Published in final edited form as:

Med Care. 2017 December ; 55(12): e120–e130. doi:10.1097/MLR.0000000000000464.

Adjustment for Variable Adherence under Hierarchical Structure: Instrumental Variable Modeling via Compound Residual Inclusion

Tyson H. Holmes, Ph.D.¹, Donna M. Zulman, M.D., M.S.^{2,3}, and Clete A. Kushida, M.D., Ph.D.¹

¹Stanford University Human Immune Monitoring Center, Institute for Immunity Transplantation and Infection, Stanford University School of Medicine, Stanford, CA, 94305, USA.

²Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, CA, 94025, USA.

³Division of General Medical Disciplines, Department of Medicine, Stanford University, Stanford, CA, 94305, USA

Abstract

Background—Variable adherence to assigned conditions is common in randomized clinical trials.

Objectives—A generalized modeling framework under **longitudinal** data structures is proposed for regression estimation of the **causal** effect of variable adherence on outcome, with emphasis upon adjustment for unobserved confounders.

Research Design—A nonlinear, **nonparametric random-coefficients** modeling approach is described. Estimates of local average treatment effects among compliers can be obtained simultaneously for all assigned conditions to which participants are randomly assigned within the trial. **Two** techniques are combined to address **time-varying and time-invariant** unobserved confounding—residual inclusion and **nonparametric random-coefficients modeling**. Together these yield a compound, two-stage-residual inclusion, instrumental variables model.

Subjects—The proposed method is illustrated through a set of simulation studies to examine small-sample bias and in application to neurocognitive **outcome** data from a large, multicenter,

Corresponding Author: Dr. Tyson H. Holmes, Stanford University Human Immune Monitoring Center, Institute for Immunity Transplantation and Infection, 299 Campus Drive, Fairchild Science Building, Stanford, CA 94305, Phone (650) 725-2925, tholmes@stanford.edu.

Tyson H. Holmes, Ph.D., B.S.

Stanford University Human Immune Monitoring Center, Institute for Immunity Transplantation and Infection, 299 Campus Drive, Fairchild Science Building, Stanford, CA, 94305, Phone (650) 725-2925, tholmes@stanford.edu

Donna M. Zulman, M.D., M.S.

Center for Innovation to Implementation, VA Palo Alto Health Care System, Division of General Medical Disciplines, Stanford University, 1265 Welch Road, MSOB MC 5475, Stanford, CA, 94305, Phone (650) 725-7747, Fax (650) 617-2736, dzulman@stanford.edu

Clete A. Kushida, M.D., Ph.D.

Stanford Sleep Medicine Center and Department of Psychiatry and Behavioral Sciences, 450 Broadway Street, Pavilion C, 2nd Floor, Redwood City, CA, 94063, Phone (650) 721-7560, clete@stanford.edu

randomized clinical trial in sleep medicine for continuous positive airway pressure treatment of obstructive sleep apnea.

Results—Results of simulation studies indicate that, relative to a standard comparator, the proposed estimator reduces bias in estimates of the causal effect of variable adherence. Bias reductions were greatest at higher levels of residual variance and when confounders were time-varying.

Conclusions—The proposed modeling framework is flexible in the distributions of outcomes that can be modeled, applicable to repeated measures **longitudinal** structures, and **provides effective reduction of bias due to** unobserved confounders.

Keywords

instrumental variable; longitudinal data; adherence; causal inference; nonparametric random-coefficients modeling

2. Introduction

2.1 Overview of problem and approach

Variable adherence to assigned condition is commonplace in randomized clinical trials (RCTs). Oftentimes, variable adherence is endogenous in nature, being driven by patient self-selection. A variety of statistical methods have been developed for the purpose of estimating consistently the cause-and-effect (causal) relationship between observed adherence and the clinical outcome of interest in the presence of unobserved confounding factors induced by endogeneity [1, 2]. Here we propose a generalized modeling framework that, for the first time, combines two-stage residual inclusion and nonparametric random-coefficients modeling to provide *comprehensive* adjustment for time-varying and time-invariant unobserved confounding within estimates of the causal effect of variable adherence for longitudinal sampling structures. This extension to longitudinal sampling structures is important because many RCTs monitor longitudinal outcomes on participants. By definition, RCTs assign participants to two or more conditions. The proposed modeling framework provides adjustment for time-varying and time-invariant unobserved confounding within each assigned condition of the RCT (Section 2.6).

2.2 Instrumental Variables

Under variable adherence, an intention-to-treat (ITT [3]) analysis gives estimates of treatment *effectiveness* (i.e., effect of treatment “as assigned” sensu [4]). However, the ITT estimate has limited clinical utility because it does not address response to the amount of treatment as actually received [5]. ITT analysis forgoes the opportunity to capitalize on a possible opportunity to estimate *dose response* from outcome data collected under variable adherence. Consistent estimation of dose response is complicated by possible confounders—factors which cause variation in adherence *and* cause variation in outcome. An example is “confounding by indication” [6] wherein a participant adjusts adherence based on current medical status per medical advice. Various methods to address confounding have been proposed, including propensity scores [7, 8, 9], G-estimation [10], and instrumental variables (IV [11]). IV methods are among those that offer a distinct advantage of permitting

consistent estimation of dose-response in the presence of *unobserved* confounders, which is an important property because few if any situations may exist where investigators can be certain that all possible confounders have been measured. For this reason, we have incorporated IVs as a central component of our proposed modeling framework.

IV methods need to satisfy certain assumptions to generate consistent estimates of the causal effect of an intervention under variable adherence (i.e., the “local average treatment effect” among compliers, LATEC [12]). These assumptions need to be considered in each specific application of IV methods. The first **assumption** is that the IV must cause exposure to intervention and the second is that IVs affect outcome only indirectly through exposure to intervention (“exclusion restriction”) [13]. A third assumption is that IV and outcome have no common causes. As Hernan and Robins [13] explain, the random assignment mechanism satisfies all three of these conditions especially, regarding exclusion restriction, under double-blind RCTs. True random assignment is unquestionably exogenous and often associated strongly with degree of adherence to assigned condition. Unusual circumstances may arise where this may not be true, such as **in the presence of “defiers”** [14], participants predisposed to adhering to whichever assigned condition they are *not* assigned, which would violate the fourth requisite IV assumption of “monotonicity” [13]. A fifth assumption is that randomized units (e.g., patients or clusters) have mutually independent potential outcomes (“stable unit treatment value assumption,” SUTVA [14, 15]). For present purposes, SUTVA is assumed to apply.

2.3 Longitudinal structure

The current literature on the development and application of IV methods to variable adherence is active and broad (e.g., [16 - 20]). Slower to develop have been models, methods of estimation and example IV applications to **hierarchical sampling structures, including** longitudinal, despite **the fact that** many RCTs collect longitudinal measurements on major outcomes **within participants** across study visits. Much existing work on longitudinal IV methods, especially earlier studies, focused on the limited circumstance in which adherence is coded as a binary presence-absence variable (e.g., [21]), which may be due to the fact that, in general, early literature on adjustment for variable adherence emphasized binary measures of adherence. In reality, adherence often varies across a gradation of levels. Bond et al. [11] allow for multi-level adherence in a longitudinal setting. They discuss application via generalized least squares (GLS) and employ generalized estimating equations (GEE [22]) in their worked example. Lamiraud et al. [23] also employed GEE.

GEE and GLS are marginal models, providing estimates of causal effects of treatment as received *on average across the population* [24]. Despite recent interest with “population health” [25], population average estimates are off-purpose for most RCTs, where the goal is to understand how individual patients respond to treatment (e.g., [26]). **In contrast to GEE and GLS, random-coefficients** models [24] are well-suited for characterizing individual patient response, because they offer estimates of longitudinal response trajectories of individual patients and of the average patient. Only under a linear model do estimates of population-average effects and average-patient effects coincide [24]; but many settings require a nonlinear modeling framework (e.g., binary outcome), which generalized linear

random-coefficients models and nonlinear **random-coefficients** models [27] accommodate. To improve upon existing marginal modeling and/or linear modeling approaches, the method presented here employs **random-coefficients** models with allowance for nonlinear functional structure (Sections 2.4 and 3.1) and participant-level random coefficients.

2.4 Random coefficients

To date, random coefficients (“random effects”) have been employed only rarely to aid estimation of the LATEC. Sitlani et al. [28] assumed that random coefficients capture latent factors that are explanatory of endogenous selection to an assigned surgical intervention. Small et al. [15] adjusted for variable adherence in the setting of a binary outcome. Lamiraud et al. [23] employed a correlated random-coefficients [29] structure. Correlated random coefficients **are** a very flexible structure because separate participant-level random coefficients could be employed for adherence and for outcome with these two random coefficients allowed to covary. A covariance structure is also specified at the level of individual observations. A shared random coefficients (SRC [29]) specification is another possibility (e.g., [28]). Within our proposed modeling framework we assume, as in previous literature [28], that random coefficients capture latent factors that are explanatory of endogenous selection. We go even further and recognize that the presence of unobserved confounders can, probably often, result in complex distributions for random coefficients. Specifically, unobserved confounders can generate a mixture distribution. We model that complexity in a first-stage equation with adherence as the outcome regressed on randomized condition as an instrument with a *nonparametric distribution* for random intercepts [30, 31]. Rather than shared or correlated random coefficient modeling, we employ the predicted values of these first-stage random intercepts as a fixed (latent effects) covariate in the second stage equations for the efficacy outcome. Using first-stage random coefficients as a second-stage fixed covariate conditions the second stage on the first and thereby eliminates the need for specification of a between-equation covariance structure, thereby simplifying model formulation. Second-stage equations regress the efficacy outcome on these predicted first-stage random intercepts and observed adherence. This facilitates adjustment for unobserved confounders (captured by the first-stage random coefficient) that are *invariant across repeated measures* as a form of two-stage residual inclusion (TSRI, [32]). Including the **same predicted random coefficient values** in both stages **should avoid** unnecessary introduction of inconsistency into the estimate of the causal effect of variable adherence [11].

2.5 Two-stage residual inclusion

TSRI has been shown to provide a consistent estimate of the causal effect of interest under linear and nonlinear model structures [30]. Indeed, TSRI addresses the concern raised by Lamiraud et al. [23] about application of IV methods to nonlinear settings. In addition to adjustment for time-invariant unobserved confounding, our method also employs the first-stage residuals as an additional fixed covariate in second-stage efficacy equations to offer adjustment for *time-varying* unobserved confounders thereby providing *compound*, TSRI, instrumental variable estimation (cTSRI-IV). Details are provided in Section 3.

2.6 Simultaneous adjustment

RCTs by definition consist of two or more conditions to which participants are assigned. The dose-response relationship between treatment as received and efficacy outcome may be potentially obscured by endogenous selection (and unobserved confounding) in some or all assigned conditions; and the precise nature of endogenous selection can easily vary among these conditions. Our system of equations (Section 3.1) permits *simultaneous* and *condition-specific* estimation of the LATEC for each and every of the study conditions to which participants are randomly assigned in the RCT. This efficient use of data from a RCT may be particularly helpful in application to comparative effectiveness studies, which are a class of RCT designs that are seeing widening use in efforts to enhance equipoise.

3. Methods

3.1 Model specification

Parameters, random variables, and observed values will be denoted by uppercase Greek letters, uppercase Roman letters, and lowercase Roman letters, respectively. Let random variable A_{kit} denote adherence to the randomly assigned study condition k for person i , $i = \{1, \dots, n\}$, at time t ; and let random variable O_{kit} denote outcome for condition k , person i at time t . Randomly assigned conditions are indexed by k , $k = \{1, \dots, \kappa\}$, $2 < \kappa$. To simplify exposition here, we assume randomization is to two assigned conditions, $\kappa \equiv 2$, which allows illustration of the basic method and simplification of some notation. (Extension to $2 < \kappa$ is straightforward.) For $1 \leq \gamma_1 < \gamma_2$, models for the expectations of A_{kit} and O_{kit} are given by the following system of equations (i.e., $k \in \{1, 2\}$ for adherence and for outcome):

$$\mu_{ki} \stackrel{\text{def}}{=} E[A_{kit}|R_{ki}] = g\left(\alpha_0 + \alpha_1(k-1) + \sum_{p=2}^{\gamma_1} \alpha_{pk} x_{pkit} + R_{ki}\right) \theta_{ki} \stackrel{\text{def}}{=} E\left[O_{kit}|A_{kit}=a_{kit}, R_{ki}=r_{ki}, \dot{R}_{ki}\right] = g_k\left(\beta_0 + \beta_{k1} a_{kit} + \sum_{p=2}^{\gamma_1} \beta_{pk} x_{pkit} + \sum_{p=\gamma_1+1}^{\gamma_2} \beta_{pk} x_{pkit} + \varepsilon_k(a_{kit} - \mu_{ki}) + \varsigma_k g^{-1}(r_{ki}) + \dot{R}_{ki}\right)$$

[1a]

where g and g_k are some continuous real-valued mappings from $\mathbb{R}$ to $\mathbb{R}$ (e.g., link functions). The respective first-stage and second-stage participant level random coefficient are R_{ki} and $\dot{R}_{ki}$. The term $a_{kit} - \mu_{ki}$ provides residual inclusion to adjust for *time-varying* endogeneity effect ε_k for the k^{th} assigned condition; ς_k provides residual inclusion to adjust for *time-invariant* endogeneity effect ζ_k for the k^{th} assigned condition, with requisite inverse transformation g^{-1} ; and x_{pkit} is the p^{th} covariate measured on the i^{th} person at time t in the k^{th} condition. Covariates may include interactions with assigned condition and with time. Typically μ_{ki} will be a single equation, while the θ_{ki} can be formulated as a separate equation for each assigned condition. All first-stage covariates are included in the second stage as main effects to avoid introduction of bias [11]. Defining $Z_{kit} \stackrel{\text{def}}{=} A_{kit} - \mu_{ki}$ and $\dot{Z}_{kit} \stackrel{\text{def}}{=} O_{kit} - \theta_{ki}$ take

$$E[R_{ki}] = E[Z_{kit}] = E[\dot{Z}_{kit}] \equiv 0 \quad [1b]$$

and

$$Var\left[\left\{R_{1i}, R_{2i}, \dot{R}_{1i}, \dot{R}_{2i}, Z_{1it}, Z_{2it}, \dot{Z}_{1it}, \dot{Z}_{2it}\right\}\right] \stackrel{def}{=} \begin{bmatrix} \sigma_{11}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{21}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{11}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{21}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{12}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{12}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{12}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{22}^2 \end{bmatrix} \quad [1c]$$

Random coefficients and residuals are independent between assigned conditions due to randomized assignment; and random coefficients and residuals are independent between stages due to residual inclusion (conditioning). Since 1c does not account for longitudinal structure, we specify

$$Var\left[\left\{Z_{1i1} \cdots Z_{1i\tau}, Z_{2i1} \cdots Z_{2i\tau}, \dot{Z}_{1i1} \cdots \dot{Z}_{1i\tau}, \dot{Z}_{2i1} \cdots \dot{Z}_{2i\tau}\right\}\right] \stackrel{def}{=} \begin{bmatrix} \sigma_{12}^2 P_1 & 0 & 0 & 0 \\ 0 & \sigma_{22}^2 P_2 & 0 & 0 \\ 0 & 0 & \sigma_{12}^2 P_1 & 0 \\ 0 & 0 & 0 & \sigma_{22}^2 P_2 \end{bmatrix} \quad [1d]$$

with

$$P_k \stackrel{def}{=} \begin{bmatrix} 1 & \rho_{k12} & \cdots & \rho_{k1\tau} \\ \rho_{k12} & 1 & \cdots & \rho_{k2\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1\tau} & \rho_{k2\tau} & \cdots & 1 \end{bmatrix}, \quad \dot{P}_k \stackrel{def}{=} \begin{bmatrix} 1 & \dot{\rho}_{k12} & \cdots & \dot{\rho}_{k1\tau} \\ \dot{\rho}_{k12} & 1 & \cdots & \dot{\rho}_{k2\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \dot{\rho}_{k1\tau} & \dot{\rho}_{k2\tau} & \cdots & 1 \end{bmatrix} \quad [1e]$$

Together, 1a through 1e constitute a *completely general* formulation to permit cTSRI-IV estimation for $\kappa \equiv 2$. The choice of mappings, covariates, any further restrictions on the

covariance structure, as well as statistical distributions for adherence, outcomes, and random coefficients will be specific to each application; and, as indicated above, models can be specified for $2 < \kappa$. Specific model structures are illustrated in Sections 3.3 and 4.

3.2 Parameter estimation

Various approaches to cTSRI-IV estimation of the parameters of model 1 are possible. Here maximum likelihood estimation is employed for its efficiency and availability. Let ψ_k denote the first-stage (adherence outcome) expectation parameters and ψ'_k denote the second-stage (efficacy outcome) expectation parameters. Respective observed adherence and efficacy outcome vectors are a_k and O_k for assigned condition k . We appropriately *condition on the random coefficients* such that the resultant *generalized* sequence of likelihood specifications are

$$\mathcal{L} \left[a | \psi_1, \psi_2, \sigma_{11}^2, \sigma_{21}^2 \right] = \prod_{k=1}^2 \prod_{i=1}^n \mathcal{L} \left[a_{kit} | \psi_k, R_{ki} \right] \mathcal{L} \left[R_{ki} | \sigma_{k1}^2 \right], \quad [2a]$$

$$\mathcal{L} \left[o_k | \psi'_k \right] = \prod_{i=1}^n \mathcal{L} \left[o_{kit} | \{ \psi_k, r_{ki}, a_{kit} \} \right] \mathcal{L} \left[R_{ki} | \sigma_{k2}^2 \right]. \quad [2b]$$

Because we wish to place no further restrictions on the distribution of the R_{ki} , estimation of 2a is via an expectation-maximization (EM) algorithm [31, 32]. The EM algorithm is initiated by drawing initial values for the random coefficients from a mixture of normal distributions, although final predictions r_{ki} resulting from application of the EM algorithm need not have a distribution that is a mixture of normal distributions. Use of an EM algorithm is intuitively appealing because it is designed to recover latent structures, in this case latent confounders. Likelihoods 2a and 2b can be combined as a product (due to independence between assigned conditions and conditioning in the second stage) for estimation purposes but that only offers an efficiency advantage in terms of reduced variance of parameter estimates if the second stage model for the efficacy outcome shares a common parameter with the first stage model for the adherence outcome—a restriction that may often not be warranted on scientific grounds. Nor do we wish to necessarily have common parameters across assigned conditions in the second stage, although this option could be explored. A sequential estimation procedure is illustrated in Sections 3.3 and 4.

The structure of model 1 can induce strong correlation among A_{kit} , Z_{kit} , and R_{ki} . This multicollinearity complicates second-stage (efficacy outcome) parameter estimation. In addition to second-stage predictors a_{kit} , Z_{kit} , and r_{ki} , this multicollinearity can also encompass the second-stage intercept. Regularization approaches such as a ridge penalty are one possibility (Section 5.1), although, without care, these may introduce bias. In Sections 3.3 through 4.5, we apply an alternative strategy that combines orthogonalization and white-noise data augmentation [33] to mitigate multicollinearity's impact on finite-sample bias and variance.

3.3 Simulation studies

A set of ninety-six simulation studies were conducted to assess small-sample bias properties of the proposed cTSRI-IV estimator. All studies assumed a RCT with 1:1 random assignment of participants to two different conditions and participants observed at six visits that were regularly-spaced in time. Each participant's dose was simulated as a linear function of a fixed intercept term $\alpha_0 = 100$, a random participant-specific intercept term

$Z_i \stackrel{iid}{\sim} N(0, 25)$, **participant's age** $\alpha_1 = \frac{1}{5}$ (*time-invariant*), assigned condition $\alpha_2 \in \{10, 50\}$

(instrument), and a *time-varying* random variable $I_{it} = B_{it} - 1$, $B_{it} \stackrel{iid}{\sim} \text{Bernoulli}(\frac{1}{2})$ of effect $\alpha_3 = 5$. Parameter α_2 was varied in value to represent weak ($\alpha_2 = 10$) and strong ($\alpha_2 = 50$) instruments. In the first assigned condition, efficacy response was simulated as a linear function of a fixed intercept term, $\beta_0 = 0$, visit $\beta_1 \in \{\frac{1}{2}, 1\}$, participant's dose $\beta_2 \in \{1, 2, 3, 4, 5, 6\}$ of that condition, participant's age $\beta_3 = \frac{1}{5}$, which served as a time-invariant confounder, and two random terms—a random participant-specific deviation Z_i from the

fixed intercept ($Z_i \stackrel{iid}{\sim} N(0, c_1 25)$, $c_1 \in \{0, 5, 10, 15, 20\}$) **and a time-varying residual**

$E_{it} \stackrel{iid}{\sim} N(0, c_2 \sigma^2)$, $c_2 \in \{1, 20\}$, where, for example, $c_2 \in \{1, 20\}$ indicates that c_2 was varied over the set of values across simulation studies (Table 1). The efficacy response in the second condition differed from the first due to the addition of I_{it} , $\beta_4 = 5$, which served as a time-varying confounder, and the removal of the age effect $\beta_3 = 0$ (time-invariant confounder) in the linear predictor. For computational efficiency, simulation studies were nested such that c_1 and β_2 varied over their each pseudo-random (Mersenne-Twister) draw of Z_i . Five hundred data sets were randomly generated at each of two sample sizes ($n \in \{50, 100\}$) for each combination of parameter values. The full set of simulations was repeated setting $\hat{g}_k(\cdot) = \exp(\cdot)$ to examine finite sample properties of the cTSRI-IV estimator under nonlinear structure.

For each simulated data set, contrasting approaches were used for fitting. The “naïve” approach entailed a single-stage regression of the efficacy outcome on dose, the dose by condition interaction, visit, and the visit by condition interaction. Fitting employed either a linear mixed-coefficients model formulation with fixed coefficients for the four covariates, normally distributed random intercepts and residuals, and estimation via restricted maximum likelihood for $\hat{g}_k(\cdot) = 1$ or, in the simulations under nonlinear structure, a generalized linear mixed-coefficients model formulation with fixed coefficients for the four covariates, normally distributed random intercepts, logarithm link, and estimation via adaptive Gaussian quadrature. The cTSRI-IV approach was in two stages. The first stage was a linear mixed-coefficients model formulation with fixed coefficients for condition, visit and their interaction; nonparametrically distributed random intercepts; normally distributed residuals; and estimation via the EM algorithm, initiated with a mixture of two normal distributions for the random intercepts. (In practice, different mixture quantities are applied and goodness of fit assessed, with the quantity of components dependent upon the structure of the time-invariant latent confounders). The second stage entailed fitting a separate fixed-coefficients generalized linear mixed model for each assigned condition. For each of these two models, the efficacy outcome was regressed on dose, visit, the participant's first-stage predicted value for the random intercept, and the participant's first-stage predicted residual values. Second-

stage residuals were assumed to be normally (or log-normally) distributed, as designed. Multicollinearity was addressed through a combination of Gram-Schmidt orthogonalization of r_{ki} with respect to a_{kit} and white-noise augmentation of the intercept vector as $1 + \lambda w$, for tuning constant $0 < \lambda$, where w is a $6 \times n$ element vector of pseudo-random independent draws from a standard normal distribution. Constant λ will need to be optimized in each application. Settings too close to zero cause failure in convergence of the optimization algorithm while settings too far from zero increase bias and variance of the estimate of the LATEC. In general, λ should be employed as an algorithm convergence tuning parameter and set to a value just large enough to provide reliable algorithm convergence. We present results for settings of $\lambda = 0.01$ for the linear model and $\lambda \in [0.01, 0.25]$ for the nonlinear model. Nonparametric fixed effects modeling was performed using the `nplmreg` package in R version 3.2 [34]. We found that the trust-region optimization algorithm, with central finite difference derivative approximation and a small initial trust region radius (e.g., < 0.01), available in the `nlmixed` procedure in SAS[®] version 9.4 (SAS[®] Institute, Cary, NC, USA), gave algorithm convergence failures below 4.7% and 13.5% during simulations for linear and nonlinear models, respectively. Second-stage random coefficients were predicted via Laplace approximation. Because neither the naïve approach nor cTSRI-IV estimation explicitly included age nor I_{it} as covariates, these served as unobserved time-invariant and time-varying confounders, respectively. Code for simulation studies is available upon request (THH).

4. Results

4.1 Simulation results

Results of the 96 simulation studies are summarized in Table 1. Compared to the naïve estimator, cTSRI-IV estimation reduced bias in estimates of adherence dosage in all but a few rare cases (gray cells in Table 1). Bias reduction was greatest for higher levels of residual variance ($c_2 = 20$). Among cTSRI-IV estimates, bias reduction was consistently greater for a time-varying confounder than the time-invariant confounder. Bias reduction was similar across linear and nonlinear models.

4.2 The APPLES trial

Application of the **cTSRI-IV estimator** is demonstrated using data from a RCT in sleep medicine. The Apnea Positive Pressure Long-term Efficacy Study (APPLES) was a multicenter, clinical trial sponsored by the National Heart, Lung and Blood Institute, in which 1,105 participants were randomized at a 1:1 ratio to either continuous positive airway pressure (CPAP) therapy for obstructive sleep apnea (OSA) or to a sham device [35]. Per the study's protocol, each participant was to complete baseline screening and diagnostic testing, be randomized to sham or active, and followed for 6 months. All participants provided informed consent. The study protocol was approved by the Institutional Review Board at each of five participating centers. Three primary neurocognitive outcomes, corresponding to each of three major domains of neurocognitive function (attention and vigilance, learning and memory, and executive/frontal lobe function), were assessed pre-randomization (baseline) and at 2-month and 6-month visits during the active intervention period: the Buschke Selective Reminding Test - Sum Recall (BSRT-SR), the Pathfinder Number Test -

Total Time (Trails A of Trail Making Test; PFNT-TT), and the primary mid-day score from the Sustained Working Memory Test – Overall Mid-Day Index (SWMT-OMI [36]). Greater neurocognitive ability is indicated by higher scores for BSRT-SR and SWMT-OMI and lower scores for PFNT-TT. **Here we restrict** analysis to the PFNT-TT as the efficacy outcome. A total of 849 participants had data at the 2-month and 6-month visits for the analysis presented here, $n_1 = 412$ for sham and $n_2 = 437$ for active. Information about the impact of missing data due to dropout is detailed in Kushida et al. [35]. All analyses described in the remainder of Section 4 were performed in SAS® version 9.4 (SAS® Institute, Cary, NC, USA) and R [34].

4.3 cTSRI-IV estimation for APPLES

Participants in APPLES were randomized to two assigned conditions, so $\kappa = 2$ as in Section 3.1. Separately for each of these three outcomes, we fit model 1 with the following tailoring to **APPLES**:

$$\begin{aligned} \mu_{ki} &\stackrel{\text{def}}{=} E[A_{kit}|R_{ki}] = \alpha_0 + \alpha_1(k-1) + \sum_{p=2}^6 \alpha_p x_{pi} + R_{ki} \\ \theta_{ki} &\stackrel{\text{def}}{=} E\left[O_{kit}|A_{kit}=a_{kit}, R_{ki}=r_{ki}, R_{ki}\right] = \beta_{k0} + \beta_{k1}a_{kit} + \sum_{p=2}^6 \beta_{pk}x_{pki} + \varepsilon_k(a_{kit} - \mu_{ki}) + \zeta_k r_{ki} + R_{ki} \end{aligned}$$

[3a]

and

$$P_k \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad P_k' \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad [3b]$$

Here $k = 1$ for sham and $k = 2$ for active. Covariates $\{x_2, x_3, x_4, x_5\}$ are those used to construct randomization strata: race (non-Caucasian $x_2 = 0$, Caucasian $x_2 = 1$), gender (female $x_3 = 0$, male $x_3 = 1$), moderately severe OSA at baseline ($x_4 = 1$, $x_4 = 0$ otherwise) and severe OSA at baseline ($x_5 = 1$, $x_5 = 0$ otherwise). Baseline age in years is also a covariate (x_6). Age is a potentially strong confounder. In our experience, adherence often tends to increase with participants' age in RCTs. Also, performance on neurocognitive measures can decline with age. For purposes of exposition, this model assumes a common LATEC across the subpopulations defined by the covariates. The common LATEC assumption can be relaxed [11] and indeed this possibility should be examined in a complete analysis, including for the APPLES trial. Random coefficients $\mu_{ki} \sim \mathcal{N}(0, \sigma^2)$ are also included in the second stage to account for possible correlation between the two repeated measurements. All parameter estimates are time-averages across the 2-month and 6-month visits. Alternative formulations that allow for interaction with time are also possible. Specification 3b assumes that repeated measurements on the same outcome (adherence or a neurocognitive measure) within a person are uncorrelated conditional upon the random

coefficient R_{ki} . We assume $Z_{kit} \sim N(0, \sigma_{k2}^2)$ and $Z'_{kit} \sim N(0, \sigma_{k3}^2)$ and $\mathbf{g}_k = \mathbf{1}, k \in \{1, 2\}$.
Likelihood formulation is

$$\mathcal{L} [a | \psi_1, \psi_2, \sigma_{11}^2, \sigma_{21}^2] = \prod_{k=1}^2 \prod_{i=1}^n \phi [a_{kit} | \psi_k, R_{ki}] f_2 [R_{ki} | \sigma_1^2], \quad [4a]$$

$$\mathcal{L} [o_k | \psi_k] = \prod_{i=1}^n \phi [o_{kit} | \left\{ \psi_k, r_{ki}, a_{kit} \right\}] \phi [R_{ki} | \sigma_{k2}^2], \quad [4b]$$

where $\phi[\cdot]$ denotes the normal density, f_2 is the density function for an unspecified distribution where the subscript indicates a two-component mixture density, and $k \in \{1, 2\}$.

4.4 Model fitting and hypothesis testing

Adherence was measured nightly **as hours of device usage per night** via a monitor built into the sham and active CPAP devices (Encore® Pro SmartCard®, Phillips Respironics® Inc., Andover, MA, USA) and formulated as the average nightly value over the prior two months. Approximately 21% and 29% of nightly adherence data were missing at 2 months and 6 months, respectively, yielding the sample size of 849 participants for fitting model 3. For present analysis, missing status was assumed to be non-informative (i.e., due to monitor malfunction) so that 2-month averages were calculated using only observed adherence values. Alternative approaches could entail some form of imputation, including setting missing values to zero assuming the participant was not using the device.

Parameters were estimated by separately optimizing each of the three likelihood equations (4a and 4b) using the trust-region optimization algorithm, as described in Section 3.3, with an initial trust region radius setting of 0.01; and we found $\lambda = 0.01$ provided reliable optimization algorithm convergence. Second-stage random coefficients were predicted via Laplace approximation. To propagate sampling error properly from the first stage into the second, we employed bootstrap resampling stratified on assigned condition. From 1,000 bootstrap resamples we calculated simple percentile 95% confidence bounds.

4.5 Findings

Table 2 reports estimates of bootstrap 95% confidence intervals from the fit of model 3 to the PFNT-TT efficacy outcome data from APPLIES. Results suggest that an increasing dose of active adherence decreases (improves) average PFNT-TT scores, a finding not detected in the absence of cTSRI-IV estimation in the original report [35]. The distribution of second-stage efficacy residuals were somewhat skewed (sample skewness = 1.46) and strongly leptokurtic (sample kurtosis = 7.89), indicating a violation of the model's assumption that these residuals are drawn from a Gaussian distribution. We suggest possible remedies in Section 5.1.

5. Discussion

5.1 Statistical implications

The proposed method addresses estimation of the causal effect of treatment as received (LATEC) under possibly nonlinear and longitudinal model structures. **Reduction in bias due to unobserved confounding** is provided through isolation and adjustment for exogenous and endogenous components of adherence via time-invariant and time-variant residual inclusion (TSRI). Adjustment for *observed* confounders, such as age and known comorbidities, is also essential. Age may be useful because adherence and health outcomes often vary among RCT participants of differing ages. Altogether, this especially comprehensive approach to adjustment for unobserved confounders should effectively mitigate inconsistency in estimates of the LATEC in RCTs of longitudinal sampling structures, as our simulation studies suggest.

A potential drawback to comprehensive adjustment for unobserved confounders is that the cTSRI-IV model is enriched in fixed parameters, especially with incorporation of interaction terms and/or spline bases for covariates. When sample sizes are modest, identification of model parameters may oftentimes require further restrictions on those parameters. One general approach would be to add a L_q -penalty term (e.g., $q = 1$ for lasso or $q = 2$ for ridge regression) to the likelihood (equation 2) or, even better, apply these penalties to derive inputs such as principal components [37] as an additional means for addressing multicollinearity. This approach comes with additional computational complexity as penalty parameters may need be estimated via sample reuse procedures [37]. Further, additional sample reuse procedures, such as bootstrapping [38], would be needed to generate confidence intervals on parameter estimates that propagate sampling error in the penalty parameter estimate into LATEC estimates. To minimize introduction of bias into the estimate of the LATEC, the LATEC parameters could be excluded from the penalty term. Penalized estimation of the LATEC under the cTSRI-IV model merits study.

The cTSRI-IV model also contains random components (random effects and residuals). The standard approach to modeling random coefficients is to assume that these coefficients follow a Gaussian distribution. We propose a nonparametric finite mixture alternative to account for the overdispersion induced by the latent confounder(s). Another nonparametric approach is generalized maximum entropy estimation [39]. Generalized maximum entropy estimation **can be** useful for small sample sizes, in the presence of multicollinearity among covariates, and can also accommodate residual distributions that are skewed or strongly leptokurtic (Section 4.5) [40]. Bayesian estimation is yet another alternative and an attractive one in a parameter-rich setting because, like generalized maximum entropy estimation, Bayesian estimation permits borrowing information about parameters from prior published studies via empirically informed priors [41]; and Bayesian estimation can employ finite mixtures to model complex (e.g. leptokurtic) residual distributions.

We employ a nonparametric bootstrap procedure for constructing confidence intervals on model parameters (Section 4.4) to accommodate the possibly nonlinear structure of the model and two-stage estimation procedure. Various standard diagnostics, bias correctives, and efficiency gains are available for the nonparametric bootstrap [42], including those that

allow for residual heteroscedasticity [43], so that, altogether the nonparametric bootstrap allows proper inference without sacrificing flexibility of model structure.

5.2 Limitations

The list of assumptions for consistent IV estimates of dose response is not short (Section 2.2). Among those assumptions is that the IV impacts outcome only indirectly through exposure to intervention (“exclusion restriction”). This should result if the RCT is effectively double-blinded [13]. APPLES was designed as a double-blind study; yet, analysis of exit interviews suggested that some participants, especially in the **sham** condition, may have been aware of their assigned condition, although overall degree of agreement between assigned and blinded guesses was poor [34]. Another assumption, SUTVA, will not hold if some participants interfere with other participants’ outcomes through social interactions [44].

5.3 Conclusions

We introduce a method for estimating the causal effects of treatment as received simultaneously for each assigned condition of a RCT within a framework that permits nonlinear modeling of **longitudinal** data structures. A comprehensive approach to adjustment for unobserved confounders is included. Future work could explore **asymptotic properties** of the proposed **estimator and** the implications of stochastic adherence [45].

Acknowledgments

Dr. Holmes was supported in part by the Department of Veterans Affairs Office of Specialty Care Transformation. Dr. Zulman was supported by a Department of Veterans Affairs HSR&D Career Development Award (CDA 12-173). APPLES was funded by contract 5U01-HL-068060 from the National Heart, Lung and Blood Institute. The APPLES pilot studies were supported by grants from the American Academy of Sleep Medicine and the Sleep Medicine Education and Research Foundation to Stanford University and by the National Institute of Neurological Disorders and Stroke (N44-NS-002394) to SAM Technology.

References

1. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. *Clin Trials*. 2007; 4:58–73. [PubMed: 17327246]
2. Sagarin BJ, West SG, Ratnikov A, Homan WK, Ritchie TD, Hansen EJ. Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychol Methods*. 2014; 19:317–333. [PubMed: 24773358]
3. Hill, AB. Principles of medical statistics. Oxford University Press; New York: 1961.
4. Ten Have TR, Normand ST, Marcus SM, et al. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatr Ann*. 2008; 38:772–783. [PubMed: 20717484]
5. Feinman RD. Intention-to-treat. what is the question? *Nutr Metabol*. 2009; 6:1.
6. Slone D, Shapiro S, Miettinen OS, et al. Drug evaluation after marketing. *Ann Intern Med*. 1979; 90:257–261. [PubMed: 443660]
7. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
8. Imbens GW. The role of the propensity score in estimating dose-response functions. *Biometrika*. 2000; 87:706–710.
9. Imai K, Van Dyk DA. Causal inference with general treatment regimes: Generalizing the propensity score. *J Am Stat Assoc*. 2004; 99:854–866.

10. Ten Have TR, Joffe M, Cary M. Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Stat Med.* 2003; 22:1255–1283. [PubMed: 12687654]
11. Bond SJ, White IR, Walker AS. Instrumental variables and interactions in the causal analysis of a complex clinical trial. *Stat Med.* 2007; 26:1473–1496. [PubMed: 16900567]
12. Huber M. Sensitivity checks for the local average treatment effect. *Econ Lett.* 2014; 123:220–223.
13. Hernán MA, Robins JM. Instruments for causal inference: An epidemiologist's dream? *Epidemiol.* 2006; 17:360–372.
14. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996; 91:444–455.
15. Small DS, Ten Have TR, Joffe MM, et al. Random effects logistic models for analysing efficacy of a longitudinal randomized treatment with non-adherence. *Stat Med.* 2006; 25:1981–2007. [PubMed: 16220487]
16. McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Stat Med.* 2009; 28:2639–2652. [PubMed: 19579227]
17. Forcina, A. Partial compliance, effect of treatment on the treated and instrumental variables.. In: Fichet, B.Piccolo, D.Verde, R., et al., editors. *Classification and multivariate analysis for complex data structures.* Springer-Verlag; Berlin: 2011. p. 317-324.
18. Miladinovic B, Kumar A, Hozo I, et al. Instrumental variable meta-analysis of individual patient data: Application to adjust for treatment non-compliance. *BMC Med Res Methodol.* 2011; 11
19. Nelson RE, Nebeker JR, Hayden C, et al. Comparing adherence to two different HIV antiretroviral regimens: An instrumental variable analysis. *AIDS Behav.* 2013; 17:160–167. [PubMed: 22869102]
20. Reardon SF, Unlu F, Zhu P, et al. Bias and bias correction in multisite instrumental variables analysis of heterogeneous mediator effects. *J Educ Behav Stat.* 2014; 39:53–86.
21. Sato T. A method for the analysis of repeated binary outcomes in randomized clinical trials with non-compliance. *Stat Med.* 2001; 20:2761–2774. [PubMed: 11523081]
22. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986; 73:13–22.
23. Lamiraud K, Moatti J, Raffi F, et al. Adherence to and effectiveness of highly active antiretroviral treatment for HIV infection: Assessing the bidirectional relationship. *Med Care.* 2012; 50:410–418. [PubMed: 22362167]
24. Diggle, PJ., Liang, KY., Zeger, SL. *Analysis of longitudinal data.* Clarendon Press; Oxford: 1996.
25. Milat AJ, King L, Newson R, et al. Increasing the scale and adoption of population health interventions: Experiences and perspectives of policy makers, practitioners, and researchers. *Health Res Policy Syst.* 2014; 12 Article 18.
26. Winchell C, Rappaport BA, Roca R, et al. Reanalysis of methamphetamine dependence treatment trial. *CNS Neurosci Therap.* 2012; 18:367–368. [PubMed: 22533722]
27. McCulloch, CE., Searle, SR. *Generalized, linear, and mixed models.* John Wiley & Sons, Inc.; New York: 2001.
28. Sitlani CM, Heagerty PJ, Blood EA, et al. Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Stat Med.* 2012; 31:1738–1760. [PubMed: 22344923]
29. McCulloch C. Joint modelling of mixed outcome types using latent variables. *Stat Meth Med Res.* 2008; 17:53–73.
30. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ.* 2008; 27:531–543. [PubMed: 18192044]
31. Aitkin M. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics.* 1999; 55:117–128. [PubMed: 11318145]
32. Einbeck J, Hinde J. A note on NPML estimation for exponential family regression models with unspecified dispersion parameter. *Austrian Journal of Statistics.* 2006; 35:233–243.
33. Conlin AK, Martin EB, Morris AJ. Data augmentation: An alternative approach to the analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems.* 1998; 44(1-2):161–173.
34. R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing; Vienna: 2015.

35. Kushida CA, Nichols DA, Holmes TH, et al. Effects of continuous positive airway pressure on neurocognitive function in obstructive sleep apnea patients: The apnea positive pressure long-term efficacy study (APPLES). *SLEEP*. 2012; 35:1593–1602. [PubMed: 23204602]
36. Kushida CA, Nichols DA, Quan SF, et al. The apnea positive pressure long-term efficacy study (APPLES): Rationale, design, methods, and procedures. *J Clin Sleep Med*. 2006; 2:288–300. [PubMed: 17561541]
37. Hastie, T., Tibshirani, R., Friedman, JH. The elements of statistical learning: data mining, inference, and prediction. second edition. Springer; New York: 2009.
38. Kazimi C, Brownstone D. Bootstrap confidence bands for shrinkage estimators. *J Econometrics*. 1999; 90:99–127.
39. Peeters LMK. Estimating a random-coefficients sample-selection model using generalized maximum entropy. *Econ Lett*. 2004; 84:87–92.
40. Golan, A., Judge, G., Miller, D. Maximum entropy econometrics: robust estimation with limited data. John Wiley & Sons; Chichester: 1996.
41. Holmes TH, Lewis DB. Bayesian immunological model development from the literature: example investigation of recent thymic emigrants. *J Immunol Methods*. 2014; 414:32–50. [PubMed: 25179832]
42. Davison, AC., Hinkley, DV. Bootstrap methods and their application. Cambridge University Press; Cambridge: 1997.
43. Davidson R, Flachaire E. The wild bootstrap, tamed at last. *J Econom*. 2008; 146:162–169.
44. Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. *Stat Meth Med Res*. 2012; 21:55–75.
45. Scosyrev E. Identification of causal effects using instrumental variables in randomized trials with stochastic compliance. *Biom J*. 2013; 55:97–113. [PubMed: 23180483]

Table 1

Results of 96 simulation studies as described in section 3.3.

Type	n	a_2	β_1	c_1	c_2	β_2	Naïve		cTSRI-IV	
							Time-invariant confounder	Time-varying confounder	Time-invariant confounder	Time-varying confounder
							$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$
L	50	10	1	10	1	1	1.41	1.49	1.20	1.09
L	50	10	1/2	0	1	2	2.25	2.37	2.19	2.08
L	50	10	1	5	1	3	3.46	3.60	3.20	3.09
L	50	10	1/2	15	1	4	4.73	4.14	4.18	4.07
L	50	10	1	20	1	5	5.77	5.09	5.20	5.08
L	50	10	1/2	10	1	6	6.42	6.50	6.18	6.07
L	50	10	1	10	20	1	2.24	2.45	1.18	1.07
L	50	10	1/2	0	20	2	2.13	2.25	2.19	2.15
L	50	10	1	5	20	3	3.70	3.87	3.28	3.16
L	50	10	1/2	15	20	4	5.71	5.96	4.30	4.06
L	50	10	1	20	20	5	7.06	7.34	5.31	5.18
L	50	10	1/2	10	20	6	7.25	7.46	6.29	6.01
L	50	50	1	10	1	1	1.34	1.50	1.15	1.05
L	50	50	1/2	0	1	2	2.21	2.35	2.14	2.04
L	50	50	1	5	1	3	3.40	3.60	3.15	3.05
L	50	50	1/2	15	1	4	4.35	4.44	4.14	4.05
L	50	50	1	20	1	5	5.67	5.16	5.15	5.05
L	50	50	1/2	10	1	6	6.34	6.50	6.14	6.04
L	50	50	1	10	20	1	2.11	2.54	1.22	1.09
L	50	50	1/2	0	20	2	2.12	2.24	2.16	2.01
L	50	50	1	5	20	3	3.65	3.93	3.16	3.03
L	50	50	1/2	15	20	4	5.53	6.09	4.08	4.05
L	50	50	1	20	20	5	6.81	7.46	5.26	5.07
L	50	50	1/2	10	20	6	7.12	7.55	6.01	6.05
L	100	10	1	10	1	1	1.42	1.50	1.20	1.09

Type	n	a_2	β_1	c_1	c_2	β_2	cTSRI-IV			
							Naïve		Time-varying confounder	
							Time-invariant confounder	$\hat{\beta}_2$	Time-invariant confounder	$\hat{\beta}_2$
L	100	10	1/2	0	1	2	2.25	2.37	2.18	2.07
L	100	10	1	5	1	3	3.47	3.61	3.20	3.08
L	100	10	1/2	15	1	4	4.76	4.10	4.19	4.07
L	100	10	1	20	1	5	5.77	5.09	5.20	5.09
L	100	10	1/2	10	1	6	6.41	6.50	6.19	6.07
L	100	10	1	10	20	1	2.26	2.47	1.17	1.01
L	100	10	1/2	0	20	2	2.13	2.24	2.21	2.07
L	100	10	1	5	20	3	3.70	3.86	3.16	3.14
L	100	10	1/2	15	20	4	5.73	5.98	4.12	4.03
L	100	10	1	20	20	5	7.07	7.35	5.17	5.05
L	100	10	1/2	10	20	6	7.26	7.47	6.20	6.11
L	100	50	1	10	1	1	1.34	1.50	1.15	1.05
L	100	50	1/2	0	1	2	2.20	2.34	2.14	2.04
L	100	50	1	5	1	3	3.40	3.60	3.15	3.05
L	100	50	1/2	15	1	4	4.34	4.44	4.14	4.04
L	100	50	1	20	1	5	5.71	5.12	5.15	5.05
L	100	50	1/2	10	1	6	6.34	6.50	6.14	6.04
L	100	50	1	10	20	1	2.11	2.54	1.13	1.07
L	100	50	1/2	0	20	2	2.12	2.24	2.13	2.08
L	100	50	1	5	20	3	3.63	3.90	3.12	3.04
L	100	50	1/2	15	20	4	5.53	6.09	4.10	4.10
L	100	50	1	20	20	5	6.82	7.47	5.14	5.06
L	100	50	1/2	10	20	6	7.12	7.55	6.17	6.07
NL	50	10	1	10	1	1	1.43	1.50	1.21	1.10
NL	50	10	1/2	0	1	2	2.26	2.38	2.21	2.09
NL	50	10	1	5	1	3	3.50	3.64	3.22	3.15
NL	50	10	1/2	15	1	4	4.71	4.19	4.18	4.08
NL	50	10	1	20	1	5	5.78	5.11	5.20	5.09

Type	n	α_2	β_1	c_1	c_2	β_2	Naïve			cTSRI-IV		
							Time-invariant confounder	Time-varying confounder	$\hat{\beta}_2$	Time-invariant confounder	Time-varying confounder	$\hat{\beta}_2$
							$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_2$
NL	50	10	1/2	10	1	6	6.47	6.56	6.19	6.09	6.19	6.09
NL	50	10	1	10	20	1	2.25	2.46	1.18	1.06	1.18	1.06
NL	50	10	1/2	0	20	2	2.14	2.26	2.15	2.03	2.15	2.03
NL	50	10	1	5	20	3	3.71	3.87	3.16	3.04	3.16	3.04
NL	50	10	1/2	15	20	4	5.73	5.98	4.17	4.08	4.17	4.08
NL	50	10	1	20	20	5	7.06	7.34	5.20	5.08	5.20	5.08
NL	50	10	1/2	10	20	6	7.25	7.45	6.17	6.03	6.17	6.03
NL	50	50	1	10	1	1	1.38	1.55	1.15	1.05	1.15	1.05
NL	50	50	1/2	0	1	2	2.25	2.41	2.14	2.05	2.14	2.05
NL	50	50	1	5	1	3	3.52	3.76	3.15	3.05	3.15	3.05
NL	50	50	1/2	15	1	4	4.51	4.68	4.14	4.05	4.14	4.05
NL	50	50	1	20	1	5	5.62	5.56	5.15	5.05	5.15	5.05
NL	50	50	1/2	10	1	6	-616.18*	-299.39*	6.14	6.05	6.14	6.05
NL	50	50	1	10	20	1	2.12	2.56	1.15	1.05	1.15	1.05
NL	50	50	1/2	0	20	2	2.12	2.23	2.14	2.04	2.14	2.04
NL	50	50	1	5	20	3	3.66	3.95	3.15	3.05	3.15	3.05
NL	50	50	1/2	15	20	4	5.59	6.17	4.14	4.05	4.14	4.05
NL	50	50	1	20	20	5	6.93	7.61	5.15	5.05	5.15	5.05
NL	50	50	1/2	10	20	6	-2905.40*	-3469.35*	6.14	6.04	6.14	6.04
NL	100	10	1	10	1	1	1.42	1.49	1.20	1.09	1.20	1.09
NL	100	10	1/2	0	1	2	2.25	2.37	2.19	2.08	2.19	2.08
NL	100	10	1	5	1	3	3.49	3.63	3.20	3.10	3.20	3.10
NL	100	10	1/2	15	1	4	4.74	4.13	4.19	4.07	4.19	4.07
NL	100	10	1	20	1	5	5.77	5.10	5.20	5.09	5.20	5.09
NL	100	10	1/2	10	1	6	6.46	6.55	6.19	6.08	6.19	6.08
NL	100	10	1	10	20	1	2.26	2.47	1.18	1.08	1.18	1.08
NL	100	10	1/2	0	20	2	2.12	2.24	2.19	2.05	2.19	2.05
NL	100	10	1	5	20	3	3.71	3.88	3.19	3.08	3.19	3.08

Type	n	a ₂	β ₁	c ₁	c ₂	β ₂	Naïve			cTSRI-IV				
							Time-invariant confounder		β ₂	Time-varying confounder		β ₂	Time-invariant confounder	Time-varying confounder
							β ₂	β ₂						
NL	100	10	1/2	15	20	4	5.73		5.98		4.18		4.07	
NL	100	10	1	20	20	5	7.16		7.43		5.21		5.10	
NL	100	10	1/2	10	20	6	7.26		7.47		6.18		6.05	
NL	100	50	1	10	1	1	1.35		1.55		1.15		1.05	
NL	100	50	1/2	0	1	2	2.25		2.40		2.14		2.05	
NL	100	50	1	5	1	3	3.50		3.75		3.15		3.05	
NL	100	50	1/2	15	1	4	4.50		4.70		4.14		4.05	
NL	100	50	1	20	1	5	5.60		5.55		5.15		5.06	
NL	100	50	1/2	10	1	6	-8196.30 [*]		-8040.75 [*]		6.14		6.05	
NL	100	50	1	10	20	1	2.15		2.60		1.15		1.05	
NL	100	50	1/2	0	20	2	2.15		2.25		2.14		2.04	
NL	100	50	1	5	20	3	3.65		3.95		3.15		3.05	
NL	100	50	1/2	15	20	4	5.60		6.15		4.14		4.05	
NL	100	50	1	20	20	5	6.95		7.65		5.15		5.06	
NL	100	50	1/2	10	20	6	-5196.85 [*]		-5543.85 [*]		6.14		6.04	

Types of models are linear = L or nonlinear = N. n = total sample size, a_2 = strength of instrument, β_1 = effect on efficacy outcome of each visit, β_2 = effect on efficacy outcome of each dose unit, c_1 = scaling factor for the variance of the time-invariant random deviation Z_i from the fixed intercept, c_2 = scaling factor for the variance of the time-varying residual E_{it} and $\hat{\beta}_2$ is the estimate of effect on efficacy outcome of each dose unit for the indicated estimator and condition. Cells in gray marks instances where bias of CTSRI-IV estimator exceeded that for the naïve estimator.

^{*} failure of algorithm convergence.

Ninety-five percent confidence bounds on parameter estimates from fit of model 3 to the PFNT-TT efficacy outcome from APPLES. RE = random effect.

Table 2

Parameter	2.5%	97.5%
Sham Intercept	21.03	24.18
Active Intercept	22.12	25.01
Sham RE	8.98	11.20
Active RE	-5.01	-2.21
Sham Residual	-5.40	-1.73
Active Residual	-0.02	1.79
Sham Age	0.49	2.78
Active Age	0.68	3.51
Sham White	-0.45	-0.08
Active White	-0.68	-0.08
Sham Male	-0.22	0.00
Active Male	-0.10	0.06
Sham Moderate	0.00	0.40
Active Moderate	0.03	0.31
Sham Severe	-0.19	0.13
Active Severe	0.15	0.69
Sham Adherence	-0.07	0.10
Active Adherence	-0.66	-0.35