

Assessing the Total Effect of Time-Varying Predictors in Prevention Research

Bethany Cara Bray,^{1,4} Daniel Almirall,² Rick S. Zimmerman,³
Donald Lynam,³ and Susan A. Murphy²

Published online: 18 February 2006

Observational data are often used to address prevention questions such as, “If alcohol initiation could be delayed, would that in turn cause a delay in marijuana initiation?” This question is concerned with the total causal effect of the timing of alcohol initiation on the timing of marijuana initiation. Unfortunately, when observational data are used to address a question such as the above, alternative explanations for the observed relationship between the predictor, here timing of alcohol initiation, and the response abound. These alternative explanations are due to the presence of confounders. Adjusting for confounders when using observational data is a particularly challenging problem when the predictor and confounders are time-varying. When time-varying confounders are present, the standard method of adjusting for confounders may fail to reduce bias and indeed can increase bias. In this paper, an intuitive and accessible graphical approach is used to illustrate how the standard method of controlling for confounders may result in biased total causal effect estimates. The graphical approach also provides an intuitive justification for an alternate method proposed by James Robins [Robins, J. M. (1998). *1997 Proceedings of the American Statistical Association, section on Bayesian statistical science* (pp. 1–10). Retrieved from <http://www.biostat.harvard.edu/robins/research.html>; Robins, J. M., Hernán, M., & Brumback, B. (2000). *Epidemiology*, *11*(5), 550–560]. The above two methods are illustrated by addressing the motivating question. Implications for prevention researchers who wish to estimate total causal effects using longitudinal observational data are discussed.

KEY WORDS: confounding; weighting; total effect; time-varying; graphical approach .

Observational data are often used to address prevention questions concerning the consequences of an adolescent’s actions on drug use. Consider the motivating question, “If alcohol initiation could be delayed, would that in turn cause a delay in marijuana initiation?” The answer to this question could

be used to anticipate whether an alcohol use prevention program implemented during adolescence might also have effects on marijuana use. The answer is provided by the total causal effect of delaying the timing of alcohol initiation (predictor) on the timing of marijuana initiation (response). As is well known, a fundamental problem in addressing this question with observational data is the presence of confounders.

Confounders are common correlates of the predictor and response, and often provide alternate explanations for the observed relation between the two. An example of a common correlate of alcohol and marijuana initiation is peer pressure resistance. Adolescents with high levels of peer pressure resistance may be less likely to initiate both alcohol and marijuana use. In this case, failing to take peer pressure resistance into account would result in a biased

¹The Methodology Center, Department of Human Development and Family Studies, The Pennsylvania State University, University Park, Pennsylvania.

²Institute for Social Research, Department of Statistics, The University of Michigan, Ann Arbor, Michigan.

³Departments of Communication and Psychology, The University of Kentucky, Lexington, Kentucky.

⁴Correspondence should be directed to Bethany Cara Bray The Methodology Center, Pennsylvania State University, 204 E. Calder Way Suite 400, State College, Pennsylvania 16801; e-mail: bcbay@psu.edu.

estimate of the total causal effect of alcohol initiation on marijuana initiation. For example, the coefficient of alcohol initiation in a simple regression of marijuana initiation on alcohol initiation would reflect a combination of (a) the compositional differences in the types of individuals who choose to initiate or not initiate alcohol use due to peer pressure resistance and (b) the true total causal effect of alcohol initiation on marijuana initiation. Thus, when confounding is not controlled, the coefficient of alcohol initiation could not be regarded as an unbiased estimate of the total causal effect of alcohol initiation on marijuana initiation.

It follows that, in order to make valid causal inference concerning the effect of a predictor on an outcome using observational data, confounders must be controlled in some way. This problem is well recognized in the social science, statistical, and econometric literatures (Cochran & Rubin, 1973; Shadish *et al.*, 2002; Winship & Morgan, 1999). Indeed, a variety of statistical adjustment procedures have been proposed to account for confounding in the non-longitudinal setting. These nonexperimental methods for estimating causal effects include standard covariate adjustment methods (Bohrnstedt & Knoke, 1982; Bollen, 1989), selection models (Heckman & Hotz, 1989; Winship & Mare, 1992), and instrumental variable approaches (Angrist *et al.*, 1996; Little & Yau, 1998), as well as propensity score methods including propensity score stratification and matching (Rosenbaum & Rubin, 1984, 1985). Standard covariate adjustment is perhaps the most commonly used method to adjust for confounding. With this method, baseline variables that are thought to be confounders of the observed effect of the predictor on the response are included as covariates in a regression model; we call this the standard method.

The problem of adjusting for confounders using observational data is more challenging in the longitudinal setting, when values of both the predictor and the set of confounders can change over time (i.e., they are time-varying). In this setting and in the presence of unmeasured confounders, the standard method of adjusting for confounders may fail to reduce bias and can cause further bias.

Robins and colleagues (Hernán *et al.*, 2000, 2001; Robins, 1998; Robins *et al.*, 2000) have developed a weighting method that allows researchers, using relatively parsimonious models and “over the counter” statistical software, to adjust for time-varying confounders without the problems associated with the standard method. Despite the weighting method’s ease of implementation and its availability for al-

most a decade, however, it has not found ready adoption among researchers outside epidemiology. The few exceptions include Barber *et al.* (2004), Agerbo (2005), and work using marginal structural models to assess neighborhood-level causal effects (e.g., Oakes, 2004).

The fundamental contribution of this paper is to elucidate for the prevention community the potential for bias in assessing the effects of time-varying predictors, and to explicate and disseminate a technique from the field of epidemiology that can be used to mitigate these biases. This is achieved by meeting the following objectives. The primary objective of this paper is the presentation of an intuitive and accessible graphical approach (Pearl, 1998) to illustrate how the standard method may produce bias in the presence of unmeasured indirect confounders. The graphical approach is used to define three types of confounders; in particular, it is used to discuss the issues surrounding unmeasured time-varying confounders. The second objective of this paper is to use the graphical approach to illustrate that, unlike the standard method, Robins’ weighting method does not produce biased effects in the presence of unmeasured *indirect* confounders. We illustrate, using the graphical approach, that both methods may produce biased effects if there are unmeasured *direct* confounders.

In addition, an application of the weighting method using real data is presented. This example provides a detailed illustration of the technique, including how to construct the weights and how to obtain robust standard errors, as well as some guidance for researchers interested in carrying out the weighting method within levels of baseline covariates (i.e., using models that condition on sex and/or race, for example). These steps and more are elaborated upon in the appendix and in an accompanying web-appendix (which can be found at <http://methodology.psu.edu/publications/typappen.html>). The web-appendix contains example hypothesized data, as well as SAS programming code for data manipulation, weight creation, and analysis.

THE TOTAL CAUSAL EFFECT

A question about the effect of a putative cause A on a response of interest B, without reference to how the effect is transmitted from A to B, refers to a question about the *total causal effect* of A on B. Total causal effects encompass the full (net) effect of A on

B, every way in which this effect can take place. This includes, for instance, direct effects of A on B, as well as indirect effects transmitted through intermediate variables that are observed or unobserved, known or unknown. The notion of a direct or indirect causal effect can arise only when additional causal hypotheses are put forth concerning intermediate variables (often called *mediators*) that are said to carry some or all of the effect of A on B. This paper does not concern itself with assessing direct or indirect effects. Instead, the answer to the motivating question of this paper is provided by the total causal effect, which encompasses the effect transmitted through all causal pathways whereby alcohol initiation may alter the timing of marijuana initiation.

It is useful to consider the total causal effect of alcohol initiation on marijuana initiation for at least two reasons. For example, if the initiation of alcohol use has a strong effect on the timing of marijuana initiation, then interventions designed to decrease marijuana initiation might fruitfully include components developed for preventing alcohol initiation. Further it is often useful to understand the derivative effects from already established alcohol misuse prevention programs; evidence that reinforces the causal effect of alcohol initiation on marijuana initiation will provide an indication that such programs have additional positive consequences.

A GRAPHICAL APPROACH TO CONFOUNDING

A graphical approach is used here to define, illustrate, and discuss three classes of confounders and their role in the estimation of total causal effects. The graphical approach used throughout resembles the one used in structural equations modeling (SEM; Bollen, 1989), but has a different purpose. The graphs are employed as a heuristic tool to discuss and highlight issues concerning time-varying confounding. The graphs are not used, as they are in SEM, to describe full structural systems of causal effects to be estimated. As we have stated, we are interested in estimating the total causal effect. The weighting method that is presented in later sections does not require estimating a full structural system of equations in order to assess this effect.

Consider studying the total causal effect of a putative cause, alcohol initiation (Alc), on a response of interest, marijuana initiation (Mj). Denote the complete set of variables existing before the measurement of Alc by \mathbf{V} . Further, within the set of variables

\mathbf{V} denote the subset of all *observed* variables by \mathbf{O} .⁵ Let the subset \mathbf{U} denote the complement of \mathbf{O} —that is, \mathbf{U} is the subset of all *unobserved* or unknown variables in \mathbf{V} . Using this notation, $\mathbf{V} = (\mathbf{O}, \mathbf{U})$. To be clear about the role of time, the temporal order of the objects just defined is $(\mathbf{V}, \text{Alc}, \text{Mj})$. In our example, the subset \mathbf{O} may include sex, race, and ability to resist peer pressure (PPress). \mathbf{U} may include variables that are either too difficult or too expensive to measure in a particular study, such as family functioning status. In addition, \mathbf{U} includes variables that are *unknown* to the scientist.

Any variable in \mathbf{V} that is related to both Alc and Mj is a confounder of the total causal effect of Alc on Mj. For example, in the context of our motivating question, peer pressure resistance is an example of a confounder (as discussed in the introduction) belonging to \mathbf{V} . Depending on whether a confounder is observed (i.e., belongs to \mathbf{O}) or unobserved (i.e., belongs to \mathbf{U}) determines whether the variable is a *measured* or *unmeasured confounder*, respectively.

This paper considers data structures like $(\mathbf{V}, \text{Alc}, \text{Mj})$ in the time-varying longitudinal setting. Particularly, we consider time event history or “survival” data in which both the putative cause (alcohol initiation) and the response (marijuana initiation) are time-varying, as are other measured and unmeasured variables. Thus, henceforth, we index the variables $(\mathbf{V}, \text{Alc}, \text{Mj})$ by time t . The survival methods used to fit the models in later sections model the probability of marijuana initiation at time t given no marijuana initiation up to and including time $t - 1$ as a function of alcohol initiation (and possibly other variables) at time t . This probability is sometimes known as the hazard probability of initiating marijuana at time t . That is, the basic model considered throughout this paper is a discrete-time survival logistic regression model of the form

$$\log \left(\frac{p_t}{1 - p_t} \right) = \alpha_t + \beta_1 \text{Alc}_t, \quad (1)$$

where p_t denotes the hazard probability of marijuana initiation at time t (Allison, 1995; Singer & Willet, 1993).⁶ This model, however, is modified throughout the paper to make several points about the

⁵In this paper, we assume that all variables in the subset \mathbf{O} are measured precisely. Consequently, we consider neither measurement error nor measurement models.

⁶This model makes the proportional hazards assumption that the effect of alcohol initiation is the same at all time points. This *modeling assumption* is maintained throughout the paper; it has no bearing on the causal issues that are discussed throughout.

inclusion of confounders; for instance, additional terms for confounders or time-invariant covariates (e.g., sex and race) may be added.

Panel A of Fig. 1 depicts an example of *measured confounding* by peer pressure resistance in the longitudinal setting, using a graphical approach. Figure 1 uses Alc_t , Mj_t , $PPress_t$, and U_t to represent alcohol initiation, marijuana initiation, peer pressure resistance, and unobserved variables at time t , respectively. For example, Alc_1 is alcohol initiation at time 1 and Alc_2 is alcohol initiation at time 2. Further, in this figure, the subvector \mathbf{O}_t has only one element, namely peer pressure resistance, and any other variables are assumed unobserved and are included in \mathbf{U}_t . $PPress_t$ is enclosed in a rectangle to indicate that it is an observed variable; the subvector \mathbf{U}_t is enclosed in an oval to indicate that it is an unobserved variable and to differentiate it from $PPress_t$.

Figure 1 includes arrows from peer pressure resistance ($PPress_t$) to both alcohol initiation (Alc_t) and marijuana initiation (Mj_t) at both times—indicating that there is an effect of peer pressure resistance on both and that peer pressure resistance is a confounder of the effect of alcohol initiation on marijuana initiation at times 1 and 2. Observe that, in Panel A, \mathbf{U}_t is not a confounder because \mathbf{U}_t does not affect both Alc_t and Mj_t .

Panels B and C of Fig. 1 illustrate examples of *unmeasured confounding* (in addition to the measured confounding by $PPress_t$). In both panels, \mathbf{U}_t is related to both Alc_t and Mj_t . An important distinction has been made in Panels B and C, however, with regard to how unmeasured confounding takes place. In Panel B, the effect of \mathbf{U}_t on Alc_t is a direct one, not passing through an intermediate variable. In Panel C, the effect of \mathbf{U}_t on Alc_t is an indirect one, passing

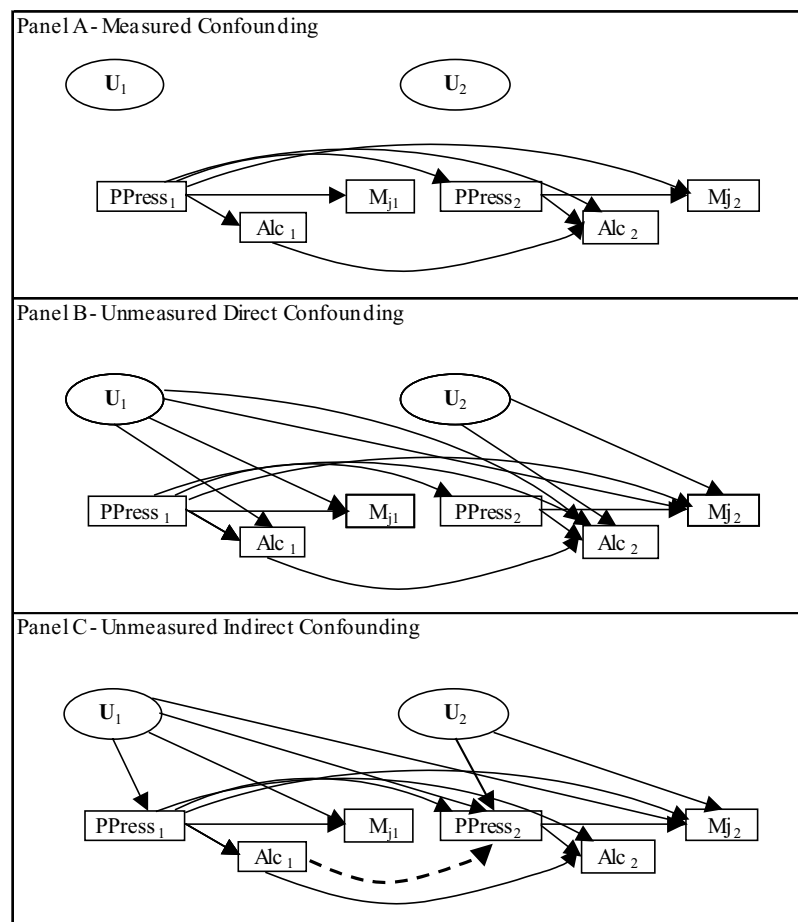


Fig. 1. Sample scenarios depicting three classes of confounding in observational studies.

through the observed variable $PPress_t$. In Panel B, U_t is referred to as an *unmeasured direct confounder*; in Panel C, U_t is referred to as an *unmeasured indirect confounder*.

Note that additional arrows may be appended to each system of relations depicted in Panels A, B, and C, which may or may not change a variable's status as a measured confounder, an unmeasured direct confounder, or an unmeasured indirect confounder. In turn, changing a variable's status may or may not change a panel's interpretation. For example, in Panel A, an additional arrow from U_t to Mj_t does not change its status as depicting measured confounding, nor does adding an arrow from U_t to $PPress_t$. Adding both of these arrows simultaneously, however, would illustrate both measured and unmeasured indirect confounding. Similarly, in Panel B, adding an additional arrow from U_t to $PPress_t$ does not change its status as depicting unmeasured direct confounding, but the panel would then illustrate both unmeasured direct and unmeasured indirect confounding. Finally, adding an additional arrow from Alc_t to Mj_t in any of the panels does not change any of their interpretations.

An important point is that in addition to including variables that are either too expensive or too difficult to measure, U_t includes *unknown* confounders of the effect of alcohol initiation on marijuana initiation. In other words, U_t may include variables related to both Alc_t and Mj_t that are part of prevention theories not yet discovered, or U_t may include undiscovered or unidentified bio-psychosocial confounding variables of importance to prevention scientists. Only to the extent that a scientist believes all possible variables related to both alcohol initiation and marijuana initiation are observed in a study, can he or she believe that a relation such as the one depicted in either Panel B or Panel C does not exist. Our position in this paper is that unknown confounding variables always exist—that is, U_t is always a nonempty set with at least one element that is related to both alcohol initiation (the predictor and putative cause) and marijuana initiation (the response). The question that remains, then, is whether U_t acts as an indirect or direct confounder, and what the consequences of that may be when assessing the total causal effect.

If we assume that any one of the three hypothetical scenarios in Panels A, B, and C (without the dotted arrow) depicts the true structural system, alcohol initiation has no true total causal effect on marijuana initiation. This is because there is *no* causal path from either Alc_1 or Alc_2 to either Mj_1

or Mj_2 . Despite the lack of a true total causal effect, a simple regression of marijuana initiation on alcohol initiation under all three scenarios would yield a nonzero coefficient for alcohol initiation. In all three panels, the apparent effects reflect an imbalance in the number of participants who initiate alcohol use due to peer pressure resistance. In addition, in Panels B and C, the apparent effects reflect an imbalance in the number of participants who initiate alcohol use due to variables that are unobserved, U_t . These apparent effects are often called spurious effects or effects due to confounding bias; they do not reflect true total causal effects of alcohol initiation on marijuana initiation.

Various statistical strategies exist that can be used to correct for spurious effects resulting from measured confounding like that depicted in Fig. 1, Panel A. When it is assumed that Panel A depicts the true structural system, the most common strategy is to include the observed confounders as additional variables in the regression of the response on the putative cause (predictor). In the context of Panel A, this strategy would involve a discrete-time survival logistic regression model such as,

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha_t + \beta_1 Alc_t + \beta_2 PPress_t. \quad (2)$$

By including $PPress_t$ in Eq. (2), the imbalance in the number of participants who initiate alcohol due to peer pressure resistance is accounted for, leaving the total causal effect of alcohol initiation on marijuana initiation untainted (in this case there would be a null effect). Provided the functional form specified on the right-hand side of Eq. (2) is correct, the parameter β_1 has a valid interpretation as the total causal effect of alcohol initiation on marijuana initiation (on the log-odds scale).⁷ We call this the *standard method* of adjusting for confounders.

As we have noted, however, confounding in Panels B and C is a result of both observed ($PPress_t$) and unobserved (U_t) variables. So, does the standard method work under the hypothetical scenarios depicted in Panels B and C, as it did in Panel A?

In the presence of unmeasured direct confounders, that is, assuming that Panel B depicts the true structural system, the standard method cannot be used to obtain a valid estimate of the total causal

⁷If an arrow from Alc_1 to $PPress_2$ were present in Fig. 1, Panel A, then β_1 in Eq. (2) would be interpreted as the direct causal effect of alcohol initiation on marijuana initiation.

effect of alcohol initiation on marijuana initiation because there is no way to use information from the measured variables, as we did in Eq. (2), to eliminate the bias due to \mathbf{U}_t . In other words, there is no way to break the relation between \mathbf{U}_t and either Alc_t or Mj_t at both times using PPress_t . Other methods such as selection models and their extensions, prominent in the econometrics literature (Heckman, 1979; Heckman & Hotz, 1989), can be used to adjust for unmeasured direct confounding. These methods, however, make additional distributional assumptions about the unmeasured (including unknown) variables. Our point here is that the *standard method* of adjusting for confounders, by itself, does not adjust for direct confounding by \mathbf{U}_t as it does in the scenario in Panel A.

Finally, assume that Panel C depicts the true structural system of relations. In the presence of unmeasured indirect confounders, as depicted here, the standard method may or may not be used to adjust for confounding by PPress_t and \mathbf{U}_t . Its use depends on the existence of the dotted arrow from Alc_1 to PPress_2 .

In the absence of the dotted arrow, the standard method can be used to adjust for confounding by PPress_t for exactly the same reasons as in Panel A, discussed above. In addition, the standard method adjusts for the spurious effects resulting from confounding by \mathbf{U}_t , in the absence of the dotted arrow. This is because confounding due to \mathbf{U}_t in Panel C acts only through the measured confounder PPress_t . By conditioning on PPress_t , as in Eq. (2), the correlation between \mathbf{U}_t and Alc_t is broken at both times, thereby eliminating \mathbf{U}_t as a confounder of the total causal effect of alcohol initiation on marijuana initiation.

In the presence of the dotted arrow from Alc_{t-1} to PPress_t (here Alc_1 to PPress_2), however, the standard method fails to adjust for confounding due to \mathbf{U}_t and may cause more bias.⁸ The dotted arrow is likely to exist in prevention settings; for example, this effect is likely to exist in our setting because participants who initiate alcohol use may be less likely to resist peer pressure in the future. Given this, it is important to understand how the standard method may fail in the presence of the dotted arrow and time-varying unmeasured indirect confounders, as in Panel C. This is the topic of the next section.

WHY THE STANDARD METHOD MAY FAIL IN LONGITUDINAL SETTINGS

Although it is possible to use the standard method when there is no unmeasured confounding (Fig. 1, Panel A), the standard method of adjusting for confounders cannot be used in the presence of unmeasured direct confounders (Fig. 1, Panel B). Thus, it is necessary to assume that unmeasured direct confounders do not exist. The purpose of this section is to show that, in addition, it may not be possible to use the standard method to adjust for confounding in the time-varying setting in the presence of unmeasured indirect confounders (Fig. 1, Panel C). Specifically, we show graphically why the standard method fails to adjust for confounding and, in fact, may cause additional bias when (1) unmeasured indirect confounders exist and (2) at least one of the observed variables through which unmeasured indirect confounding occurs is also affected by past levels of the predictor.

To understand how the standard method may fail, consider Fig. 2. Panel A of Fig. 2 shows a full structural system of relations among alcohol initiation, peer pressure resistance, marijuana initiation, and unobserved variables at two times. In this hypothetical scenario, every variable is affected by every other variable that precedes it, including the \mathbf{U}_t s, with the following two exceptions: \mathbf{U}_1 does not affect Alc_1 or Alc_2 directly, and \mathbf{U}_2 does not affect Alc_2 directly. In contrast, \mathbf{U}_1 and \mathbf{U}_2 can affect the predictors Alc_1 and Alc_2 indirectly via observed variables PPress_1 and PPress_2 . Therefore, Panel A of Fig. 2 presents a scenario in which both measured and unmeasured indirect confounding is present, but unmeasured direct confounders do not exist. In the time-varying setting, the standard method of adjusting for confounders in this scenario involves a regression model like Eq. (2), for the hazard probability of marijuana initiation at time t .

Including PPress_t in Eq. (2) is used to eliminate or help reduce the confounding bias due to the measured confounders PPress_1 and PPress_2 . Typically, β_1 is used to describe the total causal effect of alcohol initiation on marijuana initiation. However, when Conditions 1 and 2 are satisfied, conditioning on the measured confounder PPress_t may actually induce a spurious correlation between alcohol initiation and marijuana initiation, so that β_1 cannot be used to obtain the total causal effect of alcohol initiation on marijuana initiation.

⁸In addition, β_1 would be interpreted as the direct causal effect of alcohol initiation on marijuana initiation. For simplicity and reasons of space, this is not elaborated upon in this paper.

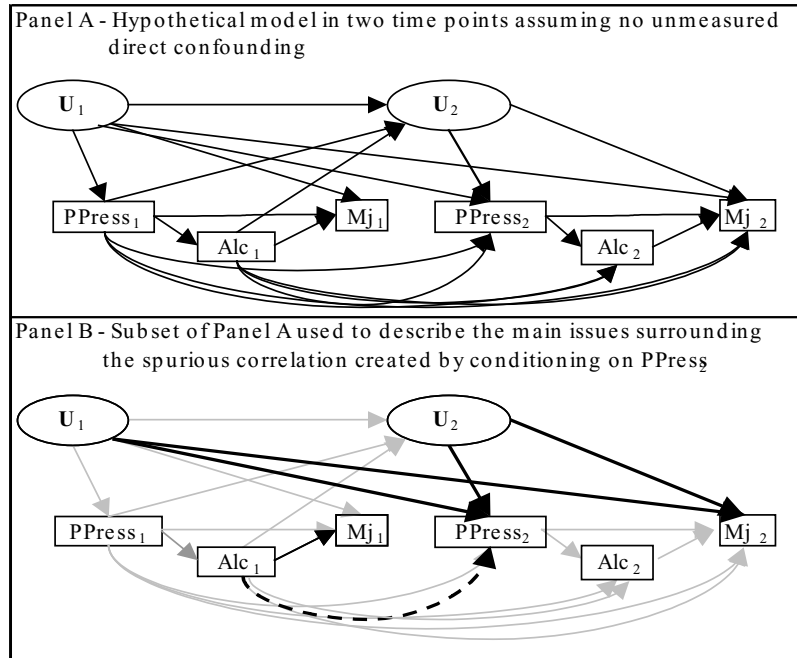


Fig. 2. The conditions under which adjusting for PPress₂ results in a spurious correlation.

To see how this spurious correlation arises, consider Panel B of Fig. 2, which shows a subset of the relations from Panel A. In particular, Panel B highlights Conditions 1 and 2, which can be seen from the solid arrows illustrating that U_1 and U_2 are unmeasured indirect confounders and the dashed arrow illustrating that Alc_1 affects the observed variable PPress₂.

Recall that the standard method includes Alc_t and PPress _{t} in a model for the hazard probability of marijuana initiation. The problem with the standard method is that by conditioning on *both* Alc_1 and PPress₂, as Eq. (2) does, additional noncausal paths from Alc_1 to Mj_2 become part of the contribution of the effect of Alc_1 on Mj_2 to the observed total causal effect of alcohol initiation on marijuana initiation, β_1 . This is seen in Panel B of Fig. 2 in that conditioning on PPress₂ has the effect of “opening” noncausal paths from Alc_1 to Mj_2 via both U_1 and U_2 . In other words, because U_1 and U_2 both affect PPress₂ (Condition 1) and Alc_1 affects PPress₂ (Condition 2), the observed effect of alcohol initiation on marijuana initiation (β_1) includes the correlations that make up the noncausal paths Alc_1 to PPress₂ to U_1 to Mj_2 and Alc_1 to PPress₂ to U_2 to Mj_2 . Therefore, the total causal effect of alcohol initiation on marijuana initiation cannot be based on β_1 . The noncausal associ-

ations (bias) incurred in β_1 as a result of conditioning on PPress _{t} is an example of Berkson’s Paradox (Berkson, 1946), from the epidemiological literature. This phenomenon is also related to Pearl’s (1998) *collider* issue.⁹

USING WEIGHTS TO ADJUST FOR TIME-VARYING CONFOUNDING

In this section, we describe a weighting method first developed by Robins (1999)¹⁰ for assessing total causal effects in the longitudinal setting. This method uses sample design weights to adjust for time-varying confounders of total causal effects.

⁹Here, PPress₂ is known as a collider in Panel B of Fig. 2 because it is affected by the U_s and Alc_1 . As Pearl notes, conditioning on colliders has the effect of inducing noncausal relations among parents of the collider (the U_s s and Alc_1) that are otherwise not related causally. Because we are unable to condition further on the unobserved or unknown variables (the U_s s) in our setting, the induced noncausal relations become part of the observed relation between the collider’s observed parent (Alc_1) and the response (Mj_2).

¹⁰The weighting method is also known as inverse-probability-of-treatment-weighting, or IPTW, where “treatment” refers to the exposure or putative cause.

Robins *et al.* (2000) and Joffe *et al.* (2004) provide a more detailed presentation and review, respectively, of the weighting method and its associated class of causal models, the *marginal structural models*.

One way to understand the weighting method is to consider a hypothetical randomized experiment—the gold standard for assessing total causal effects. In a randomized experiment, one would randomize adolescents to initiate or not initiate alcohol; the researcher would decide the probability of alcohol initiation, p_i , for each adolescent i in the sample. Additionally, the researcher is at liberty to randomize the full sample of adolescents using a common probability, say $p_i = .5$, or the researcher may choose to randomize within levels of particular baseline variables, such as sex and race: $p_i = P[\text{Alc}_i | \text{Sex}_i, \text{Race}_i]$.

Randomization would ensure balance in the alcohol initiating and noninitiating groups across all pre-alcohol initiation variables (including peer pressure resistance) across the full sample or within levels of particular baseline variables. Effectively, randomization would eliminate confounding by eliminating the relation between alcohol initiation and any pre-alcohol initiation variables. Corresponding to this design, the researcher would be happy to employ a simple model of analysis, such as a simple regression of marijuana initiation on alcohol initiation (Eq. (1)), to assess the total causal effect. (Or, in the case of randomization probabilities depending on Sex and Race, a simple regression of marijuana initiation on alcohol initiation within level combinations of Sex and Race.) That is, the researcher would not be inclined to adjust for peer pressure resistance or any other variable in a response regression model for the total effect of alcohol initiation on marijuana initiation.

The notion of a randomized experiment has a natural extension to studies with a time-varying predictor, where time is measured in discrete intervals. Essentially, the extension is an experiment in which randomization to initiate alcohol takes place at various time points. The effect of sequentially randomizing in this manner is to ensure balance in the alcohol initiating and noninitiating groups at every time point t across all covariates (observed or unobserved) occurring prior to time t .

The weighting procedure we present here can be used in observational longitudinal studies to mimic (under certain assumptions) a sequentially randomized experiment by creating balance in the alcohol

initiation and noninitiation groups at each time point. The weighting procedure achieves this, essentially, by creating a pseudo-sample in which the association between confounders and alcohol initiation is removed at every time point. To understand this, we discuss the weighting procedure using a graphical approach, before describing how to form the weights in more detail.

The goal of the weighting method is to create a pseudo-sample in which the arrows originating at PPress_t and ending at Alc_t at both times in Panel C of Fig. 1 are eliminated. That is, in our example, the goal of the weighting method is to use weights to equalize the composition of participants with different peer pressure resistance levels (and other confounders) within the two groups of initiators and noninitiators of alcohol, thereby mimicking the randomized experiment and breaking the relation between peer pressure resistance (and other confounders) and alcohol initiation. This idea is shown graphically in Panel A of Fig. 3.

Once weights have been used to create balance between the alcohol initiation groups, there is no longer a correlation between PPress_t and Alc_t at either time, just as in a randomized experiment. The result is a pseudo-sample in which peer pressure resistance (and other confounders) are not related to alcohol initiation, as illustrated in Panel B of Fig. 3. Then, given this new pseudo-sample, exhibiting the relations pictured in Panel B, the response regression model can take a simple form—the one described by Eq. (1).

In contrast to the standard method, the weighting method precludes the necessity of having to condition on time-varying measured confounders to control confounding. Instead, weights are created that balance the data in such a way that the relations between (measured and unmeasured indirect) confounders and alcohol initiation are eliminated at each time point. By eliminating this relation, it is unnecessary to include the measured confounders in the response regression model, avoiding the spurious correlation problem. That is, the researcher avoids suffering the consequences of Berkson's Paradox (Berkson, 1946); or equivalently, the researcher avoids having to condition on colliders (Pearl, 1988) that induce noncausal relationships. In sum, the weighting method adjusts for time-varying measured and unmeasured indirect confounders, whereas the standard method fails in the presence of unmeasured indirect confounders.

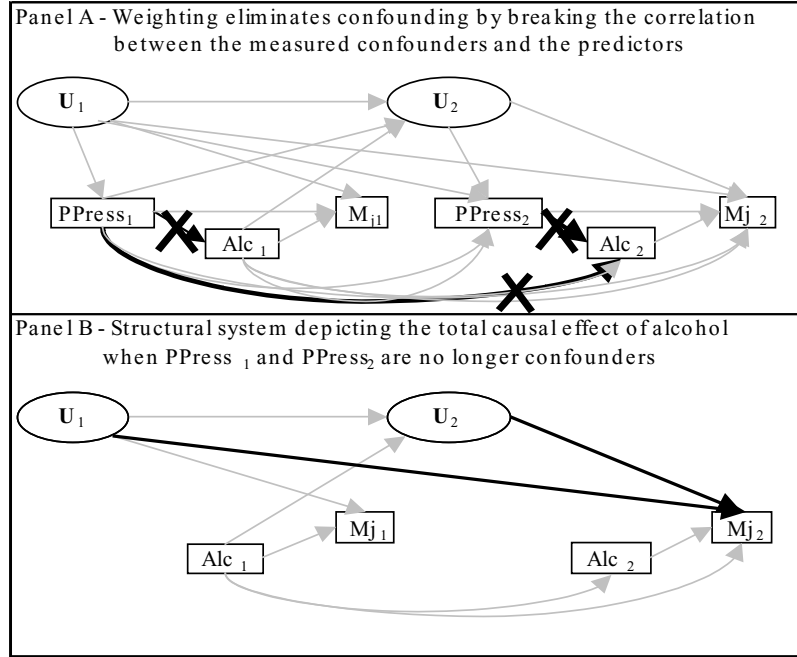


Fig. 3. Using the weighting method to control for confounding.

Creating Weights in the Time-Varying Setting

To create the weights, note that in a survival analysis setting, such as ours, each participant remains in the data set until he or she initiates marijuana use at some time t . For every time point t ($t = 1, 2, \dots, T$) that the participant is in the data set, we construct a weight that takes the form

$$W_t = \prod_{j=1}^T \frac{P[\text{Alc}_j | \overline{\text{Alc}}_{j-1}]}{P[\text{Alc}_j | \overline{\text{Alc}}_{j-1}, \overline{\text{Conf}}_j]}, \quad (3)$$

where Conf_t is a vector of time-varying covariates that are possible confounders at time t ; and the “overbars” represent history, so that $\overline{\text{Alc}}_{j-1} = (\text{Alc}_1, \text{Alc}_2, \dots, \text{Alc}_{j-1})$ and $\overline{\text{Conf}}_j = (\text{Conf}_1, \text{Conf}_2, \dots, \text{Conf}_j)$. The weights are essentially the product of ratios of propensity scores, where at each time point the propensity scores are a function of measured confounder history and alcohol initiation history. Observe that because alcohol initiation is itself a non-repeatable event, a discrete-time survival analysis model may be used to obtain the required numerator and denominator probabilities.

Following Robins (1999) and Robins *et al.* (2000), the probability of alcohol initiation given

prior alcohol initiation history is used in the numerator of W_t . According to Robins *et al.* (2000), the numerator may depend on any function of prior alcohol initiation history, including the unit function. Defining the weights this way, however, stabilizes the distribution of the W_t , which in turn, corresponds with a less variable estimator of the total causal effect. In addition, defining W_t this way provides a nice interpretation of the weights: observe that if Alc_j is independent of Conf_j (that is, Conf_j is not a confounder) at every time j up to time T , the $P[\text{Alc}_j | \overline{\text{Alc}}_{j-1}, \overline{\text{Conf}}_j] = P[\text{Alc}_j | \overline{\text{Alc}}_{j-1}]$ for every j , so that $W_t = 1$, implying that every adolescent in the sample contributes only one copy of himself or herself to the pseudo-sample, thereby leaving the sample unchanged.

The form of W_t may be modified slightly in a way that is dictated by the form of the final response model of interest, or equivalently, by the scientific question of interest. Put another way, the final form of the weights (in particular, the numerator) may be dictated by the ideal randomized trial the researcher is interested in simulating. For example, the research scientist may be interested in assessing the total causal effect of time-varying alcohol initiation on marijuana initiation within levels of Sex and Race. That is, the scientist may be interested in a

final response regression model of the form

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha_t + \gamma_1 \text{Sex} + \gamma_2 \text{Race} + \beta_1 \text{Alc}_t, \quad (4)$$

or some other linear function of the variables Alc_t , Sex , and Race , such as a model that explores how Sex alters the effect of, or interacts with, Alc_t . [Note that Eq. (4) is simply a modified version of Eq. (1) that includes Sex and Race as covariates.] In this case, the weights would take the following form:

$$W_t = \prod_{j=1}^T \frac{P[\text{Alc}_j | \overline{\text{Alc}}_{j-1}, \text{Sex}, \text{Race}]}{P[\text{Alc}_j | \overline{\text{Alc}}_{j-1}, \overline{\text{Conf}}_j, \text{Sex}, \text{Race}]}. \quad (5)$$

The effect of weighting the sample in this way is, essentially, to create a pseudo-sample in which confounding according to Conf_t is taken care of within levels of Sex and Race . That is, the weighted sample resembles data collected from a sequentially randomized trial in which the numerator probabilities are the randomization probabilities. Note that, in this case, any confounding bias due to Sex and Race is adjusted for by the response model itself. This is the strategy we employ, for example, in the next section of this paper in which we illustrate the weighting method with real data.

Observe that the final response regression model of Eq. (4) does not condition on time-varying confounders, but does condition on the baseline covariates. We may condition on the baseline covariates without suffering the effects of spurious correlations such as those described in the previous section. This is because the baseline variables are observed prior to any instance of alcohol initiation. The intuition here is that we should reweight the data to reflect the subpopulations of interest, described by the baseline covariates, in the final response regression model. Note that in Eq. (5), the effect is to balance the data with respect to an analysis that takes place within levels of combinations of Sex and Race .

AN EXAMPLE

In this section, we illustrate the weighting methodology by addressing the research question “Does delaying alcohol initiation lead to a delay in marijuana initiation?” That is, the goal is to assess the total causal effect of the timing of alcohol initiation (Alc_t) on the timing of marijuana initiation (Mj_t). Data from the Lexington Longitudinal Study,

a longitudinal study of etiological pathways to substance use, deviant behavior, and psychopathology, are used for this purpose.

Participants

The participants are a convenience sample¹¹ ($N = 210$) of a cohort who were part of the Lexington Longitudinal Study. Note that these data are for illustration purposes only and should not be used to make accurate inference about the substantive relation between alcohol and marijuana initiation. Participants were assessed via written questionnaires beginning in the 1987–88 school year, prior to starting the sixth grade (see Clayton *et al.*, 1996, for a detailed description of the initial recruitment and assessment procedures). Individuals in the current study completed this questionnaire and follow-up questionnaires on at least three of five data collection occasions (post-6th grade, 7th or 8th grade, and 9th or 10th grade), a mailed survey at age 19–20, and a laboratory protocol completed at age 20–21. Data for the current study were taken from the early school-based assessments and the most recent laboratory assessment.

Confounders

The following are the confounders (Conf_t), and their related measures, that are used in forming the weights. For detailed information on these measures see Clayton *et al.* (1991).

Peer Pressure Resistance (Time-Varying)

This 7-item scale was designed to measure the ability to resist negative peer pressure and higher scores indicate a stronger ability to resist or ignore peer pressure. Peer pressure resistance was measured six times over the course of the study; if a participant had a missing measurement, the last available measurement was carried forward to the subsequent time.¹²

¹¹The analyses presented here include only those participants who had no missing data on heart rate, performance IQ, verbal IQ, average sensation seeking, first peer pressure resistance measurement, and time of initiation of alcohol, cigarettes, conduct disorder, and other drug use. Therefore, this sample may not be representative of any subset of adolescents.

¹²Last observation carried forward (LOCF) imputation was used only for peer pressure resistance because it was the only

Conduct Disorder (Time-Varying)

Conduct disorder is a measure of the time period in which two or more different incidents of conduct problems in a single time period first occurred. Participants were asked about the presence of 14 specific conduct problems from the following four general areas: aggression against people or animals, destruction of property, deceitfulness or theft, and serious violation of rules.

Cigarette Use (Time-Varying)

Cigarette use is a measure of the time period in which the initiation of cigarette use first occurred.

Other Drug Use (Time-Varying)

Other drug use is a measure of the time period in which the initiation of any drug other than marijuana, alcohol, and tobacco such as cocaine, crack, inhalants, psychedelics, amphetamines, barbiturates, tranquilizers, heroin, or other analgesics first occurred.

Sensation Seeking

Sensation Seeking was measured using 18 items that were based on Zuckerman's (1994) 40-item sensation seeking scales. For the present analyses, scores are averaged across administrations to yield one overall Sensation Seeking score.¹³

IQ

IQ was assessed using two subtests from the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). Scores on the Vocabulary subtest served as indicators of Verbal IQ, whereas scores on

the Block Design subtest served as indicators of Performance IQ.

Heart Rate

Resting heart rate was obtained during the extensive laboratory assessment. Note that heart rate is believed to be related to both the predictor and response (i.e., a confounder) because, although resting heart rate has not been studied often in substance use, it is one of the most consistent correlates of antisocial behavior (Raine, 1993). It can be argued that lower resting heart rate is an index of fearlessness, which would predispose individuals to any number of dangerous or harmful activities or that low resting heart rate reflects autonomic underarousal, which might facilitate stimulation-seeking behavior. Thus, heart rate is likely a confounder of the relation between alcohol and marijuana initiation.

Variables may be considered confounders for a variety of theoretical reasons, such as the example of heart rate above. Although not necessary, it is possible to check if a set of variables is a common correlate of both the predictor and the response. Bray *et al.* (2003) demonstrate that the above list of variables is indeed a common correlate of both the timing of alcohol and marijuana initiation.

Time

Time is measured every third of a school year (fall, winter, summer), beginning in sixth grade for a total of 30 intervals. The variables that are time-varying are indicated by a subscript t . The predictor and response take on the value 0 prior to initiation, and take on the value 1 in the period of initiation and remain 1 thereafter. This is a survival analysis scenario in the sense that each participant has one person-period for each time until he or she initiates marijuana use or the study ends, whichever occurs first. That is, this is time-to-event data where the event of interest is the initiation of marijuana use, after which a participant is no longer at-risk for initiating marijuana use and drops out of the study.

Analyses

Three methods of estimating the total causal effect are considered. Each of these methods uses a modified version of Eq. (1) for the response

time-varying variable for which there was missing data among the participants in these analyses.

¹³Sensation Seeking is conceived as a relatively stable personality trait (Zuckerman, 1994), thus, the average score across assessments is used in these analyses. Empirically, sensation seeking is quite stable; in the larger sample from which these data are drawn, the 1-year stabilities for sensation seeking approach the maximum correlation possible given the reliabilities of the scales (average 1-year stability = .70).

regression model and is explained in detail below. The first method is the *Naïve Method*. This response regression model includes sex, race, alcohol initiation, and an intercept term for each school year, but omits confounders. Logistic regression is used to model the binary response, M_{jt} . The probability of initiating marijuana use at time t among those without previous marijuana use, p_t , is the quantity of interest. The naïve model is

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha \text{Schyr} + \beta_1 \text{Sex} + \beta_2 \text{Race} + \beta_3 \text{Alc}_t, \quad (6)$$

and the intercept term αSchyr represents

$$\alpha \text{Schyr} = \alpha_1 \text{Schyr}_1 + \alpha_2 \text{Schyr}_2 + \dots + \alpha_{10} \text{Schyr}_{10}. \quad (7)$$

The second method is the *Standard Method*. This response regression model also includes sex, race, alcohol initiation, and an intercept term for each school year. Additionally, the response regression model also includes all measured confounders as covariates. The standard model is

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha \text{Schyr} + \beta_1 \text{Sex} + \beta_2 \text{Race} + \beta_3 \text{Alc}_t + \gamma' \text{Conf}_t, \quad (8)$$

where the confounders are represented by

$$\text{Conf}_t = [\text{PPress}_t, \text{Odg}_t, \text{Hr}, \text{PIQ}, \text{VIQ}, \text{Asss}, \text{Cig}_t, \text{CD}_t]', \quad (9)$$

and the 8-vector of the confounder regression coefficients is represented by γ .

The third method is the *Weighting Method*. This response regression model is the same as the one used with the naïve method, but adjusts for confounders by weighting the response regression model (i.e., the model is fit via weighted regression). The form of the weights used with the weighting method is discussed in the previous section. A detailed explanation of the weight creation in this particular example is provided in the appendix.

Standard errors normally calculated using standard weighted-GLM software treat weights as fixed constants rather than as predicted values. To account for the sampling variation due to our estimation of the weights, we utilize the robust (or so-called, Huber-White or sandwich) standard errors (White, 1982) for the total causal effect estimates in the fi-

nal response model.¹⁴ This is easily implemented in SAS, using the REPEATED and SCWGT options in PROC GENMOD (Robins *et al.*, 2000).

Results

The convenience sample contains 121 female and 41 non-White participants, and 5,729 person-periods. The results obtained using the convenience sample are for illustration only and are *not* meant to be used for accurate inference about the substantive relation between alcohol and marijuana initiation. Descriptive statistics for peer pressure resistance, sensation seeking, IQ, and heart rate are displayed in Table 1. Table 2 displays coefficient estimates and the estimated odds of marijuana initiation, among those without prior marijuana initiation, for those who have initiated alcohol use. For simplicity, the coefficients for the intercepts and baseline variables are omitted in Table 2. An $\alpha = .05$ level of significance is used for all analyses. The p values reported for the coefficient estimates are based on one-tailed tests, consistent with the hypothesis stemming from the research question.

Examining Table 2, it is clear that, depending on the method used, answers to the question “Does delaying alcohol initiation lead to a delay in marijuana initiation?” are different. The naïve method implies that the estimated odds of marijuana use is significant and that prior-initiators of alcohol are more likely to initiate marijuana use than non-initiators (odds ratio = 5.10, $p < .001$); the odds of initiating marijuana for prior alcohol initiators is roughly five times higher than that for noninitiators of alcohol. However, the naïve method is almost certainly biased because it does not control for confounders such as peer pressure resistance.

Examining the results from the standard method, the odds are also highly significant (odds ratio = 2.28, $p = .002$). Yet, although it is the convention, the use of the standard method may produce bias because of the spurious correlations discussed earlier.

The weighting method is not subject to the spurious correlations discussed earlier, but it does control for confounding. Comparing the results of the

¹⁴Robins (1998) shows that confidence intervals according to the robust standard errors calculated in this way have coverage probability of at least 95%; that is, confidence intervals using this method are (possibly) conservative.

Table 1. Descriptive Statistics of Non-Time-Varying Variables and Peer Pressure Resistance

Variable	Frequency	Percent	Mean	SD	Minimum	Maximum
Sex						
Male	89	42.4				
Female	121	57.6				
Race						
White	169	80.5				
Non-White	41	19.5				
Heart Rate			70.3	11.8	48	111
Performance IQ			32.8	10.2	10	50
Verbal IQ			42.4	12.6	11	70
Average Sensation Seeking			50.4	9.4	24	78
Peer Pressure Resistance						
Prior to 6th Grade			40.2	6.5	9	45
Post-6th Grade			40.9	5.6	13	45
7th or 8th Grade			38.3	7.3	9	45
9th or 10th Grade			35.8	8.6	9	45
Mailed Survey			34.2	9.4	9	45
Laboratory Protocol			32.5	9.4	9	45

Note. Sample size for each variable in this table is $N = 210$.

weighing method with the other methods, the naïve method appears to overestimate the total causal effect of alcohol initiation on marijuana initiation and the standard method appears to underestimate the total causal effect. With the weighting method, among those without prior marijuana use, the odds of initiating marijuana for prior alcohol-initiators is estimated at roughly 3 and a third times higher than for noninitiators of alcohol (odds ratio = 3.36, $p < .001$). A significant alcohol coefficient is estimated with all three methods, but we believe that the desired interpretation of the total causal effect of alcohol initiation on marijuana initiation is best obtained from the weighting method. It is purely coincidental that the result from the weighting method falls in between the results from the naïve and standard methods. The weighting method may result in coefficients and p values of different magnitude when compared to the standard or naïve methods. It is important to note that the method of using sample weights is not meant to increase significance, but rather to construct an unbiased estimator of the total effect of delaying the timing of a predictor on the timing of a response. Therefore, the significance of a coefficient may change in either direction when examining weighted and unweighted coefficients. When the confounders are time-varying outcomes of past predictors, the estimate from the weighting method is the only one that is unbiased and is always preferred.

DISCUSSION

Using a graphical approach we have illustrated how the standard method may produce biased estimates of the total causal effect in longitudinal settings. Specifically, we have shown how the standard method may fail in the presence of unmeasured indirect confounders when past levels of the predictor are related to future levels of confounding variables. Further, we have illustrated the use of an alternate

Table 2. Response Regression Models with Alcohol Initiation as the Predictor and Marijuana Initiation as the Response

	Naïve	Standard	Weighted
Predictor			
Alcohol	1.63***	0.82**	1.21***
Odds	5.10	2.28	3.36
Time-varying confounders			
Conduct disorder		0.36	
Cigarettes		0.41	
Other drug use		1.30**	
Peer pressure resistance		-0.05***	
Non-time-varying confounders			
Heart rate		-0.01	
Verbal IQ		-0.03**	
Performance IQ		-0.01	
Average Sensation Seeking		0.02	

Note. Coefficients for intercepts and baseline variables are omitted. The Naïve and Standard Models do not include confounders by definition, see text.

** $p < .01$. *** $p < .001$ (one-tailed test).

approach to estimating total causal effects in longitudinal studies, namely the weighting method. The graphical approach was used to illustrate how the weighting method is able to estimate the total causal effect of a time-varying predictor in the presence of time-varying measured and unmeasured indirect confounders. Finally, an example application with comparisons of total causal effect estimates from three methods was presented. Several concerns about the weighting method, however, should be addressed.

Altering the Data by Weighting

Justifiably, many substantive researchers who collect their own data are unhappy with any method that appears to alter the data. It is essential to realize that the weighting method presented here does not utilize or alter the response, nor does the weighting method alter the predictor–response relation of any particular individual. That is, the true predictor–response relation is preserved in the pseudo-sample created by the weighting method. The weighting method changes the composition of participants so that certain predictor–response pairs have a higher weight and others have a lower weight. This is done to equalize the composition of types of participants between the predictor levels. As mentioned earlier, in an experimental study, randomization of participants to predictor levels equalizes the composition of participants between predictor levels on measured, unmeasured direct, and unmeasured indirect confounders. The weighting method attempts to mimic the effect of randomization.

Functional Forms and the “Task” of the Response Model

The weighting method does not presume a more complex functional form, as compared to the standard method, for the conditional mean of the response given the putative cause and confounder history. In fact, a response model using the weighting method is necessarily more parsimonious; the response model used with the weighting method resembles the more parsimonious model that would be used in the ideal randomized setting. With the weighting method, the weights are used to adjust for confounding bias, whereas the response model focuses only on modeling the total causal effect of interest. In contrast, the standard method “asks

more” of the response model—the response model is used both to adjust for confounding and to model the causal phenomena of interest. This is a powerful advantage of the weighting method over the standard method. Researchers are able to focus on adjusting for confounding and modeling in two separate stages of analysis.

Including Non-Confounders in the Weights

It is possible to check if a measured covariate by itself is a confounder by studying the relation between (a) the covariate and the putative cause (alcohol initiation) and (b) the covariate and the response (marijuana initiation) at each time point. On the basis of these preliminary bivariate analyses, the researcher may then decide whether or not to include the measured covariate in the weights. Even if the researcher failed to do this, however, including a non-confounder in the weights has no effect on the final analysis of the effect of the predictor on the response. The non-confounder will, essentially, make no contribution (in terms of participant-specific copies) to the pseudo (weighted) sample. For more on this property of the weights, see Barber *et al.* (2004).

The Weighting Method Versus Propensity Score Methods of Rosenbaum–Rubin

Propensity scores refer to the probability of being exposed (in our context, the probability of initiating alcohol) given past confounder history. Propensity scores are useful as a dimension reduction tool for causal inference that summarizes the effect of confounders on exposure to one score. Although the probabilities that make up the weights of the method presented here may be regarded as propensity scores, the weighting method differs in one important respect when compared to the popular propensity score based methods of Rosenbaum and Rubin (1984, 1985). The propensity score stratification and matching methods of Rosenbaum and Rubin are applicable only in the non-time-varying setting, and have no obvious extension to the time-varying setting (Imbens, 2000).

Implications for Prevention

The weighting method requires the assumption of no unmeasured direct confounders. In contrast,

the standard method requires the complete absence of unmeasured confounders (both direct and indirect).¹⁵ The main implications of these facts, and of this research more generally, are twofold. First, because the set of potential unmeasured confounders includes unknown variables (including perhaps variables representing scientific theories not yet discovered), we should guard against these confounders by using the weighting method when analyzing the effects of time-varying predictors in the presence of known time-varying confounders.

The second implication has a more proactive, design-related recommendation. Although it is impossible, in the absence of randomization, to ensure that the *no unmeasured direct confounders* assumption of the weighting method is met, we should try to collect data so as to ameliorate the effects of confounding bias due to unmeasured direct confounders. That is, we should not only collect the measures of variables that are thought to causally affect the response, but also the measures of variables that might causally affect the predictor. Because the predictor is time-varying, we can expect that many such variables will be time-varying as well. Doing so increases the likelihood of satisfying the main assumption of the weighting method by decreasing the set of potential unmeasured direct confounders.

APPENDIX: WEIGHT COMPUTATION

Complete details and explanation of generic SAS programming code that can be used for the weight creation, naïve method, standard method, and weighted method are available at <http://methodology.psu.edu/publications/tvpappen.html>. Also provided are two simulated datasets that allow for practice and analysis in conjunction with a review of the SAS code. At each measured time point, t , where a participant is at risk for response initiation (e.g., marijuana initiation) a weight component is created. Each weight component is the ratio of two predicted probabilities. The numerator is the predicted probability of a participant's observed predictor initiation or noninitiation in period t , given past predictor initiation status (e.g., alcohol initiation or noninitiation) and baseline variables (e.g., sex and race), for those still at risk of response initiation. The denominator is the predicted probability of a participant's

observed predictor initiation or noninitiation in period t , given past predictor initiation status, baseline variables, and confounders, for those still at risk of response initiation. Thus, the numerator and denominator models only differ in that the confounders are present in the denominator predicted probability. Note that if alcohol initiation (i.e., predictor initiation) occurs prior to time t , but before marijuana initiation (i.e., response initiation), the ratio at time t is 1 because both the numerator and denominator predicted probabilities are 1 (i.e., the predicted probability of initiating alcohol use after the initiation of alcohol use is 1 for any values of the confounders and baseline variables). Hence, the weight component does not need to be computed after predictor or response initiation (whichever occurs first). Because the numerator and denominator probabilities are computed only for those still at risk of predictor and response initiation, conditioning on $\overline{\text{Alc}}_{i-1}$ and $\overline{\text{Mj}}_{i-1}$ (complete past predictor and response patterns, respectively), as shown in Eq. (5), is not necessary. Note that conditioning on $\overline{\text{Mj}}_{i-1}$ in Eq. (5) is implicit because this is a survival analysis setting in which a participant is no longer in the dataset if the participant has initiated marijuana use; hence $\overline{\text{Mj}}_{i-1} = \mathbf{0}$ for all participants in the dataset at time i . Thus, the equations below do not include past Alc or Mj. The model for the numerator regression model is

$$\log\left(\frac{\text{numpr}_{ii}}{1 - \text{numpr}_{ii}}\right) = \alpha \text{Schyr} + \beta_1 \text{Sex}_i + \beta_2 \text{Race}_i, \quad (\text{A.1})$$

whereas the model for the denominator regression model is

$$\log\left(\frac{\text{denpr}_{ii}}{1 - \text{denpr}_{ii}}\right) = \alpha \text{Schyr} + \beta_1 \text{Sex}_i + \beta_2 \text{Race}_i + \theta \text{Conf}_{ii}. \quad (\text{A.2})$$

Thus, the weight component for participant i before alcoholinitiation is

$$\frac{1 - \text{numpr}_{ii}}{1 - \text{denpr}_{ii}}, \quad (\text{A.3})$$

and the weight component for participant i at alcohol initiation is

$$\frac{\text{numpr}_{ii}}{\text{denpr}_{ii}}. \quad (\text{A.4})$$

The weight at time t , W_t , is the product of these weight components up to time t . If at time $t - 1$

¹⁵These assumptions, concerning unmeasured variables, are by definition not testable.

participant i has yet to initiate alcohol use then the weight at time $t - 1$ is

$$W_{t-1} = \left(\frac{1 - \text{numpr}_{it}}{1 - \text{denpr}_{it}} \right)_{t-1} \dots \left(\frac{1 - \text{numpr}_{it}}{1 - \text{denpr}_{it}} \right)_1. \quad (\text{A.5})$$

If at time t participant i initiates alcohol use then the form of the weight at time t is

$$W_t = \left(\frac{\text{numpr}_{it}}{\text{denpr}_{it}} \right)_t \left(\frac{1 - \text{numpr}_{it}}{1 - \text{denpr}_{it}} \right)_{t-1} \dots \left(\frac{1 - \text{numpr}_{it}}{1 - \text{denpr}_{it}} \right)_1. \quad (\text{A.6})$$

The weight W_s for all times, s larger than t , remains equal to W_t as each weight component is now equal to 1. Each participant has a weight for each time point until either the participant initiates the response or the study ends.

ACKNOWLEDGMENTS

Preparation of this paper was supported by grant P50-DA-10075 from the National Institute on Drug Abuse to the Methodology Center at the Pennsylvania State University, by grant T32-DA-017629 to the Methodology Center and the Prevention Research Center for the Promotion of Human Development at the Pennsylvania State University, and by the National Institute on Drug Abuse award K02-DA-15674-01.

REFERENCE

- Agerbo, E. (2005). Effect of psychiatric illness and labour market status on suicide: A healthy worker effect? *Journal of Epidemiology and Community Health, 59*(7), 598–602.
- Allison, P. D. (1995). *Survival analysis using the SAS system: A practical guide*. Cary, NC: SAS Institute.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 434, 444–455.
- Barber, J. S., Murphy, S. A., & Verbitsky, N. (2004). Adjusting for time-varying confounding in survival analysis. *Sociological Methodology, 34*, 163–192.
- Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometric Bulletin, 2*, 47–53.
- Bohrnstedt, G. W., & Knoke, D. (1982). *Statistics for social data analysis*. Itasca, IL: P. E. Peacock Publishers.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bray, B. C., Zimmerman, R. S., Lynam, D., & Murphy, S. (2003). *Assessing the total effect of time-varying predictors in prevention research* (Technical Report 03-59). University Park, PA: The Methodology Center, Pennsylvania State University.
- Clayton, R. R., Cattarello, A., Day, L. E., & Walden, K. P. (1991). Persuasive communications and drug prevention: An evaluation of the D.A.R.E. program. In L. Donohew, H. E. Sypher, & W. J. Bukoski (Eds.), *Persuasive communication and drug abuse prevention* (pp. 279–294). Hillsdale, NJ: Erlbaum.
- Clayton, R. R., Cattarello, A. M., & Johnstone, B. M. (1996). The effectiveness of drug abuse resistance education (project DARE): 5-year follow-up results. *Preventive Medicine, 25*, 307–318.
- Cochran, W., & Rubin, D. (1973). Controlling bias in observational studies. *Sankya—The Indian Journal of Statistics, Series A, 35*(Dec), 417–446.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153–161.
- Heckman, J. J., & Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs—The case of manpower training. *Journal of the American Statistical Association, 84*, 408, 862–874.
- Hernán, M., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology, 11*(5), 561–570.
- Hernán, M., Brumback, B., & Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association, 96*, 454, 440–448.
- Imbens, G. (2000). The role of the propensity score in estimating dose–response functions. *Biometrika, 87*(3), 706–710.
- Joffe, M. M., Ten Have, T. R., Feldman, H. I., & Kimmel, S. E. (2004). Model selection, confounder control, and marginal structural models: Review and new applications. *American Statistician, 58*(4), 272–279.
- Little, R. J., & Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods, 3*(2), 147–159.
- Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science and Medicine, 58*(10), 1929–1952.
- Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods and Research, 27*, 226–284.
- Raine, A. (1993). *The psychopathology of crime: Criminal behavior as a clinical disorder*. San Diego, CA: Academic Press.
- Robins, J. M. (1998). Marginal structural models. *1997 proceedings of the American Statistical Association, section on Bayesian statistical science* (pp. 1–10). Retrieved from <http://www.biostat.harvard.edu/~robins/research.html>
- Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese, 121*(1/2), 151–179.
- Robins, J. M., Hernán, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*(5), 550–560.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 387, 516–524.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*(1), 33–38.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Singer, J. D., & Willet, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Education Statistics, 18*, 155–195.
- Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler Adult Intelligence Scale. *Contemporary Educational Psychology, 6*(2), 82–85.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–25.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, *18*, 327–350.

Winship, C., & Morgan, S. L. (1999). The estimation of causal ef-

fects from observational data. *Annual Review of Sociology*, *25*, 659–707.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge: Cambridge University Press.