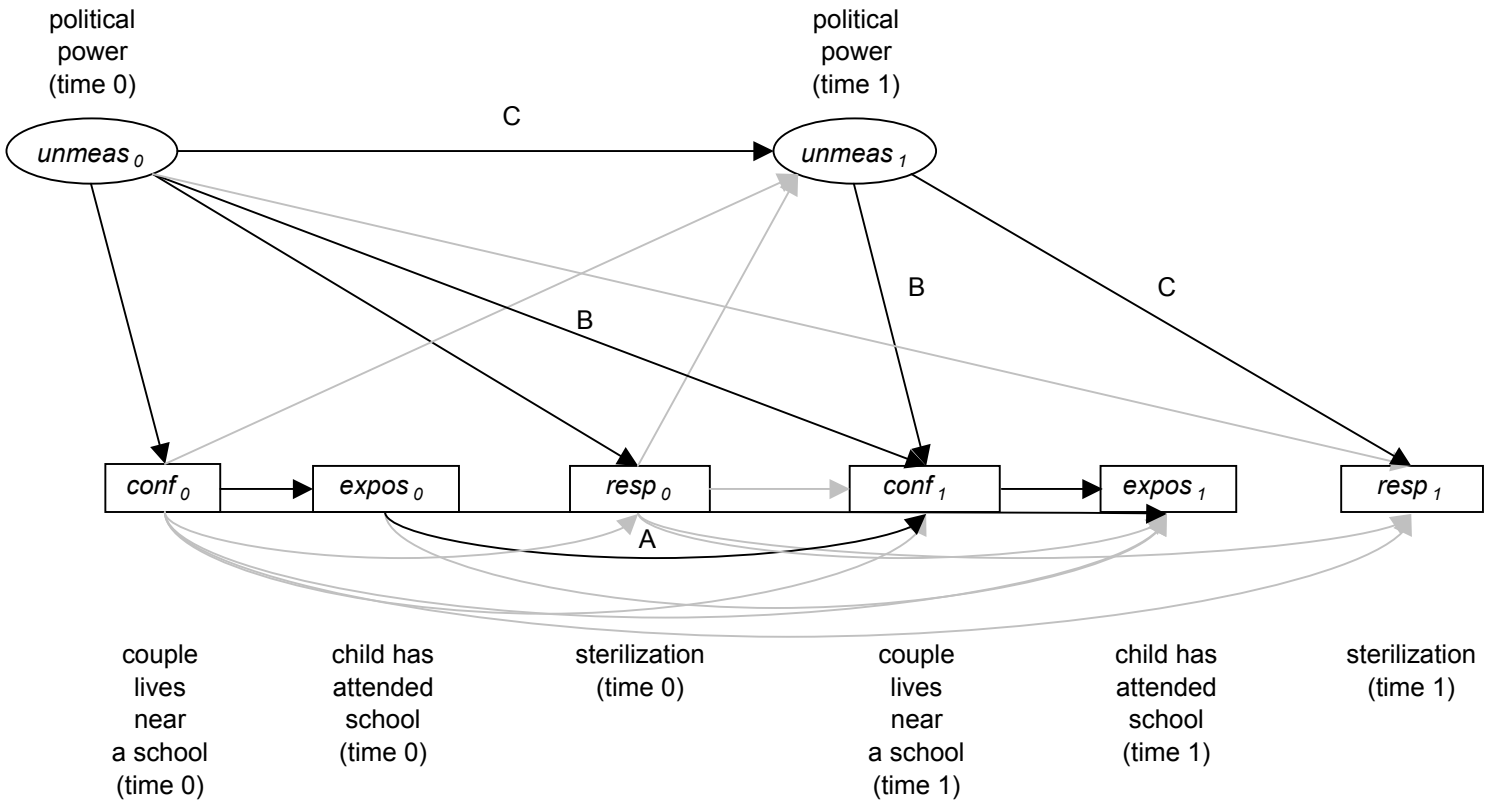

Figure 1
Graphical representation of the multivariate distribution of *unmeas* , *conf* , *expos* , and *resp*

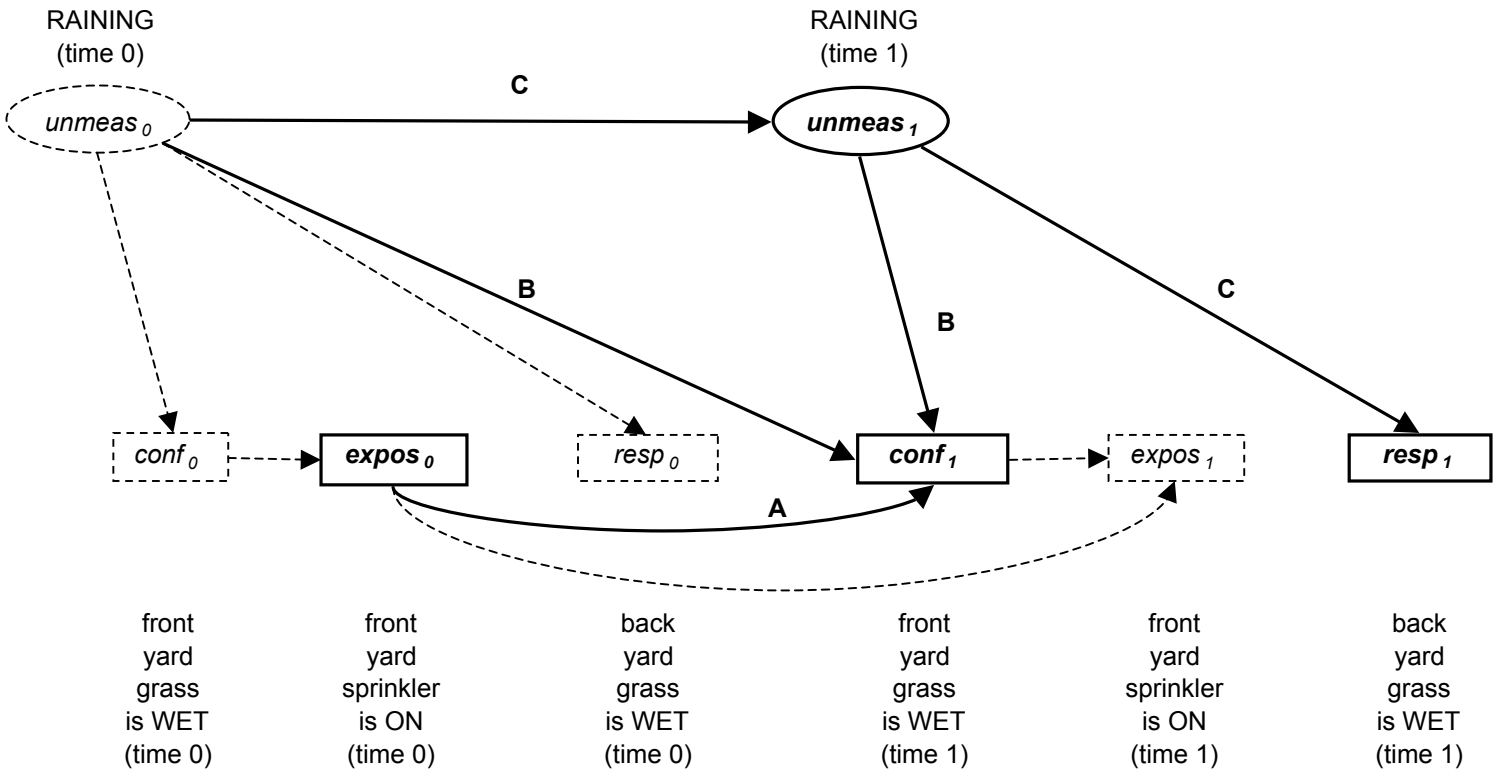

Figure 2
Graphical representation of the multivariate distribution of *unmeas* , *conf* , *expos* , and *resp*

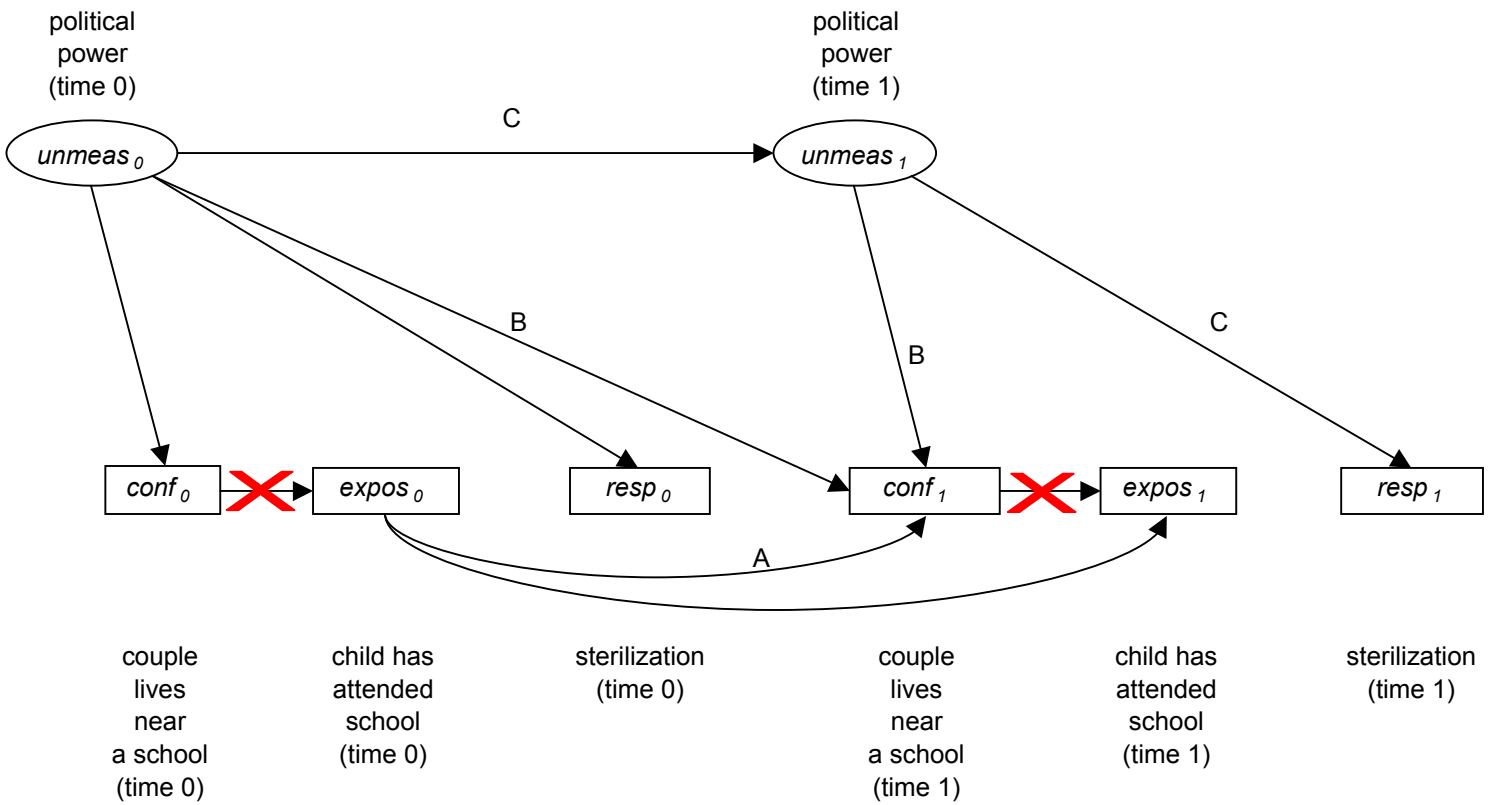

Figure 3
Graphical representation of the multivariate distribution of *unmeas* , *conf* , *expos* , and *resp* in the weighted sample

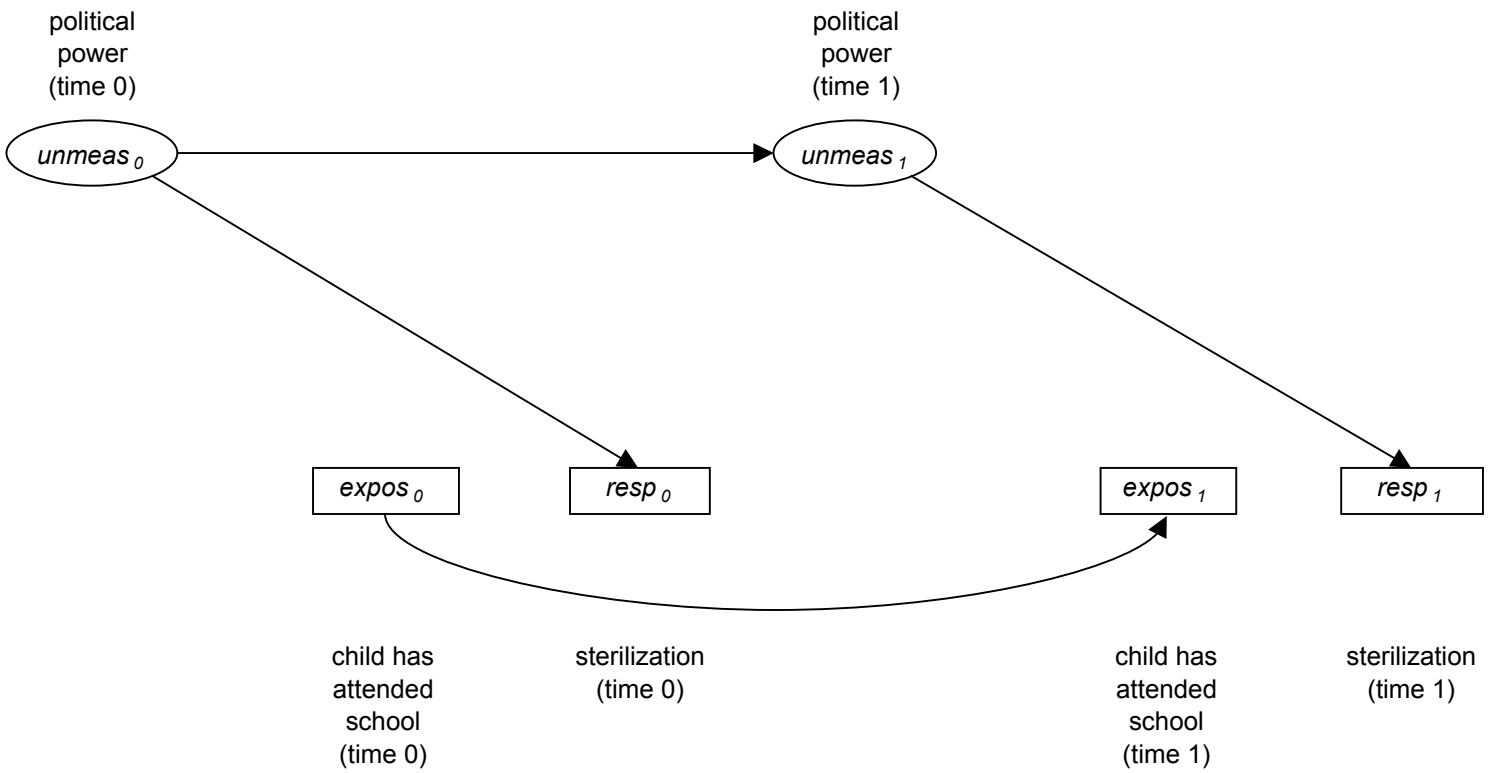

Figure 4
Graphical representation of the multivariate distribution of *unmeas* , *expos* , and *resp* in the weighted sample

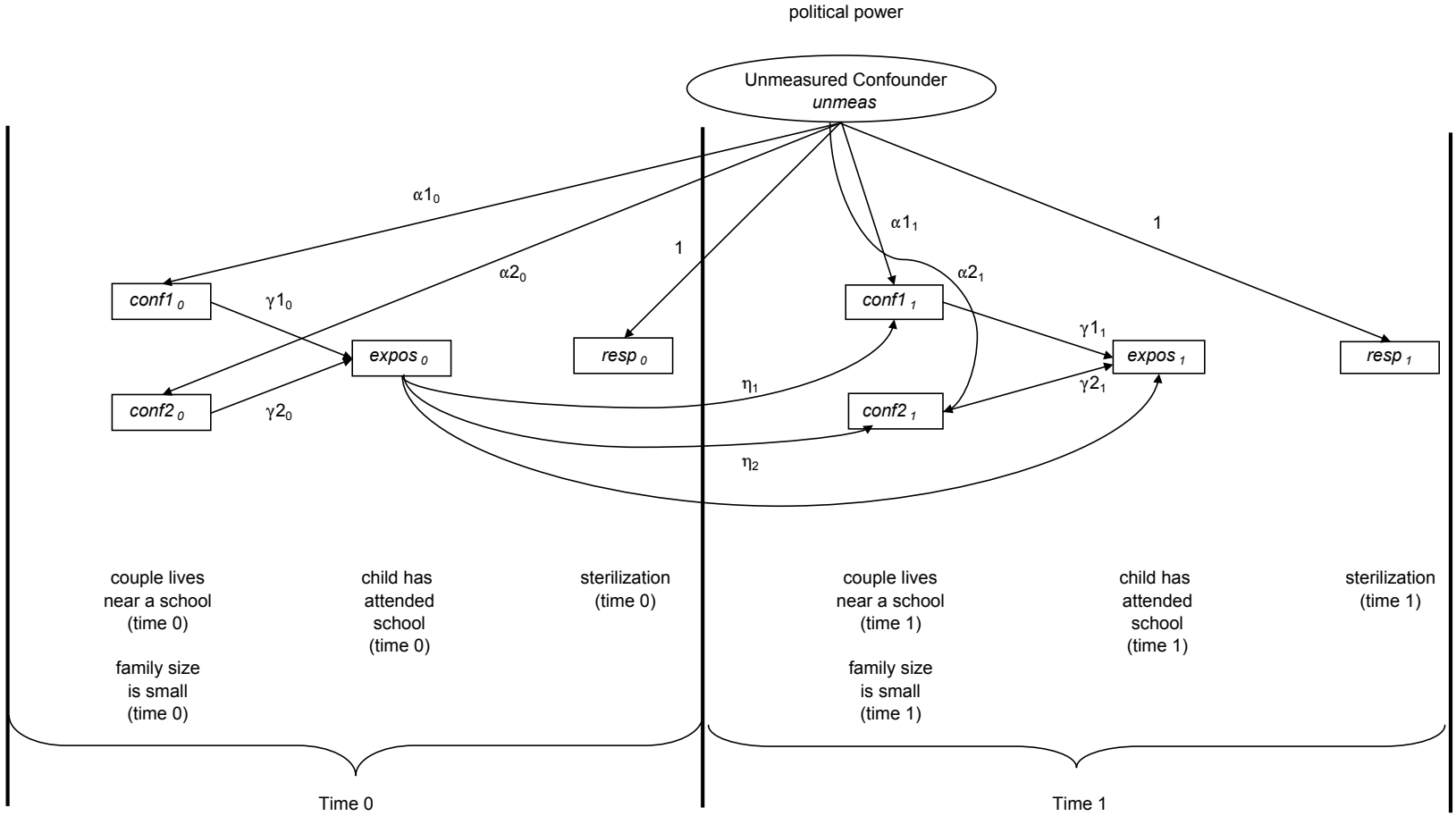

Figure 5
Multivariate distribution of all variables in simulation

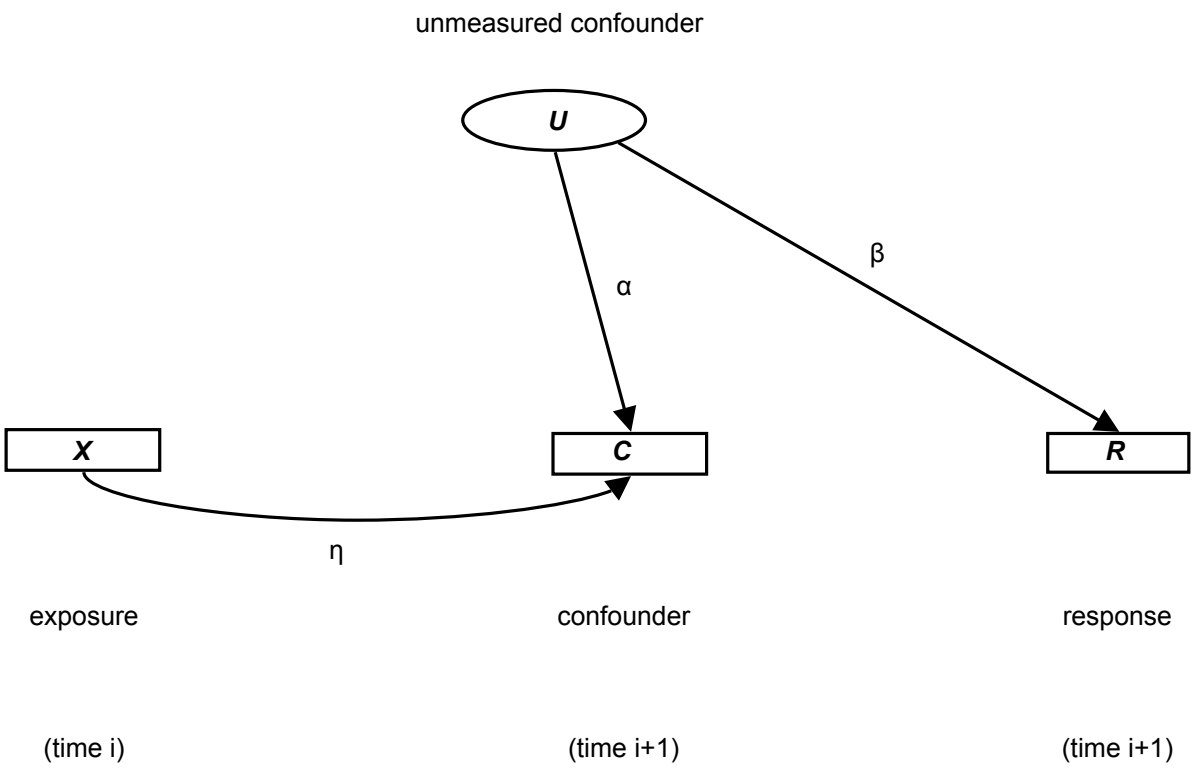

Figure 6
A family of multivariate distributions for X, C, U, R

TABLE 4
Means and Standard Deviations of Variables used in Analyses of Chitwan Valley Family Study Data

Variable	Mean	Std. Error	Min.	Max.
Couple has been sterilized	.61		0	1
Any child has ever attended school	.73		0	1
School is present within 5 minutes walk	.42		0	1
Family size				
Couple has 1 child	.15		0	1
Couple has 2 or 3 children	.46		0	1
Couple has 4 or more children	.37		0	1
Mother ever attended school	.34		0	1
Husband's years of education	4.51	4.66	0	16
Mother lived near school during childhood	.79		0	1
Birth cohort				
1962-1971 (age 25 - 34)				
1952-1961 (age 35 - 44)	.35		0	1
1942-1951 (age 45 - 54)	.23		0	1
Ethnic Group				
Upper Caste Hindu	.47		0	1
Low Caste Hindu	.11		0	1
Newar	.07		0	1
Hill Tibeto-Burmese	.17		0	1
Terai Tibeto-Burmese	.18		0	1
Miles to nearest town	8.60	3.96	.02	17.70
Years since first child turned age 6	11.02	8.13	0	36

TABLE 5
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*** (.10)	.74*** (.11)	.68*** (.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 (.09)	-.12 (.10)	-.07 (.10)
Husband's years of education	.02* (.01)	.02* (.01)	.02* (.01)
Mother lived near school during childhood	.26** (.10)	.30** (.10)	.33** (.12)
Birth cohort[b]			
1952-1961 (age 35 - 44)	-.61*** (.09)	-.65*** (.09)	-.56*** (.10)
1942-1951 (age 45 - 54)	-1.24*** (.12)	-1.30*** (.12)	-1.22*** (.13)
Ethnic group[c]			
Low Caste Hindu	-.24 (.13)	-.18 (.13)	-.23 (.15)
Newar	.13 (.14)	.14 (.14)	.13 (.14)
Hill Tibeto-Burmese	-.11 (.11)	-.06 (.11)	-.02 (.12)
Terai Tibeto-Burmese	-.84*** (.13)	-.88*** (.14)	-.84*** (.14)
Miles to nearest town	-.001 (.001)	-.001 (.001)	-.001 (.001)
Years since first birth	.0003 (.01)	-.03** (.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 6
Means and Standard Deviations of Variables used in Analyses of Intergenerational Panel Study Data

Variable	Mean	Std. Error	Min.	Max.
Had first birth	.58		0	1
Dropped out of school (6 months or more)	.96		0	1
Educational attainment (years after age 15)	4.71	2.31		
Marital status				
Cohabiting	.12		0	1
Married	.55		0	1
Ever went steady (by age 18)	.75		0	1
Mother's completed family size	3.80	1.49	1	10
Family financial assets (at age 0)	$633.87	680.00	0	$4,400
Mother worked outside the home (age 15)	.60		0	1
Mother Catholic	.53		0	1
Mother's experiences with divorce/remarriage				
Mother divorced and remarried	.11		0	1
Mother divorced and not remarried	.10		0	1
Age	27.76	4.45	15	32.17
Female	.51		0	1

Note: Statistics pertaining to time-varying measures refer to the final period of observation – either the period in which the event occurred, or censor.

TABLE 7
Logistic regression estimates (with robust standard errors) of hazard of first birth on dropping out of school

	Naive	Standard	Weighted
	(1)	(2)	(3)
Dropped out of school (6 months or more)	1.46* (.64)	1.20*** (.20)	1.51** (.64)
Educational attainment		-.01 (.03)	
Marital status[a]			
Cohabiting		1.34*** (.14)	
Married		2.50*** (.11)	
Ever went steady (by age 18)	.84 (1.17)	.40*** (.13)	.85 (1.17)
Mother's completed family size	.28* (.13)	.17*** (.03)	.28* (.13)
Family financial assets (at age 0)	-.26 (.21)	-.21** (.08)	-.26 (.21)
Mother worked outside the home (age 15)	-.29 (.24)	-.19* (.10)	-.28 (.24)
Mother Catholic	.14 (.24)	-.16 (.10)	.14 (.24)
Mother's experiences with divorce/remarriage[c]			
Mother divorced and remarried	-.24 (.50)	-.32* (.16)	-.23 (.50)
Mother divorced and not remarried	.20 (.46)	.07 (.17)	.19 (.46)
Age	.10*** (.03)	-.04** (.02)	.10*** (.03)
Female	.77** (.33)	.07 (.10)	.81** (.33)
N (individuals)	883	883	883
N (person-months)	133,393	133,393	133,393

Notes:
[a] Reference group is single.
[b] Reference group is not working.
[c] Reference group is continuously married mothers.
* $p < .05$, ** $p < .01$, *** $p < .001$, one tailed tests except where noted.

TABLE 8
Definitions of variables used to create simulated data

Variable	Formula used
unmeas	p(*unmeas* = 0) = 1/3 p(*unmeas* = 1) = 1/3 p(*unmeas* = 2) = 1/3
$resp_0$	if *unmeas* = 0, $resp_0 = 0$ if *unmeas* = 1, $resp_0 = 0$ if *unmeas* = 2, $resp_0 = 1$
$resp_1$	if *unmeas* = 0, $resp_1 = 0$ if *unmeas* = 1, $resp_1 = 1$
$expos_0$	$p(expos_0 = 1) = \frac{\exp(\gamma 0 + \gamma 1_0 * conf1_0 + \gamma 2_0 * conf2_0)}{1 + \exp(\gamma 0 + \gamma 1_0 * conf1_0 + \gamma 2_0 * conf2_0)}$
$expos_1$	$p(expos_1 = 1 \mid expos0 = 0) = \frac{\exp(\gamma 0 + \gamma 1_1 * conf1_1 + \gamma 2_1 * conf2_1)}{1 + \exp(\gamma 0 + \gamma 1_1 * conf1_1 + \gamma 2_1 * conf2_1)}$ $p(expos_1 = 1 \mid expos0 = 1) = 1$
$conf1_0$	$p(conf1_0 = 1) = \frac{\exp(\alpha 0 + \alpha 1_0 * unmeas)}{1 + \exp(\alpha 0 + \alpha 1_0 * unmeas)}$
$conf1_1$	$p(conf1_1 = 1) = \frac{\exp(\eta 0 + \eta 1_1 * expos_0 + \alpha 1_1 * unmeas)}{1 + \exp(\eta 0 + \eta 1_1 * expos_0 + \alpha 1_1 * unmeas)}$
$conf2_0$	$p(conf2_0 = 1) = \frac{\exp(\alpha 0 + \alpha 1_0 * unmeas)}{1 + \exp(\alpha 0 + \alpha 1_0 * unmeas)}$
$conf2_1$	$p(conf2_1 = 1) = \frac{\exp(\eta 0 + \eta 2_1 * expos_0 + \alpha 2_1 * unmeas)}{1 + \exp(\eta 0 + \eta 2_1 * expos_0 + \alpha 2_1 * unmeas)}$

TABLE 9

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

	Naive			Standard			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; η_0 = 0.0; $\eta_1 = \eta_2$ = 1.5

TABLE 10

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

	Naive		Standard		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; η_0 = 0.0; $\eta_1 = \eta_2$ = 1.5

TABLE 11

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

	Naive			Standard			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; η_0 = 0.0; $\boldsymbol{\eta_1 = \eta_2 = 1.0}$									
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; η_0 = 0.0; $\boldsymbol{\eta_1 = \eta_2 = 0.5}$									
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.

TABLE 12

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

	Naive			Standard (model includes less important confounder, *conf2*)			Standard (model includes more important confounder, *conf1*)			Standard (model includes both measured confounders, *conf1* and *conf2*)			Partially Weighted (weights computed using less important confounder, *conf2*)			Partially Weighted (weights computed using more important confounder, *conf1*)			Weighted (weights computed using both measured confounders, *conf1* and *conf2*)		
	(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	.20	.31	.16	-.11	.19	.12	-.42	.07	.02	-.74	-.04	.14	.14	.20	.10	.05	.07	.004	.01	.01
	(.15)	(.19)	(.28)	(.16)	(.20)	(.29)	(.16)	(.25)	(.39)	(.16)	(.27)	(.40)	(.15)	(.19)	(.28)	(.14)	(.19)	(.28)	(.14)	(.19)	(.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; $\boldsymbol{\alpha 1_0 = \alpha 1_1 = 2.25}$; $\boldsymbol{\alpha 2_0 = \alpha 2_1 = .75}$; $\gamma_0 = 0.0$; $\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1 = 0.5$; $\eta_0 = 0.0$; $\eta_1 = \eta_2 = 1.5$

TABLE 13

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

	Naive			Standard			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	.43 (.14	-.41 (.26)	1.39 (.29)	.08 (.15)	-.59 (.31)	-.25 (.36)	.12 (.15)	-.72 (.43)	.90 (.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: $\boldsymbol{\alpha_0 = 2.0}$; $\alpha 1_0 = \alpha 1_1 = \alpha 2_0 = \alpha 2_1 = 1.5$; $\boldsymbol{\gamma_0 = -15.0}$; $\boldsymbol{\gamma 1_0 = \gamma 1_1 = \gamma 2_0 = \gamma 2_1 = 8.0}$; $\eta_0 = 0.0$; $\eta_1 = \eta_2 = 1.5$

APPENDIX A: COMPUTING THE WEIGHTS

```
data sasuser.atrisk;
set sasuser.simpy;

if expos0 = 1 and i = 1 then delete; /* if child sent to school in
previous period */

run;
```

Note that i denotes time period. The statement above creates the "at risk" data set. It deletes all person-periods after exposure (sent child to school, in our example) has occurred. This data set represents all person-periods where the respondent is at risk for exposure and has not yet experienced the outcome. In our example, this data set represents the person-periods of couples who are at risk of sending a child to school and who have not yet become sterilized.

```
* numerator calculation;

data numprob;        /* numerator probability */
set sasuser.atrisk;

     proc logistic out=numest data=numprob;
     model expos=i;
     output out=numpest prob=probn;
     by dtst;
run;

data numpest;             /* numerator probability estimate */
```

```
set numpest;
     if a=1 then probn=1-probn;
run;
```

The logistic regression with expos (sending a child to school, in our example) as the response using the "atrisk" data set (described above) estimates the numerator for the weights. The numerator is the estimated probability of the actual exposure received, among those still at risk of the outcome (resp) and who have not experienced the exposure (expos) prior to this time period. For couples who are exposed (expos=1) in period i, the numerator is the obtained estimated probability. For couples who are not exposed in period i, the numerator is one minus the estimated probability. Thus, couples who sent a child to school are assigned the probability of sending a child to school, and couples who didn't send a child to school are assigned the probability of not sending a child to school.

```
* denominator calculation;

data denprob;
set sasuser.atrisk;
     proc logistic out=denest data=denprob;
     model expos=conf1 conf2 i;
     output out=denpest prob=probd;
     by dtst;
run;

data denpest;
set denpest;
     if expos=1 then probd = 1-probd;
run;
```

The denominator calculation is similar to the numerator calculation, but the logistic regression model *includes the confounders* (conf1 and conf2).

```
* merge the numerator and denominator data sets;

proc sort data=denpest; by dtst i j; run;
proc sort data=numpest; by dtst i j; run;

data wcomp;
merge denpest numpest;
     by dtst i j;
run;
```

This merges the numerator and denominator components into a single data set so we can compute the weights. The data sets are merged by the dataset ID (dtst) (because our simulation uses 1000 datasets; see Appendix D), the observation period (i), and the couple (j).

```
* add the weight to the person period data set;

proc sort data=wcomp; by dtst i j; run;
proc sort data=sasuser.simpy; by dtst i j; run;

data sasuser.simpy;
     merge sasuser.simpy wcomp;
     by dtst i j;
run;
```

This merges the weight components into the person-period data set.

```
* compute weight;

data sasuser.simpy;
set sasuser.simpy;

if i eq 0 then do;
     weight=probn/probd;
     retain weight;
end;

if i eq 1 and expos0 eq 0 then weight=weight*(probn/probd);
if i eq 1 and expos0 eq 1 then weight=weight*1;

run;
```

This computes the weight for each person-period. If the period is 0, the weight is simply the ratio of the numerator to the denominator. This weight must also be 'retained' in SAS if it is the initial period (i=0). This is so that the next period (i=1) can use the weight from the initial period. If exposure in the initial period is zero (expos=0), the next period weight is the weight from the initial period multiplied by the numerator-denominator ratio for the next period. However, if exposure in the initial period is one (expos=1), then the next period weight is simply the weight from the initial period.

APPENDIX B: QUANTITATIVE REASONING UNDERLYING THE USE OF WEIGHTS TO ELIMINATE BIAS CAUSED BY OBSERVED CONFOUNDERS

In order to most clearly present the quantitative reasoning behind weighting, we first consider the simple case of estimating a mean response when there is time-independent confounding. Note that in this case the standard approach of including the confounders in a regression of the response on exposure eliminates bias; however, we include this case to introduce the ideas. Next we consider estimating a mean response when there is time-varying confounding and time-varying exposure. To simplify the arguments, assume all random variables are discrete. The justification for using weights to control confounding in survival analysis regressions is similar; however, the technical nature increases substantially. Thus, we omit this argument.

Estimation of a Mean Response: Suppose we wish to estimate the mean of Y following exposure X. But, we realize that the effect of X on Y may be confounded; in particular, individuals used their knowledge of a background variable S to decide whether they would be exposed, and S is predictive of the response (S is a confounder). This is the causal inference setting as exemplified in Winship and Morgan (1999). They use Rubin's causal model to illustrate the issues; we do, as well. Thus, we posit a response $Y(1)$ that would be observed if X= 1 and a response $Y(0)$ that would be observed if $X = 0$. Because the exposure X is either 1 or 0, we observe Y(0) or Y(1), but only one of these responses. The observed response can be expressed as $Y = XY(1) + (1 - X)Y(0)$, or even more compactly by $Y = Y(X)$. Our goal is to estimate the mean response under each exposure; that is, $EY(0)$ and $EY(1)$. In this setting, the assumption of ignorability is the assumption that X is

independent of ($Y(0)$, $Y(1)$) conditional on S.[1] Denote the probability that $X = x$ given $S = s$ by $p(x|s)$. Below we use the following equality:

P[Y=y|X=1, S=s] = P[Y(X)=y| X=1,S=s] because Y=Y(X).

= P[Y(1)=y|X=1, S=s] because Y(X)=Y(1) when X=1.

= P[Y(1)=y|S=s] because X is independent of Y(1) given S.

To estimate $EY(1)$, we write this expectation in terms of the probability distributions. Then we replace the probability distributions by averages over the sample to form an estimator of $EY(1)$. Now,

$$EY(1) = \sum_{y} yP[Y(1) = y]$$
$$= \sum_{y,s} yP[Y(1) = y|S = s]P[S = s]$$
$$= \sum_{y,s} yP[Y(1) = y|X = 1, S = s]P[S = s]$$

because P[Y(1)=y|S=s] = P[Y(1)=y|X=1, S=s]. Continuing on,

$$E[Y(1)] = \sum_{s,y} yP[Y = y|X = 1, S = s]P[S = s]$$

because Y(1)=Y when X=1. And because x can only take values 1 or 0,

$$E[Y(1)] = \sum_{s,y,x} yxP[Y = y|X = x, S = s]P[S = s].$$

Because, by assumption, p(x|s)>0, we have that

$$E[Y(1)] = \sum_{s,y,x} y \frac{x}{p(x|s)} P[Y = y|X = x, S = s]p(x|s) P[S = s].$$

[1]This ignorability assumption can be weakened; we use this version to simplify the explanation.

Finally by definition of the expectation,

$$E[Y(1)] = E\left[Y\frac{X}{p(X|S)}\right].$$

So to estimate the mean of Y(1) we can use the weighted average,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \frac{X_i}{p(X_i|S_i)}.$$

We could use a similar estimator for the mean of Y(0). If p(x|s) is known, then this weighted average is unbiased in small samples. If we must estimate p(x|s), then we model this probability (because x is binary, a logistic or probit model would be natural) and estimate the parameters in the model. In this case, the weighted average using the estimated weights is unbiased for large samples.

Note that the arguments above do not assume that S is a scalar; frequently S is a vector. This formula can be generalized to permit multivariate X (in this case the density p(x|s) is a multivariate density).

Estimation of a Time-Varying Mean: Justification of the weighting method for the combination of time-varying exposures, time-varying confounding, and time-varying responses (as in survival analysis) is more technical. However the basic ideas are the same as above. To illustrate the ideas , we generalize the above example to include two time points. Subscripts are used to denote time. The exposure patterns are (0,0) indicating no exposure at both times 1 and 2; (1,0), (0,1) and (1,1) have similar interpretations. We wish to estimate the mean of Y at time points 1 and 2 for each of four exposure patterns. S is a time varying confounder. Prior to exposure we observe S_1. After exposure at time 1 we observe Y_1 and then S_2, and after exposure at time 2, we observe Y_2. To summarize, assume the order of occurrence is S_1, X_1, Y_1, S_2, X_2, Y_2. It is possible that individuals may have used

their knowledge of (S_1,S_2, X_1, Y_1) to select their exposure X_2 at time 2. Thus we may have time-varying confounding because S_2 may have been used in selecting X_2; furthermore, this confounding may be endogenous as S_2 may have been affected by X_1.

Again, we use Rubin's causal model. Consider one person. This person has 6 potential responses and two potential intermediate outcomes. These are $Y_1(1)$, $Y_1(0)$ (responses if person were to be exposed or not, respectively, at time 1), $S_2(1)$, $S_2(0)$ (intermediate outcomes according to whether the person would be exposed at time 1) and $Y_2(0,0)$, $Y_2(0,1)$, $Y_2(1,0)$, $Y_2(1,1)$ (potential responses at time 2 to each of the 4 exposure patterns). Of course we only observe a subset of the potential outcomes. The data for this person is S_1, X_1, $Y_1(X_1)$, $S_2(X_1)$, X_2, $Y_2(X_1, X_2)$ (so $Y_1=Y_1(X_1)$, $S_2=S_2(X_1)$, $Y_2=Y_2(X_1,X_2)$). In this context, the assumption of sequential ignorability is the assumption that

$(Y_1(1),Y_1(0),S_2(1),S_2(0),Y_2(0,0), Y_2(0,1), Y_2(1,0), Y_2(1,1))$ is independent of X_1 given S_1

and

$(Y_2(0,0), Y_2(0,1), Y_2(1,0), Y_2(1,1))$ is independent of X_2 given $(S_1,X_1,S_2,Y_1,)$.

Denote the probability that $X_1=x_1$ given $S_1=s_1$ by $p_1(x_1|s_1)$ and denote the probability of $X_2=x_2$ given $S_1=s_1$, $X_1=x_1$, $Y_1=y_1$, $S_2=s_2$ by $p_2(x_2|s_1,x_1,s_2,y_1)$. From our previous example, we see that an unbiased estimator of $EY_1(1)$ is the weighted average,

$$\frac{1}{n}\sum_{i=1}^{n} Y_{1i}\frac{X_{1i}}{p(X_{1i}|S_{1i})}.$$

We use the same types of arguments to construct an unbiased estimator of $EY_2(0,1)$ as follows. In the process we use several facts that follow from the sequential ignorability assumption. Because $(Y_1(1),Y_1(0),S_2(1),S_2(0),Y_2(0,0), Y_2(0,1), Y_2(1,0), Y_2(1,1))$ is independent of X_1 given S_1, we have that

$P[Y_2(0,1)=y|\ S_1=s_1,\ Y_1(0)=y_1,\ S_2(0)=s_2]$

$= P[Y_2(0,1)=y|\ S_1=s_1,\ X_1=0,\ Y_1(0)=y_1,\ S_2(0)=s_2]$

$= P[Y_2(0,1)=y|\ S_1=s_1,\ X_1=0,\ Y_1=y_1,\ S_2=s_2]$

where the last equality follows because $Y_1=Y_1(0)$ when $X_1=0$ and similarly for S_1.

However, because $(Y_2(0,0),\ Y_2(0,1),\ Y_2(1,0),\ Y_2(1,1))$ is independent of X_2 given (S_1,X_1,S_2,Y_1), we obtain

$= P[Y_2(0,1)=y|\ S_1=s_1,\ X_1=0,\ Y_1=y_1,\ S_2=s_2,\ X_2=1]$

$= P[Y_2=y|\ S_1=s_1,\ X_1=0,\ Y_1=y_1,\ S_2=s_2,\ X_2=1]$

because $Y_2(0,1)=Y_2$ when $X_1=0$ and $X_2=1$. Again $(Y_1(1),Y_1(0),S_2(1),S_2(0),Y_2(0,0),\ Y_2(0,1),\ Y_2(1,0),\ Y_2(1,1))$ independent of X_1 given S_1 implies that

$P[Y_1(0)=y_1,\ S_2(0)=s_2|\ S_1=s_1]$

$= P[Y_1(0)=y_1,\ S_2(0)=s_2|\ S_1=s_1\ X_1=0]$

$= P[Y_1(X_1)=y_1,\ S_2(X_1)=s_2|\ S_1=s_1\ X_1=0]$

$= P[Y_1=y_1,\ S_2=s_2|\ S_1=s_1\ X_1=0].$

In summary, we have demonstrated that

$$P[Y_2(0,1)=y_2|\ S_1=s_1,\ Y_1(0)=y_1,\ S_2(0)=s_2]=P[Y_2=y_2|\ S_1=s_1,\ X_1=0,\ Y_1=y_1,\ S_2=s_2,\ X_2=1]$$

and that

$$P[Y_1(0)=y_1,\ S_2(0)=s_2|S_1=s_1]=P[Y_1=y_1,\ S_2=s_2|S_1=s_1,\ X_1=0]$$

hold assuming sequential ignorability. We can use these results to restate $EY_2(0,1)$ as

$$\begin{aligned}E[Y_2(0,1)] = &\sum_{s_1,x_1,y_1,s_2,x_2,y_2} y_2(1-x_1)x_2 P[Y_2=y_2|S_1=s_1,X_1=x_1,Y_1=y_1,S_2=s_2,X_2=x_2]\\ &\times P[Y_1=y_1,S_2=s_2|S_1=s_1,X_1=x_1]P[S_1=s_1]\\ = &\sum_{s_1,x_1,y_1,s_2,x_2,y_2} y_2 \frac{(1-x_1)x_2}{p_1(x_1|s_1)p_2(x_2|s_1,x_1,y_1,s_2)} P[Y_2=y_2|S_1=s_1,X_1=x_1,Y_1=y_1,S_2=s_2,X_2=x_2]\\ &\times p_2(x_2|s_1,x_1,y_1,s_2)P[Y_1=y_1,S_2=s_2|S_1=s_1,X_1=x_1]\\ &\times p_1(x_1|s_1)P[S_1=s_1]\end{aligned}$$

where the last equality follows from the results proved using sequential ignorability. Now we use the fact that x_1 and x_2 are 0, 1 valued:

$$
\begin{aligned}
E[Y_2(0,1)] &= \sum_{y_2} y_2 P[Y_2(0,1) = y_2] \\
&= \sum_{s_1,y_1,s_2,y_2} y_2 P[Y_2(0,1) = y_2 | S_1 = s_1, Y_1(0) = y_1, S_2(0) = s_2] \\
&\qquad \times P[Y_1(0) = y_1, S_2(0) = s_2 | S_1 = s_1] P[S_1 = s_1] \\
&= \sum_{s_1,y_1,s_2,y_2} y_2 P[Y_2 = y_2 | S_1 = s_1, X_1 = 0, Y_1 = y_1, S_2 = s_2, X_2 = 1] \\
&\qquad \times P[Y_1 = y_1, S_2 = s_2 | S_1 = s_1, X_1 = 0] P[S_1 = s_1]
\end{aligned}
$$

Thus by definition of the expectation we have,

$$E[Y_2(0,1)] = E\left[Y_2 \frac{(1-X_1)X_1}{p_1(X_1|S_1)p_2(X_2|S_1,X_1,Y_1,S_2)}\right].$$

Replacing expectations by sample averages we see that an unbiased estimator of $EY_2(0,1)$ is

$$\frac{1}{n}\sum_{i=1}^{n} Y_{2i} \frac{(1-X_{1i})X_{2i}}{p_1(X_{1i}|S_{1i})p_2(X_{2i}|S_{1i},X_{1i},Y_{1i},S_{2i})}.$$