

Running head: TIME-VARYING SUBGROUPS ANALYSIS

Subgroups Analysis when Treatment and Moderators are Time-varying

Daniel Almirall

Institute for Social Research

University of Michigan

Daniel F. McCaffrey, Rajeev Ramchand

RAND Corporation

Susan A. Murphy

Department of Statistics and

Institute for Social Research

University of Michigan

Address correspondence to:

Daniel Almirall

Institute for Social Research

University of Michigan

426 Thompson Street, Suite 2204

Ann Arbor , MI 48104-2321

e-mail: [dalmiral@umich.edu](mailto:dalmiral@umich.edu)

## Abstract

Prevention scientists are often interested in understanding characteristics of participants that are predictive of treatment effects because these characteristics can be used to inform the types of individuals who benefit more or less from treatment or prevention programs. Often, effect moderation questions are examined using subgroups analysis or, equivalently, using covariate  $\times$  treatment interactions in the context of regression analysis. This article focuses on conceptualizing and examining causal effect moderation in longitudinal settings in which both treatment and the putative moderators are time-varying. Studying effect moderation in the time-varying setting helps identify which individuals will benefit more or less from additional treatment services on the basis of both individual characteristics and their evolving outcomes, symptoms, severity, and need. Examining effect moderation in these longitudinal settings, however, is difficult because moderators of future treatment may themselves be affected by prior treatment (for example, future moderators may be mediators of prior treatment). This article introduces moderated intermediate causal effects in the time-varying setting, describes how they are part of Robins' Structural Nested Mean Model, discusses two problems with using a traditional regression approach to estimate these effects, and describes a new approach (a 2-stage regression estimator) to estimate these effects. The methodology is illustrated using longitudinal data to examine the time-varying effects of receiving community-based substance abuse treatment as a function of time-varying severity (or need).

## Subgroups Analysis when Treatment and Moderators are Time-varying

Rich longitudinal data in which treatments (exposures, or primary predictors) and their moderators, mediators, and outcomes are time-varying provide an opportunity for scientists to examine more interesting scientific questions than can be examined using cross-sectional data. Longitudinal treatment data allows scientists to examine the timing, duration, and sequencing effects of treatments on subsequent health outcomes. Further, this type of data allows scientists to examine how time-varying treatments exhibit their effects (time-varying causal effect mediation), and allows them to examine the types of subjects for whom time-varying treatments have stronger, weaker, opposing, or null effects (time-varying causal effect moderation). This article focuses specifically on the issue of conceptualizing and examining causal effect moderation in settings in which both treatment and the putative moderators are time-varying. To illustrate what we mean by time-varying causal effect moderation, consider our motivating data example which has measures available (a) on whether subjects do or do not receive community-based treatment for substance use over different time-intervals, (b) on symptom severity (or need for treatment) at baseline and at the end of each time-interval, and (c) on a primary end-of-study outcome, such as a measure of environmental risk for substance use. Treatment is expected to reduce environmental risk. Using these data, we are interested in examining sets of questions concerning the moderated time-varying effects of treatment on environmental risk, such as: “What is the impact of receiving treatment during months 1-3 on end-of-study environmental risk outcomes

as a function of baseline severity?” and “What is the impact of receiving treatment during months 7-9 (versus not receiving treatment) on end-of-study environmental risk outcomes as a function of baseline severity, treatment received between 1-3 months, and severity during months 4-6?”. These questions begin to address the distal and proximal incremental effects of additional substance use treatment conditional on the changing needs/severity of the subject. Examining these questions inform clinical practice by shedding light on whether to continue to provide substance-use treatment as a function of the changing needs or evolving symptomatology of the subject.

Studying effect moderation essentially involves examining the impact of treatment within different “subgroups” of participants defined on the basis of one or more covariates and, because of this, it is sometimes referred to as “subgroups analysis”. The focus of this article is to describe how to carry out subgroups analysis in settings in which subjects move in and out of treatment and subgroup composition changes over time (i.e., putative moderators are also time-varying).

Examining causal effect moderation in the time-varying setting is difficult because moderators of subsequent treatment may themselves be affected by prior treatment. For example, we would like to examine how symptom severity or the need for treatment between months 4-6 moderates the impact of treatment between months 7-9 on end-of-study environmental risk; however, severity or need during months 4-6 may itself depend on having received treatment during months 1-3. In these settings, traditional regression methods that adjust naively for time-varying moderators fail to estimate the time-varying causal effects of interest (Bray et al., 2006; Murphy et al., 2001; Robins et al., 2000; Robins, 1987, 1994).

In order to examine time-varying causal effect moderation properly, it is necessary to carefully define the quantities of scientific interest. In this article, we accomplish this by acknowledging the explicit temporal ordering between treatments/exposures, moderators, and outcomes, by relying on the language of potential outcomes to define causal effects, and by considering Robins' Structural Nested Mean Model (SNMM; Robins, 1994), described below, which serves as a guide for incorporating time-varying moderators in the regression framework.

The specific aims of this article are (1) to describe how the SNMM is used to frame scientific questions concerning causal effect moderation when both treatments and their putative moderators are time-varying, (2) to illustrate an application of the simple-to-use 2-stage regression estimator of the SNMM in the context of a substance abuse example of possible interest to prevention scientists, (3) to compare this estimator to the traditional regression estimator, and (4) to explicate these ideas in a simple time-varying setting, with just two time points, one binary treatment, one binary putative moderator at each time point, and a continuous end-of-study outcome.

### Moderated Intermediate Causal Effects

#### *Temporal Set-up and Notation*

Time is divided into 3-month intervals over the course of a year.  $a_1$  denotes receiving substance abuse treatment ( $a_1 = 1$ ) or not ( $a_1 = 0$ ) at some point during the first 3 months of the study;  $a_2$  denotes receiving treatment or not at some point during months 7-9. Therefore,  $(a_1, a_2)$  denotes one of four fixed treatment sequences an individual could receive during the 1-3 month interval and the 7-9 month

interval, respectively. For example,  $(a_1, a_2) = (1, 0)$  means treatment was received during months 1-3, but no treatment was received during months 7-9. For each individual  $i$ , let  $S_{0i}$  represent a binary indicator variable denoting high baseline (month 0) severity ( $S_{0i} = 1$ ) or low baseline severity ( $S_{0i} = 0$ ). Similarly, let  $S_{1i}(a_1)$  represent a binary indicator variable denoting severity during the 4-6 month interval. Let  $Y_i(a_1, a_2)$ —the primary outcome—denote a continuous, end-of-study (10-12 month interval) outcome measuring environmental risk under the treatment sequence  $(a_1, a_2)$ ; by end-of-study we mean occurring after  $a_2$ . Each subject  $i$  has four potential end-of-study outcomes  $\{Y_i(0, 0), Y_i(0, 1), Y_i(1, 0), Y_i(1, 1)\}$ , and two intermediate severity outcomes  $\{S_{1i}(0), S_{1i}(1)\}$ . For example,  $Y_i(1, 0)$  is the potential environmental risk value had subject  $i$  received treatment during the first 3 months, but not received treatment during the 7-9 month interval.  $S_{1i}(a_1)$  is similarly indexed by  $a_1$  because severity during the 4-6 month interval may be potentially affected by treatment during the first 3 months.

### *Defining the Moderated Intermediate Causal Effects*

In the potential outcomes framework (Holland, 1986; Rubin, 1974), causal effects are defined as contrasts between the potential outcomes  $Y_i(a_1, a_2)$  at different values of  $(a_1, a_2)$ . We define two sets of moderated intermediate causal effects of interest, one set of effects at each time point (for simplicity, the subscript  $i$  is sometimes omitted). The first set of causal effects is defined as  $\mu_{1i}(S_{0i}, a_1) = E(Y(a_1, 0) - Y(0, 0) \mid S_{0i}) = a_1 \times E(Y(1, 0) - Y(0, 0) \mid S_{0i})$ , the average effect of treatment sequence  $(a_1, 0)$  relative to treatment sequence  $(0, 0)$  within subgroups of  $S_{0i}$  (note that this effect is zero if  $a_1$  is zero). Since there are two subgroups (levels of  $S_{0i}$ ), we can express  $\mu_{1i}$  using a 2-dimensional vector of parameters  $\beta_1 = (\beta_{10}, \beta_{11})$

as  $\mu_{1i}(S_{0i}, a_1; \beta_1) = a_1 \times (\beta_{10} + \beta_{11}S_{0i}) = \beta_{10}a_1 + \beta_{11}a_1S_{0i}$ . This is the familiar covariate  $\times$  treatment interaction notation (Baron & Kenny, 1986):  $\beta_{10}$  is the mean of the individual causal effects  $Y_i(1, 0) - Y_i(0, 0)$  among the subgroup of subjects having low baseline severity ( $S_{0i} = 0$ ), whereas  $\beta_{10} + \beta_{11}$  is the mean of the individual causal effects among the subgroup having high baseline severity ( $S_{0i} = 1$ ). Both differences compare mean end-of-study environmental risk scores had all subjects within the given subgroup received treatment during the first 3 months and not received any during the 7-9 month interval  $(a_1, a_2) = (1, 0)$  versus had all subjects within the given subgroup not received treatment in either time interval  $(a_1, a_2) = (0, 0)$ . Note that if  $\beta_{11} = 0$  (no covariate-treatment interaction), then  $S_{0i}$  is not a moderator of the impact of  $a_1$  on  $Y(a_1, 0)$  because in this case, the average effect of  $(a_1, a_2) = (1, 0)$  versus  $(a_1, a_2) = (0, 0)$  does not differ by subgroups of  $S_{0i}$ .

The second set of causal effects is similarly defined as  $\mu_{2i}(S_{0i}, a_1, S_{1i}(a_1), a_2) = E(Y(a_1, a_2) - Y(a_1, 0) \mid S_{0i}, a_1, S_{1i}(a_1)) = a_2 \times E(Y(a_1, 1) - Y(a_1, 0) \mid S_{0i}, a_1, S_{1i}(a_1))$ , the average effect of treatment sequence  $(a_1, a_2)$  relative to  $(a_1, 0)$  within subgroups of  $(S_{0i}, a_1, S_{1i}(a_1))$  (as before, this effect is zero if  $a_2$  is zero). Since there are  $2^3 = 8$  subgroup combinations of  $(S_{0i}, a_1, S_{1i}(a_1))$ , it can be expressed as

$$\begin{aligned} \mu_{2i}(S_{0i}, a_1, S_{1i}(a_1), a_2; \beta_2) = \\ a_2 \times \left( \beta_{20} + \beta_{21}S_{0i} + \beta_{22}S_{1i}(a_1) + \beta_{23}S_{0i}S_{1i}(a_1) \right. \\ \left. + \beta_{24}a_1 + \beta_{25}S_{0i}a_1 + \beta_{26}a_1S_{1i}(a_1) + \beta_{27}S_{0i}a_1S_{1i}(a_1) \right) \quad (1) \end{aligned}$$

using an 8-dimensional vector of parameters  $\beta_2 = (\beta_{20}, \dots, \beta_{27})$  and covariate  $\times$  treatment notation. Specific linear combinations of  $\beta_2$  returns the mean of the individual causal effects  $Y_i(a_1, 1) - Y_i(a_1, 0)$  for specific subgroups of  $(S_{0i}, a_1, S_{1i}(a_1))$ . For example, among subjects with low severity at baseline

( $S_0 = 0$ ), who had not received treatment during the first three months ( $a_1 = 0$ ), and who had low severity during the 4-6 month interval under no previous treatment ( $S_{1i}(0) = 0$ ),  $\beta_{20}$  is the difference in mean end-of-study environmental risk scores had they all received treatment during the 7-9 month interval ( $a_2 = 1$ ) versus had they not ( $a_2 = 0$ ). As another example,  $\beta_{20} + \beta_{21}$  makes a similar comparison, but among the subgroup of subjects that had high severity at baseline.

### *The Structural Nested Mean Model*

Robins' Structural Nested Mean Model (SNMM; 1994) formally relates  $\mu_1$  and  $\mu_2$ , the causal effects of interest, to the conditional mean of  $Y(a_1, a_2)$  given  $(S_{0i}, S_{1i}(a_1))$ ; this will be important when considering how to estimate the  $\mu_t$ 's in the regression context. The SNMM is expressed as a telescoping sum as follows:

$$E(Y_i(a_1, a_2) \mid S_{0i}, S_{1i}(a_1)) = \beta_0 + \epsilon_{1i}(S_{0i}) + \mu_{1i}(S_{0i}, a_1; \beta_1) + \epsilon_{2i}(S_{0i}, a_1, S_{1i}(a_1)) + \mu_{2i}(S_{0i}, a_1, S_{1i}(a_1), a_2; \beta_2), \quad (2)$$

where the intercept  $\beta_0 = E(Y_i(0, 0))$  is the mean under no treatment,  $\epsilon_{1i}(S_{0i}) = E(Y_i(0, 0) \mid S_{0i}) - E(Y_i(0, 0))$  is the association between  $S_{0i}$  and the outcome had no subjects received treatment, and  $\epsilon_{2i}(S_{0i}, a_1, S_{1i}(a_1)) = E(Y_i(a_1, 0) \mid S_{0i}, S_{1i}(a_1)) - E(Y_i(a_1, 0) \mid S_{0i})$  is the association between  $S_{1i}$  and the outcome had subjects with characteristics  $(a_1, S_{0i})$  received no treatment during the 7-9 month interval. The  $\epsilon_{ti}$ 's are not necessarily causal because they are not contrasts in the potential outcomes of interest  $Y(a_1, a_2)$ . Rather, by definition, they define both causal and non-causal relationships between the  $S_t$  and the outcome, including associations between  $S_t$  and the outcome due to other variables (e.g., social support) related to both  $S_t$  and the outcome. Finally, the  $\epsilon_{ti}$ 's are, by definition, mean zero given the



past: that is,  $E(\epsilon_{2i}(S_{0i}, a_1, S_{1i}(a_1)) \mid S_{0i}) = E(E(Y_i(a_1, 0) \mid S_{0i}, S_{1i}(a_1)) - E(Y_i(a_1, 0) \mid S_{0i}) \mid S_{0i}) = E(E(Y_i(a_1, 0) \mid S_{0i}, S_{1i}(a_1)) \mid S_{0i}) - E(Y_i(a_1, 0) \mid S_{0i}) = 0$ ; similarly,  $E(\epsilon_{1i}(S_{0i})) = 0$ . As we describe below, this property of the  $\epsilon_t$ 's motivates the development of the 2-stage regression with residuals estimator.

### Estimation

Before presenting a 2-stage regression with residuals approach to estimating  $\mu_1$  and  $\mu_2$  in the SNMM, we first introduce the observed data and a pair of identifying assumptions, and we describe two problems with the traditional regression approach.

#### *Observed Data and Identifying Assumptions*

The observed data, in temporal order, is  $\{S_{0i}, A_{1i}, S_{1i}, A_{2i}, Y_i\}$ . For subject  $i$ , let  $A_{1i}$  and  $A_{2i}$  denote binary observed treatment. Both the traditional and 2-stage estimators below assume that apart from the putative observed moderator  $S_{0i}$ , there exist no other variables that directly impact both  $A_{1i}$  and  $Y_i$ ; and apart from the history of moderators and prior treatment  $(S_{0i}, A_1, S_{1i})$ , there are no other variables that directly impact both  $A_{2i}$  and  $Y_i$ . Under this untestable No Direct Confounders Assumption (and the Consistency Assumption; see Appendix A), which we assume henceforth, it is possible to identify  $\mu_1$  and  $\mu_2$  using the observed data. In a sequential multiple assignment randomized trial (e.g., Murphy, 2005), the No Direct Confounders Assumption is met by design; whereas in an observational study, such as with our illustrative data example, meeting this assumption requires substantive knowledge concerning the joint predictors of treatment and outcome. In observational study settings, additional baseline and time-varying, continuous and discrete, covariates could be included in  $S_t$  to achieve this assumption, but in our

example we omit doing this for expositional simplicity.

*Two Problems with The Traditional Regression Estimator*

We consider a traditional regression estimator that includes the main effects for each of the predictors ( $S_{0i}, A_{1i}, S_{1i}, A_{2i}$ ) and all possible interactions in a model for the conditional mean of  $Y$ . Since there are 4 binary predictors (2 dummy-coded treatment variables and 2 dummy-coded moderator variables), the saturated model has  $2^4 = 16$  parameters in total. It can be expressed as:

$$\begin{aligned} E(Y_i | S_{0i}, A_{1i}, S_{1i}, A_{2i}) = & \beta_0^* + \eta_1^* S_{0i} + A_{1i}(\beta_{10}^* + \beta_{11}^* S_{0i}) + S_{1i}(\eta_2^* + \eta_3^* S_{0i} + \eta_4^* A_{1i} + \eta_5^* S_{0i} A_{1i}) \\ & + A_2(\beta_{20}^* + \beta_{21}^* S_{0i} + \beta_{22}^* S_{1i} + \beta_{23}^* S_{0i} S_{1i} \\ & + \beta_{24}^* A_{1i} + \beta_{25}^* S_{0i} A_{1i} + \beta_{26}^* A_{1i} S_{1i} + \beta_{27}^* S_{0i} A_{1i} S_{1i}) \quad (3) \end{aligned}$$

Unfortunately, however, the regression terms  $A_1(\beta_{10}^* + \beta_{11}^* S_{0i})$  may not correctly describe how baseline severity moderates the causal effect of receiving treatment during the first 3 months on environmental risk ( $\mu_{1i}$ ); that is,  $\beta_1^* = (\beta_{10}^*, \beta_{11}^*)$  may not be equal to the causal parameters  $\beta_1 = (\beta_{10}, \beta_{11})$ . The problem arises because receiving treatment during the first 3 months of the study ( $A_{1i}$ ) is likely to impact (e.g., reduce, on average) whether or not a subject has high severity during the 4-6 month interval ( $S_{1i}$ ). For example,  $S_{1i}$  may be a mediator of the effect of  $A_{1i}$ .

Specifically, adjusting naively for  $S_{1i}$  (a possible outcome of whether or not the subject received treatment during the first 3 months) causes at least two problems with the interpretation of  $\beta_1^*$  in the traditional regression framework (Barber, Murphy, & Verbitsky, 2004; Bray et al., 2006; Greenland, Pearl, & Robins, 1999; Murphy et al., 2001; Pearl, 1998; Robins et al., 2000; Robins, 1987, 1994; Rosenbaum, 1984): **Problem 1.** First, by adjusting for  $S_{1i}$  in Equation 3, any

indirect effect of  $A_{1i}$  on  $Y_i$  that occurs through  $S_{1i}$  is not captured in  $\beta_1^*$ . This is a problem because our primary effect of interest is the total causal effect of  $A_{1i}$  on  $Y_i$  (any way in which it occurs) conditional on  $S_1$ . **Problem 2.** Secondly, by adjusting for  $S_{1i}$  in Equation 3, the estimated parameters  $\beta_1^*$  will include spurious associations that arise from non-causal associations (that is, not on the causal pathway) between  $S_{1i}$  and  $Y_i$ . This non-causal association is due to common causes of both  $S_{1i}$  and  $Y_i$ .

Importantly, these two problems are not due to confounders, which are defined as variables that influence both treatment receipt (the  $A_{ti}$ 's) and the outcome. Indeed, these two problems occur even in settings in which  $A_1$  and  $A_2$  are randomized (Robins, 1987, 1989, 1994, 1997).

Following Bray et al. (2006) and Barber et al. (2004), the hypothetical scenario in Figure 1 is intended to serve as a heuristic tool to better explain how the two problems listed above might manifest themselves in the context of our motivating example. (Figure 1 is a heuristic tool to help clarify ideas. We have purposefully left out certain arrows and variables from this figure (for example, we have left out the so-called “direct effect” of  $A_{1i}$  on  $Y_i$ ). The lack or presence of arrows in this figure does not necessarily imply assumptions we are making in our illustrative data analysis in the subsequent section.) In Figure 1, receiving substance abuse treatment during the first three months is beneficial in terms of reducing the probability of high severity between 4-6 months (negative path **a**). The figure also shows that having high severity during 4-6 months leads to higher levels of environmental risk between 10-12 months (the positive path **b**). Together, paths **a** and **b** imply an overall beneficial (negative) effect of  $A_{1i}$  on  $Y_i$ . Adjusting for  $S_{1i}$  as in Equation 3, however, essentially removes path **b** from the overall initial

treatment effect; this causes a bias toward zero, making the apparent, observed effect less negative.

To explain the second problem listed above, we consider a scenario where another variable that is not of particular interest (or is unknown, unmeasured, or unobserved and thus is not included) in our analysis, such as social support, is related to both  $S_{1i}$  and  $Y_i$ . (Note that for purposes of this explanation, we assume that social support is neither a moderator of interest nor a confounder of the impact of treatment received on environmental risk.) Suppose, as in Figure 1, that higher levels of social support leads to lower probability of high severity between 4-6 months (negative path **c**); and higher levels of social support also leads to lower levels of environmental risk between 10-12 months (negative path **d**). Naively adjusting for  $S_{1i}$  has the effect of “opening” or allowing the non-causal “backdoor path” (Pearl, 1998) between  $A_{1i}$  and  $Y_i$  via social support (path **a-c-d**; that is, the path  $A_{1i} \xrightarrow{a(-)} S_{1i} \xleftarrow{c(-)} \text{social support} \xrightarrow{d(-)} Y_i$ ), creating an apparent, overall positive spurious association between receiving treatment during the first 3 months  $A_{1i}$  and end-of-study environmental risk  $Y_i$ . To obtain the sign of the spurious association, we multiply the product of the signs along the backdoor path times  $-1$  since one of the paths involves a collider ( $S_{1i}$ ); the path analysis rules are given in Barber et al. (2004). Hence, in this example, naively adjusting for  $S_{1i}$  further compounds the downward bias that is due to the first problem listed above.

For a conceptual explanation of this spurious effect, we begin by describing how adjusting for  $S_{1i}$  creates a spurious negative correlation between  $A_{1i}$  and social support: Subjects with high 4-6 month severity ( $S_{1i} = 1$ ) who receive treatment ( $A_{1i} = 1$ ) are more likely to have had lower social support. This is likely the case

because treatment would otherwise have yielded low 4-6 month severity. Hence, high severity among treated cases must be due to low social support since it is the only other factor influencing severity. This implies that early treatment and social support will appear to be negatively associated (within levels of severity during 4-6 months). Since receiving treatment is spuriously associated with lower levels of social support (as a consequence of adjusting for  $S_{1i}$ ), and lower levels of social support imply higher levels of environmental risk, it follows that (in the data), receiving treatment is spuriously associated with higher levels of environmental risk. Hence, a spurious positive association between  $A_{1i}$  and  $Y_i$ .

These two problems cause bias in traditional regression estimates of the  $\mu_{1i}$  parameters. However, estimates of the parameters in  $\mu_{2i}$  are unaffected by these two problems because the regression does not adjust for post- $A_{2i}$  covariates (e.g., severity scores possibly impacted by  $A_{2i}$ ). More generally, in settings with more than two time points, these two problems affect all but the final set of parameters.

#### *A Simple-to-Use 2-Stage Estimator of the SNMM*

In the previous section, we discussed problems with the traditional regression estimator, Equation 3. In this section, we return to the SNMM, which, as we have developed previously, is the appropriate model to consider when there is interest in estimating the  $\mu_t$ 's in the context of a model relating  $Y_i$  and the covariates  $(S_{0i}, A_{1i}, S_{1i}, A_{2i})$ . In particular, we present a 2-stage regression estimator of the SNMM which, under the assumptions in Appendix A, provides unbiased estimates of the causal effects of interest, the  $\mu_t$ 's.

The key to the proposed 2-stage regression approach is the use of the residuals  $\delta_{S_{0i}} = S_{0i} - E(S_{0i})$  and  $\delta_{S_{1i}} = S_{1i} - E(S_{1i} | S_{0i}, A_{1i})$  to create models for  $\epsilon_1$  and  $\epsilon_2$ ,

respectively, which satisfy their conditional mean zero property under the SNMM. (We note that under the assumptions in Appendix A, the SNMM property that  $E(\epsilon_{2i}(S_{0i}, a_1, S_{1i}(a_1)) | S_{0i}) = 0$ , translates to  $E(\epsilon_{2i}(S_{0i}, A_{1i}, S_{1i}) | S_{0i}, A_{1i}) = 0$  in terms of the observed data.)

Stage 1 of the 2-stage approach is to estimate the residuals. Since in our motivating example, the putative moderators  $S_0$  and  $S_1$  are binary, the expectations in  $\delta_{S_{0i}}$  and  $\delta_{S_{1i}}$  are probabilities. That is,  $\delta_{S_{1i}} = S_{1i} - p_{1i}$ , where  $p_{1i} = Pr(S_{1i} = 1 | S_{0i}, A_{1i})$ ; similarly,  $\delta_{S_{0i}} = S_{0i} - p_0$ , where  $p_0$  is the proportion of subjects with  $S_{0i} = 1$ . Correspondingly, in our example,  $p_{1i}$  may take up to four values; and  $p_0$  is just one value. The probabilities  $p_{1i}$  can be estimated from a logistic regression such as  $\log(p_{1i}/(1 - p_{1i})) = \gamma_{20} + \gamma_{21}S_{0i} + \gamma_{22}A_{1i} + \gamma_{23}S_{0i}A_{1i}$  or, in our example, can be read-off from the  $2 \times 2 \times 2$  frequency table of  $(S_0, A_1, S_1)$ .

Stage 2 is to estimate the SNMM by fitting the following regression model:

$$\begin{aligned} E(Y_i | S_{0i}, A_{1i}, S_{1i}, A_{2i}) = & \beta_0 + \eta_1 \widehat{\delta}_{S_{0i}} + A_{1i}(\beta_{10} + \beta_{11}S_{0i}) + \widehat{\delta}_{S_{1i}}(\eta_2 + \eta_3 S_{0i} + \eta_4 A_{1i} + \eta_5 S_{0i}A_{1i}) \\ & + A_{2i}(\beta_{20} + \beta_{21}S_{0i} + \beta_{22}S_{1i} + \beta_{23}S_{0i}S_{1i} \\ & + \beta_{24}A_{1i} + \beta_{25}S_{0i}A_{1i} + \beta_{26}A_{1i}S_{1i} + \beta_{27}S_{0i}A_{1i}S_{1i}) \quad (4) \end{aligned}$$

where  $\widehat{\delta}_{S_{0i}}$  and  $\widehat{\delta}_{S_{1i}}$  are the estimated residuals taken from Stage 1. From Equation 4, we can see that  $\eta_1 \delta_{S_{0i}}$  forms the model for  $\epsilon_1$ ; and  $\delta_{S_{1i}}(\eta_2 + \eta_3 S_{0i} + \eta_4 A_{1i} + \eta_5 S_{0i}A_{1i})$  forms the model for  $\epsilon_2$ . Further, these are appropriate models for  $\epsilon_1$  and  $\epsilon_2$  since both  $E(\delta_{S_{1i}}(\eta_2 + \eta_3 S_{0i} + \eta_4 A_{1i} + \eta_5 S_{0i}A_{1i}) | S_{0i}, A_{1i})$  and  $E(\eta_1 \delta_{S_{0i}})$  equal zero. Intuitively, the 2-stage estimator works by purging (residualizing) the association of  $S_{ti}$  with its past (in particular  $A_{1i}$ , in the case of  $S_{1i}$ ) to get around the problems with the traditional regression estimator.

Note the similarities between the traditional regression estimator, Equation 3, and the 2-stage regression estimator, Equation 4: they differ only in that  $S_{ti}$  in Equation (3) is replaced by the residual  $\delta_{S_{ti}}$  in certain terms in models for the  $\epsilon_t$ 's. Not all  $S_{ti}$ 's need to be replaced by a residual; only those needed to ensure that the models for the  $\epsilon_t$ 's satisfy the conditional mean zero property. To see this, note that in Equation 3,  $E\left(S_{1i}(\eta_2 + \eta_3 S_{0i} + \eta_4 A_{1i} + \eta_5 S_{0i} A_{1i}) \mid S_{0i}, A_{1i}\right)$  may not be zero (e.g., if  $A_{1i}$  is associated with  $S_{1i}$ ); whereas we replace only  $S_{1i}$  by the residual  $\delta_{S_{1i}}$  in the analogous term in Equation 4, which as described above, ensures it has mean zero.

### Illustrative Data Example

The methodology is illustrated using observational study data ( $n = 2870$ ) pooled from a number of adolescent treatment studies funded by the Substance Abuse and Mental Health Services Administration's (SAMHSA's) Center for Substance Abuse Treatment (CSAT) involving adolescents entering community-based substance abuse treatment programs. All data points were collected using the Global Appraisal of Individual Needs (GAIN; Dennis et al., 2002), a structured clinical interview of client characteristics and functioning administered at admission (baseline) and every 3 months for the duration of a year. For the illustrative analyses, time-varying treatment  $A_{ti} = 1$  if a subject reported receiving inpatient treatment, outpatient treatment, or both at some point during the three-month interval; and  $A_{ti} = 0$  otherwise. For the time-varying moderator, subjects were categorized as  $S_{ti} = 1$  ("high severity") if continuous measures of environmental risk and substance use frequency were large based on a median cut-off;  $S_{ti} = 0$  ("low severity") otherwise. The continuous end-of-study outcome  $Y_i$

was based on the environmental risk measure which measures the amount of time spent with people in living, vocational, or social environments that used alcohol/drugs, were involved in illegal activity, who argued, who were not in school or work and who have never been in treatment. These analyses are meant to demonstrate the methods and issues in a simplified context. In fact, the No Direct Confounders Assumption is likely not met. In a more developed, albeit more complex, example, we would need to consider potential time-varying confounders and conduct a thorough investigation of assumptions.

### *Descriptives and Exploratory Data Analysis*

Of the  $n = 2870$  subjects in our sample, 87% reported receiving some treatment between 1-3 months, whereas 23% reported receiving treatment between 7-9 months. 41% were classified as having high severity (or “in need”) at baseline; this proportion diminished to 26% between 7-9 months. The overall mean end-of-study environmental risk was 34.5 (SD=21; median=30; range=(0,84)).

In order to explore  $\mu_{1i}$  using the data, we created Figure 2 using the subset of subjects not receiving any treatment between 7-9 months. (Recall that  $\mu_{1i}$  represents the effect of receiving some treatment between 1-3 months versus no treatment between 1-3 months, under no treatment between 7-9 months.) The exploratory data analysis (EDA) shown in Figure 2 suggests that some treatment (versus none) is beneficial in terms of lowering environmental risk. The effect appears to be strongest among the subgroup of participants with higher baseline severity. For example, within this subgroup, the EDA suggests a  $36.9 - 44.2 = -7.3$  point mean difference due to receiving some treatment.

The EDA for  $\mu_{2i}$  shown in Figure 3 (in this figure we use all of the data)



suggests a beneficial effect of receiving treatment between 7-9 months in terms of lowering environmental risk in all subgroups with the exception of the subgroup of subjects who had an increase in their severity between 4-6 months after experiencing low baseline severity and no treatment during 1-3 months. In this subgroup, treatment appears to be harmful on average, although the distributions show heavy overlap toward the high end of  $Y_i$ .

### *The Moderated Effects of Community-based Substance Treatment*

Table 1 reports estimates using the traditional regression estimator (Equation 3) and estimates of the SNMM using the 2-stage regression estimator (Equation 4). (For the 2-stage estimator, stage 1 estimates of the  $\gamma_2$  were as follows:  $(\gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (-1.51, 1.14, 0.17, -0.49)$ , suggesting that early treatment has the beneficial effect of lowering the probability of intermediate severity among the subgroup of subjects with high baseline severity.) As expected, the traditional estimator and the 2-stage estimator produce identical estimates for the  $t = 2$  moderated intermediate causal effects (and these results are also identical to the results reported in Figure 3). Within six of the 8 subgroups, receiving treatment during months 7-9 (versus not receiving treatment) lead to lower levels of environmental risk; within 3 of these subgroups,  $(S_0, A_1, S_1) = (0, 1, 0), (1, 1, 0)$  and  $(0, 1, 1)$ , the effect was statistically significant at the  $\alpha = 0.05$  level (inference based on the 2-stage estimator). The most beneficial effect was among  $(S_0, A_1, S_1) = (0, 1, 1)$ : effect size =  $(\beta_{20} + \beta_{22} + \beta_{24} + \beta_{26})/SD(Y) = -10.07/21 = -0.48; p < 0.01$ .

Estimates of the  $t = 1$  moderated causal effects from the traditional regression were biased toward zero relative to the estimates from the 2-stage estimator (showing a smaller beneficial effect of early treatment), as might happen due to the

types of biases explained above. The 2-stage regression estimates of the  $t = 1$  effects are more in-line with the EDA shown in Figure 2; for example, the 2-stage estimator reports an estimated effect of  $\beta_{10} + \beta_{11} = -2.69 + (-4.92) = -7.6$  (effect size =  $-0.36$ ;  $p < 0.01$ ) among the subgroup of participants with high baseline severity which is like the  $-7.3$  from the EDA, but very different from the estimate of  $-2.99 - 1.16 = -4.15$  reported by the traditional regression approach. Since the EDA in Figure 2 does not rely on a model that adjusts for  $S_{1i}$ , this comparison further illustrates that bias can occur when naively adjusting for  $S_{1i}$  using the traditional regression approach. The reason the 2-stage estimates of the  $t = 1$  effects are not identical to the EDA in Figure 2 is because the 2-stage estimator uses all of the study data to estimate the  $t = 1$  moderated causal effect, whereas the EDA shown in Figure 2 uses only subjects not receiving treatment between 7-9 months. This illustrates another advantage of the 2-stage estimator, which is that it allows us to use all of the study data to estimate the  $t = 1$  moderated effects, which in some cases may lead to more efficient estimates.

## Conclusions and Discussion

An overarching goal of this article was to shed light on how to frame the problem of subgroups analysis, or causal effect moderation, in settings in which the treatment or exposure is time-varying and so are the covariates hypothesized to moderate the effect of treatment. We discussed moderated intermediate causal effects in the context of Robins' Structural Nested Mean Model (SNMM). We discussed two problems associated with the traditional regression approach; we described an alternative simple-to-use 2-stage regression estimator, introduced

previously (Almirall, Ten Have, & Murphy, 2009); and we illustrated the methodology in the context of a substance use example.

To serve as an illustration of the methods and description of the key issues, we presented the 2-stage estimator in a simple setting with just two time points, one end-of-study outcome variable, and one binary moderator and one binary treatment variable at each time point. Indeed, we considered a saturated, or cell-means, SNMM throughout this article requiring no functional-form modeling assumptions. However, the 2-stage estimator is applicable in settings more general than those considered here (see Almirall et al., 2009), including settings with more than two time points, with categorical or continuous time-varying moderator or treatment variables, with longitudinal outcomes  $Y_{ti}$ , settings where prior levels of the primary longitudinal outcome itself is a moderator of the impact of subsequent treatment on future outcomes. Technically-oriented readers should reference Robins (1994) for more general information on, and other applications of, the SNMM.

Further, for expositional simplicity, we did not consider any moderators and treatments measured concurrently during the same time-interval, but this is not a limitation of the methodology. For example, using our data we could have considered a 3 time-point SNMM defining the effect of treatment during months 1-3 conditional on baseline measures ( $\mu_1$ ), the effect of treatment during months 4-6 conditional on moderators and treatment through the end of month 3 ( $\mu_2$ ), and the effect of treatment during months 7-9 conditional on moderators and treatment through the end of month 6 ( $\mu_3$ ). In general, when setting up SNMMs for examining time-varying effect moderation, an important step is to ensure that for each  $\mu_t$ , the moderators precede treatment, which, in turn, precede the outcome.

When using the 2-stage approach to examine a non-saturated SNMM, such as when  $S_t$  is continuous, additional parametric modeling assumptions are necessary concerning the true functional form of the  $\epsilon_t$ 's and  $\mu_t$ 's in order to obtain unbiased estimates of the causal effects  $\mu_t$ . Concerning the form of  $\epsilon_t$ , this requires first making assumptions about the distribution of  $S_t$  given the past, and second making assumptions concerning which features of this distribution to include in  $\epsilon_t$  (subject to the conditional mean zero property). In non-saturated SNMMs, Rubin's (1994) semi-parametric G-estimator provides unbiased estimates of the  $\mu_t$  if either models for the  $\epsilon_t$  are correctly specified or if models for the distribution of  $A_t$  given the past are correctly specified (true models for the form of the  $\mu_t$  are still required). The double-robustness property of the G-estimator may come at the price of higher variance, however, as suggested in simulation experiments in Almirall et al. (2009).

In practice, there may exist other time-varying variables that are not moderators of interest, but may be confounders of the effect of treatment. For instance, in our illustrative data example, there may exist time-varying variables, such as emotional or mental health, that directly influence both treatment receipt and environmental risk; in the presence of such variables, our estimates of the causal effects of treatment are biased due to confounding. One approach is to consider these additional variables as time-varying moderators in the SNMM; due to space and in order to keep the model simple, we did not attempt this approach in our example. This strategy, however, may not be feasible if the number of time-varying confounders is large. In forthcoming work, we describe how to address this issue in the context of the SNMM in settings where the number of time-varying confounders may far exceed the number of moderators of interest.

The 2-stage estimator can be implemented easily using standard statistical software packages (e.g., SPSS, SAS, R). Bootstrap methods can be used to obtain standard errors; asymptotic standard errors (ASE) have also been derived (Almirall et al., 2009). An R (R Development Core Team, 2009) function that executes the 2-stage estimator (in a setting more general than the one considered in this article) and provides estimates of the ASE is available at: <http://methcenter.psu.edu>.

The 2-stage estimator does not necessarily rely on sample sizes as large as what we used in our data example ( $n = 2870$ ). Because the 2-stage approach is similar to traditional multiple regression analysis, we conjecture that the power of tests would be similar to multiple regression.

The methodology described here, and its extensions, will be useful in both theoretical and applied prevention research in which there is interest in understanding the effects of time-varying exposures or treatments. In theoretical empirical research, the SNMM can be used, for example, to examine the effects of manipulable (and proximal) time-changing factors on distal outcomes—and to understand if and how these effects are strengthened or weakened over time given individual level changes or changes in context or environment. This research could inform conceptual models to guide the development of preventive interventions. In applied research, the SNMM can be used to learn about time-varying features that could be used to guide the timing of transitions from one prevention intervention component to the other, or the duration or dosage of interventions over time. Features found to be moderators using the SNMM could be used in the design of a sequential multiple assignment trial (Murphy, 2005), for example, for developing more optimal, adaptive, prevention interventions.

## References

- Almirall, D., Ten Have, T., & Murphy, S. (2009). Structural nested mean models for assessing time-varying effect moderation. *Biometrics*, *66*(1), 131-139.
- Barber, J., Murphy, S., & Verbitsky, N. (2004, November). Adjusting for time-varying confounding in survival analysis. *Sociological Methodology*, *34*(1), 163-192.
- Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.
- Bray, B., Almirall, D., Zimmerman, R., Lynam, D., & Murphy, S. (2006, March). Assessing the total effect of time-varying predictors in prevention research. *Prevention Science*, *7*(1), 1-17.
- Dennis, M. L., Titus, J. C., White, M. K., Unsicker, J. I., & Hodgkins, D. (2002). *Global Appraisal of Individual Needs: Administration guide for the GAIN and related measures*. Bloomington, IL. (<http://www.chestnut.org/li/gain>)
- Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*, 37-48.
- Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945-970.
- Murphy, S. (2005, May). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, *24*(10), 1455-1481.
- Murphy, S., Laan, M. J. van der, Robins, J. M., & CPPRG. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, *96*, 1410-1423.

- Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods and Research*, 27, 226-284.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. <http://www.R-project.org>. Vienna, Austria.
- 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 control of confounding by intermediate variables. *Statistics in Medicine*, 8, 679-701.
- Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics, Theory and Methods*, 23, 2379-2412.
- Robins, J. M. (1997). Latent variable modeling and applications to causality. In M. Berkane (Ed.), *Causal inference from complex longitudinal data* (p. 69-117). New York: Springer.
- Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5), 550-560.
- Rosenbaum, P. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, A*, 147(5), 656-666.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688-701.

## Appendix

## Identifying Assumptions

*No Direct Confounders Assumption*

Robins' (1994, 1997) "No Direct Confounders Assumption" is stated formally in two parts as follows:

1. the set of potential outcomes

$\{Y(a_1, a_2) : a_1, a_2 \in (0, 1)\} = \{Y(0, 0), Y(0, 1), Y(1, 0), Y(1, 1)\}$  is independent of  $A_1$  within levels of  $S_{0i}$ ; and

2. the set  $\{Y(a_1, a_2) : a_1, a_2 \in (0, 1)\}$  is independent of  $A_2$  within levels of  $S_{0i}$ ,  $A_{1i}$ , and  $S_{1i}$ .

*Consistency Assumption*

Robins' (1994, 1997) "Consistency Assumption" states that, for each subject  $i$ , the observed  $S_{1i}$  (observed under  $A_{1i}$ ) agrees with (or, is consistent with) the potential outcome indexed by the same treatment; that is,

$$S_{1i} = S_{1i}(A_{1i}) = A_{1i}S_{1i}(1) + (1 - A_{1i})S_{1i}(0). \quad (5)$$

Similarly, the consistency assumption states that, for each subject  $i$ , the observed outcome  $Y_i$  under  $A_{1i}$  and  $A_{2i}$  agrees with the potential outcome indexed by the same treatment values; that is,

$$\begin{aligned} Y_i = Y_i(A_{1i}, A_{2i}) &= A_{1i}A_{2i}Y_i(1, 1) + A_{1i}(1 - A_{2i})Y_i(1, 0) \\ &+ (1 - A_{1i})A_{2i}Y_i(0, 1) + (1 - A_{1i})(1 - A_{2i})Y_i(0, 0). \quad (6) \end{aligned}$$



Author Note

Funding for this work was provided by the following grants: R01-DA-015697 (McCaffrey), R01-DA-017507 (Ramchand), R01-MH-080015 (Murphy), and P50-DA-010075 (Murphy). The authors would like to thank Andrew R. Morral, Beth Ann Griffin, and Scott N. Compton for comments and suggestions, Cha-Chi Fan for guidance with the data, and three anonymous reviewers and the associate editor for helpful comments and suggestions.

Table 1

*Estimates of the parameters making up the moderated effects of time-varying substance abuse treatment on environmental risk as a function of time-varying severity.*

Term	Parameter	Traditional Estimator		2-Stage Estimator	
		EST	SE	EST	SE
Intercept	$\beta_0$	33.89	1.51	39.72	1.34
$A_1$	$\beta_{10}$	-2.99	1.65	-2.69	1.58
$A_1 S_0$	$\beta_{11}$	-1.16	3.25	-4.92	3.04
$A_2$	$\beta_{20}$	-6.69	3.69	-6.69	4.28
$A_2 S_0$	$\beta_{21}$	2.74	7.00	2.74	5.72
$A_2 S_1$	$\beta_{22}$	22.39	11.07	22.39	17.24
$A_2 S_0 S_1$	$\beta_{23}$	-28.16	14.56	-28.16	19.14
$A_2 A_1$	$\beta_{24}$	1.29	3.96	1.29	4.44
$A_2 S_0 A_1$	$\beta_{25}$	-0.25	7.33	-0.25	6.05
$A_2 A_1 S_1$	$\beta_{26}$	-27.06	11.54	-27.06	17.60
$A_2 S_0 A_1 S_1$	$\beta_{27}$	35.75	15.24	35.75	19.81

## Figure Captions

*Figure 1.* Figure used to illustrate two problems with adjusting naïvely for  $S_{1i}$ , a possible outcome of  $A_{1i}$ , in a traditional regression framework when interested in estimating the causal effect of  $A_{1i}$  on  $Y_i$ .

*Figure 2.* Exploratory Data Analysis: Box and whisker plots of environmental risk at the end-of-study by treatment during 1-3 months conditional on baseline severity. Filled discs denote the median. Filled triangles denote the mean; mean values are reported under each triangle.

*Figure 3.* Exploratory Data Analysis: Box and whisker plots of environmental risk at the end-of-study by treatment during 7-9 months conditional on baseline severity, treatment during 1-3 months, and severity during 7-9 months. Filled discs denote the median. Filled triangles denote the mean; mean values are reported under each triangle.





