Estimating Time-Varying Causal Effect Moderation for Micro-Randomized Trials with Binary Outcomes

Tianchen Qian[1], Hyesun Yoo[2], Pedrag Klasnja[3], Daniel Almirall[4], Susan Murphy[1]

[1]Department of Statistics, Harvard University. [2]Department of Statistics, University of Michigan. [3]School of Information, University of Michigan. [4]Institute for Social Research, University of Michigan

Summary

Binary outcome is common in mobile health studies. We focus on estimating the time-varying causal effect moderation for data from micro-randomized trials with binary outcomes. We give the definition of moderated treatment effect in this setting, and provide two estimation methods. One estimation method is for the proximal treatment effect conditional on the entire history, and the estimator is semiparametric locally efficient. The other estimation method, based on weighted and centered least squares, is for the proximal treatment effect marginal but conditional only on a subset of variables in history. Both estimators are robust in the sense that they do not require a correct model for the outcome process. The methods are illustrated by simulation studies.

Notation

Observed data:

- X_t: covariate information prior to t-th decision point
- A_t: treatment assignment at t-th decision point
- Y_{t+1}: proximal outcome after t-th decision point

Potential outcome:

- $X_t(\bar{a}_{t-1})$: potential covariate under treatment history $\bar{a}_{t-1}$
- $Y_{t+1}(\bar{a}_t)$: potential covariate under treatment history $\bar{a}_t$
- $H_t(\bar{A}_{t-1}) = \{X_1, A_1, Y_2(A_1), \ldots, A_{t-1}, Y_t(\bar{A}_{t-1}), X_t(\bar{A}_{t-1})\}$

Treatment effect:

- Proximal treatment effect conditional on the entire history:

$$\log \frac{E\{Y_{t+1}(\bar{A}_{t-1}, 1) \mid H_t(\bar{A}_{t-1})\}}{E\{Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_t(\bar{A}_{t-1})\}}. \tag{1}$$

- Proximal treatment effect conditional on a subset of variables in history:

$$\log \frac{E\{Y_{t+1}(\bar{A}_{t-1}, 1) \mid S_t(\bar{A}_{t-1})\}}{E\{Y_{t+1}(\bar{A}_{t-1}, 0) \mid S_t(\bar{A}_{t-1})\}}, \tag{2}$$

  where S_t is a vector of summary variables chosen from H_t. For example, S_t would typically include variables in H_t that may modify the treatment effect, and it likely also includes deterministic entries such as 1 for the intercept and terms in time, t.

Many mobile health interventions are designed to affect an individual proximally in time [2]. For example, activity suggestions in HeartSteps encourage an individual to walk immediately after s/he receives a push notification. Suppose $0 \in \mathcal{A}$ denotes the control, i.e., no intervention delivery. Equation (1) defines the proximal effect of intervention $A_t = a$ on Y_{t+1} given history H_t. It denotes the log of the relative risk had an individual received the intervention versus if the individual had not received the intervention at decision point t, conditional on his/her history. In other words, (1) represents whether (and by how much) delivering intervention at time t is increases the individual's response as opposed to no intervention, given an individual's current context. Suppose the intervention is intended to increase the probability of $Y_{t+1} = 1$, then testing for a positive (1) tells us whether the intervention achieves its intended effect.

Equation (2) is different from (1) in that the conditioning is only over a subset of variables S_t in history H_t. Scientifically, this is of interest in primary analyses, for example, where the treatment effect is marginal over alla variables in H_t (i.e., setting $S_t = \emptyset$).

Causal Assumptions

To express the proximal treatment effect in terms of the observed data, we make the following assumptions:

1. **Consistency.** The observed data equals the potential outcome under observed treatment assignment. In particular, $Y_2 = Y_2(A_1)$, $X_2 = X_2(A_1)$, $A_2 = A_2(A_1)$, and for each subsequent $t \le T$, $Y_t = Y_t(\bar{A}_{t-1})$, $X_t = X_t(\bar{A}_{t-1})$, $A_t = A_t(\bar{A}_{t-1})$, and lastly, $Y_{T+1} = Y_{T+1}(\bar{A}_T)$.
2. **Positivity.** If $\Pr(H_t = h_t) > 0$, then $\Pr(A_t = a \mid H_t = h_t) > 0$ for all $a \in \mathcal{A}$.
3. **Sequential ignorability.** For $1 \le t \le T$, the potential outcomes $\{Y_{t+1}(\bar{a}_t), X_{t+1}(\bar{a}_t), A_{t+1}(\bar{a}_t), \ldots, Y_{T+1}(\bar{a}_T)\}$ is independent of A_t conditional on H_t.

In an MRT, because the treatment is sequentially randomized and there is always some exploration of different treatment options under each context, Assumptions 2 and 3 always hold by design. Assumption 1 may fail to hold if there are interactions between individuals; for example, in mobile health interventions with social media components, one individual's treatment can impact another individual's outcome. In practice it is critical to ensure these assumptions by a proper design.

Under these assumptions, the treatment effects in (1) and (2) can be rewritten in terms of the observed data:

$$\log \frac{E\{Y_{t+1}(\bar{A}_{t-1}, 1) \mid H_t(\bar{A}_{t-1})\}}{E\{Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_t(\bar{A}_{t-1})\}} = \log \frac{E\{Y_{t+1} \mid H_t, A_t = 1\}}{E\{Y_{t+1} \mid H_t, A_t = 0\}}.$$

and

$$\log \frac{E\{Y_{t+1}(\bar{A}_{t-1}, 1) \mid S_t(\bar{A}_{t-1})\}}{E\{Y_{t+1}(\bar{A}_{t-1}, 0) \mid S_t(\bar{A}_{t-1})\}} = \log \frac{E[E\{Y_{t+1} \mid H_t, A_t = 1\} \mid S_t]}{E[E\{Y_{t+1} \mid H_t, A_t = 0\} \mid S_t]}.$$

Semiparametric Locally Efficient Estimator for Treatment Effect (1)

For simplicity, suppose a linear model for $1 \le t \le T$:

$$\log \frac{E(Y_{t+1} \mid H_t, A_t = 1)}{E(Y_{t+1} \mid H_t, A_t = 0)} = S_t^T \beta^*, \tag{3}$$

where β^* is the true value of a p-dimensional parameter β. Suppose $g_t(H_t)^T \alpha$ is a working model for $\log E(Y_{t+1} \mid H_t, A_t = 0)$. The estimating equation for (α, β) is as follows:

$$m_1(\alpha, \beta) := \sum_{t=1}^{T} \frac{e^{-A_t S_t^T \beta}(Y_{t+1} - e^{g_t(H_t)^T \alpha + A_t S_t^T \beta})}{(1 - e^{g_t(H_t)^T \alpha})p_t + (e^{-S_t^T \beta} - e^{g_t(H_t)^T \alpha})(1 - p_t)} \times \begin{bmatrix} g_t(H_t) \\ (A_t - p_t) S_t \end{bmatrix}. \tag{4}$$

Asymptotic result. Suppose $p_t = \Pr(A_t = 1 \mid H_t)$ is known. Let $\dot{m}_1$ be the derivative of $m_1(\alpha, \beta)$ with respect to (α, β). The solutions to the estimating equation $\mathbb{P}_n m_1(\alpha, \beta) = 0$ yields an estimator $(\hat{\alpha}, \hat{\beta})$ for which $\sqrt{n}(\hat{\beta} - \beta^*)$ is asymptotically normal with mean zero and variance-covariance matrix consistently estimated by the lower block diagonal $(p \times p)$ entry of the matrix $(\mathbb{P}_n \dot{m}_1(\hat{\alpha}, \hat{\beta}))^{-1} \mathbb{P}_n m_1(\hat{\alpha}, \hat{\beta})^{\otimes 2} (\mathbb{P}_n \dot{m}_1(\hat{\alpha}, \hat{\beta}))^{-1^T}$.

Remark. The estimating function $m_1(\alpha, \beta)$ is related to the multiplicative structural nested mean model by Robins [4]. The estimator $\hat{\beta}$ is consistent under misspecified $E(Y_{t+1} \mid H_t, A_t = 0)$. When $\exp\{g(H_t)^T \alpha\}$ is a correct model for $E[Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_t]$, $\hat{\beta}$ achieves the semiparametric efficiency lower bound.

Weighted and Centered Least Squares Estimator for Treatment Effect (2)

There are settings where (3) doesn't hold. For example, a primary analysis of MRT data may focus on the treatment effect that is marginal over all variables in H_t (i.e., setting $S_t = \emptyset$). In such settings, we make the following assumption on the marginalized treatment effect. For $1 \le t \le T$, suppose

$$\log \frac{E[E\{Y_{t+1} \mid H_t, A_t = 1\} \mid S_t]}{E[E\{Y_{t+1} \mid H_t, A_t = 0\} \mid S_t]} = S_t^T \beta^*. \tag{5}$$

To estimate β in (5), we use the following estimating equation

$$m_2(\alpha, \beta) = \sum_{t=1}^{T} e^{-A_t S_t^T \beta}\{Y_{t+1} - e^{g_t(H_t)^T \alpha + A_t S_t^T \beta}\} \times \frac{\tilde{p}_t(A_t \mid S_t)}{p_t(A_t \mid H_t)} \begin{bmatrix} g_t(H_t) \\ \{A_t - \tilde{p}_t(1 \mid S_t)\} S_t \end{bmatrix}, \tag{6}$$

where $p_t(A_t \mid H_t) = \Pr(A_t \mid H_t)$ is the true randomization probability and $\tilde{p}_t(A_t \mid S_t)$ is an arbitrary probability as long as it depends on H_t only via S_t. $g_t(H_t)$ is a vector of features constructed from H_t, and $\exp\{g_t(H_t)^T \alpha\}$ is a working model for $E(Y_{t+1} \mid H_t, A_t = 0)$.

Asymptotic result. Suppose $p_t = \Pr(A_t = 1 \mid H_t)$ is known. Let $\dot{m}_2$ be the derivative of $m_2(\alpha, \beta)$ with respect to (α, β). The solutions to the estimating equation $\mathbb{P}_n m_2(\alpha, \beta) = 0$ yields an estimator $(\hat{\alpha}, \hat{\beta})$ for which $\sqrt{n}(\hat{\beta} - \beta^*)$ is asymptotically normal with mean zero and variance-covariance matrix consistently estimated by the lower block diagonal $(p \times p)$ entry of the matrix $(\mathbb{P}_n \dot{m}_2(\hat{\alpha}, \hat{\beta}))^{-1} \mathbb{P}_n m_2(\hat{\alpha}, \hat{\beta})^{\otimes 2} (\mathbb{P}_n \dot{m}_2(\hat{\alpha}, \hat{\beta}))^{-1^T}$.

Remark. The estimating function $m_2(\alpha, \beta)$ generalizes the weighted and centered least squares method in Boruvka et al. [1] to binary outcome. The estimator $\hat{\beta}$ is consistent under misspecified $E(Y_{t+1} \mid H_t, A_t = 0)$, and the choice of $\tilde{p}_t$ doesn't affect its consistency as long as it depends on H_t only through S_t. Such robustness is due to the use of weighting (by J_t) and centering (by $\tilde{p}_t$). When the randomization probability p_t is constant, one can set $\tilde{p}_t = p_t$ and $J_t \equiv 1$.

Standard Error and Small Sample Correction

When sample size is small, the sandwich estimators for the variance in previous sections can be anti-conservative. To address this, we adopt the small sample correction technique in Mancl and DeRouen [3] to modify the term $\mathbb{P}_n m(\hat{\alpha}, \hat{\beta})^{\otimes 2}$ in the variance estimator. In particular, we premultiply the vector of each individual's residuals, $\{Y_{t+1} - e^{g_t(H_t)^T \hat{\alpha} + A_t S_t^T \hat{\beta}}\}_{1 \le t \le T}$, by the inverse of the identity matrix minus the leverage for this individual. Also, as in [1], we use critical values from a t distribution. In particular, for a known p-dimensional vector c, to test the null hypothesis $c^T \beta = 0$ or to form confidence intervals, we use the critical value $t_{n-p-q}^{-1}(1 - \xi)$, where p, q are the dimensions of β, α, respectively, and ξ is the significance level.

References

[1] Boruvka, A., D. Almirall, K. Witkiewitz, and S. A. Murphy (2017). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association* (just-accepted).

[2] Heron, K. E. and J. M. Smyth (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology 15*(1), 1–39.

[3] Mancl, L. A. and T. A. DeRouen (2001). A covariance estimator for gee with improved small-sample properties. *Biometrics 57*(1), 126–134.

[4] Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods 23*(8), 2379–2412.

Simulation Study

Let "**Eff**" denote the semiparametric locally efficient estimator for treatment effect (1), and let "**WCLS**" denote the weighted and centered least squares estimator for treatment effect (2).

Generative model. We consider a covariate X_t that takes three values $0, 1, 2$ with equal probability. The treatment indicator A_t is binary with randomization probability η, where η is a constant (i.e., randomization probability doesn't depend on history). The outcome Y_{t+1} is generated from Bernoulli distribution with mean

$$E(Y_{t+1} \mid H_t, A_t) = E(Y_{t+1} \mid H_t, A_t = 0) \exp\{A_t(\beta_0 + \beta_1 X_t)\},$$

where $E(Y_{t+1} \mid H_t, A_t = 0) = \gamma_0 \mathbb{1}(X_t = 0) + \gamma_1 \mathbb{1}(X_t = 1) + \gamma_2 \mathbb{1}(X_t = 2)$. We choose the true parameter values to be $\eta = 0.2$, $\gamma_0 = 0.2$, $\gamma_1 = 0.5$, $\gamma_2 = 0.4$, $\beta_0 = 0.1$, $\beta_1 = 0.3$.

Simulation study 1. We set $S_t = \emptyset$ in the analysis model for treatment effect, and we impose the working model $\exp\{g_t(H_t)^T \alpha\} = \exp(\alpha_0 + \alpha_1 X_t)$ for $E(Y_{t+1} \mid H_t, A_t = 0)$. The generative model implies that

$$\log \frac{E(Y_{t+1} \mid H_t, A_t = 1)}{E(Y_{t+1} \mid H_t, A_t = 0)} = \beta_0 + \beta_1 X_t.$$

Therefore, setting $S_t = \emptyset$ implies that the treatment effect model (3) for Eff is wrong, and it is expected to be inconsistent in this case. For WCLS, the marginal treatment effect model (5) is correct. Hence, we expect the WCLS estimator to be consistent. For log-linear GEE, since its consistency requires a correct model for $E(Y_{t+1} \mid H_t, A_t)$ but our working model for $E(Y_{t+1} \mid H_t, A_t = 0)$ is wrong (the true model is not linear in X_t), we expect the estimator to be inconsistent. These conjectures are supported by the following table.

Table 1: Comparison of three estimators of the marginal treatment effect ($S_t = \emptyset$), when the treatment effect conditional on the full history depends on X_t.

Estimator	Sample size	Bias	SD	RMSE	CP (unadj)	CP (adj)
Eff	30	**0.047**	0.070	0.084	**0.92**	0.94
	50	**0.047**	0.057	0.073	**0.88**	**0.89**
	100	**0.051**	0.040	0.064	**0.78**	**0.79**
WCLS	30	0.000	0.072	0.072	0.95	0.96
	50	-0.001	0.058	0.058	**0.93**	0.94
	100	0.002	0.041	0.041	0.94	0.94
GEE	30	**0.040**	0.068	0.080	**0.90**	**0.92**
	50	**0.040**	0.055	0.068	**0.87**	**0.88**
	100	**0.043**	0.039	0.058	**0.77**	**0.78**

True parameter value is 0.477. The three estimators are Eff (locally efficient estimator), WCLS (weighted and centered least squares), GEE (log linear GEE). SD: standard deviation. RMSE: root mean squared error. CP: 95% confidence interval coverage probability, before (unadj) and after (adj) small sample correction. Boldface indicates whether Bias or CP are significantly different, at the 5% level, from 0 or 0.95, respectively.

Simulation study 2. Here we focus on the treatment effect modification by setting $S_t = X_t$. With such choice of S_t, the analysis model of the treatment effect is correct for Eff, WCLS and BRM, because the generative model implies that

$$\log \frac{E(Y_{t+1} \mid H_t, A_t = 1)}{E(Y_{t+1} \mid H_t, A_t = 0)} = \log \frac{E\{E(Y_{t+1} \mid H_t, A_t = 1) \mid S_t\}}{E\{E(Y_{t+1} \mid H_t, A_t = 0) \mid S_t\}} = \beta_0 + \beta_1 S_t.$$

Therefore, we expect both estimators to be consistent for $\beta_0 = 0.1$ and $\beta_1 = 0.3$. The working model for $E(Y_{t+1} \mid H_t, A_t = 0)$ we use is again $\exp(\alpha_0 + \alpha_1 X_t)$, which is misspecified. Regarding the relative efficiency between Eff and WCLS, although Eff only achieves the semiparametric efficiency bound when the working model for $E(Y_{t+1} \mid H_t, A_t = 0)$ is correct, we may still expect it to be more efficient than WCLS here. The log-linear GEE estimator is expected to be inconsistent for the same reason as in the first simulation. These conjectures are supported by the following table.

Table 2: Comparison of the four estimators of the treatment effect modification ($S_t = X_t$), when the treatment effect conditional on the full history is correctly specified.

		β_0					β_1				
Estimator	Sample size	Bias	RMSE	SD	CP (unadj)	CP (adj)	Bias	RMSE	SD	CP (unadj)	CP (adj)
Eff	30	0.00	0.19	0.19	0.95	0.96	0.00	0.12	0.12	**0.92**	**0.93**
	50	0.00	0.14	0.14	0.96	0.96	0.00	0.09	0.09	**0.93**	0.94
	100	0.00	0.11	0.11	0.95	0.95	0.00	0.07	0.07	**0.92**	**0.93**
WCLS	30	-0.01	0.21	0.21	0.95	0.96	0.01	0.14	0.14	0.94	0.95
	50	-0.01	0.16	0.16	0.95	0.96	0.01	0.10	0.10	0.96	0.96
	100	0.00	0.12	0.12	0.95	0.95	0.00	0.08	0.08	0.94	0.95
GEE	30	-0.01	0.18	0.18	0.96	0.96	0.01	0.12	0.12	0.96	0.96
	50	0.00	0.14	0.14	0.95	0.96	0.00	0.09	0.09	0.96	0.96
	100	**0.15**	0.17	0.09	**0.56**	**0.56**	**-0.12**	0.13	0.05	**0.32**	**0.33**

True parameter value is $\beta_0 = 0.1$, $\beta_1 = 0.3$. The four estimators are Eff (locally efficient estimator), WCLS (weighted and centered least squares), GEE (log linear GEE). SD: standard deviation. RMSE: root mean squared error. CP: 95% confidence interval coverage probability, before (unadj) and after (adj) small sample correction. Boldface indicates whether Bias or CP are significantly different, at the 5% level, from 0 or 0.95, respectively.

Contact Information

- Email: qiantianchen@fas.harvard.edu
- Phone: +1 (609) 216-4600