NIH Public Access
Author Manuscript
*Stat Med*. Author manuscript; available in PMC 2015 September 28.

Published in final edited form as:
*Stat Med*. 2014 September 28; 33(22): 3905–3918. doi:10.1002/sim.6199.

# Causal Inference for Community-based Multi-layered Intervention Study

**P. Wu**[1], **D. Gunzler**[2], **N. Lu**[3,5], **T. Chen**[3], **P. Wymen**[4], and **X.M. Tu**[3,4,5]

[1]Value Institute, Christiana Care Health System, John H Ammon Medical Education Center, 4755 Ogletown-Stanton Road, Newark, DE 19718

[2]Center for Health Care Research & Policy, Case Western Reserve University at MetroHealth Medical Center, 2500 MetroHealth Drive, Cleveland, OH 44109-1998

[3]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642

[4]Department of Psychiatry, University of Rochester, Rochester, NY 14642

[5]Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, NY 14424

## Summary

Estimating causal treatment effect for Randomized Controlled Trials (RCTs) under post-treatment confounding, i.e., noncompliance and informative dropouts, is becoming an important problem in intervention/prevention studies when the treatment exposures are not completely controlled. When confounding is present in a study, the traditional Intention-to-treat (ITT) approach could under-estimate the treatment effect due to insufficient exposure of treatment. In the recent two decades, many papers have been published to address such confounders to investigate the causal relationship between treatment and outcome of interest based on different modeling strategies. Most of the existing approaches, however, are suitable only for standard experiments. In this paper, we propose a new class of structural functional response model (SFRM) to address post-treatment confounding in complex multi-layered intervention studies within a longitudinal data setting. The new approach offers robust inference and is readily implemented. We illustrate and assess the performance of the proposed SFRM using both real and simulated data.

### Keywords

Causal treatment effect; Noncompliance; Functional Response Models; Randomized Controlled Trials; Missing Data

## 1 Introduction

Although Randomized Controlled Trials (RCTs) remain as a benchmark for clinical research and practice, observational studies and semi-RCTs (trials that initiate treatment dynamically when needed) have become more popular, especially in studies in the behavioral and social sciences, health policy and health economics, because of the large amount of data generated by new web technologies and social media. Even within the confine of RCTs, we see more

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

community-based, multi-layered and multi-modal, dynamic interventions to take advantage of both static (e.g., genetic traits) and dynamic (e.g., treatment response) information during the treatment. In this paper, we focus on community-based multi-layered RCTs and introduce a new class of structural functional response models to address complex treatment noncompliance issues when evaluating treatment effects.

The proposed approach is motivated by a community-based multi-layered RCT–the Rochester Resilience Project (RRP), where post-treatment noncompliance arises from both the primary (subject) and supportive (support group) layer. The RRP is designed to promote behaviorally and emotionally healthy trajectories in 1st–3rd grade urban children who are showing aggressive-disruptive and school socialization problems, a group at elevated risk for future mental health disorders, substance abuse problems, reduced educational outcomes and costly services. The study involved 401 children randomized to the intervention and control groups. In addition, the study interventionists also worked with parents to teach children a set of skills to strengthen emotion self-regulation, adaptive social behavior and classroom conduct. Parent visits focus first on identifying parent goals for the child, then on introducing and preparing parents to use activity sets that teach and reinforce children's use of emotion self-regulation skills and incorporating those skills into their everyday relationship.

Our initial intention-to-treat (ITT) analyses failed to show any treatment effect for the primary behavior outcomes. Since ITT estimates are defined based on treatment assignment at randomization, rather than what actually goes on during the trial, such estimates completely ignore issues pertaining to violations of treatment protocols such as treatment noncompliance. For example, had only a small fraction of subjects in the intervention condition taken the treatment as prescribed, ITT would unduly underestimate the effect of receiving the intervention. However, child participation over 18 months was, as expected, high due to skill lessons being delivered in the school setting; 97% of children in the intervention condition completed all 14 lessons in the first year, and 81% completed all 10 lessons in the second year. Of the 39 non-completers, 33 were children relocating to non-study schools. Non-participation was unrelated to any baseline outcome measure.

Parent participation, however, was significantly lower; as shown in Table 1, with only 63.4% of parents (128 of 203 enrolled) participating in one or more intervention visits and few completing the 15 scheduled sessions. Under this condition of lower participation, ITT analyses are less informative about the true causal effects of parent involvement in the program, especially if the effect of treatment on child outcomes is achieved in part through parental participation.

A number of approaches for addressing treatment noncompliance in RCTs have been developed based on the counterfactual outcome framework, such as the instrumental variable [1], principal stratification [2] and structural mean models [3]. None of the available methods address treatment noncompliance in multi-layered intervention studies. In this paper, we develop a new approach to extend the principles in these approaches to this new setting with treatment noncompliance from multiple layers of the intervention. In Section 2, we briefly review the counterfactual outcome based causal framework and introduce a class

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of structural functional response model (SFRM) to address both pre- and post-treatment confounding. In Section 3, the SFRM is extended to address treatment noncompliance in multi-layered interventions within a longitudinal study setting. Simulation studies are presented in Section 4 to evaluate the performance of the proposed SFRM. In Section 5, we apply the approach to address the variability in parent participation in the two-layered RRP study. We conclude with a discussion in Section 6.

## 2 Structural Functional Response Models for Causal Inference

### 2.1 Counterfactual Outcomes

The concept of *counterfactual outcome*, the underpinning of the modern causal inference paradigm, addresses the fundamental question of causal treatment effect [4]. Under this framework, associated with every patient is a *potential outcome* for each treatment condition, and the treatment effect is defined by the difference between the outcomes in response to the respective treatments from the same individual, thereby free of any confounding effect and providing a conceptual basis for causal effect without relying on the notation of randomization.

For example, if the two potential outcomes for the *i*th child in the RRP Study are $y_{i1}$ and $y_{i0}$ for the intervention and control condition, the difference $\Delta_i = y_{i1} - y_{i0}$ is the treatment effect for the child. Since this difference is based on the outcomes from the same child, it must be the result of the intervention. Unfortunately, since only the outcome from the treatment condition actually assigned is observed, this difference is unobservable. A large part of the causal inference literature centers on how to estimate the *average*, or *population-level*, causal treatment effect, $\Delta = E(y_{i1} - y_{i0})$.

In RCTs, treatment assignment is independent of potential outcomes, i.e., $y_{ik} \perp z_i$, where $z_i$ denotes a binary indicator for treatment assignment and $\perp$ denotes stochastic independence. In this case, the average causal effect $E(y_{i1} - y_{i0})$ can be estimated by the difference between the two sample means from the intervention and control group:

$$\hat{\Delta} = \acute{\text{O}}\acute{\text{r}}_{\cdot 1} - \acute{\text{O}}\acute{\text{r}}_{\cdot 0}, \acute{\text{O}}\acute{\text{r}}_{\cdot 1} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{i_1 1}, \acute{\text{O}}\acute{\text{r}}_{\cdot 0} = \frac{1}{n_0}\sum_{i=1}^{n_0} y_{i_0 0}, \quad (1)$$

where $n_k$ denotes the number of subjects assigned to the *k*th treatment group such that $n = n_1 + n_0$ and $i_k$ denotes the *i*th subject within the *k*th treatment group. Note that $y_{i_k k}$ refers to the observed outcome for the $i_k$th subject in the assigned *k*th treatment, while $y_{ik}$ denotes the potential outcome corresponding to the *k*th treatment.

The above shows that standard statistical models such as linear regression and mixed-effects models can be applied to RCTs to infer causal treatment effects. Randomization is key to the transition from the unobserved individual level difference, $y_{i1} - y_{i0}$, to the estimable average treatment effect by the computable sample means in (1). For non-randomized trials such as most epidemiological studies, exposure to treatment or agent is non-random, in which case (1) generally does not estimate the average causal effect $\Delta = E(y_{i1} - y_{i0})$. Thus, associations found in observational studies generally do not imply causation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

### 2.2 Structural Functional Response Models

Since only one of the potential outcomes $y_{ik}$ is observable, we cannot model the $y_{ik}$'s directly using conventional regression models. One way around this is to model the observed outcomes such as $y_{i_k k}$ as in the preceding section. Alternatively, we can circumvent this difficulty by constructing an observable response based on the unobserved $y_{ik}$ and relate the response created to the mean of $y_{ik}$ as follows:

$$E\left(\frac{z_i^k(1-z_i)^{1-k}y_{ik}}{\pi^k(1-\pi)^{1-k}}\right)=\mu_k, E(z_i)=\pi, z_i=0,1, 1\le i\le n, k=0,1, \quad (2)$$

where $\mu_k = E(y_{ik})$ is the mean of potential outcome $y_{ik}$, since it is readily checked that

$$E\left(\frac{z_i^k(1-z_i)^{1-k}y_{ik}}{\pi^k(1-\pi)^{1-k}}\right)=\frac{1}{\pi^k(1-\pi)^{1-k}}E\left(z_i^k(1-z_i)^{1-k}y_{ik}\right)=\mu_k.$$

Although $y_{ik}$ are not both observed, the *functional response*,

$f(y_{i0}, y_{i1}, z_i)=\frac{z_i^k(1-z_i)^{1-k}y_{ik}}{\pi^k(1-\pi)^{1-k}}$, in (2) is still well defined. If $\pi$ is known as in most RCTs, it is unnecessary to model $z_i$ and (2) reduces to the first equation.

The model in (2) is not a conventional regression model such as the generalized linear or non-linear models, since $f(y_{i0}, y_{i1}, z_i)$ is not a single linear response such as $y_{ik}$ or $z_i$. Rather, this model is a member of the following class of functional response models (FRM):

$$E\left[f\left(y_{i_1},\ldots,y_{i_q},\boldsymbol{\theta}\right)\middle|\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_q}\right]=h(\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_q};\boldsymbol{\theta}),(i_1,\ldots,i_q)\in C_q^n, \quad (3)$$

where $f(\cdot)$ is some function, $h(\cdot)$ is some smooth function (e.g., continuous second-order derivatives), $y_i$ and $x_i$ denote some response and explanatory variables, $C_q^n$ denotes the set of $\binom{n}{q}$ combinations of $q$ distinct elements $(i_1, \ldots, i_q)$ from the integer set $\{1, \ldots, n\}$ and $\theta$ a vector of parameters. The response $f(y_{i1}, \ldots, y_{iq}, \theta)$ in (3) for the general FRM can be quite a complex function of multiple outcomes (e.g., $y_{ik}$, $z_i$ in (2)) from different subjects as well as unknown parameters $\theta$ (e.g., $\pi$ in (2)). By generalizing the response variable in this fashion, (3) provides a general framework for modeling a broad set of problems involving higher-order moments and between-subject attributes. The FRM has been applied to a range of methodological issues involving multi-subject responses such as extensions of the Mann-Whitney-Wilcoxon rank sum test to longitudinal and causal inference settings [5, 6], social network analysis [7, 8, 9], gene expression analysis [10], reliability coefficients [11, 12, 13, 14, 15, 16, 17] and complex response functions such as models for population mixtures [18] and structural equation models [19].

Because of its relationship to (3), the model in (2) will be referred to as the Structural FRM (SFRM):

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

$$E(f_{ik}(y_{i0}, y_{i1}, z_i))=h_{ik}(\boldsymbol{\theta}), f_{i1}=\frac{z_i y_{i1}}{\pi}, f_{i2}=\frac{(1-z_i)y_{i0}}{1-\pi}, f_{i3}=z_i, h_{i1}(\boldsymbol{\theta})=\mu_1, h_{i2}(\boldsymbol{\theta})=\mu_0, h_{i3}(\boldsymbol{\theta})=\pi, \quad (4)$$

where $\theta = (\mu_1, \mu_0, \pi)^\top$ denotes the collection of the parameters for this SFRM. Before adding more complexity to this SFRM to address treatment noncompliance within our context, let us first extend it to address selection bias in observational studies.

**2.2.1 Selection Bias by Pretreatment Confounders—**If subjects are not randomized with respect to the treatment condition (or exposure) as in observational studies (e.g., survey, epidemiologic studies), $y_{ik} \perp z_i$ is generally not true. In the presence of such *selection bias*, if $\mathbf{w}_i$ is a vector of covariates containing all sources of confounding such that the *ignorability condition* [20], $y_{ik} \perp z_i \mid \mathbf{w}_i$, holds, then we have:

$$E\left(\frac{z_i^k(1-z_i)^{1-k}y_{ik}}{\pi(\mathbf{w}_i)^k(1-\pi(\mathbf{w}_i))^{1-k}}\right)=E\left[E\left(\frac{z_i^k(1-z_i)^{1-k}y_{ik}}{\pi(\mathbf{w}_i)^k(1-\pi(\mathbf{w}_i))^{1-k}}|\mathbf{w}_i\right)\right]=\mu_k. \quad (5)$$

where $\pi(\mathbf{w}_i) = E(z_i \mid \mathbf{w}_i)$. We may model $z_i$ using a generalized linear model such as logistic regression:

$$E(z_i|\mathbf{w}_i)=\pi(\mathbf{w}_i;\boldsymbol{\eta}), \mathrm{logit}(\pi(\mathbf{w}_i;\boldsymbol{\eta}))=\boldsymbol{\eta}^\top\mathbf{w}_i, 1 \le i \le n. \quad (6)$$

By combing (5) and (6), we have the following SFRM to provide valid inference about $\theta = (\mu_1, \mu_0, \boldsymbol{\eta}^\top)^\top$ under selection bias:

$$\begin{aligned} &E(f_{ik}(y_{i0}, y_{i1}, z_i, \mathbf{w}_i)|\mathscr{F}_k) \\ &\quad =h_{ik}(\boldsymbol{\theta}), f_{i1} \\ &\quad =\frac{z_i y_{i1}}{\pi(\mathbf{w}_i;\boldsymbol{\eta})}, f_{i2} \\ &\quad =\frac{(1-z_i)y_{i0}}{1-\pi(\mathbf{w}_i;\boldsymbol{\eta})}, f_{i3} \\ &\quad =z_i, h_{i1}(\boldsymbol{\theta}) \\ &\quad =\mu_1, h_{i2}(\boldsymbol{\theta})=\mu_0, h_{i3}(\mathbf{w}_i;\boldsymbol{\theta}) \\ &\quad =\pi(\mathbf{w}_i;\boldsymbol{\eta})), \mathrm{logit}(\pi(\mathbf{w}_i;\boldsymbol{\eta})) \\ &\quad =\boldsymbol{\eta}^\top\mathbf{w}_i, \mathscr{F}_1 \\ &\quad =\mathscr{F}_2=\{0\}, \mathscr{F}_3=\{\mathbf{w}_i\}, 1 \le i \le n, \end{aligned} \quad (7)$$

where $\mathscr{F}_k = \{0\}$ ($k = 1, 2$) denotes the sigma field generated by the constant 0 and $\mathscr{F}_3 = \mathbf{w}_i$ denotes the sigma field generated by $\mathbf{w}_i$. Note that $E(f_{ik}(y_{i0}, y_{i1}, z_i, \mathbf{w}_i) \mid \mathscr{F}_k) = E(f_{ik}(y_{i0}, y_{i1}, z_i, \mathbf{w}_i))$, since $\mathscr{F}_k$ is contained in $\mathscr{F}_3$ for $k = 1, 2$ (e.g., see Kowalski and Tu [12]).

**2.2.2 Treatment Noncompliance as Post-treatment Confounders—**In many RCTs, even well-planned and executed ones, treatment effect may be significantly modified by levels of exposure of intervention (e.g., compliance or dosage) due to treatment noncompliance. One popular approach for addressing this primary post-treatment confounder is the structural mean model (SMM)[3]. Other competing approaches also address treatment noncompliance such as the Instrumental Variable[1] and Principal

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Stratification methods[2]. However, only SMM models treatment compliance on a continuous scale, which is more appropriate for session attendance within our context. We first frame this model within the FRM framework and then discuss its extensions to accommodate multi-layered interventions and missing data in Section 3.

Consider a randomized medication vs. placebo study and let $d_{i1}$ denote a continuous potential outcome of medication use, if the $i$th subject is assigned to the medication condition. The SMM models the dose effect on treatment difference as follows:

$$E(y_{i1} - y_{i0}|d_{i1}) = g(d_{i1}), \quad (8)$$

where $g(\cdot)$ is known up to a set of parameters (i.e., only the functional form of $g(d_{i1})$ is specified). However, the above model cannot be fit directly using conventional statistical methods, since only one of the potential outcomes ($y_{i1}$, $y_{i0}$) is observed. For RCTs, $y_{i1}, y_{i0} \perp z_i$ and it follows that:

$$E(y_{i1}|d_{i1}, z_i = 1) = g(d_{i1}) + E(y_{i0}|d_{i1}, z_i = 0). \quad (9)$$

By conditioning on the assigned treatment $z_i = k$, $y_{ik}$ in (9) represents the observed outcome from the $k$th treatment group ($k = 0, 1$). Thus, $E(y_{i0} \mid d_{i1}, z_i = 0)$ cannot be modeled directly, since $d_{i1}$ is not observed for the subjects assigned to the placebo condition.

If treatment compliance is tracked for the subjects in the placebo group, then $d_{i0}$, the potential outcome of placebo use if the subject is assigned to the placebo condition, is observed. Because of randomization and the fact that subjects cannot distinguish between medication and placebo, $d_{i0}$ has the same distribution as $d_{i1}$. Thus, we may replace $d_{i1}$ by $d_{i0}$ in $E(y_{i0} \mid d_{i1}, z_i = 0)$ to re-express (9) as:

$$E(y_{i1}|d_{i1}, z_i = 1) = g(d_{i1}) + E(y_{i0}|d_{i0}, z_i = 0). \quad (10)$$

Under this *treatment compliance explainable* condition, we will be able to model the right side to obtain estimates of dose-response relationships $g(d_{i1})$ [21].

Although applicable to medication studies, the SMM in (10) in general does not apply to psychosocial research. Many psychosocial intervention studies do offer attention or information controls and subjects in such control groups may also be tracked for their session attendance. However, unlike medication studies, compliance observed in the control group $d_{i0}$ generally does not have the same distribution as $d_{i1}$. For example, consider a HIV prevention intervention study for teenage girls at high risk for HIV infection, in which the intervention condition contains information on HIV infection, condom use and safe sex, while the control condition consists of nutritional and dietary information. Subjects with high compliance in the intervention group are generally different from their counterparts in the control condition; sexually active girls may form a majority of those with high attendance in the intervention group, while such girls might have low attendance rates, had they been assigned to the control condition. Thus, when assessing the effect of prevention intervention using outcomes of HIV risk behavior such as number of unprotected vaginal

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

sex over the past month, it is not meaningful to compare compliant subgroups between the two treatment conditions.

Thus for psychosocial research studies, we cannot simply replace $d_{i1}$ in $E(y_{i0} \mid d_{i1}, z_i = 0)$ by a measure of treatment compliance such as session attendance in the control group $d_{i0}$ as in medication trials. In many studies, it is reasonable to assume that there is sufficient information to predict $d_{i1}$, i.e., given a set of covariates $\mathbf{x}_i$, $d_{i1}$ is independent of $y_{i0}$. For example, if $\mathbf{x}_i$ contains information on sexuality and other information on a subject's interest to attend sessions in the intervention condition of the HIV study example above, $y_{i0}$ may no longer depend on $d_{i1}$ given $\mathbf{x}_i$. In this case, $E(y_{i0} \mid d_{i1}, \mathbf{x}_i, z_i = 0) = E(y_{i0} \mid \mathbf{x}_i, z_i = 0)$. Thus, under this *ignorability condition*, $y_{i0} \perp d_{i1} \mid \mathbf{x}_i$, (9) becomes:

$$E(y_{i1}|\mathbf{x}_i, d_{i1}, z_i=1)=g(d_{i1})+E(y_{i0}|\mathbf{x}_i, z_i=0). \quad (11)$$

Note that the SMM in this case is essentially the same as the Principal Stratification Model, except that it requires neither discretization of $d_{i1}$ nor parametric distribution models for $y_{ik}$, since (11) only specifies the conditional mean of $y_{ik}$ given $d_{i1}$, $\mathbf{x}_i$ and $z_i$.

By modeling $E(y_{i0} \mid \mathbf{x}_i, z_i = 0)$ and casting (11) in the form of FRM, we obtain the following SFRM for modeling treatment compliance measured by a continuous dose variable $d_{i1}$ (for the intervention condition only):

$$\begin{aligned} &f_{i1}=\frac{z_i y_{i1}}{\pi}, f_{i2}=\frac{(1-z_i)y_{i0}}{1-\pi}, f_{i3}=z_i, 1 \le i \le n, \\ &h_{i1}(\mathbf{x}, \boldsymbol{\beta})=h(\mathbf{x}_i, \boldsymbol{\beta}), h_{i2}(\mathbf{x}_i, d_{i1}, \boldsymbol{\theta})=g(d_{i1}, \gamma)+h(\mathbf{x}_i, \boldsymbol{\beta}), \boldsymbol{\theta}=(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \pi)^\top, \end{aligned} \quad (12)$$

where $h(\mathbf{x}, \boldsymbol{\beta})$ ($g(d, \boldsymbol{\gamma})$) is some function of $\mathbf{x}$ ($d$) parameterized by $\boldsymbol{\beta}$ ($\boldsymbol{\gamma}$). As before, $n$ is the sample size of the study, i.e., the sample size of the intervention plus the control group. Although for RCTs it is not necessary to include $\pi$ as a parameter, the general SFRM in (12) allows us to extend this model to observational studies. For example, for non-randomized studies, $y_{ik} \perp z_i$ in general is not true. If $y_{ik} \perp z_i$ holds conditional on a set of covariates $\mathbf{w}_i$ (possibly overlapping with $\mathbf{x}_i$), then by modeling $\pi$ as a function of $\mathbf{w}_i$ as in (6), the following SFRM still provides consistent estimates in the face of selection bias:

$$\begin{aligned} &f_{i1}=\frac{z_i y_{i1}}{\pi(\mathbf{w}_i;\boldsymbol{\eta})}, f_{i2}=\frac{(1-z_i)y_{i0}}{1-\pi(\mathbf{w}_i;\boldsymbol{\eta})}, f_{i3}=z_i, 1 \le i \le n, \\ &h_{i1}=h_{i1}(\mathbf{x}_i, \boldsymbol{\beta}), h_{i2}(\mathbf{x}_i, d_{i1}, \boldsymbol{\beta}, \boldsymbol{\gamma})=g(d_{i1}, \boldsymbol{\gamma})+h_{i1}(\mathbf{x}_i, \boldsymbol{\beta}), \\ &h_{i3}=\pi(\mathbf{w}_i;\boldsymbol{\eta}), \mathrm{logit}(\pi(\mathbf{w}_i;\boldsymbol{\eta}))=\boldsymbol{\eta}^\top \mathbf{w}_i, \boldsymbol{\theta}=\left(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\eta}^\top\right)^\top. \end{aligned} \quad (13)$$

We can model $h(\mathbf{x}_i, \boldsymbol{\beta})$ and $g(d_{i1}, \boldsymbol{\gamma})$ in various ways. For example, we may model both as a linear function: $h_1(\mathbf{x}_i, \boldsymbol{\beta})=\mathbf{x}_i^\top \boldsymbol{\beta}$ and $g(d_{i1}, \boldsymbol{\gamma}) = d_{i1}$. By specifying an appropriate form for $g(d_{i1}, \boldsymbol{\gamma})$, we may also extend (12) to non-continuous dose variables such as categorical variables. Further, by appropriately specifying $h_1(\mathbf{x}_i, \boldsymbol{\beta})$ and $h_2(\mathbf{x}_i, d_{i1}, \boldsymbol{\beta})$, we can also generalize (12) to non-continuous responses. For example, for a binary $y_i$, we may specify $h_1(\mathbf{x}_i, \boldsymbol{\beta})$ and $h_2(\mathbf{x}_i, d_{i1}, \boldsymbol{\beta})$ as follows:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$h_1(\mathbf{x}_i,\boldsymbol{\beta})=\text{logit}^{-1}\left(\mathbf{x}_i^\top\boldsymbol{\beta}\right), h_2(\mathbf{x}_i,d_{i1},\boldsymbol{\beta})=\text{logit}^{-1}(g(d_{i1},\boldsymbol{\gamma})+h_1(\mathbf{x}_i,\boldsymbol{\beta})).$$

**2.2.3 Inference for Structural Functional Response Models—**We focus on inference about $\theta = (\beta^\top, \gamma^\top, \eta^\top)^\top$ for the SFRM in (13), from which (7) and (12) follow as a special case. Let

$$\mathbf{f}(\mathbf{y}_i;z_i)=(f_{i1},f_{i2},f_{i3})^\top, \mathbf{h}_i(\boldsymbol{\theta})=(h_{i1},h_{i2},h_{i3})^\top, 1\le i\le n,$$

where $f_{ik}$ and $h_{ik}$ are defined in (13). Then, consistent estimates of $\theta$ are readily obtained by using the Generalized Estimating Equations (GEE) for FRM [18, 12, 22]:

$$\mathbf{U}(\boldsymbol{\theta})=\sum_{i=1}^{n}D_iV_i^{-1}S_i=\mathbf{0}, S_i=\mathbf{f}_i-\mathbf{h}_i, D_i=\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{h}_i,\ V_i=A_i^{\frac{1}{2}}R(\boldsymbol{\alpha})A_i^{\frac{1}{2}}, A_i=diag_t(\,Var(f_{it}|\mathscr{F}_{it})), \quad (14)$$

where $R(\alpha)$ denotes a choice of working correlation matrix.

The choice of $R(\alpha)$ and associated properties for the GEE estimate of $\theta$ have been extensively discussed in the literature, which are stated for ease of reference without justifications [23, 24]. In particular, the GEE estimate may not be consistent in the presence of time-varying covariates under working correlation structures other than the working independence model [23]. Thus, the working independence model may be used in general to ensure valid inference. Although this simple working correlation structure may incur some loss of efficiency for time-dependent covariates [24] and thus other models such as the uniform compound symmetry matrix may be used in some specific applications to improve power, it suffices for the purpose of illustrating the proposed approach. We focus on the working independence model in what follows unless otherwise stated.

## 3 Extension to Complex Studies

We first extend the SFRM in Section 2 to longitudinal data and then to multi-layered intervention studies.

### 3.1 Longitudinal Data with Missing Values

Let $\mathbf{y}_{it} = (y_{it1}, y_{it0})^\top$ ($\mathbf{x}_{it}$) denote the potential outcomes of $y_{it}$ (a vector of explanatory variables) of interest with $i$ ($t$) indexing the subject (assessment time) for $1 \le i \le n$ and $1 \le t \le T$. By applying (13) to each time point, we obtain a longitudinal version of the SFRM:

$$\begin{aligned}&\mathbf{f}_i=\left(\mathbf{f}_{i1}^\top,\ldots,\mathbf{f}_{iT}^\top,z_i\right)^\top, \mathbf{h}_i=\left(\mathbf{h}_{i1}^\top,\ldots,\mathbf{h}_{iT}^\top,\pi_i\right)^\top, E(\mathbf{f}_i|\mathbf{x}_i)=\mathbf{h}(\mathbf{x}_i,\boldsymbol{\theta}), 1\le i\le n,\\ &\mathbf{f}_{it}=(f_{it1},f_{it2})^\top, f_{it1}=\frac{z_i}{\pi_i}y_{it1}, f_{it2}=\frac{1-z_i}{1-\pi_i}y_{it0}, \mathbf{h}_{it}=(h_{it1},h_{it2})^\top,\\ &h_{it1}=h_1(\mathbf{x}_{it},\boldsymbol{\beta}), h_{it2}=g_t(d_{i1},\boldsymbol{\gamma})+h_{it1}, \pi_i=\text{logit}^{-1}\left(\boldsymbol{\eta}^\top\mathbf{w}_i\right), \boldsymbol{\theta}=\left(\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\eta}^\top\right)^\top.\end{aligned} \quad (15)$$

Inference for the FRM above is based on the following GEE for FRM [18, 12, 22]:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\mathbf{U}(\boldsymbol{\theta})=\sum_{i=1}^{n} D_i V_i^{-1} S_i=\mathbf{0}, S_i=\mathbf{f}_i-\mathbf{h}_i, D_i=\frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{h}_i,\ V_i=A_i^{\frac{1}{2}} R(\boldsymbol{\alpha}) A_i^{\frac{1}{2}}, A_i=diag_t(Var(\mathbf{f}_{it}|\mathbf{x}_{it})), \quad (16)$$

where $D_i$ and $V_i$ are readily computed given (15) and $R(\boldsymbol{\alpha})$ denotes a choice of working correlation matrix.

Missing data is a common issue in longitudinal studies. The GEE in (16) generally yields biased estimates under the missing at random (MAR) mechanism [25, 26, 27]. The *weighted generalized estimating equations* (WGEE), a common approach for addressing this issue, has been extended to the FRM [18, 22]. We adapt this approach to the current context, with an alternative implementation to simplify the inference procedure. As in the literature, we assume Monotone Missing Data Patterns (MMDP) to facilitate inference [18, 22, 25, 26, 27].

Let $y_{it}$ denotes the observed potential outcome, i.e., $y_{it} = y_{itk}$ if the subject is assigned the $k$th treatment. Let

$$\mathbf{y}_{it^-}=\left(y_{i1},\ldots,y_{i(t-1)}\right)^\top, \mathbf{x}_{it^-}=\left(\mathbf{x}_{i1}^\top,\ldots \mathbf{x}_{i(t-1)}^\top\right)^\top, 1\le t\le m,$$

denoting the all individual responses ($\mathbf{y}_{it^-}$) and explanatory variables ($\mathbf{x}_{it^-}$) prior to time $t$. Let

$$r_{it}=\begin{cases} 1 & \text{if } i\text{th subject is observed at time } t \\ 0 & \text{otherwise} \end{cases},$$

$$p_{it}=\begin{cases} 1 & \text{if } t=1 \\ E(r_{it}=1|r_{i(t-1)}=1,\mathbf{x}_{it^-},\mathbf{y}_{it^-}) & \text{if } t>1 \end{cases}, \text{logit}^{-1}(p_{it})=\xi_{0t}+\boldsymbol{\xi}_{xt}^\top\mathbf{x}_{it^-}+\boldsymbol{\xi}_{yt}^\top\mathbf{y}_{it^-}, \quad (17)$$

$$\Psi_{it}=\left(\prod_{s=1}^{t} p_{it}\right)^{-1} r_{it}\mathbf{I}_2, \Psi_i(\boldsymbol{\xi})=diag_t(\Psi_{it}), \quad \boldsymbol{\xi}_t=(\xi_{0t},\boldsymbol{\xi}_{xt}^\top,\boldsymbol{\xi}_{yt}^\top)^\top, \quad \boldsymbol{\xi}=(\boldsymbol{\xi}_2^\top,\ldots,\boldsymbol{\xi}_T^\top)^\top.$$

We assume no missing data at baseline such that $r_{i1} \equiv 1$ ($1 \le i \le n$). Under this, MAR and MMDP assumptions, $p_{it}$ in (17) is well defined for $1 \le t \le T$. By integrating the weights $\Psi_i$ into the GEE in (16), we obtain the following WGEE for inference about $\boldsymbol{\beta}$:

$$\mathbf{U}(\boldsymbol{\theta},\boldsymbol{\xi})=\sum_{i=1}^{n} D_i V_i^{-1}\Psi_i S_i=\mathbf{0}. \quad (18)$$

In the extant literature, an estimate $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$, obtained from a separate set of estimating equations, is substituted into the WGEE and (18) is then solved for $\boldsymbol{\theta}$ to obtain the WGEE estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Since $\hat{\boldsymbol{\theta}}$ is conditional upon $\hat{\boldsymbol{\xi}}$, its asymptotic variance is then adjusted to account for the sampling variability of $\hat{\boldsymbol{\xi}}$. If $\boldsymbol{\alpha}$ is $\sqrt{n}$-consistent and $\hat{\boldsymbol{\xi}}$ is asymptotically normal, the WGEE estimate $\hat{\boldsymbol{\theta}}$ obtained from (17) is consistent and asymptotically normal [18, 22, 27]. The procedure for adjusting the sampling variability of $\hat{\boldsymbol{\xi}}$ in the asymptotic variance is quite complex and thus we discuss an alternative approach to estimate $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$ simultaneously.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Let

$$\mathbf{f}_i=(\mathbf{f}_{i1}^\top,\ldots,\mathbf{f}_{iT}^\top,z_i,r_{i2},\ldots,r_{iT})^\top,\mathbf{h}_i=(\mathbf{h}_{i1}^\top,\ldots,\mathbf{h}_{iT}^\top,\pi_i,p_{i2},\ldots,p_{iT})^\top,\ \boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\boldsymbol{\eta}^\top,\boldsymbol{\gamma}^\top)^\top,1\le i\le n,1\le t\le T, \quad (19)$$

where $\mathbf{f}_{it}$, $\mathbf{h}_{it}$ and $\pi_i$ are defined in (15), and $r_{it}$ and $p_{it}$ are defined in (17). Consider the WGEE in (18), but with $D_i$ and $\Psi_i$ redefined as follows to provide estimates for both $\theta$ and $\xi$:

$$\begin{aligned}
&D_i=\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{h}_i,V_i=\begin{pmatrix}V_{i11}&0&0\\0&V_{i22}&0\\0&0&V_{i33}\end{pmatrix},V_{i11}=A_i^{\frac{1}{2}}R(\boldsymbol{\alpha})A_i^{\frac{1}{2}},V_{i22}=\pi_i(1-\pi_i),\\
&V_{i33}=\begin{pmatrix}p_{i2}(1-p_{i2})&\cdots&0\\&\ddots&\vdots\\&\cdots&p_{iT}(1-p_{iT})\end{pmatrix},\Psi_i=\begin{pmatrix}\Psi_{i11}&0&0\\0&\Psi_{i22}&0\\0&0&\Psi_{i33}\end{pmatrix},\\
&\Psi_{i11}=diag(\Psi_{it}),\Psi_{it}=r_{it}\left(\prod_{s=1}^{t}p_{is}\right)^{-1}\mathbf{I}_2,\Psi_{i22}=1,\Psi_{i33}=\begin{pmatrix}r_{i1}&\cdots&0\\&\ddots&\vdots\\&&r_{i(T-1)}\end{pmatrix},
\end{aligned} \quad (20)$$

where $A_i$ is defined in (17). Unlike (18), the WGEE in (19) makes joint inference about $\theta$ and $\xi$. Thus, no adjustment is necessary for the asymptotic variance of the WGEE estimate of $\theta$ to account for the sampling variability of $\hat{\xi}$ as in the standard approach above.

### 3.2 Multi-layered Intervention Study

We now extend the SFRM above to multi-layered interventions to address treatment noncompliance from different intervention layers, such as the child and parent layers of the RRP. For notational brevity, we focus on two-layered interventions, since extensions to multi-layered interventions with more than two layers are straightforward.

Consider a two-layered intervention study and let $u_{i1}$ denote some (continuous) treatment compliance measure for the second layer. By taking into account both compliance measures $d_{i1}$ and $u_{i1}$, we obtain from (11) the following dose-response relationship:

$$E(y_{i1}|\mathbf{x}_i,d_{i1},u_{i1},z_i=1)=g(d_{i1},u_{i1})+E(y_{i0}|\mathbf{x}_i,z_i=0). \quad (21)$$

We assume that the covariates $\mathbf{x}_i$ sufficiently explain treatment compliance patterns for both the primary and secondary layers of the multi-layered intervention, i.e., $d_{i1}, y_{i0} \perp \mathbf{x}_i$ and $u_{i1}, y_{i0} \perp \mathbf{x}_i$. In some studies, treatment noncompliance may be limited to some intervention layers, in which case $\mathbf{x}_i$ is only required to explain the affected layers. For example, in the RRP, noncompliance is a major issue only for the second parent support layer and the ignorability condition only needs to be assumed for parent participation.

By formulating (21) as an FRM as in the case of single-layered intervention study, we obtain the following SFRM for modeling the effect of treatment noncompliance on the outcome in a two-layered intervention study:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$f_{i1}=\frac{z_i y_{i1}}{\pi(\mathbf{w}_i;\boldsymbol{\eta})}, f_{i2}=\frac{(1-z_i)y_{i0}}{1-\pi(\mathbf{w}_i;\boldsymbol{\eta})}, f_{i3}=z_i, 1\le i\le n,$$
$$h_{i1}=h_i(\mathbf{x}_i,\boldsymbol{\beta}), h_{i2}(\mathbf{x}_i,d_{i1},u_{i1})=g(d_{i1},u_{i1},\boldsymbol{\gamma})+h_1(\mathbf{x}_i,\boldsymbol{\beta}), \quad (22)$$
$$\pi_i=\pi(\mathbf{w}_i;\boldsymbol{\eta}), E(z_i|\mathbf{x}_i,d_{i1},u_{i1},\boldsymbol{\theta})=\pi_i, \boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\eta}^\top)^\top.$$

where $1 \le i \le n$. The above has essentially the same form as the single-layered SFRM, except that the treatment effect $g(d_{i1}, u_{i1}, \boldsymbol{\gamma})$ is a function of compliance from both the primary and secondary intervention layers. Note that (22) applies to observational studies well, in which case $\mathbf{w}_i$ is assumed to account for all sources of selection bias.

We can model treatment effect $g(d_{i1}, u_{i1}, \boldsymbol{\gamma})$ to reflect treatment compliance in both layers. For example, we may specify an additive effect function, $g(d_{i1}, u_{i1}, \boldsymbol{\gamma}) = \gamma_1 d_{i1} + \gamma_2 u_{i1}$ or we may also include a between-layer treatment compliance interaction $d_{i1}u_{i1}$. If the treatment effect is moderated by some covariate $c_i$, we may also include treatment moderating effect by setting $g(d_{i1}, u_{i1}, c_i, \boldsymbol{\gamma}) = c_i(\gamma_1 d_{i1} + \gamma_2 u_{i1})$. If the moderating effect only occurs to one of the intervention layers, we may model $g(d_{i1}, u_{i1}, c_i, \boldsymbol{\gamma})$ as $\gamma_1 c_i d_{i1} + \gamma_2 u_{i1}$ or $\gamma_1 d_{i1} + \gamma_2 c_i u_{i1}$, depending on whether the moderating effect operates at the primary or secondary layer of the intervention.

As in the case of single-layered intervention study, the cross-sectional SFRM in (22) is readily extended to longitudinal studies. For example, by replacing the treatment effect function $g_t(d_{i1}, \boldsymbol{\gamma})$ in (15) by $g_t(d_{i1}, u_{i1}, \boldsymbol{\gamma})$ in (22), the SFRM in (15) can be applied to model the effect of treatment compliance for two-layered observational studies. As well, by modeling the missing data under MAR using (17), we can make joint inference about $\boldsymbol{\theta}$ in (22) and $\boldsymbol{\xi}$ for the missing data model using a WGEE akin to (18), but with $D_i$, $V_i$, $\Psi_i$ and $S_i$ in (20) redefined based on (22).

In the above, we have assumed that both $d_{i1}$ and $u_{i1}$ are continuous. The models are easily extended to non-continuous compliance variables, if either $d_{i1}$ or $u_{i1}$ or both are non-continuous.

## 4 Simulation Studies

We carried out a series of simulation studies to assess the performance of the proposed SFRM for multi-layered intervention studies for the most general case under both pre-treatment and post-treatment confounders. Since our RRP is a two-layered intervention study, we only considered this special case for the simulation study. We assessed the performance of the models under both cross-sectional and longitudinal data.

We considered continuous and binary outcomes $y_i$ for both cross-sectional and longitudinal data settings, with a continuous treatment noncompliance variable for both the primary and secondary layer. For space consideration, we only report results for two sample sizes $n = 50$ and 200 for a continuous response in cross-sectional data (Model I) and $n = 100$ and 400 for a binary response in longitudinal data (Model II). The increase in sample size for the binary outcome is to achieve more reliable estimates because of data sparseness in this binary response case, especially in the presence of missing data in the longitudinal data setting. All

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

simulations were performed with a Monte Carlo (MC) sample of 1,000. All analyses were carried out using codes developed by the authors for implementing the models considered using the R software platform [28].

For the cross-sectional data scenario, let $y_{ik}$ ($k = 0, 1$) be a continuous outcome in Model I and let $d_i$ ($u_i$) denote a continuous treatment noncompliance variable for the primary (secondary) intervention layer. Model I for the continuous $y_{ik}$ is defined as follows:

$$\text{Model I} - \text{Continuous } y_{ik} \text{ for Cross-sectional Data} \quad \begin{aligned} y_{i0}| \quad & x_i, b_i = \mu(x_i;\boldsymbol{\beta}) + b_i + e_{i0}, \mu(x_i;\boldsymbol{\beta}) = \beta_0 + \beta_1 x_i, \\ y_{i1}| \quad & \begin{cases} d_i, u_i, x_i, b_i = g_1(d_i, u_i, x_i;\boldsymbol{\gamma}) + \mu(x_i;\boldsymbol{\beta}) + b_i + e_{i1} \\ d_i, u_i, x_i, c_i, b_i = g_2(d_i, u_i, x_i, c_i;\boldsymbol{\gamma}) + \mu(x_i;\boldsymbol{\beta}) + b_i + e_{i1} \end{cases}, \\ g_1(d_i, u_i, x_i;\boldsymbol{\gamma}) = \quad & \gamma_0 d_i + \gamma_1 u_i + \gamma_2 u_i d_i, g_2(d_i, u_i, x_i, c_i;\boldsymbol{\gamma}) = c_i g_1(d_i, u_i, x_i;\boldsymbol{\gamma}), \\ \pi_i = \quad & \text{logit}^{-1}(\eta_0 + \eta_1 x_i), d_i, u_i \sim U(0, 5), x_i, c_i \sim N(0, 1), \\ b_i \sim \quad & (\chi_1^2 - 1)\sqrt{\sigma_b^2/2}, e_{i1}, e_{i0} \sim (\chi_1^2 - 1)\sqrt{\sigma^2/2}, \\ \boldsymbol{\beta} = \quad & (\beta_0, \beta_1)^{\mathsf{T}} = (5, 2), \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^{\mathsf{T}} = (0.5, 0.5, 0.4), \\ \boldsymbol{\eta} = \quad & (\eta_0, \eta_1)^{\mathsf{T}} = (0, -1), \sigma_b^2 = \sigma^2 = 1, \boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\eta}^\top\right)^\top, \end{aligned} \quad (2$$

where $z_i$ is the indicator of treatment assignment, $x_i$ is a confounding variable (for both pre- and post-treatment), $c_i$ is a treatment moderator, $g_1$ ($g_2$) is a function modeling the effect of treatment noncompliance without (with) the treatment moderator, $U$ ($a$, $b$) denotes a uniform over the interval between $a$ and $b$, and $\chi_p^2$ denotes a $\chi^2$ distribution with $p$ degrees of freedom. Since ($y_{i0}$, $y_{i1}$) share the same random effect $b_i$, they are not independent. Note that to demonstrate robustness of the SFRM, both the random effect $b_i$ and model error $e_{ik}$ followed non-normal distributions. In (23), we considered two treatment effect functions, $g_1(d_i, u_i, x_i; \boldsymbol{\gamma})$ and $g_2(d_i, u_i, x_i, c_i; \boldsymbol{\gamma})$, with the latter including a moderating effect of the former by a treatment moderator $c_i$. This moderator $c_i$ can be associated with either the primary or secondary layer of the multi-layered intervention.

Shown in Table 2 are the estimates of $\boldsymbol{\theta}$, along with their model-based (Mod. S.E.) and empirical (Emp. S.E.) standard errors for Model I. The model-based standard errors were computed based on the estimated asymptotic variance, while their empirical counterparts were calculated from the MC replicates. At the larger sample size $n = 200$, all parameter estimates were quite close to the true values of the respective parameters. The model-based standard errors also matched their empirical counterparts quite well. Although the difference all increased between the parameter estimates and their true values and between the model-based and empirical standard errors for the smaller sample size $n = 50$, the SFRM still performed quite well.

For the longitudinal data, as noted earlier, we only report results for a binary response. We extended both the mean for the control group, $\mu_t$ ($x_i$; $\boldsymbol{\beta}$), and the treatment effect function, $g_t$ ($d_i$, $u_i$, $x_i$, $c_i$; $\boldsymbol{\gamma}$), in the cross-sectional case to include a temporal trend. In addition, to reflect the treatment noncompliance patterns in the RRP study, where treatment noncompliance only occurred in the supportive parent layer, we only considered treatment noncompliance in the second layer. As in the cross-sectional data setting, we also included a treatment moderator $c_i$ in $g_t$ ($d_i$, $u_i$, $x_i$, $c_i$; $\boldsymbol{\gamma}$). For notational brevity, we only considered one treatment

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

effect function and two assessments, with $t = 1$ (2) denoting the baseline (follow-up). We created about 22% missing data at the follow-up.

We discussed two approaches for longitudinal data analysis. The first employs the conventional WGEE that conditions on the estimates of the missing data model and adjusts the variance estimates of parameter estimates to account for the sampling variability in the estimates of the missing data model. Since the adjustment part is quite complex, we also discussed an alternative that utilized the flexibility of FRM to model both missing data and treatment effect simultaneously. We used this latter approach in the simulation study.

For the binary response $y_{ik}$, the SFRM is given by:

$$\text{ModelII}-\text{Binary}\,y_{ik}\,\text{for Longitudinal Data Setting}\quad\begin{array}{ll} y_{it0}| & x_i=\text{logit}^{-1}(\mu_t(x_i;\boldsymbol{\beta})),\mu_t(x_i;\boldsymbol{\beta})=\beta_0+\beta_1 t+\beta_2 x_i+\beta_3 x_i t,\\ y_{it1}| & d_i,u_i,x_i,c_i=\text{logit}^{-1}(g_t(d_i,u_i,x_i,c_i;\boldsymbol{\gamma}))+\mu_t(x_i;\boldsymbol{\beta})\},\\ \pi_i= & \text{logit}^{-1}(\eta_0+\eta_1 x_i),g_t(d_i,u_i,x_i,c_i;\boldsymbol{\gamma})=\gamma_0 u_i t+\gamma_2 c_i u_i t,\\ p_i= & \text{logit}^{-1}(\xi_0+\xi_1 y_{i0}^o),d_i,u_i\sim U(0,4),x_i,c_i\sim N(0.1)\\ \boldsymbol{\beta}= & (\beta_0,\beta_1,\beta_2,\beta_3)^{\mathsf{T}}=(-1,1,1,-1),\boldsymbol{\gamma}=(\gamma_0,\gamma_1)^{\mathsf{T}}=(1,1),\\ \boldsymbol{\eta}= & (\eta_0,\eta_1)^{\mathsf{T}}=(0,-1),\boldsymbol{\xi}=(\xi_0,\xi_1)^{\mathsf{T}}=(1,1),\boldsymbol{\theta}=\left(\boldsymbol{\beta}^{\top},\boldsymbol{\gamma}^{\top},\boldsymbol{\eta}^{\top},\boldsymbol{\xi}^{\top}\right)^{\top}\end{array}. \quad (24)$$

where $p_i = E(r_{i1} = 1 \mid y_{i1})$ is the probability of missing data at the follow-up $t = 2$ for both the treatment and control groups. For the control group, we included a time as well as a time by covariate interaction. As indicated earlier, the treatment effect function $g_t$ ($d_i$, $u_i$, $x_i$, $c_i$; $\gamma$) also included a treatment moderator $c_i$. Since the probability of missing response at post-treatment $p_i$ depends on the baseline $y_{i1}$, the missing data mechanism follows the MAR. Under the specified $\xi$, there was about 22% missing data. The correlated $y_{itk}$ were created by the copula methods [29, 30]. The correlation between the two potential outcomes with each assessment time as well as between two assessments within the same potential outcome in our simulation study was set at about 0.5, uncontrolled for any of the explanatory variables.

Shown in Table 3 are the estimates of $\theta$, along with their model-based (Mod. S.E.) and empirical standard (Emp. S.E.) errors for Model II. In comparison to the cross-sectional data case, Table 3 contains estimates for the additional parameters $\xi = (\xi_0, \xi_1)^{\mathsf{T}}$ for the missing data model. As in the case of cross-sectional data, both the parameter estimates and model-based standard errors were quite good when compared to their true values or empirical counterparts.

## 5 Rochester Child Resilience Study

The Rochester Resilience Project (RRP) is a randomized two-layered intervention study with significant treatment noncompliance by the parent, whose participation forms the second supportive layer of the intervention. The study's enrollment began in Fall 2006, with data collection for the final cohort completed by June 2011. There were 401 students from first up to third grade from Rochester City School District elementary schools. The study examines how children with a higher risk of developing behavioral problems in the intervention condition improve as compared to the control condition over a 30-month period. Each child was assessed at baseline, and 6, 18, and 30 months post baseline.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Since treatment compliance was quite good for the children in the study, we only considered variability in the parent participation. In order to apply the proposed SFRM to analyze the data in this study, we first examined the baseline covariates to see if any of these variables effectively predicted the patterns of treatment noncompliance. We treated the second-layer noncompliance measure, $u_i$, the number of session attendance by the parent, as a continuous variable and applied linear regression.

Shown in Table 4 are the estimated coefficients, standard errors and p-values of the variables that significantly predicted the number of session attendance $u_i$ from the linear regression model. The variable School Number represents the different schools which the children attended. The variable PNC stands for the Perceived Need for Care scale, assessing frequency over past six months that parent viewed her child as needing help for behavior or emotional problems, including from communication with others about child [31]. The DomEX Baseline denotes the baseline value of the subscale of the Dominic Interactive self-report, assessing symptoms of three externalizing (oppositional defiant, conduct problems, and ADHD) problems [32]. The results from the regression show that session participation was significantly different across the different schools and children with different PNC and DomEX baseline values. In addition, parent age also significantly predicted the session attendance.

For our illustrations of the model, we focused on two primary behavior outcomes of the study, the Teacher ratings of aggressive behavior (AthAcc) and Parent rating of internalizing behavior problem (PIntD). For both outcomes, higher values indicate fewer problems. For each of these behavior outcomes $y_{it}$, let $y_{it1}$ and $y_{it0}$ denote the potential outcomes of $y_{it}$ at baseline ($t = 1$) and each of the three follow-ups ($2 \le t \le 4$). We modeled the causal treatment effect as a function of treatment compliance from the parent layer using a SFRM as follows:

$$\begin{aligned} &E\left(\frac{1-z_i}{1-\pi}y_{it0}|u_i\right)=\mu_{it}, E(\frac{z_i}{\pi}y_{it1}|u_i)=g_{it}+\mu_{it}, E(z_i)=\pi, \\ &\mu_{it}=\beta_0+\beta_1 t+x_{i1}\beta_2+\beta_3 x_{i1}t+\beta_4 x_{i2}+\beta_5 x_{i3}+\beta_6 x_{i4}+\beta_7 x_{i5}+\beta_8 x_{i6}+\beta_9 x_{i7}+\beta_{10} x_{i8}, \\ &g_{it}=\gamma u_i t, 1 \le t \le 4, \end{aligned} \tag{25}$$

where $z_i$ is the indicator variable of treatment assignment with $z_i = 1$ (0) for intervention (control), $x_{i1}$ denotes the age of the child at baseline, $x_{i2} - x_{i5}$ denote the four binary indicators of School 19, 22, 30, 45, and $x_{i6}$, $x_{i7}$ and $x_{i8}$ denote the PNC, DomEXT Baseline and Parent Age variables, respectively. In addition, we included Age and Age by time interaction, since our theory as well as preliminary analyses show that these behavioral outcomes have different trajectories for children of different ages.

Prior to fitting the SFRM, we examined the missing data mechanism using logistic regression to determine whether missing data at each of the follow-up times, 6, 18, and 30 months post-baseline, depended on the observed outcomes at prior assessment times. Results indicated that missing data was not associated with the observed data for any of the two behavior outcomes considered. Thus, we assumed the dropouts for these two behavior outcomes in this RRP study followed the Missing Complete at Random (MCAR)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

mechanism. The MCAR mechanism was also consistent with the excellent treatment compliance observed for the study subjects (children), since unlike parent participation both the intervention and assessment were performed during the regular school time.

Shown in Table 5 are the estimates (Est.), standard errors (S.E.) and p-values (p-value) for the parameter $\gamma$ in the treatment effect function $g_{it}$ in (25) for the two behavior outcomes analyzed. Within the context of the study, this parameter $\gamma$ measures the rate of change of the behavior outcome per month for each additional session attended by the parent. The results show that for both behavior outcomes $\gamma$ was quite significant, with the positive estimate indicating that the intervention improved the child's behaviors and reduced the risk for future mental disorder and substance abuse. With the SFRM in (25), causal treatment effect is given by $\gamma u_i$. For example, if the parent of the child attended all the planned 15 sessions, then $u_i = 15$ and the causal effect is $\beta_4 u_i = 0.25$ per month time in the scale of the AthAcc outcome. Thus, in 18 months post-baseline, for instance, the intervention will on average increase the child AthAcc outcome by 4.32 points.

For comparison purposes, we also performed the intent-to-treat (ITT) analysis for the two behavior outcomes by setting $u_i = 1$ in $g_{it}$ of the SFRM in (25). The estimated $\gamma$, standard errors (S.E.) and p-values (p-value) are shown in Table 5 under the column "ITT Effect". As seen, $\gamma$ was not significant for either outcome. Thus, parent support played a significant role in improving the two child behavior outcomes in this two-layered intervention study.

## 6 Discussion

We developed an approach to address treatment noncompliance in multi-layered intervention studies. This approach extends the structural mean model (SMM) to multi-layered intervention and longitudinal data settings. We selected the SMM to develop our approach because of the need to model treatment noncompliance on a continuous scale. Other competing approaches such as the Principal Stratification method characterize variability in treatment noncompliance using categorical outcomes. However, within the context of multi-layered intervention study, such methods yield a large number of noncompliance categories, limiting their applications. For example, if a 4-level categorical outcome is used to characterize treatment noncompliance for each layer of a 2-layered intervention, we will need a 16-level categorical outcome to understand treatment noncompliance when considering interactions of noncompliance patterns between the two intervention layers. The larger number of levels of a categorical outcome may cause problems for fitting models, if there are a limited number of subjects in one or more strata (defined by the levels of the categorical outcome). With the freedom to choose a continuous or categorical noncompliance measure as in the SMM and proposed SFRM, we can consider between-layer interactions in a more parsimonious and reliable fashion.

We also adopted the distribution-free framework of SMM for inference for our proposed model. Using the framework of FRM, we are able to provide robust inference about model parameters like the SMM and accommodate noncompliance from multiple intervention layers as well as missing data under MAR. Our simulation studies show that the proposed approach perform quite well even for a sample size as small as 50 (for combined

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

intervention and control groups). As well, applications of the proposed model to the Rochester Resilience Project demonstrate the importance to consider treatment noncompliance from the supportive parent layer in this two-layered intervention study.

## References


1. Angrist J, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with discussion). Journal of the American Statistical Association. 1996; 91:444–472.
2. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58:21–29. [PubMed: 11890317]
3. Robins JM. Correcting for noncompliance in randomized trials using structural nested mean models. Communications in Statistics. 1994; 23:2379–2412.
4. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974; 66:688–701.
5. Chen R, Chen T, Lu N, Zhang H, Wu P, Feng C, Tu XM. Extending the Mann-Whitney-Wilcoxon rank sum test to longitudinal data analysis with covariates. Applied Statistics. in press.
6. Wu P, Han Y, Chen T, Tu XM. Causal inference for Mann-Whitney-Wilcoxon rank sum and other nonparametric statistics. Statistics in Medicine. 2014; 33(8):1261–1271. [PubMed: 24132928]
7. El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent based modeling in social epidemiology. Epidemiologic Perspectives and Innovations. 2012; 9:1–9. [PubMed: 22296660]
8. Lu, N.; White, AM.; Wu, P.; He, H.; Hu, J.; Feng, C.; Tu, XM. Social Network Endogeneity and Its Implications for Statistical and Causal Inferences. In: Lu, N.; White, AM.; Tu, XM., editors. Social Networking: Recent Trends, Emerging Issues and Future Outlook. New York: Nova Science; 2013.
9. Yu Q, Tang W, Kowalski J, Tu XM. Multivariate U-Statistics: A tutorial with applications. Wiley Interdisciplinary Reviews – Computational Statistics. 2011; 3:457–471.
10. Kowalski J, Powell J. Nonparametric inference for stochastic linear hypotheses: Application to high-dimensional data. Biometrika. 2004; 91(2):393–408.
11. King TS, Chinchilli VM. A generalized concordance correlation coefficient for continuous and categorical data. Statistics in Medicine. 2001; 20:2131–2147. [PubMed: 11439426]
12. Kowalski, J.; Tu, XM. Modern Applied U Statistics. New York: Wiley; 2007.
13. Tu XM, Feng C, Kowalski J, Tang W, Wang H, Wan C, Ma Y. Correlation analysis for longitudinal data: Applications to HIV and psychosocial research. Statistics in Medicine. 2007; 26:4116–4138. [PubMed: 17342700]
14. Ma Y, Tang W, Feng C, Tu XM. Inference for Kappas for longitudinal study data: Applications to sexual health research. Biometrics. 2008; 64:781–789. [PubMed: 18047535]
15. Ma Y, Tang W, Yu Q, Tu XM. Modeling concordance correlation coefficient for longitudinal study data. Psychometrika. 2010; 75:99–119.
16. Ma Y, Alejandro GD, Hui Z, Tu XM. A U-statistics based approach for modeling Cronbach Coefficient Alpha within a longitudinal data setting. Statistics in Medicine. 2011; 29(6):659–670.
17. Lu N, Chen T, Wu P, Gunzler D, Zhang H, He H, Tu XM. Functional response models for intraclass correlation coefficients. Applied Statistics. in press.
18. Yu Q, Chen R, Tang W, He H, Gallop R, Crits-Christoph P, Hu J, Tu XM. Distribution-free models for longitudinal count responses with over-dispersion and structural zeros. Statistics in Medicine. 2013; 32:2390–2405. [PubMed: 23239019]
19. Gunzler D, Tang W, Lu N, Wu P, Tu XM. A class of distribution-free models for longitudinal mediation analysis. Psychometrika. in press.
20. Rubin DB. Bayesian inference for causal effects: The Role of Randomization. Annals of Statistics. 1978; 6:34–58.
21. Fischer K, Goetghebeur E. Structural mean effects of noncompliance. Journal of the American Statistical Association. 2004; 99(468):918–928.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


22. Gunzler D, Tang W, Lu N, Wu P, Tu XM. A class of distribution-free models for longitudinal mediation analysis. Psychometrika. in press.
23. Pepe MS, Anderson GL. A Cautionary Note on Inference for Marginal Regression Models with Longitudinal Data and General Correlated Response Data. Communication in Statistics-Simulation. 1994; 23:939–951.
24. Fitzmaurice GM. A caveat concerning independence estimating equations with multiple multivariate binary data. Biometrics. 1995; 51:309–317. [PubMed: 7766784]
25. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association. 1995; 90:106–121.
26. Lu N, Tang W, He H, Yu Q, Crits-Christoph P, Zhang H, Tu XM. On the impact of parametric assumptions and robust alternatives for longitudinal data analysis. Biometrical Journal. 2009; 51:627–643. [PubMed: 19688758]
27. Wu P, Tu XM, Kowalski J. On Assessing Model Fit for Distribution-Free Longitudinal Models under Missing Data. Statistics in Medicine. 2014; 33(1):143–157. [PubMed: 23897653]
28. R Development Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2010. R: A language and environment for statistical computing. ISBN 3-900051-07-0, URL http://www.R-project.org.
29. Nelsen, RB. An introduction to Copulas. New York: Springer; 2006.
30. Zhang H, Lu N, Feng C, Thurston SW, Xia Y, Tu XM. On Fitting Generalized Linear Mixed-effects Models for Binary Responses using Different Statistical Packages. Statistics in Medicine. 2011; 30:2562–2572.
31. Meadows G, Burgess P, Fossey E, Harvey C. Perceived need for mental health care, findings from the Australian National Survey of Mental Health and Wellbeing. Psychological Medicine. 2000; 30:645–656. [PubMed: 10883719]
32. Valla JP, Bergeron L, Smolla N. The Dominic-R: A pictorial interview for 6- to 11-year old children. Journal of the American Academy of Child and Adolescent Psychiatry. 2000; 39:85–93. [PubMed: 10638071]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 1**

Child Resilience Complete dataset

| Total sessions attended by parents of children in intervention | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Sessions** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9+** | **total** |
| Frequency | 74 | 38 | 26 | 11 | 9 | 10 | 3 | 1 | 2 | 30 | 202 |
| Percent | 36.6 | 18.8 | 12.9 | 5.4 | 4.5 | 5.0 | 1.5 | .5 | 1 | 15 | 100 |
| Cumulative Percent | 36.6 | 55.4 | 68.3 | 73.8 | 78.2 | 83.2 | 84.7 | 85.1 | 86.1 | 100.0 | 100 |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 2**

Parameter estimates and standard errors for Model I with a cross-sectional continuous response.

| **Parameter estimates and standard errors for Model I with treatment effect functions $g_1/g_2$** | | | | | | |
|---|---|---|---|---|---|---|
| | ***n* = 50** | | | ***n* = 200** | | |
| **Parameter** | **Est.** | **Mod. S.E.** | **Emp. S.E.** | **Est.** | **Mod. S.E.** | **Emp. S.E.** |
| $\gamma_0 = 0.5$ | 0.475/0.514 | 0.629/0.566 | 0.744/0.758 | 0.496/0.495 | 0.318/0.274 | 0.326/0.289 |
| $\gamma_1 = 0.5$ | 0.459/0.535 | 0.655/0.553 | 0.808/0.760 | 0.505/0.515 | 0.315/0.271 | 0.323/0.300 |
| $\gamma_2 = 0.4$ | 0.428/0.377 | 0.369/0.313 | 0.453/0.438 | 0.402/0.395 | 0.179/0.163 | 0.196/0.175 |
| $\beta_0 = 5$ | 5.107/4.981 | 0.289/0.304 | 0.338/0.394 | 4.998/5.012 | 0.158/0.157 | 0.175/0.172 |
| $\beta_1 = 2$ | 1.995/2.013 | 0.341/0.348 | 0.402/0.509 | 2.002/1.976 | 0.189/0.189 | 0.201/0.195 |
| $\eta_0 = 0$ | 0.033/−0.003 | 0.325/0.326 | 0.329/0.343 | 0.001/−0.003 | 0.150/0.157 | 0.158/0.150 |
| $\eta_1 = -1$ | −1.086/−1.107 | 0.394/0.395 | 0.422/0.461 | −1.017/−1.016 | 0.189/0.189 | 0.188/0.199 |

*Stat Med*. Author manuscript; available in PMC 2015 September 28.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 3**

Parameter estimates and standard errors for Model II with a longitudinal binary response.

| Parameter estimates and standard errors for Model II | | | | | | |
|---|---|---|---|---|---|---|
| | ***n* = 100** | | | ***n* = 400** | | |
| **Parameter** | **Est.** | **Mod. S.E.** | **Emp. S.E.** | **Est.** | **Mod. S.E.** | **Emp. S.E.** |
| $\gamma_0 = 1$ | 0.873 | 0.497 | 0.508 | 1.047 | 0.290 | 0.321 |
| $\gamma_1 = 1$ | 0.964 | 0.512 | 0.586 | 1.070 | 0.284 | 0.317 |
| $\beta_0 = -1$ | −1.067 | 0.468 | 0.491 | −1.022 | 0.216 | 0.213 |
| $\beta_1 = 1$ | 1.128 | 0.461 | 0.525 | 1.025 | 0.205 | 0.217 |
| $\beta_2 = 1$ | 1.089 | 0.589 | 0.605 | 1.016 | 0.264 | 0.279 |
| $\beta_3 = -1$ | −1.182 | 0.583 | 0.690 | −1.044 | 0.261 | 0.275 |
| $\eta_{l0} = 0$ | 0.021 | 0.209 | 0.224 | −0.001 | 0.108 | 0.109 |
| $\eta_{l1} = -1$ | −1.058 | 0.291 | 0.303 | −1.008 | 0.144 | 0.148 |
| $\xi_0 = 1$ | 1.022 | 0.273 | 0.292 | 1.010 | 0.131 | 0.135 |
| $\xi_1 = 1$ | 1.088 | 0.689 | 0.702 | 1.033 | 0.325 | 0.346 |

*Stat Med*. Author manuscript; available in PMC 2015 September 28.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## Table 4

Estimates, standard errors and p-values for significant predictors of parent participation for the Rochester Resilience Project from generalized linear models.

| Significant Predictors for Parent Participation | | | |
|---|---|---|---|
| **Explanatory Variable** | **Estimate** | **Standard Error** | **p-value** |
| PNC | 0.9191 | 0.2698 | 0.0008 |
| Parent Age | 0.0882 | 0.0293 | 0.0030 |
| DomEX Baseline | 0.9127 | 0.0373 | 0.0154 |
| School Number | | | <.0001 |
| School 19 | −4.1065 | 0.8446 | <.0001 |
| School 22 | −3.3860 | 0.9122 | 0.0003 |
| School 30 | −3.1342 | 0.9873 | 0.0018 |
| School 45 | −4.3626 | 0.8440 | <.0001 |
| School 50 | 0.0000 | | |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 5**

Child Resilience Complete dataset

| Estimation Results of Treatment Time Effect (γ) | | | | | | |
|---|---|---|---|---|---|---|
| | Causal Effect | | | ITT Effect | | |
| | Est. | S.E. | p-value | Est. | S.E. | p-value |
| AthAcc | 0.0167 | 0.0014 | <0.0001 | 0.0053 | 0.0069 | 0.2235 |
| PIntD | 0.1640 | 0.0163 | <0.0001 | 0.0476 | 0.0663 | 0.2365 |