

Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a Repeated Binary Outcome

James M. Robins¹

Harvard School of Public Health
Boston, MA

Sander Greenland²

UCLA School of Public Health
Los Angeles, CA

Fu-Chang Hu³

College of Public Health, National Taiwan University
Taipei, Taiwan, ROC⁴

SUMMARY

We provide sufficient conditions for estimating from longitudinal data the causal effect of a time-dependent exposure or treatment on the marginal probability of response for a dichotomous outcome. We then show how one can estimate this effect under these conditions using the G-computation algorithm of Robins (1986). We also derive the conditions under which some current approaches to the analysis of longitudinal data, such as the generalized estimating equations (GEE) approach of Zeger and Liang (1986), the feedback model techniques of Liang and Zeger (1991), and within-subject conditional methods, can provide valid tests and estimates of causal effects. We use our methods to estimate the causal effect of maternal stress on the marginal probability of a child's illness from the Mothers' Stress and Children's Morbidity data and compare our results with those previously obtained by Zeger and Liang (1986) using a generalized estimating equations (GEE) approach.

Keywords: Causal effects; Time-dependent covariates; Longitudinal data; G-computation algorithm; generalized estimating equations (GEE); Markov chains.

¹James M. Robins is Professor of Epidemiology and Biostatistics, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, robins@epinet.harvard.edu.

²Sander Greenland is Professor of Epidemiology, UCLA School of Public Health, Los Angeles, CA, 90095.

³Fu-Chang Hu is Professor, College of Public Health, National Taiwan University, 1, Jen-Ai Road, Sec. 1, Taipei, Taiwan, ROC.

⁴The authors are very grateful to Professor Cheryl Alexander at The Johns Hopkins University for kindly offering the Mothers' Stress and Children's Morbidity data, to Mr. Robert M. Byers and Ms. Monique Pambrun for their initial computing work, and to Professor Nan M. Laird for her comments on an early draft of this paper, which was the first part of the Sc.D. dissertation of the third author in the Department of Biostatistics at the Harvard School of Public Health under the advice of the first author.

The Mothers' Stress and Children's Morbidity Study (MSCM) is a longitudinal observational study of the causal effect of maternal stress on childhood illness (Zeger and Liang, 1986, pp. 125-128). In the MSCM data, the daily prevalence of childhood illness was 14%. Consider the following causal questions. How would the prevalence change if an on-going, fully effective stress reduction intervention program were instituted? How would prevalence change if conditions worsen and all mothers were subjected to substantial stress on a daily basis? To attempt to answer these questions, we shall use a formal model for causal effects in longitudinal studies introduced in Robins (1986, 1987ab). This model extends Rubin's (1974, 1978) causal model for "point" treatment studies to longitudinal studies with time-varying treatments and confounders. We show that the methods for causal inference developed by Robins provide a better justified basis for answering the above causal questions in longitudinal data in general and in the MSCM study in particular than do methods based on generalized estimating equations (GEEs).

The paper is organized as follows. In an observational study one can, in general, identify causal effects only when there are no unmeasured confounders. We use a counterfactual causal model to provide a formal definition of the assumption of no unmeasured confounders. Under this no-confounding assumption, we show, in Section 3, how to construct tests and estimators of the effect of stress on the marginal probability (i.e. prevalence) of illness using the G-computation algorithm of Robins, and, in Section 5, we reanalyze the mothers' stress and morbidity data with these methods. In Section 4, we show that the generalized estimating equation approach of Liang and Zeger could be used to test for or estimate causal effects only if past illness history is not a confounder for the effect of maternal stress on subsequent illness. In Section 5, we provide evidence that past illness history is a confounder in the MSCM data.

We also show that, if past illness history is the only time-dependent confounder, one can test for an effect of stress on the marginal probability of illness by testing whether illness at time m is conditionally independent of past stress history, given past illness history and baseline confounders. However, one must use the G-computation algorithm to estimate the magnitude of the effect. We show, in Section 6, that if there are additional time-dependent confounding factors, one cannot validly test the null hypothesis of no causal effect of stress on childhood illness by ordinary regression techniques, including the feedback model techniques of Liang and Zeger (1991), regardless of whether or not one adjusts for past confounder history in the analysis. Again, in this setting, valid tests and consistent estimates can be constructed using the G-computation algorithm. In Section 7, we derive conditions under which one can validly estimate the causal effect of stress on childhood illness by using within mother-child pair conditional likelihood methods. Robins has developed several other methods that complement the g-computation algorithm approach for estimating the causal effect of a time-varying exposure or treatment on the marginal mean of a dichotomous outcome. In Sec. 8, we briefly contrast the g-computation algorithm with two of these alternative methods: g-estimation of structural nested models and inverse probability of treatment weighted estimation of marginal structural models.

2.1. Causal Effect of a Time-Dependent Exposure

Our goal will be to estimate the effect of maternal stress on the prevalence of childhood illness at each of the M times at which the outcome was measured. Therefore, let $m = 1, 2, \dots, M$ index the discrete times at which data on the outcome $Y_{m,i}$ and the exposure $A_{m,i}$ were obtained, and let $i = 1, 2, \dots, N$ index the children in the MSCM study. In the context of the MSCM study, $Y_{m,i}$ was 1 if child i was ill on day m and was zero otherwise; $A_{m,i}$ was 1 if child i 's mother was suffering from stress on day m and was zero otherwise. Let X_i be a vector of time-independent baseline covariates recording household size, child's race, parental, employment, and marital status, and let $L_{m,i}$ be a vector of additional time-dependent covariates.

Rubin (1974, 1978) and Holland (1986) argued that a formal treatment of causal effects required explicit reference to possibly counterfactual outcomes. See Lewis (1973) for philosophical foundations of this approach and Rubin (1990) for a review of its history in statistics. We define $Y_{m,i}^{(0)}$ to be what the outcome of child i at time m was or would have been if the child had not been exposed to mother's stress at any time during follow-up. Similarly, $Y_{m,i}^{(1)}$ is what the outcome of the same child was or would have been if that child had been continuously exposed to mother's stress beginning at start of follow-up. In other words, $Y_{m,i}^{(0)}$ and $Y_{m,i}^{(1)}$ refer to the same child under different conditions; at least one and possibly both of these conditions (never exposed, always exposed) did not occur (i.e., is counterfactual), so that at most one and possibly neither equals $Y_{m,i}$ (the actual outcome). Our notation implicitly incorporates Cox's (1958) "no interaction between units" assumption that child i 's counterfactual outcomes do not depend on the treatment (i.e., maternal stress) received by any other child. Rubin (1978) refers to this assumption as the stable unit treatment value assumption.

For any process Z , let $\bar{Z}_m = (Z_1, \dots, Z_m)$ for $m \geq 1$ and zero if $m \leq 0$. We assume that the N data strings $(\bar{Y}_{M,i}^{(0)}, \bar{Y}_{M,i}^{(1)}, \bar{Y}_{M,i}, \bar{A}_{M,i}, \bar{L}_{M,i}, X_i)$, $i = 1, \dots, N$ are realizations of N independently and identically distributed random vectors and will henceforth suppress their dependence on the subject index i . Here we have adopted the notational convention that for any process Z , $\bar{Z}_m = (Z_1, \dots, Z_m)$ is the history of the process until time m . For notational convenience, we define $\bar{Z}_m \equiv 0$ for $m \leq 0$.

Define

$$\theta_m^{(0)}(x) \equiv E \left[Y_m^{(0)} \mid X = x \right], \quad \theta_m^{(1)}(x) \equiv E \left[Y_m^{(1)} \mid X = x \right].$$

Thus, for example, $\theta_m^{(0)}(x)$ is the prevalence of childhood illness at time m had, contrary to fact, no exposure to maternal stress occurred during the follow-up period. Note that, because at most one of $Y_m^{(0)}$ and $Y_m^{(1)}$ may be observed on the same subject and not all subjects at a given $X = x$ have constant exposure, neither $\theta_m^{(0)}(x)$ nor $\theta_m^{(1)}(x)$ are identifiable without further assumptions. We define the average causal effect of continuous exposure versus no exposure on the marginal expectation of Y at time m for subjects with baseline covariates $X = x$ as

$$\delta_m(x) \equiv \theta_m^{(1)}(x) - \theta_m^{(0)}(x) = E \left[Y_m^{(1)} \mid X = x \right] - E \left[Y_m^{(0)} \mid X = x \right].$$

That is, $\delta_m(x)$ is the marginal probability given $X = x$ of illness at time m had all the children

been continuously exposed to mother’s stress minus the marginal probability of illness at time m had all children never been exposed to mother’s stress since start of follow-up. We call $\delta_m(x)$ the extended causal risk difference at time m for subjects with baseline covariate values x (cf. Robins, 1989).

Although the actual study lasted just 30 days, we might be interested in what the value of $\delta_m(x)$ would be for m larger than 30 days but less than 60 days. Over longer periods, one would need to account for the variation in illness rates with season of the year. In this paper, we are mainly interested in estimating $\theta_m^{(1)}(x)$, $\theta_m^{(0)}(x)$, and the causal parameters $\delta_m(x)$ for m between 20 and 60.

We have made two assumptions that may well not hold in the MSCM data. First, our assumption of independence between subjects may be false. For example, if (i) the data on all 167 subjects was collected over the same 30 day interval, (ii) spells of cold and rainy weather lead to increased rates of childhood illness, and (iii) meteorological data were not recorded for data analysis, there would be a positive correlation between the variables $Y_{m,i}$ and $Y_{m,j}$ on subjects i and j attributable to the omitted weather variables. Second, the assumption of no interaction between units may be false. For example, if a number of the studied children attended the same day care center, then illness could spread from one child to the next. It is well known that the assumption of no interaction between units fails for infectious diseases. Even though possibly false, we have chosen to impose these two assumptions because (i) they were implicit in Zeger and Liang’s (1986) GEE analysis of the MSCM data, because whenever the assumption of no interaction between units is false, the assumption of statistical independence between subjects will also fail (ii) our principle objective is to derive, under their assumptions, the conditions under which a GEE analysis would have a causal interpretation and to propose an alternative analysis strategy when these conditions do not hold, and (iii) in most longitudinal data sets, the two assumptions are warranted. Failure to impose the two assumptions would have greatly complicated the analysis, seriously lengthened the paper, and drawn attention away from our main points.

2.2. Assumptions for Identification of Causal Parameters

We may formalize the notion of no uncontrolled confounders with the assumption that, for each time k , exposure A_k is independent of the vector of future potential outcomes $\left\{ \left(Y_m^{(0)}, Y_m^{(1)} \right); m = k + 1, k + 2, \dots, M \right\}$, given the observed exposure history \bar{A}_{k-1} through time $k - 1$, the observed history of the outcome $\bar{Y}_k = (Y_1, \dots, Y_k)$ through time k , the history of other controlled time-dependent potential confounding factors $\bar{L}_k = (L_1, \dots, L_k)$ through time k , and baseline covariates X . That is,

$$\left\{ \left(Y_m^{(0)}, Y_m^{(1)} \right); m = k + 1, k + 2, \dots, M \right\} \amalg A_k \mid \bar{A}_{k-1}, \bar{Y}_k, \bar{L}_k, X \quad (1)$$

for all $k \geq 1$, where $X \amalg Y \mid Z$ means “ X is conditionally independent of Y given Z .” Assumption (1) says that the random variable denoting actual exposure (i.e., maternal stress) at time k is independent of the future counterfactual random variables conditional on the observed past. Robins (1986, 1987a) has shown that Assumption 1 is sufficient for identification of the parameters $\theta_m^{(1)}(x)$, $\theta_m^{(0)}(x)$, $\delta_m(x)$ from the observed data provided the conditional probabili-

ity of exposure and of non-exposure at each time k given the observed past $(\bar{A}_{k-1}, \bar{Y}_k, \bar{L}_k, X)$ exceeds zero.

Under our counterfactual model, the latent variables

$$\left\{ \left(Y_m^{(0)}, Y_m^{(1)} \right); m = k + 1, k + 2, \dots, M \right\}$$

are not influenced by the exposure A_k actually received at time k (Robins, 1986; Robins, 1992, 1994). In this respect, $Y_m^{(0)}$ and $Y_m^{(1)}$ are like baseline covariates. It follows that if, at each time k , mother’s stress is assigned to children at random by the flip of a coin, then assumption (1) would be true even if the probability that the coin lands heads depends upon past histories $\bar{A}_{k-1}, \bar{Y}_k, \bar{L}_k$ and baseline covariates X . Because such physical randomization guarantees assumption (1), most people accept that valid causal inferences can be obtained from a randomized trial. See Rubin (1978), Holland (1986), Greenland and Robins (1986) and Robins (1986) for further discussion. Because in an observational study, assumption (1) cannot be guaranteed to hold and cannot be empirically tested, drawing causal inferences from observational data is much more controversial.

3. TESTING AND ESTIMATION WHEN THERE ARE NO TIME-VARYING CONFOUNDERS

Suppose that the no-confounding assumption (1) remains true when \bar{L}_k is deleted. That is, suppose

$$\left\{ \left(Y_m^{(0)}, Y_m^{(1)} \right); m = k + 1, k + 2, \dots, M \right\} \amalg A_k \mid \bar{A}_{k-1}, \bar{Y}_k, X \text{ for all } k. \quad (2)$$

Assumption 2 formalizes the idea that there is no confounding for the effect of current exposure on outcome by time-dependent or other uncontrolled covariates.

When the observed data are $(\bar{Y}_M, \bar{A}_M, X)$, the no-confounding assumption (2) allows us to test the following sharp causal null hypothesis that mother’s stress has no effect on child’s illness.

$$Y_m^{(0)} = Y_m^{(1)} = Y_m \text{ for all } m. \quad (3)$$

This sharp causal null hypothesis implies that each subject’s observed and latent outcomes are identical. Given assumption (2), the restriction on the observable random variables $(\bar{Y}_M, \bar{A}_M, X)$ implied by hypothesis (3) is that

$$(Y_{k+1}, \dots, Y_M) \amalg A_k \mid \bar{A}_{k-1}, \bar{Y}_k, X \text{ for all } k, \quad (4)$$

which states that outcomes observed after time k are independent of exposure at k , given exposure up to k , outcomes up through k , and the fixed covariates. It is a purely associational hypothesis; because it involves only observed variables, it is testable without further assumptions. Therefore, given the no-confounding assumption (2), rejection of the associational null hypothesis (4) logically entails rejection of the causal null hypothesis (3). Consequently, an α -level test of the associational null hypothesis (4) will also be an α -level test of the causal null hypothesis (3).

We also have the following result, which is a special case of the G-null theorem of Robins (1986, Sec. 6).

Proposition 1 (Testing): The associational null hypothesis (4) is true if and only if, at all times m ,

$$E [Y_m | \bar{Y}_{m-1}, \bar{A}_{m-1}, X] = E [Y_m | \bar{Y}_{m-1}, X]; \quad (5)$$

that is, hypothesis 4 holds if and only if the conditional expectation of Y_m does not depend on past exposure history given past outcomes and baseline. Proposition 1 implies that, under the no-confounding assumption (2), an α -level test of the regression null hypothesis (5) is an α -level test of the associational null hypothesis (4), and hence is an α -level test of the sharp causal null hypothesis (3).

We may test hypothesis (5) by specifying a parametric statistical model $P_m(\beta)$ for its left-hand side:

$$E [Y_m | \bar{Y}_{m-1}, \bar{A}_{m-1}, X] \equiv \Pr [Y_m = 1 | \bar{Y}_{m-1}, \bar{A}_{m-1}, X] = P_m(\beta) . \quad (6)$$

An example of such a model is, for $m \geq 4$,

$$\begin{aligned} \text{logit} [P_m(\beta)] = & \\ & \beta_0 + \beta_1 A_{m-1} + \beta_2 A_{m-2} + \beta_3 A_{m-3} + \beta_4 Y_{m-1} + \beta_5 Y_{m-2} + \beta_6 \text{Avg} A_{m-4} \\ & + \beta_7 \text{Avg} Y_{m-3} + \beta_8 m + \beta_9 (m A_{m-1}) + \beta_{10} (m Y_{m-1}) + \beta_{11}^T X \end{aligned} \quad (7)$$

where $\beta^T \equiv (\beta_0, \beta_1, \dots, \beta_{11}^T)^T$; $\text{Avg} A_{m-4} \equiv \sum_{k=1}^{m-4} A_k / (m-4)$ if $m \geq 5$, zero otherwise; $\text{Avg} Y_{m-3} \equiv \sum_{k=1}^{m-3} Y_k / (m-3)$ if $m \geq 4$, zero otherwise; and $m A_{m-1}$ and $m Y_{m-1}$ are time \times stress and time \times illness interaction terms, respectively.

We have restricted attention to times $m \geq 4$, because A_{m-3} is unknown when $m < 4$. To aid our exposition, we assume until Section 5 that model (7) is correct for all m . Models such as (7) that condition on the entire past have been studied by Robins (1986) for causal analysis of longitudinal data. Such models are common in the time-series and panel-data literature in econometrics (e.g. Chamberlain, 1984) and have been considered by Zeger and Liang (1991).

Given the no-confounding assumption (2), Proposition 1 implies that an α -level test of the hypothesis $\beta_a \equiv (\beta_1, \beta_2, \beta_3, \beta_6, \beta_9)^T = \mathbf{0}$ in model (7) is an α -level test of the sharp causal null hypothesis (3). Hence, we have obtained an α -level test of the sharp causal null hypothesis (4) by modeling the conditional probability of Y_m given past outcome history \bar{Y}_{m-1} , past exposure history \bar{A}_{m-1} , and baseline covariates X .

We may fit the model $P_m(\beta)$ specified in formula 7 by maximizing the likelihood

$$\prod_{i=1}^N \prod_{m=1}^M [P_{m,i}(\beta)]^{Y_{m,i}} [1 - P_{m,i}(\beta)]^{1 - Y_{m,i}} \quad (8)$$

The resulting maximum likelihood estimator (MLE) is asymptotically normal and unbiased for β . Assuming the model is correct, its asymptotic variance is consistently estimated by the inverse of the observed information matrix evaluated at $\hat{\beta}$. By a standard martingale argument,

the M contributions of a given subject i to the score equations are uncorrelated. Thus, the MLE $\widehat{\beta}$ and the observed information matrix can be obtained using any standard statistical software for logistic regression by regarding the $Y_{m,i}$ $\{m = 1, \dots, M, i = 1, \dots, N\}$ as realizations of $M \times N$ independent Bernoulli variates with probabilities of illness $P_{m,i}(\beta)$. (All of our statements concerning consistency and asymptotic normality of estimators refer to an asymptotic framework in which the number N of mother-child pairs increases to infinity, while the number of times M remains fixed.)

Estimation: Consider $\theta_m^{(1)}(x)$ and $\theta_m^{(0)}(x)$, the counterfactual means under continuous exposure and no exposure, from which we obtain estimates of the causal parameters $\delta_m(x) = \theta_m^{(1)}(x) - \theta_m^{(0)}(x)$. These can be estimated using the G-computation algorithm described by Robins (1986). In particular, we use a fundamental theorem of Robins (1986, Theorem 4.1, p. 1423) specialized to the present context.

Proposition 2 (Estimation): Let $\bar{y}_{m-1} = (y_1, \dots, y_{m-1})$ denote a possible value for \bar{Y}_{m-1} . If the no-confounding assumption (2) holds, then for $j = 0, 1$

$$\Pr \left[Y_m^{(j)} = 1 \mid \bar{Y}_{m-1}^{(j)} = \bar{y}_{m-1}, X \right] = \Pr \left[Y_m = 1 \mid \bar{Y}_{m-1} = \bar{y}_{m-1}, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X \right] \quad (9)$$

for all m , where $\mathbf{j}^{[m-1]}$ is an $m - 1$ -vector of 0's if $j = 0$ and is an $m - 1$ -vector of 1's if $j = 1$.

The left-hand side of Eq. (9) is a conditional probability involving possibly unobserved (counterfactual) variables, whereas the right-hand side depends only on the joint distribution of the observables and so is estimable in the usual sense. Thus, the no-confounding assumption (2) allows estimation of the quantities of interest.

By the law of total probability and the product rule for conditional probability, we have for $j = 0, 1$

$$\begin{aligned} \theta_m^{(j)}(X) &\equiv \Pr \left[Y_m^{(j)} = 1 \mid X \right] = \\ &\sum_{\text{all } \bar{y}_{m-1}} \Pr \left[Y_m^{(j)} = 1 \mid \bar{Y}_{m-1}^{(j)} = \bar{y}_{m-1}, X \right] \prod_{k=1}^{m-1} \Pr \left[Y_k^{(j)} = y_k \mid Y_{k-1}^{(j)} = \bar{y}_{k-1}, X \right] \end{aligned} \quad (10)$$

where the sums are over the 2^{m-1} possible values for \bar{Y}_{m-1} . With the right hand side of (9) substituted for the probabilities in (10), the sum in (10) is a special case of the g-computation algorithm formula of Robins (1986, 1987b). Given the no-confounding assumption 2, we can use Proposition 2 to substitute estimable probabilities for the probabilities involving counterfactuals in Eq. 10. Thus, the no-confounding assumption (2) renders $\theta_m^{(j)}(X)$ identifiable. For example, given model 7, the MLE of $\theta_m^{(j)}(X)$ may be obtained by replacing the probabilities in Eq. 10 by the corresponding MLE fitted values for

$$\Pr \left[Y_k = y_k \mid \bar{Y}_{k-1} = \bar{y}_{k-1}, \bar{A}_{k-1} = \mathbf{j}^{[k-1]}, X \right] \text{ and } \Pr \left[Y_m = 1 \mid \bar{Y}_{m-1} = \bar{y}_{m-1}, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X \right].$$

Since the number of summands in Eq. (10) is 2^{m-1} , the sum cannot be easily evaluated for large m , even with the aid of a computer. A Monte-Carlo approximation to the sum in Eq. (10), the Monte-Carlo G-computation algorithm (Robins, 1989) can be used for large m , say $m > 20$. Alternatively, as described in the Appendix, Hu and Robins (1993) show that for large m , the sum can be approximated by a simple analytic estimator based on a limit theorem for Markov chains under certain Markov and stationarity assumptions. We will use this approximation in our reanalysis of the Mothers' Stress and Children's Morbidity data in Section 5 below.

4. CONDITIONS FOR VALIDITY OF GEE METHODS

Zeger and Liang (1986) effectively assumed that the expectation of Y_m given the entire exposure history \bar{A}_M and baseline covariates depended only on the past exposure history \bar{A}_{m-1} and baseline covariates, i.e.,

$$E [Y_m | \bar{A}_M, X] = E [Y_m | \bar{A}_{m-1}, X] \quad (11)$$

They then modeled the expectation of Y_m given the past exposure history \bar{A}_{m-1} and X as

$$E [Y_m | \bar{A}_{m-1}, X] \equiv \Pr [Y_m = 1 | \bar{A}_{m-1}, X] = P_m(\alpha), \text{ where} \quad (12)$$

$$\text{logit} [P_m(\alpha)] = \alpha_0 + \alpha_1 A_{m-1} + \alpha_2 A_{m-2} + \alpha_3 A_{m-3} + \alpha_4^T X,$$

$\alpha \equiv (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4^T)^T$ and X is a vector of covariates recording household size, child's race, employment, and marital status. Zeger and Liang (1986) and Liang and Zeger (1986) studied the GEE estimator $\hat{\alpha}_{GEE}(V)$, defined as the solution of the estimating equations

$$\mathbf{U}_{GEE}(V, \alpha) \equiv \sum_{i=1}^N \{ \partial \mathbf{P}_i(\alpha) / \partial \alpha \} [V_i(\alpha)]^{-1} \boldsymbol{\varepsilon}_i(\alpha) = \mathbf{0},$$

where

$$\mathbf{P}(\alpha) \equiv (P_1(\alpha), \dots, P_M(\alpha))^T, \boldsymbol{\varepsilon}(\alpha) \equiv (\boldsymbol{\varepsilon}_1(\alpha), \dots, \boldsymbol{\varepsilon}_M(\alpha))^T, \boldsymbol{\varepsilon}_m(\alpha) \equiv Y_m - P_m(\alpha)$$

is the residual from model (12), and $V(\alpha)$ is a working covariance matrix (selected by the investigator) with the diagonal element in row m given by $P_m(\alpha) [1 - P_m(\alpha)]$. In particular, they showed that $\hat{\alpha}_{GEE}(V)$ is consistent and asymptotically normal for α under Eq. (11) and (12). In addition, they proposed a consistent robust asymptotic variance estimator for $\hat{\alpha}_{GEE}(V)$ that properly accounted for the correlations between Y_k and Y_m , $k \neq m$. Throughout, we will assume that the associational model (12) is correct for the marginal expectation $E [Y_m | \bar{A}_{m-1}, X]$.

Consider now the associational null hypothesis that past exposures do not predict the expected outcome, once fixed covariates are included:

$$E [Y_m | \bar{A}_{m-1}, X] = E [Y_m | X] \text{ for all } m. \quad (13)$$

This hypothesis is identifiable without further assumptions, but unlike the earlier null hypothesis (5), it omits \bar{Y}_{m-1} and hence is neither equivalent to the earlier associational hypothesis (4) nor implied by the no-confounding assumption (2) and the causal null hypothesis (3). Thus, it cannot be used to test the causal null hypothesis (3) without assumptions beyond assumption 2. This Hypothesis 13 is equivalent to the hypothesis that $\alpha_1 = \alpha_2 = \alpha_3 = 0$ under the GEE model 12. Hence, in general, the causal null hypothesis (3) cannot be validly tested using GEEs.

In Propositions 3 and 4 below, we derive some simple identifiable conditions under which the earlier associational null hypothesis (5) implies the null hypothesis 13. Under these conditions, rejection of equation 13 logically entails rejection of equation 5, which in turn entails rejection of the causal null hypothesis 3 under the no-confounding assumption 2. Propositions 3 and 4 are special cases of Theorems F1-F4 of Robins (1986).

Proposition 3 (Testing): Suppose that, for all m , either

$$E [Y_m | \bar{Y}_{m-1}, \bar{A}_{m-1}, X] = E [Y_m | \bar{A}_{m-1}, X] \quad (14)$$

(the expectation of Y_m does not depend on past outcomes, once past exposures and baseline covariates are known), or

$$E [A_m | \bar{Y}_m, \bar{A}_{m-1}, X] = E [A_m | \bar{A}_{m-1}, X] \quad (15)$$

(the expectation of current exposure A_m does not depend on past or current outcome once past exposures and baseline covariates are known). Then the associational null hypothesis (5) implies the associational null hypothesis (13).

Proposition 3 implies that if model 12, the no-confounding assumption 2, and either Eq. 14 or 15 are true, then an α -level test of the hypothesis that $(\alpha_1, \alpha_2, \alpha_3) = \mathbf{0}$ in the GEE model (12) is an α -level test of the causal null hypothesis (3).

If either Eqs. (14) or (15) is true, we say that child's past illness history \bar{Y}_m is not a confounder for the causal effect of mother's stress on the marginal probability of child's illness. Following Kalbfleisch and Prentice (1980, p. 123), when equation 15 is true we say that the time-dependent dichotomous exposure A_m is an external covariate (of the ancillary type) and there exists no feedback from output to input.

Proposition 4 (Estimation): Suppose there is no confounding by time-dependent or uncontrolled covariates (assumption 2 is true), and no confounding by illness history (either equation 14 or 15 is true). Then

$$\theta_m^{(j)}(x) \equiv E [Y_m^{(j)} | X = x] = E [Y_m | \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X = x].$$

If, in addition, the GEE logistic model 12 is correct,

$$\text{logit} [\theta_m^{(1)}(x)] - \text{logit} [\theta_m^{(0)}(x)] = \alpha_1 + \alpha_2 + \alpha_3$$

and the causal parameter $\delta_m(x) \equiv \theta_m^{(1)}(x) - \theta_m^{(0)}(x)$ equals

$$\text{expit} (\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4^T x) - \text{expit} (\alpha_0 + \alpha_4^T x)$$

where $\text{expit} (z) = e^z / (1 + e^z)$ is the logistic transform. Thus, given model 12, assumption 2, and either equation 14 or 15, we may consistently estimate causal effects by consistently estimating the parameter α of model 12. The following Propositions 5 and 6 provide conditions under which the GEE estimator of α will be consistent and are proved in the Appendix.

Proposition 5: If the exposure A_m is an external covariate (i.e., equation 15 holds), then equation 11 holds.

Proposition 5 states that, when equation 15 holds, assumption (11) of Zeger and Liang is fulfilled. Thus, given equation 15 and model 12, the parameters α in model 12 can be consistently estimated by the GEE estimator $\hat{\alpha}_{GEE}(V)$. With the no-confounding assumption 2 added, $\alpha_1 + \alpha_2 + \alpha_3$ has the causal interpretation given above.

When equation 11 does not hold, the results in Zeger and Liang (1986) do not apply, and

we will need the following result.

Proposition 6: Suppose model 12 is correct but equation 11 is not. Then the GEE estimator of α will be consistent if one chooses the independence working covariance matrix (a diagonal matrix). If, in addition, equation 14 holds, this independence GEE estimator is the maximum likelihood estimator (MLE) of α .

Proposition 6 is a statement about the joint distribution of the observables and makes no reference to the distribution of the counterfactuals $Y_m^{(0)}$ and $Y_m^{(1)}$. It is needed because, if equation 11 is false, a GEE estimator of α based on a dependent structure (nondiagonal working covariance matrix) may be inconsistent. In Section 5, we will show that equation 11 appears to be contradicted by the Mothers’ Stress and Children’s Morbidity data. Hence, even if model 12 correctly describes marginal associations in that study, estimators based on the working covariance matrices other than the “independence” matrix appear to be inappropriate for inferences on α . When Eqs. (11) and (14) are false, model 12 is not a “feedback” model in the sense of Zeger and Liang (1991), since model 12, in contrast to their feedback models, does not condition on past outcome history \bar{Y}_{m-1} .

Results in this subsection and the following one that are wholly concerned with the observable random variables are closely related to the results that have been independently derived in the econometric literature (see Chamberlain, 1984). This literature did not, however, explicitly discuss counterfactual outcomes.

To summarize, Propositions 4-6 show that, given assumption (2) (no time-dependent or unmeasured confounders), if Eqs. (14) or (15) is true, the parameters $(\alpha_1, \alpha_2, \alpha_3)$ of model 12 have a causal interpretation and can be consistently estimated from the observables.

Note that since models (6) and (12) are statistical models for different conditional distributions, they both could be correct. If neither (14) nor (15) is true, the standard GEE logistic model (12) may be a correct associational model for $E[Y_m | \bar{A}_{m-1}, X]$, yet have no causal interpretation. For example, if neither (14) nor (15) is true, even if the no-confounding assumption 2 holds, an α -level test of the hypothesis $(\alpha_1, \alpha_2, \alpha_3) = 0$ in model 12 is not a valid α -level test of the causal null hypothesis 3, and $\text{logit} [\theta_m^{(1)}(x)] - \text{logit} [\theta_m^{(0)}(x)] \neq \alpha_1 + \alpha_2 + \alpha_3$.

5. REANALYSIS OF MOTHERS’ STRESS AND CHILDREN’S MORBIDITY DATA

To illustrate the methods discussed in previous sections, we reanalyze the data from the Mothers’ Stress and Children’s Morbidity Study (Zeger and Liang, 1986, pp. 125-128). This study examined the relation between maternal stress and childhood illness. Observations based on mothers’ self-reports were made daily for $M = 30$ days on $N = 167$ mother-child pairs. Table 1 provides a summary of the observations used here. In this Section, we shall suppose that there is no confounding by time-dependent or uncontrolled covariates (Eq. 2).

Analysis 1: Testing Eq. 11: We begin our reanalysis by testing whether GEE estimates of the parameters of the model (12) used by Zeger and Liang (1986) are consistent if a non-diagonal working covariance matrix is specified. To do so, we tested whether $E[Y_m | \bar{A}_M, X] = E[Y_m | \bar{A}_{m-1}, X]$ (Eq. 11) is true by adding the covariate $\alpha_5 A_m$ to model 12. If equation 11 is correct, $\alpha_5 = 0$. Table 2 summarizes the results of fitting the model using an independence working covariance matrix. This analysis yielded $\hat{\alpha}_5 = 0.643$ ($p < 0.0001$), so Eq. 11 is un-

tenable. It follows from Proposition 6 that we should use only a diagonal working covariance matrix in GEE estimation of the parameters of 12.

Analysis 2: Testing Eqs. 14 and 15: Our next goal is to determine whether the parameters $(\alpha_1, \alpha_2, \alpha_3)$ of model (12) can be interpreted causally.. Given assumption 2, if either Eq. 14 or 15 is correct, then, by Propositions 4 the parameters $(\alpha_1, \alpha_2, \alpha_3)$ of model 12 would have a causal interpretation.

According to Proposition 5, Eq. 15 (past or current disease does not predict current exposure given past exposure and covariates) implies Eq. 11, so our test of Eq. 11 is also a test of Eq. 15. The former test yielded $p < 0.0001$, so Eq. 15 also appears untenable. We can also directly test Eq. 15 by assuming the following logistic regression model

$$\begin{aligned} \text{logit} \{ \Pr [A_m = 1 \mid \bar{A}_{m-1}, X] \} = \\ \gamma_0 + \gamma_1 A_{m-1} + \gamma_2 A_{m-2} + \gamma_3 A_{m-3} + \gamma_4 \text{Avg} A_{m-4} + \gamma_5 X + \gamma_6 Y_m . \end{aligned} \quad (16)$$

Eq. 15 and model 16 imply $\gamma_6 = 0$. Table 3 summarizes the results of fitting model (16); we obtained $\hat{\gamma}_6 = 0.554$ ($p = 0.0002$) and thus Eq. 15 appears untenable given model 16. Note that if Eq. 15 is false, the time-dependent exposure A_m is not an external covariate process and there is “feedback” from the output Y_m to the input A_m .

To test 14, we suppose that following model is correct for $m > 4$.

$$\begin{aligned} \text{logit} \{ \Pr [Y_m = 1 \mid \bar{Y}_{m-1}, \bar{A}_{m-1}, X] \} = \\ \beta_0 + \beta_1 A_{m-1} + \beta_2 A_{m-2} + \beta_3 A_{m-3} + \beta_4 Y_{m-1} + \\ \beta_5 Y_{m-2} + \beta_6 \text{Avg} A_{m-4} + \beta_7 \text{Avg} Y_{m-3} + \beta_8^T X \end{aligned} \quad (17)$$

where $\text{Avg} A_{m-4} \equiv \sum_{k=1}^{m-4} A_k / (m-4)$, $\text{Avg} Y_{m-3} \equiv \sum_{k=1}^{m-3} Y_k / (m-3)$, and X contains household size, child’s race, employment, and marital status. (Zeger and Liang (1986) also included the covariate “mother’s general stress” in X , but this covariate was not in the data made available to us.) Model 17 implies that the regression of current child’s illness on mother’s stress more than 3 days earlier and child’s illness more than 2 days earlier can be summarized by average past illnesses and average stress history respectively. It also assumes that there is no dependence on time and no interactions (product terms) among the covariates, or with time; that is, there are no covariates such as $m, mA_{m-1}, mY_{m-1}, A_{m-j}A_{m-k}, A_{m-j}Y_{m-k}$, or $A_{m-j}X$ in the model. In some analyses not shown here, we searched for such interactions, but none of them had p values less than .10.

Given that model 17 is correct, Eq. 14 (past disease does not predict current disease given past exposure and covariates) implies that $\beta_4 = 0$ in the following logistic regression submodel of model 17

$$\begin{aligned} \text{logit} \{ \Pr [Y_m = 1 \mid Y_{m-1}, \bar{A}_{m-1,i}, X] \} = \\ \beta_0 + \beta_1 A_{m-1} + \beta_2 A_{m-2,i} + \beta_3 A_{m-3,i} + \beta_4 Y_{m-1} + \beta_5 \text{Avg} A_{m-4,i} + \beta_6^T X \end{aligned} \quad (18)$$

Table 4 summarizes the results of fitting model 18; for model 18 we obtained $\hat{\beta}_4 = 2.58$ ($p < 0.0001$), and thus Eq. 14 appears untenable, given model 18.

To summarize, Eqs. 14 and 15 appear to be untenable for the Mothers’ Stress and Children’s

Morbidity study. We thus infer that child’s past illness history is a time-dependent confounder for the causal effect of time-varying mother’s stress on the marginal probability of child’s illness. Thus, even if we assume that model 12 and the no-confounding assumption 2 are correct, we would infer that an α -level test of $(\alpha_1, \alpha_2, \alpha_3) = 0$ in model 12 is not a valid α -level test of the causal null hypothesis (3) and that $\text{logit} \left[\theta_m^{(1)}(x) \right] - \text{logit} \left[\theta_m^{(0)}(x) \right] \neq \alpha_1 + \alpha_2 + \alpha_3$. Consequently, we cannot use the GEE modelling methods of Zeger and Liang to obtain valid causal inferences from these data.

Analysis 3: Testing the causal null hypothesis: Given the no-confounding assumption (2) and model 17, Proposition 1 implies that a valid α -level test of the causal null hypothesis (3) may be obtained by testing $(\beta_1, \beta_2, \beta_3, \beta_6) = \mathbf{0}$ in model 17. A four degree of freedom likelihood-ratio test of this joint null hypothesis yields $p < 0.01$, so the causal null hypothesis (3) appears untenable.

Analysis 4: Estimating the parameters $\theta_m^{(1)}(x)$ and $\theta_m^{(0)}(x)$: Using backward elimination with an α -level of 0.15, we discarded coefficients β_6 and β_7 in model 17. Thus, subsequent analyses are based on the following logistic regression model for $m \geq 4$,

$$\Pr \left[Y_m = 1 \mid \bar{Y}_{m-1}, \bar{A}_{m-1,i}, X \right] = \text{expit} \left[\beta_0 + \beta_1 A_{m-1} + \beta_2 A_{m-2} + \beta_3 A_{m-3} + \beta_4 Y_{m-1} + \beta_5 Y_{m-2} + \beta_6^T X \right] \quad (19)$$

Because model 19 satisfies the Markov and stationarity assumptions defined in the Appendix, we used the Markov-chain approach developed in Hu and Robins (1993) to obtain large- m analytic approximations to the MLEs $\hat{\theta}_m^{(1)}(x)$ and $\hat{\theta}_m^{(0)}(x)$ based on model 19. To illustrate, suppose $X = (1, 0, 0, 0)$, which corresponds to a white, married, unemployed mother with a household size of 2-3. For $20 < m < 60$, we obtained $\hat{\theta}_m^{(1)}(x) = 0.224$, $\hat{\theta}_m^{(0)}(x) = 0.122$ and thus $\hat{\delta}_m(x) = \hat{\theta}_m^{(1)}(x) - \hat{\theta}_m^{(0)}(x) = 0.102$ ($p < .01$). Computational details are given in the Appendix.

For comparative purposes, we also estimated the parameter using model 12 in a GEE approach with an independence (diagonal) working covariance matrix. Proposition 6 implies that if model 12 is correct, the resulting estimator is consistent for α even if it has no causal interpretation. The (biased) GEE estimator of $\delta_m(x)$ was

$$\tilde{\delta}_m(x) \equiv \text{expit} \left(\hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 + \hat{\alpha}_4^T X \right) - \text{expit} \left(\hat{\alpha}_1 + \hat{\alpha}_4^T X \right) \approx .133,$$

which was significant at the $p < .01$ level. Because Eqs. 14 and 15 are untenable, we may view the difference of 0.031 between $\tilde{\delta}_m(x) = 0.133$ and our consistent estimator $\hat{\delta}_m(x) = 0.102$ as a measure of the magnitude of bias (confounding) attributable to the failure of $\tilde{\delta}_m(x)$ to adjust for the fact that a child’s past illness history predicted both the child’s future illness and mother’s subsequent stress, conditional on past stress and on fixed covariates. The apparent bias of 0.031 in $\tilde{\delta}_m(x)$ is in the direction that we would expect, given that a child’s past illness should increase the mother’s stress and also be positively associated with subsequent illness.

6.1. Theoretical results

Estimation of the causal parameter $\theta_m(x)$ in Section 5 relied on assumption 2, which says roughly that all confounding is controllable given only the data $(\bar{Y}_{M,i}, \bar{A}_{M,i}, X_i)$, $i = 1, \dots, N$. The assumption implies that, given illness history through any day k , stress history through $k - 1$, and the baseline covariates X , those mothers stressed on day k have the same distribution of subsequent potential (counterfactual) children's outcomes as mothers without stress on day k . Assumption 2 would be false if, among children ill on day k ($Y_k = 1$) whose mothers had been unstressed through day $k - 1$ ($\bar{A}_{k-1} = 0$), those children with more severe illness were both (a) more likely to remain ill on day $k + 1$ ($Y_{k+1} = 1$) even if mother's stress on day k was eliminated (i.e., even if A_k is forced to be zero, say by an effective intervention program), and (b) more likely to have their mothers become stressed on day k (i.e., more likely to have $A_k = 1$). In this case, $\theta_m(x)$ would not be identifiable in the absence of additional data on the severity of child's illness over time.

We now derive results using the less restrictive no-confounding assumption 1, rather than the more restrictive assumption 2. We henceforth assume the available data on each study subject are $(\bar{Y}_M, \bar{A}_M, \bar{L}_M, X)$, where L_m represents one or more time-dependent potential confounding variables. In the next subsection, we will use the generalized results to reanalyze the Mothers' Stress and Children's Morbidity data in the special case in which L_m is a measure of illness severity.

Given only the less restrictive no-confounding assumption 1, the restriction on the observable random variables $(\bar{Y}_m, \bar{A}_m, \bar{L}_m, X)$ implied by the causal null hypothesis 3 is

$$\{Y_m; m = k + 1, k + 2, \dots, M\} \perp\!\!\!\perp A_k \mid \bar{A}_{k-1}, \bar{Y}_k, \bar{L}_k, X \text{ for all } k \geq 1. \quad (20)$$

Because it conditions on the \bar{L}_k , this associational null hypothesis does not imply the associational null hypothesis (4) or the regression null hypothesis (5); thus, an α -level test of the regression null hypothesis need not be an α -level test of the causal null hypothesis (3). Similarly, hypothesis 20 does not imply that $E[Y_m \mid \bar{A}_{m-1}, X] = E[Y_m \mid X]$, even if both Eqs. 14 and 15 hold; thus, an α -level test of the hypothesis $(\alpha_1, \alpha_2, \alpha_3) = \mathbf{0}$ in model 12 will not be an α -level test of the causal null hypothesis (3) even if Eqs. 14 and 15 hold. Finally, hypothesis 20 does not even imply the following natural extension of the regression hypothesis 5 to time-dependent covariates,

$$E[Y_m \mid \bar{Y}_{m-1}, \bar{A}_{m-1}, \bar{L}_{m-1}, X] = E[Y_m \mid \bar{Y}_{m-1}, \bar{L}_{m-1}, X] . \quad (21)$$

Thus an α -level test of hypothesis 21 need not be an α -level test of the causal null hypothesis (3).

It follows that the feedback model techniques of Zeger and Liang (1991) cannot in general be used to test the causal null hypothesis (3). An α -level test of the hypothesis that the outcome Y_m is independent of the preceding exposure A_{m-1} given the entire past $(\bar{A}_{m-2}, \bar{L}_{m-1}, \bar{Y}_{m-1}, X)$ is an α -level test of (3). This test may have very poor power, however. For example, if exposure takes two time periods (days) to affect the outcome Y , then the power of the test will equal its

α -level.

Nonetheless, under the additional conditions described in the following proposition, our previous results remain applicable.

Proposition 7: If, for all m , either

$$E [Y_m | \bar{Y}_{m-1}, \bar{A}_{m-1}, \bar{L}_{m-1}, X] = E [Y_m | \bar{Y}_{m-1}, \bar{A}_{m-1}, X] \quad (22)$$

or

$$\Pr [A_m = 1 | \bar{Y}_m, \bar{A}_{m-1}, \bar{L}_m, X] = \Pr [A_m = 1 | \bar{Y}_m, \bar{A}_{m-1}, X] , \quad (23)$$

then (i) the less restrictive no-confounding assumption (1) implies the more restrictive no-confounding assumption (2) and (ii) the associational null hypothesis (20) implies the associational null hypothesis (4) and the regression-null hypothesis (5).

Proposition 7 is a special case of Theorems F1-F4 in Robins (1986). Given the less restrictive no-confounding assumption 1 and either Eq. 22 or 23, any α -level test of the regression null hypothesis 5 is also an α -level test of the causal null hypothesis 3, although an α -level test of the regression hypothesis 21 is still not an α -level test of hypothesis 3. Thus Proposition 7 allows us to apply all our earlier results to the present situation (in which assumption 1 holds), provided Eq. 22 or 23 holds. In particular, given assumption 1, Eq. 22 or 23, and model 17, we have (i) an α -level test of $(\beta_1, \beta_2, \beta_3, \beta_6) = \mathbf{0}$ in model (17) is an α -level test of the causal null hypothesis (3), (ii) the estimates $\hat{\theta}_m^{(j)}$ and $\hat{\delta}_m$ of Section 3 remain consistent for $\theta_m^{(j)}$ and δ_m , and (iii) given Eqs. (14) or (15), an α -level test of Eq. (3), for example, testing whether $(\alpha_1, \alpha_2, \alpha_3) = \mathbf{0}$ in model (12) would be an α -level test of the causal null hypothesis (3). Thus, if either Eq. 22 or 23 is correct, we say covariate history \bar{L}_M is a non-confounder.

Unfortunately, the conditions (a) and (b) discussed in the beginning of this section imply violations of Eqs. 22 and 23, respectively. We thus need to consider the estimation of the causal parameter $\delta_m(x)$ under the less restrictive no-confounding assumption (1) when Eqs. 22 and 23 are both false. For notational convenience, we will assume L_m is discrete.

The following is a special case of the corollary to Theorem 1 in Robins (1987b).

Proposition 8 (Estimation): Given the no confounding assumption (1), for all j and m ,

$$\begin{aligned} \theta_m^{(j)}(x) &\equiv \Pr [Y_m^{(j)} = 1 | X = x] = \\ &\sum_{\text{all } \bar{y}_{m-1}, \bar{\ell}_{m-1}} \Pr [Y_m = 1 | \bar{Y}_{m-1} = \bar{y}_{m-1}, \bar{L}_{m-1} = \bar{\ell}_{m-1}, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X = x] \times \\ &\quad \prod_{k=1}^{m-1} \Pr [Y_k = y_k | \bar{Y}_{k-1} = \bar{y}_{k-1}, \bar{L}_{k-1} = \bar{\ell}_{k-1}, \bar{A}_{k-1} = \mathbf{j}^{[k-1]}, X = x] \times \\ &\quad \Pr [L_k = \ell_k | \bar{Y}_k = \bar{y}_k, \bar{L}_{k-1} = \bar{\ell}_{k-1}, \bar{A}_{k-1} = \mathbf{j}^{[k-1]}, X = x] \end{aligned} \quad (24)$$

where, given a particular $\bar{\ell}_{m-1}$, both ℓ_k and $\bar{\ell}_{k-1}$ are determined by the initial segment $\bar{\ell}_k$ of $\bar{\ell}_{m-1}$ through k and the sum is over all possible realizations \bar{y}_{m-1} and $\bar{\ell}_{m-1}$ of the histories \bar{Y}_{m-1} and \bar{L}_{m-1} .

Proposition 8 implies that, given the less restrictive no-confounding assumption (1), we can estimate $\theta_m^{(j)}(x)$ by estimating the unknown conditional probabilities on the right-hand side of Eq. (24). We can also test the causal null hypothesis (3) by testing whether $\delta_m(x) = 0$ for all m based on our estimates of the $\theta_m^{(j)}(x)$ parameters. However, (3) cannot be tested by ordinary

regression techniques, regardless of whether we adjust for the confounder history \bar{L}_m in the analysis.

The sum in (24) is a special case of the g-computation algorithm formula. As described in Robins (1986, 1987b), a Monte Carlo approximation to the sum in (24) is required for large m . However, in the Appendix, we show that this sum can be approximated by a simple analytic estimator under certain Markov and stationarity assumptions. We will apply this approximation below.

6.2. Further Data Analyses:

To illustrate the methods discussed in Section 6.1, we reanalyzed the Mothers' Stress and Children's Morbidity data under the less restrictive assumption (1) applied to a dichotomous L_m which took the value "1" if child i was seriously ill at time m and was "0" otherwise. We note that (Y_m, L_m) had a degenerate joint distribution, in that (i) $Y_m = 0$ implied $L_m = 0$ (a child who was not ill could not be seriously ill) and (ii) $L_m = 1$ implied $Y_m = 1$ (a severely ill child was ill). Fortunately, the theory and methods developed in Section 6.1 are applicable to any confounding factors L_m regardless of whether or not Y_m and L_m have a degenerate joint distribution. The goal of the analysis remains the estimation of the causal parameter $\delta_m(x)$. That is, we are still interested in estimating a causal parameter representing the effect of mother's stress on the marginal probability of child's illness (regardless of the severity of the illness).

Analogous to model (17) in Section 3, we assumed the following two logistic regression models were correct for $m > 4$

$$\Pr [Y_m = 1 \mid \bar{Y}_{m-1}, \bar{L}_{m-1}, \bar{A}_{m-1}, X] = P_m(\beta_1), \quad (25)$$

$$\Pr [L_m = 1 \mid \bar{Y}_m, \bar{L}_{m-1}, \bar{A}_{m-1}, X] = Y_m P_m(\beta_2) \quad (26)$$

where, for $d = 1, 2$,

$$P_m(\beta_d) = \text{expit} \left[\begin{array}{c} \beta_{0d} + \beta_{1d}A_{m-1} + \beta_{2d}A_{m-2,i} + \beta_{3d}A_{m-3} + \\ \beta_{4d}Y_{m-1} + \beta_{5d}Y_{m-2} + \beta_{6d}L_{m-1} + \beta_{7d}L_{m-2} + \\ \beta_{8d}AvgA_{m-4} + \beta_{9d}AvgY_{m-3} + \beta_{10d}AvgL_{m-3} + \beta_{11d}^T X \end{array} \right]$$

$$AvgA_{m-4} \equiv \sum_{k=1}^{m-4} A_k / (m-4), \quad AvgY_{m-3} \equiv \sum_{k=1}^{m-3} Y_k / (m-3), \quad AvgL_{m-3} \equiv \sum_{k=1}^{m-3} L_k / (m-3),$$

and X contains household size, child's race, employment, and marital status. Because (Y_m, L_m) had a degenerate joint distribution, we only needed to fit the $P_m(\beta_2)$ part of model (26) to the group of children who were ill at time m , i.e. $Y_m = 1$. This is why we pre-multiplied $P_m(\beta_2)$ by the indicator function Y_m in model (26).

It is important to emphasize that the parameters in models (25) - (26) do not have a direct causal interpretation as the causal effect of the exposure history \bar{A}_{m-1} on Y_m under our assumptions if Eqs. (22) and (23) are false. For example, the coefficient β_{21} of A_{m-2} in model (25) fails

to have a causal interpretation because, in model (25), we are adjusting for L_{m-1} , which itself can be affected by the earlier exposure A_{m-2} (Rosenbaum, 1984; Robins, 1987b). As described previously, the coefficient β_{1d} of A_{m-1} in (25) has an interpretation as the direct causal effect of A_{m-1} on Y_m ; however, β_{1d} may be zero, even if stress history \bar{A}_{m-1} affects Y_m . In general, it is only the entire g-computation algorithm formula (24) which can be interpreted as the causal effect of the exposure history \bar{A}_{m-1} on the outcome Y_m . The individual components of the sum in (24) have no causal interpretation in themselves.

Analysis 5: Testing Eqs. (22) and (23): If model (25) is correct, Eq. (22) implies that parameters $\beta_{6,1}$, $\beta_{7,1}$, and $\beta_{10,1}$ are 0. A three degree of freedom likelihood ratio χ^2 test of the joint hypothesis that $(\beta_{6,1}, \beta_{7,1}, \beta_{10,1}) = \mathbf{0}$ yielded $p = .65$. Note that this does not imply Eq. 22 is correct; it means only that we have no strong evidence that it is false.

To test Eq. (23), we fit the following logistic regression model:

$$\Pr [A_m = 1 \mid \bar{Y}_m, \bar{A}_{m-1}, \bar{L}_m, X] = \text{expit} \left[\begin{array}{c} \beta_0 + \beta_1 A_{m-1} + \beta_2 A_{m-2} + \beta_3 A_{m-3} + \\ \beta_4 Y_m + \beta_5 Y_{m-1} + \beta_6 Y_{m-2} + \beta_7 \text{Avg} A_{m-4} + \beta_8 \text{Avg} Y_{m-3} + \\ \beta_9^T X + \beta_{10} L_m \end{array} \right]$$

We obtained $\hat{\beta}_{10} = -0.1584$ with $p = 0.57$ and thus could not reject Eq. (23). Note that this does not mean that Eq. 23 is correct; it means only that we have no strong evidence that it is false.

If either of Eqs. 22 or 23 holds, by Proposition 7, the analyses described in Section 3 retain the causal interpretation that we previously ascribed to them. However, because our tests do not enable us to conclude with certainty that L_m was a non-confounder, it is of interest to determine how close the estimates of $\theta_m^{(j)}$ based on Eq. (24) (which adjust for L_m) are to the estimates obtained in Section 3 (which ignore L_m).

Analysis: Estimating the parameters $\theta_m^{(j)}(x)$ from Eq. (24): Using backward elimination with $\alpha = 0.15$, we deleted β_{8d} , β_{9d} , and β_{10d} in models (25) and (26) for $d = 1, 2$, and our subsequent analyses were based on the reduced models. These models satisfied the Markov and stationarity assumptions defined in the Appendix. Hence, we used the Markov chain approach of Hu and Robins (1993) to approximate the MLEs $\hat{\theta}_m^{(j)}$ under these models and Eq. (24). For $x = (1, 0, 0, 0)$ and $20 < m < 60$, we obtained $\hat{\theta}_m^{(1)}(x) = 0.224$, $\hat{\theta}_m^{(0)}(x) = 0.122$, and $\hat{\delta}_m(x) = 0.102$. Thus, $\hat{\theta}_m^{(j)}(x)$ and $\hat{\theta}_m(x)$ obtained controlling for L_m are to three-digit accuracy the same as those obtained in the analyses of Section 5.

7. CONDITIONAL METHODS

A referee argued that our focus on estimating the effect of maternal stress on the (marginal) prevalence of childhood illness within the levels of the baseline covariates X was misguided. The referee suggested that the causal parameter of interest should be a comparison of the illness probabilities of each child when the mother was and was not treated, a conditional mother-child-pair-specific causal parameter. The referee further added that not only was the within-pair causal parameter of greater interest but, in addition, estimation methods that use only within mother-child pair comparisons, such as conditional likelihood methods, automatically control

for between-mother and between-child differences.

We disagree with the referee on both points. Some of the following arguments are similar to those given in Sec. 6 of Neuhaus, Kalbfleisch, and Hauck (1991). We will first show that conditional methods cannot control for between-pair differences without making additional strong assumptions that are not fully testable. To focus on the fundamental issue, suppose X , L_m , A_m , and Y_m are discrete random variables with finite support. Then, under our original no confounding assumption (1), we obtain consistent non-parametric estimates of the marginal effects $\theta_m^{(j)}(x)$ at time m by replacing the unknown probabilities on the RHS of Eq. (24) by their empirical estimates. The Markov and stationarity assumptions encoded in our models make the evaluation of the sum on the RHS of (24) easy by allowing us to exploit limit theorems for Markov chains (as discussed in the Appendix); nonetheless, these Markov and stationary assumptions are not essential and can be relaxed, provided the sample size N is large relative to m .

Now suppose, following the argument of the referee, the fundamental assumption (1) was false, but becomes true if we condition on a further unobserved maternal-child-specific covariate. That is, in lieu of (1), we assume that (1) is true when modified to add a mother-child pair specific covariate effect, which we treat as a random effect σ , to the conditioning event. Then the conditional effect of maternal stress $\theta_m^{(j)}(x, \sigma) = pr \left[Y_m^{(j)} = 1 \mid X = x, \sigma \right]$ is given by the g-computation algorithm formula (24) modified by adding σ to each of the conditioning events.

Let $F(\sigma \mid x)$ be the unknown CDF of the random effects σ given $X = x$. Consider the question of whether we can construct consistent non-parametric estimators of the marginal causal effect $\theta_m^{(j)}(x) = \int \theta_m^{(j)}(x, \sigma) dF(\sigma \mid x) = E \left[\theta_m^{(j)}(x, \sigma) \mid X = x \right]$, the conditional effects $\theta_m^{(j)}(x, \sigma)$, or the conditional distribution of $\theta_m^{(j)}(x, \sigma)$ given $X = x$, i.e., $F_m^{(j)}(t \mid X) = pr \left[\theta_m^{(j)}(x, \sigma) < t \mid X = x \right]$. Because σ is unobserved and $F(\sigma \mid x)$ is unspecified, none of $F(\sigma \mid x)$, $\theta_m^{(j)}(x, \sigma)$, and $F_m^{(j)}(t \mid x)$ is even identified, much less consistently estimable, from the joint distribution of the observed data $\{\bar{Y}_{M,i}, \bar{L}_{M,i}, \bar{A}_{M,i}; i = 1, \dots, N\}$. Furthermore, the conditional effects $\theta_m^{(j)}(x, \sigma)$ themselves cannot be consistently estimated. We conclude that, when our goal is to control for between pair differences by only assuming that the modified assumption (1) holds, no method, including conditional likelihood methods, can appropriately adjust for these pair-specific differences without additional strong assumptions that cannot be fully checked from the data. As a consequence, if we go ahead and choose to impose such assumptions, our inferences concerning $\theta_m^{(j)}(x)$, $F_m^{(j)}(t \mid x)$ and $\theta_m^{(j)}(x, \sigma)$ will depend on the particular assumption imposed. Even if the original assumption (1) holds in addition to the modified assumption (1), it is still the case that neither the distribution $F_m^{(j)}(t \mid x)$ nor the conditional effects $\theta_m^{(j)}(x, \sigma)$ themselves are consistently estimable without additional assumptions that cannot be fully checked.

As examples of such assumptions, suppose we model the unknown conditional probabilities in the modified formula (24) by models (25) and (26), themselves modified only in that on the right-hand side of $P_m(\beta_d)$, we replace the fixed intercept term β_{0d} with the random intercept term σ_d . Here we view the random effect $\sigma = (\sigma_1, \sigma_2)$ as a bivariate random effect, so that the effect σ_1 of a particular mother-child pair in predicting Y_m can differ from its effect σ_2 in predicting L_m . The meaning of the remaining parameters β_d in the modified logistic models (25) - (26) differs from their meaning in the original fixed effects models, because logistic regression is neither linear nor loglinear. Nonetheless, as discussed in Sec. 6.2, the individual components

of β_1 still do not have causal interpretations as the within-pair effect of maternal stress history \bar{A}_{m-1} on childhood illness Y_m . Rather, it follows from the corollary to Theorem 1 in Robins (1987b) that only the modified g-computation algorithm integral (24), i.e., $\theta_m^{(j)}(x, \sigma)$, can have such an interpretation.

Suppose the modified conditional models (25) and (26) were correct, and we could devise a method that yields consistent estimates of the parameters $\beta_d, d = 1, 2$ of these models. Then, for each value of the random effect σ , we could obtain a consistent estimate $\hat{\theta}_m^{(j)}(x, \sigma)$ of the conditional effects by substituting our estimates based on the modified models (25) and (26) for the unknown conditional probabilities on the RHS of the modified formula (24). Suppose in addition, that either (i) we could devise a method to consistently estimate $F(\sigma | x)$ or (ii) the unmodified assumption (1) was also true, so that we would have available consistent estimates of both the marginal effect parameter $\theta_m^{(j)}(x)$ as well as the conditional parameter $\theta_m^{(j)}(x, \sigma)$. In that case, the question arises as to which parameter is of greater interest. We argue that the marginal effect parameter is of greater public health interest.

Now suppose an official is considering a maternal stress reduction program in hopes of decreasing the current prevalence of absences from day care centers due to childhood illness. In planning this program, the official might want to know the (marginal) prevalence of childhood illness that could be expected to result from a fully successful intervention program, which is the marginal effect parameter $\theta_m^{(0)}(x)$, provided the mother-child pairs in the day care center of interest are similar to (i.e., exchangeable with) the mother-child pairs in the MSCM study population. In general, public health officials will not be interested in the within-pair effect on the pairs actually studied, although the pairs' pediatricians may be. Nonetheless, advocates of the conditional effect approach argue that our ability to estimate the conditional effects $\theta_m^{(j)}(x, \sigma)$ rather than simply the marginal effect $\theta_m^{(j)}(x)$ is quite important, because it allows us to extrapolate the results of our study to alternative target populations (e.g., other day care centers) that differ from our study population in their distribution of random effects.

Therefore, consider a target population whose random effect distribution $F^*(\sigma | x)$ differs from the distribution $F(\sigma | x)$ of our study population. If this target population and our study population have identical conditional effect parameters $\theta_m^{(j)}(x, \sigma)$, then we will be able to consistently estimate the marginal effect in the target population by $\hat{\theta}_m^{(j)*}(x) = \int \hat{\theta}_m^{(j)}(x, \sigma) dF^*(\sigma | x)$. Although correct, this point is only useful if we have some means of estimating $F^*(\sigma | x)$ in our target population. Now, if σ were a real physical quantity, such as a particular genetic marker, we could estimate $F^*(\sigma | x)$ by taking a random sample of subjects in the target population and measuring the marker. However, our σ is not a real measurable physical quantity. In summary, if (i) we did have a way to estimate $F^*(\sigma | x)$, (ii) our models (25) and (26) were correct, and (iii) our target and study populations had the same conditional effect parameters $\theta_m^{(j)}(x, \sigma)$, we could then successfully extrapolate and obtain the marginal effect in the new target population. However, with no way to estimate the distribution $F^*(\sigma | x)$ or to fully check assumptions (ii) and (iii) from the data, any extrapolation based on conditional effects should be interpreted with caution.

We now return to the question of how one might estimate the conditional effects $\theta_m^{(j)}(x, \sigma)$ under the assumption that the modified models (25) and (26) were correctly specified. The referee suggested estimating the parameters of these models by conditional likelihood methods,

without having to specify models for either (i) the conditional distribution of A_m given both the random effect of σ and the past $(\bar{A}_{m-1}, \bar{L}_m, \bar{Y}_m, X)$ or (ii) the distribution $F(\sigma | x)$ of the random effect. The usual conditional logistic likelihood associated with model (25) is the conditional logistic likelihood that, for each mother-child pair i , conditions on both the total number of “successes” $\sum_{m=1}^M Y_{m,i}$ plus the regressors at each occasion m . Unfortunately, in our setting, one cannot use this conditional likelihood to obtain consistent estimates of the parameters β_1 of the modified model (25). Indeed the associated conditional score function is fixed given the conditioning event and thus cannot be used to estimate β_1 , since $Y_{m,i}$ is not only the outcome at time m but also a regressor in the models for $Y_{k,i}, k = m, m+1, m+2, m+3$. Therefore, a natural alternative would be to use a marginal likelihood approach in which (i) one specifies an additional parametric model for the conditional distribution of A_m given the random effect σ and the past, (ii) then integrates with respect to the law of σ given X , and (iii) finally maximizes the marginal (integrated) likelihood with respect to the unknown parameters. The integration could be with respect either to a parametric model for $F(\sigma | x)$ or one could choose to leave $F(\sigma | x)$ non-parametric and use methods of Laird (1978).

When M is large, an alternative, less parametric, conditional likelihood approach is available. First, in the modified models (25) and (26), we allow both the β_d and the σ_d to vary arbitrarily between subjects. We then estimate $(\beta_{di}, \sigma_{di}), d = 0, 1$, by maximum likelihood using only the M observations on subject i (Heckman, 1981, pg. 185-187). Next we estimate the subject-specific parameters $\theta_m^{(j)}(x, \sigma)$ by plugging the ML estimates of $(\beta_{di}, \sigma_{di})$ into the modified g-computation algorithm formula (24). Finally, we estimate the marginal parameter $\theta_m^{(j)}(x)$ as the sample average of the subject-specific estimates among subjects with $X = x$. This approach to estimating causal effects substitutes the functional form and stationarity assumptions (that the parameters β_{di} and σ_{di} are time-invariant) encoded in the random coefficient version of models (25) and (26) for the between-pair comparability assumption (1).

8. ALTERNATIVE APPROACHES TO ESTIMATION

Under assumption (1) of no unmeasured confounders, Robins has proposed two other methods for estimating the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. The first method is semiparametric g-estimation of a multiplicative structural nested mean model (SNMM) (Robins, 1994, 1997, 1998a). This method extends g-estimation of structural nested failure time models (Robins, 1992; Robins et al., 1992) to non-failure time repeated measure outcomes. The second method is inverse probability of treatment weighted estimators of marginal structural models (MSMs), introduced in Robins (1998ab). To describe these models, let $\bar{a} = \bar{a}_M = (a_0, \dots, a_M)$ be a deterministic treatment history through end of follow-up, and let $Y_m^{(\bar{a})}$ be a subject’s outcome at time m if, possibly contrary to fact, the subject received exposure history \bar{a} rather than their observed exposure history \bar{A} . One simple logistic MSM assumes that

$$\text{logit } E \left[Y_m^{(\bar{a})} | X \right] = \eta_{0,m} + \eta_{1,m} d(\bar{a}_{m-1}) + \eta'_2 X \quad (27)$$

where $d(\cdot)$ is a known function and the η 's are unknown parameters; for example, $d(\bar{a}_m)$ could be cumulative exposure $\sum_{k=0}^m \bar{a}_k$. This MSM model implies that the effect of treatment on the mean of $Y_m^{(\bar{a})}$ is a linear logistic function of $d(\bar{a}_{m-1})$.

To describe a multiplicative SNMM, let $(\bar{A}_k, 0)$ be the treatment history that agrees with a subject's observed treatment history \bar{A}_M through time k and is 0 thereafter. A simple multiplicative SNMM assumes that a final brief blip of exposure of magnitude A_k at time k among subjects with past history $\bar{L}_k, \bar{A}_k, \bar{Y}_k, X$ multiplies the mean of $Y_m^{(\bar{A}_{k-1,0})}$ by $\exp(\psi_m A_k)$. That is,

$$E \left[Y_m^{(\bar{A}_{k,0})} \mid \bar{L}_k, \bar{A}_k, \bar{Y}_k, X \right] = \exp(\psi_m A_k) E \left[Y_m^{(\bar{A}_{k-1,0})} \mid \bar{L}_k, \bar{A}_k, \bar{Y}_k, X \right]. \quad (28)$$

Both our MSM (27) and our multiplicative SNM model (28) can be used to estimate contrasts between $\theta_m^{(1)}(x)$ and $\theta_m^{(0)}(x)$. Specifically, $\text{logit } \theta_m^{(1)}(x) - \text{logit } \theta_m^{(0)}(x) = m\eta_{1,m}$ and $\log \left[\theta_m^{(1)}(x) / \theta_m^{(0)}(x) \right] = m\psi_m$. Furthermore, the parameters η of the MSM (27) and the parameters ψ of the multiplicative SNMM (28) can be estimated from the observed data using inverse probability of treatment weighted estimation and g-estimation respectively, provided one can correctly specify a model for $\text{pr} [A_m = 1 \mid \bar{A}_{m-1}, \bar{L}_m, \bar{Y}_m, X]$.

Estimation of SNMMs and MSMs complement estimation based on the g-computation algorithm in the sense that (i) estimation based on the g-computation algorithm does not require one to model $\text{pr} [A_m = 1 \mid \bar{A}_{m-1}, \bar{L}_m, \bar{Y}_m, X]$ but does require one model the joint distribution of (Y_m, L_m) given $\bar{A}_m, \bar{L}_{m-1}, \bar{Y}_{m-1}, X$. A drawback of SNMMs is that (i) multiplicative SNMMs require the response probabilities to be small because the multiplicative model does not respect the fact that probabilities are bounded above by one, while (ii) logistic SNMMs, which do respect this bound, cannot be estimated by the method of g-estimation (Robins, Rotnitzky, and Scharfstein, 1999).

An important advantage of an approach based on MSMs or SNMMs, compared to methods based on the g-computation algorithm formula, is that MSMs and SNMMs have parameters that take the value zero whenever the null hypothesis of no treatment effect is true. For example, when treatment has no effect on the marginal mean of the Y_m 's, $\eta_{1,m} = 0$ and $\psi_m = 0$ for all m in the above MSM and SNMM model. Analysis of the MSCSM data using MSMs and SNMMs will be reported elsewhere.

9. DISCUSSION

As argued in Robins (1986) and Rosenbaum (1984), we should not expect assumption 2 to hold if mother's stress A_k on day k could affect her child's health status Y_k on the same day k , since then Y_k could be a consequence of A_k . Zeger and Liang (1991) made a similar argument. When mother's stress and child's illness occurred together on a particular day, we had no way of determining which preceded the other temporally. However, we suspect that, as required for assumption 2, it would take at least one day for the effect of mother's stress to have health consequences for her child. Assumption 2 is also consistent with the more likely possibility that a child's illness on day k could cause a mother to become stressed on that day.

To determine the sensitivity of our statistical inferences to these substantive assumptions,

we reanalyzed the data under the opposite assumption on the temporal order. We assumed that mother’s stress on day k could cause child’s illness on that day but that illness on day k would not affect mother’s stress until day $k + 1$. Under this substantive assumption, assumption 2 is no longer tenable; however, the alternative assumption that

$$\left\{ \left(Y_m^{(0)}, Y_m^{(1)} \right); m = k, k + 1, \dots, M \right\} \prod A_k \mid \bar{A}_{k-1}, \bar{Y}_{k-1}, X \text{ for all } k \geq 1 \quad (29)$$

is reasonable. To reanalyze the data under this modified no-confounding assumption (29), it suffices to recode illness occurring on day k as having occurred on day $k + 1$ and then follow the steps described previously for analyzing the data under the original assumption (2). When we did so, we obtained an estimate 0.102 for $\delta_m(x)$, the same to three significant digits as our previous estimate. We thus conclude that our inferences appear robust to our assumptions concerning the causal ordering of mother’s stress and child’s illness on a single day. However, in an observational study, we can never know whether our assumption of no unmeasured confounders (1) is true. Thus, it is of interest to conduct a sensitivity analysis to quantify how our inferences concerning the causal effect of a treatment or exposure on an outcome vary as a function of the magnitude of confounding by unmeasured factors. Robins, Rotnitzky, and Scharfstein (1999) develop appropriate methods. These methods do not require the stationarity or functional form assumptions of the conditional methods described in Sec. 7 above.

Finally, one might be concerned that it was fruitless to attempt to estimate the causal effect of self-reported mother’s stress on self-reported child’s illness, because the exposure (self-reported stress) and the outcome (self-reported illness) were subjective and ill-defined. Nevertheless, the methods developed in this paper will also be necessary to estimate the effect of an objective well-defined time varying exposure on the marginal mean of an objective well-defined dichotomous outcome.

APPENDIX A: PROOFS OF PROPOSITIONS 5 AND 6

Proof of Proposition 5

Note that

$$\begin{aligned} f [A_m, \dots, A_M \mid Y_m, \bar{A}_{m-1}, X] &= \prod_{k=m}^M f [A_k \mid Y_m, \bar{A}_{k-1}, X] = \\ &= \prod_{k=m}^M f [A_k \mid \bar{A}_{k-1}, X] = f [A_m, \dots, A_M \mid \bar{A}_{m-1}, X] \end{aligned} \quad (A.1)$$

where the second equality uses Eq. (15). Eq. (11) is equivalent to Eq. (A.1). Hence, Eq. (15) implies Eq. (11). Chamberlain cited other references for this result.

Proof of Proposition 6

If one chooses the working covariance matrix $V(\alpha) = V_{ind}(\alpha)$, then $\mathbf{U}_{GEE}(V_{ind}, \alpha)$ is a linear combination of terms of the form

$$U_m(\alpha) \equiv g(\bar{A}_{m-1}, X; \alpha) \varepsilon_m(\alpha)$$

with

$$g(\bar{A}_{m-1}, X; \alpha) = \{\partial P_m(\alpha) / \partial \alpha\} \{P_m(\alpha) [1 - P_m(\alpha)]\}^{-1}$$

and

$$\varepsilon_m(\alpha) \equiv Y_m - P_m(\alpha).$$

Hence, we have

$$E\{\mathbf{U}_{GEE}(V_{ind}, \alpha)\} = E\left\{\sum_{i=1}^n \{\partial \mathbf{P}_i(\alpha) / \partial \alpha\} [V_{i,ind}(\alpha)]^{-1} \varepsilon_i(\alpha)\right\} = \mathbf{0}$$

under model (12) and thus $\hat{\alpha}_{GEE}(V_{ind})$ will be consistent for α .

However, if $V(\alpha)$ has non-zero off-diagonal elements, then the linear combination $\mathbf{U}_{GEE}(V, \alpha)$ contains terms of the form $q(\bar{A}_{m-1}, X; \alpha) \varepsilon_k(\alpha)$ with $k < m$ for some $q(\cdot, \cdot; \alpha)$. Since Eq. (11) is not true and model (12) does not imply that such terms have mean zero, $\hat{\alpha}_{GEE}(V)$ may not be consistent. Note that, in general, $Cov\{U_k(\alpha), U_m(\alpha)\} \neq 0$ for $k < m$ even with $V(\alpha) = V_{ind}(\alpha)$, since model (12) does not imply that $E[U_k(\alpha)U_m(\alpha)] = 0$, due to the fact that $U_k(\alpha)$ contains $\varepsilon_k(\alpha) \equiv Y_k - P_k(\alpha)$. The second part of Proposition 6 follows immediately from the fact that when Eq. (14) and model (12) are both true, the likelihood function $\mathcal{L}(\alpha)$ can be written as

$$\prod_{i=1}^n \left\{ f(X_i) \prod_{m=1}^M f[Y_{m,i} | \bar{A}_{m-1,i}, X_i; \alpha] f[A_{m,i} | \bar{Y}_{m,i}, \bar{A}_{m-1,i}, X] \right\}.$$

Thus, under model 12,

$$\partial \ln \mathcal{L}(\alpha) / \partial \alpha = \mathbf{U}_{GEE}(V_{ind}, \alpha).$$

APPENDIX B: “MARKOV CHAIN” ESTIMATION OF $\theta_m^{(j)}(x)$

Consider the general set-up of Sec. 6.1 under the assumption of (1) of no unmeasured confounders. Results of earlier sections are obtained as a special case. We shall require the following definitions. Let Y_m^* be a random variable denoting the joint levels of Y_m and L_m , with realizations y^* taking values in the set $\{1, \dots, K\}$.

In our MSCM data set with L_m denoting illness severity, Y_m^* had only three levels, i.e. $K = 3$, as follows: $Y_m^* \equiv 0$ if $Y_m = 0$ and $L_m = 0$; $Y_m^* \equiv 1$ if $Y_m = 1$ and $L_m = 0$; and $Y_m^* \equiv 2$ if $Y_m = 1$ and $L_m = 1$. Hence $Y_m = I(Y_m^* \neq 0)$ and $L_m = I(Y_m^* = 2)$.

Definition 3. We say that Y_m^* is q-Markov if for all $m \geq q$ and $j = 0$ or 1 ,

$$\Pr[Y_m^* = y^* | \bar{Y}_{m-1}^*, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X] = \Pr[Y_m^* = y^* | Y_{m-1}^*, \dots, Y_{m-q}^*, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X]$$

Definition 4. We say that Y_m^* is q-stationary if it is q-Markov and, for $j = 0$ or 1 and $m \geq q$,

$$\Pr \left[Y_m^* = y^* \mid (Y_{m-1}^*, \dots, Y_{m-q}^*) = r^{[q]}, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X = x \right] \equiv P_{r^{[q]}}^{(j)}(y^*, x)$$

is independent of time m , where $r^{[q]}$ is a q-vector of possible realizations of $(Y_{m-1}^*, \dots, Y_{m-q}^*)$. In particular, if models (25)-(26) are correct and $\beta_{8d} = \beta_{9d} = \beta_{10d} = 0$, Y_m^* is 2-stationary.

Next, note that Eq. (24) can be rewritten as

$$\theta_m^{(j)}(x) = \sum_{\{y^*; y^* \neq 0\}} \Delta_m^{(j)}(y^*, x) \quad (\text{B1})$$

where

$$\begin{aligned} \Delta_m^{(j)}(y^*, x) = & \sum_{\text{all } \bar{Y}_{m-1}^*} \{ \Pr [Y_m^* = y^* \mid \bar{Y}_{m-1}^* = \bar{y}_{m-1}^*, \bar{A}_{m-1} = \mathbf{j}^{[m-1]}, X = x] \bullet \\ & \prod_{k=1}^{m-1} \Pr [Y_k^* = y_k^* \mid \bar{Y}_{k-1}^* = \bar{y}_{k-1}^*, \bar{A}_{k-1} = \mathbf{j}^{[k-1]}, X = x] \}. \end{aligned} \quad (\text{B2})$$

The form of B2 implies that if Y_m^* is q-stationary and the $P_{r^{[q]}}^{(j)}(y^*, x)$ are all non-zero, then, the limit $\Delta^{(j)}(y^*, x) \equiv \lim_{m \rightarrow \infty} \Delta_m^{(j)}(y^*, x)$ of the $\Delta_m^{(j)}(y^*, x)$ is equal to the limiting marginal probability $\Pr(Z_m = y^*)$ in a K -state qth-order irreducible and ergodic Markov chain $\{Z_1, Z_2, \dots\}$ with stationary transition probabilities $P_{r^{[q]}}^{(j)}(y^*, x)$ for $j = 0$ or 1 . Thus, if we define

$$\theta^{(j)}(x) \equiv \sum_{\{y^*; y^* \neq 0\}} \Delta^{(j)}(y^*, x), \quad (\text{B3})$$

then, by Eq. (B1), $\theta^{(j)}(x)$ is the limit as $m \rightarrow \infty$ of the $\theta_m^{(j)}(x)$. Our estimates $\hat{\theta}_m^{(j)}(x)$ of the $\theta_m^{(j)}(x)$ in the text were the maximum likelihood (ML) estimates under models (25)-(26) with $\beta_{8d} = \beta_{9d} = \beta_{10d} = 0$ of the limiting probability $\theta^{(j)}(x)$. Hu and Robins (1993) show that, in this example, the ML estimates of $\theta^{(j)}(x)$ and $\theta_m^{(j)}(x)$ differ from one another only in the third significant digit for $m \geq 20$, justifying our approach.

In order to calculate the ML estimate of $\theta^{(j)}(x)$, we use the fact that if $\{Z_1, Z_2, \dots\}$ is a stationary K -state qth-order Markov chain, then $\{T_1, T_2, \dots\}$ with $T_m = (Z_{m-q+1}, \dots, Z_m)$ is a stationary K^q -state 1st order Markov chain for $m > q - 1$ whose $K^q \times K^q$ (stationary) transition probability matrix \mathbf{P} is a function of the stationary transition probabilities $P_{r^{[q]}}^{(j)}(y^*, x)$ of the original Markov chain $\{Z_1, Z_2, \dots\}$. The limiting stationary (marginal) probability distribution $H \equiv (h_0, h_1, \dots, h_{K^q-1})$ of the transformed Markov chain $\{T_1, T_2, \dots\}$ allows us to calculate the limiting marginal probability $\Delta^{(j)}(y^*, x)$ of the original Markov chain $\{Z_1, Z_2, \dots\}$.

We can calculate the stationary distribution H using the following results for first-order Markov chains (Cox and Miller, 1965; Karlin and Taylor, 1975; Ross, 1983). Let A be the $K^q \times K^q$ matrix $(\mathbf{P}^T - I)$ except with its last row replaced by a row-vector of K^q 1's and let C be the $K^q \times 1$ column vector $(0, 0, \dots, 1)^T$. Then

$$H^T = A^{-1}C \quad (\text{B4})$$

We fit models (25)-(26) with $\beta_{8d} = \beta_{9d} = \beta_{10d}$ set to zero by maximum likelihood to obtain estimates of the stationary transition probabilities $\hat{P}_{r^{[2]}}^{(j)}(y^*, x)$, where $r^{[2]} = (y_1^*, y_2^*)$. Separately for $j = 0$ and 1, we then used $\hat{P}_{r^{[2]}}^{(j)}(y^*, x)$ to obtain the ML estimate $\hat{\mathbf{P}}$ of the $3^2 \times 3^2$ (stationary) transition probability matrix \mathbf{P} of the transformed 3²-state 1st order Markov chain $\{T_1, T_2, \dots\}$. Next, we computed the estimate \hat{H} of the corresponding stationary (marginal) probability distribution H of Eq. (B4) by using the matrix \hat{A} calculated from our estimate $\hat{\mathbf{P}}$, and then we calculated $\hat{\Delta}^{(j)}(y^*, x)$ from \hat{H} . Finally, we used Eq. (B3) to obtain $\hat{\theta}^{(j)}(x)$.

For some outcomes of interest, stationarity may not be tenable. For example, suppose the outcome of interest was an indicator of whether an HIV infected subject's CD4 count exceeded 200 on day m . In that case, the stationarity assumption would not hold due to a downward trend in CD4-counts over time in HIV infected persons. In contrast, children's illnesses usually lasted for only a few days and there was no trend in child's illness rates over the time period of interest, i.e. 30-60 days, so that the q-stationarity assumption was quite reasonable in our example, and was consistent with the data.

References

- Chamberlain, G. (1984), "Panel data," in *Handbook of Econometrics*, Volume II, Z. Griliches and M.D. Intriligator (eds.), Amsterdam: North-Holland, pp. 1247-1318.
- Cox, D.R. (1958), *The design and planning of experiments*, London: Chapman and Hall.
- Cox, D.R. and Hinkley, D.V. (1974), *Theoretical Statistics*, London: Chapman and Hall.
- Cox, D. R. and Miller, H. D. (1965), *The Theory of Stochastic Processes*, London: Methuen.
- Greenland, S. and Robins, J. (1986), "Identifiability, exchangeability, and epidemiological confounding," *International Journal of Epidemiology*, 15, 413-419.
- Heckman, J.J. (1981), "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," in *Structural Analysis of Discrete Data with Econometric Applications*, Manski, C.F., McFadden, D. (eds.), Cambridge, MA: MIT Press, pp. 179-195.
- Holland, P. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945-961.
- Hu, F.-C. and Robins, J.M. (1993), "Using limit laws for Markov chains to estimate the causal effect of a time-varying exposure on a repeated binary outcome," in F.-C. Hu, unpublished Sc.D. Dissertation, Harvard School of Public Health, Department of Biostatistics, Boston, MA.
- Kalbfleisch, J.D., and Prentice, R.L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley.
- Karlin, S. and Taylor, H. M. (1975), *A First Course in Stochastic Processes* (2nd ed.). San Diego, CA: Academic Press.
- Laird N. (1978), "Nonparametric maximum likelihood estimation of a mixing distribution," *Journal of the American Statistical Association*, 73, 805-811.
- Lewis D.K. (1973), *Counterfactuals*, Cambridge, MA: Harvard University Press.
- Liang K-Y. and Zeger S.L. (1986), "Longitudinal Data Analysis using Generalized Linear Models," *Biometrika*, 73, 13-22.
- Neuhaus, J.M., Kalbfleisch, J.D., and Hauk, W.W. (1991), "A comparison of cluster-specific and population-average approaches for analyzing correlated binary data," *International Statistical Review*, 59, 25-35.

Robins, J.M. (1986), "A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods – Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7, 1393-1512.

Robins, J.M. (1987a), "A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods," *Journal of Chronic Disease* (40, Supplement), 2, 139s-161s.

Robins, J.M. (1987b), "Addendum to 'A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods – Application to Control of the Healthy Worker Survivor Effect'," *Computers and Mathematics with Applications*, 14, 923-945.

Robins, J.M. (1989a), "The Control of Confounding by Intermediate Variables," *Statistics in Medicine*, 8, 679-701.

Robins, J.M. (1989b), "The Analysis of Randomized and Non-randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in *Health Service Research Methodology: A Focus on AIDS*, Sechrest, L., Freeman, H., and Mulley, A. (eds.), NCHSR, U.S. Public Health Service, pp. 113-159.

Robins, J.M. (1992), "Estimation of the Time-dependent Accelerated Failure Time Model in the Presence of Confounding Factors," *Biometrika*, 79, 321-334.

Robins, J.M., Blevins D, Ritter G, and Wulfsohn M. (1992), "G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients," *Epidemiology*, 3, 319-336.

Robins, J.M. (1994). "Correcting for non-compliance in randomized trials using structural nested mean models," *Communications in Statistics*, 23, 2379-2412.

Robins, J.M. (1997), "Causal inference from complex longitudinal data," in *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)*, Berkane, M. (ed.), NY: Springer Verlag, pp. 69-117.

Robins, J.M. (1998a), "Correction for non-compliance in equivalence trials," *Statistics in Medicine*, 17, 269-302.

Robins, J.M. (1998b), "Marginal structural models," *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, pp. 1-10.

Robins, J.M., Rotnitzky, A. and Scharfstein, D. (1999), "Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models," *Statistical Models in Epidemiology*, E. Halloran (ed.), NY: Springer-Verlag (to appear).

Rosenbaum, P.R. (1984), "The consequences of adjustment for a concomitant variable that has been affected by the treatment," *Journal of the Royal Statistical Society, Series A, General*, 147, 656-666.

Rosenbaum, P. R. (1984), "From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment," *Journal of the American Statistical Association*, 79, 41-48.

Ross, S. M. (1983), *Stochastic Processes*, New York: John Wiley & Sons.

Rubin, D.B. (1974), "Estimating Causal Effects of Treatment in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Rubin, D.B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34-58.

Rubin, D.B. (1990), "Comment: Neyman (1923) and causal inference in experiments and observational studies," *Statistical Science*, 5, 472-480.

Zeger, S.L., and Liang, K-Y. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121-130.

Zeger, S.L. and Liang, K-Y. (1991), "Feedback models for discrete and continuous time series," *Statistica Sinica*, 1, 51-64.

Table 1: Proportion of Positive Responses over All Days and Subjects in Mothers' Stress and Children's Morbidity Study (from Zeger Liang, 1986)

Variable	Proportion Positive
Child ill	0.14
Mother stressed at: $t - 1$	0.14
$t - 2$	0.15
$t - 3$	0.16
Household size > 3	0.66
Non-white	0.55
Employed	0.33
Married	0.50

Table 2: Testing Whether Equation 11 Was Correct

Response Variable: Illness (t)

Covariate	ML Estimate	Standard Error	Prob.
Intercept	-2.523	0.141	0.000
► Stress (t)	0.643	0.139	0.000
Stress ($t - 1$)	0.374	0.145	0.010
Stress ($t - 2$)	-0.0063	0.154	0.968
Stress ($t - 3$)	0.331	0.141	0.0189
Married	0.441	0.114	0.000
Employed	0.168	0.122	0.168
Non-white	0.530	0.114	0.000
Household > 3	-0.682	0.112	0.000

Table 3: Testing Whether Equation 15 Was Correct**Response Variable: Stress(t)**

Covariate	ML Estimate	Standard Error	Prob.
Intercept	-2.944	0.153	0.000
Stress ($t - 1$)	0.963	0.133	0.000
Stress ($t - 2$)	0.367	0.144	0.012
Stress ($t - 3$)	0.372	0.143	0.009
Avg_Stress ($\leq t - 4$)	2.042	0.241	0.000
► Illness (t)	0.554	0.148	0.000
Married	0.197	0.117	0.091
Employed	0.169	0.124	0.174
Non-white	-0.080	0.114	0.487
Household > 3	-0.108	0.119	0.366

Table 4: Testing Whether Equation 14 Was Correct**Response Variable: Illness(t)**

Covariate	ML Estimate	Standard Error	Prob.
Intercept	-3.056	0.163	0.000
Stress ($t - 1$)	0.227	0.168	0.177
Stress ($t - 2$)	-0.289	0.182	0.112
Stress ($t - 3$)	0.454	0.163	0.005
Avg_Stress ($\leq t - 4$)	0.532	0.293	0.069
► Illness ($t - 1$)	2.576	0.126	0.000
Married	0.336	0.128	0.009
Employed	0.145	0.137	0.291
Non-white	0.324	0.129	0.012
Household > 3	-0.415	0.127	0.001