Supplementary Material for

The Micro-Randomized Trial for Developing Digital Interventions:

Experimental Design and Data Analysis Considerations

Tianchen Qian, Ashley E. Walton, Linda M. Collins, Predrag Klasnja, Stephanie T. Lanza, Inbal Nahum-Shani, Mashfiqui Rabbi, Michael A. Russell, Maureen A. Walton, Hyesun Yoo, Susan A. Murphy

Appendix A

GEE and MLM Can Be Biased When Estimating Causal Excursion Effects in MRTs

MRTs produce intensive longitudinal data (Schafer, 2006), as individuals are randomized among intervention options repeatedly during the MRT, and outcomes and covariates are assessed in tandem with randomization. Repeated measurement of the same individuals over time means that the repeated observations are likely dependent. *Generalized estimating equations* (GEE; Liang & Zeger, 1986) and *multi-level models* (MLM; Laird & Ware, 1982; Raudenbush & Bryk, 2002), the latter also known as mixed models or random effects models, have been used widely in analyzing longitudinal data. However, as we illustrate below, inappropriate application of them to MRT data may result in biased estimates of the causal excursion effects when *endogenous time-varying covariates* are included in the model. A time-varying covariate is *endogenous* if it can depend on previous outcomes or previous treatments, which commonly occurs in MRTs. For example, in analyzing the effect of activity suggestion in the subsequent 30-minute step count in HeartSteps, one may want to control for the 30-minute step count prior to each decision point to reduce noise. Because the 30-minute step count prior to a decision point can be correlated with past step counts (i.e., past outcomes), it is endogenous.

When a time-varying covariate is not endogenous, it is called *exogenous*. Examples of *exogenous time-varying covariates* include time, weather, and anything that cannot be impacted by previous treatments or previous outcomes.

Inappropriate Use of GEE and MLM Can Result in Biased Causal Excursion Effect Estimates in the Presence of Endogenous Time-Varying Covariates

Pepe & Anderson (1994) demonstrated that, in the presence of endogenous time-varying covariates, parameter estimates from GEE may be biased unless certain conditions, described below, are met. Such bias is also shown in subsequent research through simulation studies and analytic calculations (Diggle et al., 2002; Pan et al., 2000; Schildcrout & Heagerty, 2005; Tchetgen et al., 2012; Vansteelandt, 2007). For completeness we provide a brief explanation of the bias here. Consider a simplified version of the HeartSteps MRT, where there are two decision points for each individual and the feasible component options are always not restricted. Suppose the observed data for individual *i* is $(X_{i1}, A_{i1}, Y_{i2}, X_{i2}, A_{i2}, Y_{i3})$, where X_{it} denotes the 30-minute step count prior to decision point t (an endogenous time-varying covariate), A_{it} is the indicator of whether an activity suggestion is delivered at decision point t (so A_{it} has .6 probability to be 1), and $Y_{i,t+1}$ is the 30-minute step count following decision point *t*. The researcher chooses $S_{it} = X_{it}$ in equation (2): they want to assess whether the effect of the activity suggestion is moderated by the prior 30-minute step count. The researcher may then choose to impose the following linear model on the mean of the proximal outcome given the treatment and the covariate at decision point t:

$$E\left(Y_{i,t+1} \middle| A_{it}, X_{it}\right) = \alpha_0 + \alpha_1 X_{it} + A_{it}(\beta_0 + \beta_1 X_{it}), \quad (1)$$

and use GEE to estimate the coefficients $\alpha_0, \alpha_1, \beta_0, \beta_1$.[1] Often a non-independent working correlation structure is used in GEE, aiming for efficiency gain (i.e., smaller standard error of the estimated coefficients compared to GEE with working independence correlation structure).

It is well known that GEE produces consistent estimates regardless of the choice of the working correlation structure, as long as equation (1) holds; however, this is only true when all covariates are exogenous. In this above example with two decision points, Pepe & Anderson (1994) demonstrated that to guarantee the consistency of the GEE estimates, one of the following conditions needs to hold:

(i) $E\left(Y_{i,t+1}\middle|A_{it}, X_{it}\right) = E\left(Y_{i,t+1}\middle|A_{i1}, X_{i1}, A_{i2}, X_{i2}\right)$ for $t = 1,2$; or

(ii) a working independence correlation structure is used.

Condition (i) is usually violated when X_{it} is endogenous: In this particular example, X_{i2} can be correlated with Y_{i2}, so that $E(Y_{i2}|A_{i1}, X_{i1}) \neq E(Y_{i2}|A_{i1}, X_{i1}, A_{i2}, X_{i2})$. This means that unless the independent working correlation structure is used, GEE can produce biased estimates even if equation (7) holds.

The same bias can occur when MLM is used instead of GEE. In general, for each MLM there is a corresponding GEE with a non-independent correlation structure that produces the same estimated coefficients. For example, an MLM resembling equation (7) is $Y_{i,t+1} = \alpha_0 + \alpha_1 X_{it} + A_{it}(\beta_0 + \beta_1 X_{it}) + u_i + \epsilon_{it}$, where $u_i \sim \text{Normal}(0, \sigma_u^2)$ is a random intercept and $\epsilon_{it} \sim \text{Normal}(0, \sigma_\epsilon^2)$ is the error term. This corresponds to a GEE with compound symmetric (also

[1] In this particular example in which the randomization probability is constant, the β_0, β_1 in the proximal treatment effect term in (7) equals the β_0', β_1' in $E[E(Y_{t+1} \mid A_t = 1, H_t) - E(Y_{t+1} \mid A_t = 0, H_t)|S_t] = \beta_0' + \beta_1' S_t$. Therefore, if one can obtain consistent estimates for β_0, β_1, one obtains consistent estimate for the proximal treatment effect defined in (2). In general, however, when the randomization probability can depend on H_t, the β_0, β_1 in (7) no longer equals the β in (2) due to the marginalization over S_t. This is another reason, in addition to the reason that will be presented in the next paragraph of the paper, why inappropriate use of GEE results in biased proximal treatment effect estimates.

called exchangeable) working correlation structure. Given this equivalency, MLM can produce biased estimates if the covariate X_{it} is endogenous.

A Few Scenarios Where GEE or MLM Provides Consistent Causal Excursion Effect Estimates from MRT Data

GEE builds upon a marginal mean model (i.e., the relationship between the mean of the proximal outcome, the covariates, and the treatment assignments, such as (7)). If no endogenous time-varying covariates are included in the model, the feasible component options are always not restricted, and the randomization probability is constant, GEE with any working correlation structure gives consistent estimates as long as the marginal mean model is correct. If there are endogenous time-varying covariates in the model, the feasible component options are always not restricted, and the randomization probability is constant, GEE with independent working correlation structure still gives consistent estimates as long as the marginal mean model is correct, but GEE with other working correlation structure does not.

Because an MLM always corresponds to a GEE with some non-independent working correlation structure, MLM provides consistent causal excursion effect estimates if no endogenous time-varying covariates are included in the model, the feasible component options are always not restricted, and the randomization probability is constant. However, although the estimated coefficients from an MLM will generally be biased for the causal excursion effect when there are endogenous time-varying covariates, those estimated coefficients can have a different, individual-specific interpretation under a rather strong assumption. As shown in Qian, Klasnja, & Murphy (2020), if the endogenous time-varying covariates can be safely assumed to only depend on the random effect through the observed previous outcomes and previous covariates, then the fitted results from standard linear mixed models can be interpreted as a

causal effect that is conditional on the random effect (i.e., individual-specific rather than population-average) and conditional on the entire history H_t (rather than conditional only on S_t). An example where this strong assumption holds is when the endogenous time-varying covariates are previous proximal outcomes (e.g., the endogenous time-varying covariate at decision point t is the proximal outcome at decision point $t-1$).

A Mathematical Demonstration of the Bias from Inappropriate Application of GEE When There are Endogenous Time-Varying Covariates

For clarity we consider the case where each participant is in the MRT for two decision points. The data for the *i*-th participant is $(X_{i1}, A_{i1}, Y_{i2}, X_{i2}, A_{i2}, Y_{i3})$, where X_{it} is the covariate, A_{it} is the treatment assignment, and Y_{it+1} is the continuous outcome. The covariate X_{it} is endogenous time-varying, in the sense that it can depend on previous treatment and previous outcome.

The model on the marginal mean of Y_{t+1} is $E(Y_{t+1}|A_t, X_t) = \alpha_0 + \alpha_1 X_t + A_t(\beta_0 + \beta_1 X_t)$. The corresponding GEE solves the following estimating equation:

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \begin{bmatrix} 1 & 1 \\ X_{i1} & X_{i2} \\ A_{i1} & A_{i2} \\ A_{i1}X_{i1} & A_{i2}X_{i2} \end{bmatrix} V^{-1} \begin{bmatrix} Y_{i2} - \alpha_0 - \alpha_1 X_{i1} - A_{i1}(\beta_0 + \beta_1 X_{i1}) \\ Y_{i3} - \alpha_0 - \alpha_1 X_{i2} - A_{i2}(\beta_0 + \beta_1 X_{i2}) \end{bmatrix} = 0. \qquad (2)$$

Here, n denotes the number of participants, and V is a 2×2 working covariance matrix. Examples of V include the following:

- Working independence: $V = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$
- Compound symmetry: $V = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$
- Autoregressive (in the special case of two decision points, autoregressive is the same as compound symmetry): $V = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$.

In this setting, the result in Pepe & Anderson (1994) implies that GEE is guaranteed to produce consistent $\alpha_0, \alpha_1, \beta_0, \beta_1$ if either

(i) $E(Y_{t+1}|A_t, X_t) = E(Y_{t+1}|A_1, X_1, A_2, X_2)$ for $t = 1,2$, or

(ii) a working independence correlation structure is used,

and they provided simulation results to show that GEE can produce biased estimates when neither condition holds. In the following, we rephrase the intuitive argument given in Pepe and Anderson (1994) in this particular setting to show why GEE can be biased if neither condition holds.

We write $V^{-1} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$ and write the residual $r_{it} = Y_{it+1} - \alpha_0 - \alpha_1 X_{it} - A_{it}(\beta_0 + \beta_1 X_{it})$. A summand (for fixed i) in equation (8) becomes

$$\begin{bmatrix} 1 & 1 \\ X_{i1} & X_{i2} \\ A_{i1} & A_{i2} \\ A_{i1}X_{i1} & A_{i2}X_{i2} \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} r_{i1} \\ r_{i2} \end{bmatrix} = \begin{bmatrix} (w_{11} + w_{21})r_{i1} + (w_{12} + w_{22})r_{i2} \\ (w_{11}X_{i1} + w_{21}X_{i2})r_{i1} + (w_{12}X_{i1} + w_{22}X_{i2})r_{i2} \\ (w_{11}A_{i1} + w_{21}A_{i2})r_{i1} + (w_{12}A_{i1} + w_{22}A_{i2})r_{i2} \\ (w_{11}A_{i1}X_{i1} + w_{21}A_{i2}X_{i2})r_{i1} + (w_{12}A_{i1}X_{i1} + w_{22}A_{i2}X_{i2})r_{i2} \end{bmatrix}. \quad (3)$$

Because $E(Y_{t+1}|A_t, X_t) = \alpha_0 + \alpha_1 X_t + A_t(\beta_0 + \beta_1 X_t)$, we have

$$E[r_{it}] = E[X_{it}r_{it}] = E[A_{it}r_{it}] = E[A_{it}X_{it}r_{it}] = 0.$$

Therefore, all the terms with $w_{11}r_{i1}$ and $w_{22}r_{i2}$ (such as $w_{11}r_{i1}X_{i1}$; i.e., terms that are multiplied with the diagonal elements of V^{-1}) in (9) have expectation zero, and what is left are the terms with $w_{21}r_{i1}$ and $w_{12}r_{i2}$ (i.e., terms that are multiplied with the off-diagonal elements of V^{-1}). In other words, the expectation of (3) equals

$$\begin{bmatrix} 0 \\ w_{21}X_{i2}r_{i1} + w_{12}X_{i1}r_{i2} \\ w_{21}A_{i2}r_{i1} + w_{12}A_{i1}r_{i2} \\ w_{11}A_{i2}X_{i2}r_{i1} + w_{12}A_{i1}X_{i1}r_{i2} \end{bmatrix}. \quad (4)$$

Mathematical theory for GEE tells us that GEE outputs consistent $\alpha_0, \alpha_1, \beta_0, \beta_1$ when (9) has expectation zero; i.e., when (10) equals zero.

If condition (i) holds, we have $E[X_{i2}r_{i1}] = E[A_{i2}r_{i1}] = E[A_{i2}X_{i2}r_{i1}] = 0$, and similarly $E[X_{i1}r_{i2}] = E[A_{i1}r_{i2}] = E[A_{i1}X_{i1}r_{i2}] = 0$. Therefore, (10) equals 0 with any choice of V^{-1}, and GEE estimators are consistent.

If condition (ii) holds, we have $w_{21} = w_{12} = 0$. Hence (10) equals 0 and GEE estimators are consistent.

When neither condition holds, it's likely that (10) does not equal zero. For example, suppose $X_{i2} = Y_{i2}$. Then the term $X_{i2}r_{i1}$ equals

$$Y_{i2}\{Y_{i2} - \alpha_0 + \alpha_1 X_{i1} + A_{i1}(\beta_0 + \beta_1 X_{i1})\}, \quad (5)$$

which is the residual multiplied with the outcome itself. Because the residual and the outcome at the same time point are correlated, (11) likely does not equal zero. Therefore, (10) likely does not equal zero. This means GEE can be biased when neither conditions hold, i.e., when endogenous time-varying covariates are included and non-independent working correlation structure is used.

Appendix B

A General Form of the WCLS Estimator for the Causal Excursion Effect That Allows the Randomization Probability to Vary Over Time

We assume a linear model for the causal excursion effect: $\beta(t, s) = s^{\mathrm{T}}\beta$. Suppose $Z_t^{\mathrm{T}}\alpha$ is a working model for the conditional mean of Y_{t+1} given no treatment at decision point t and history H_t, $E(Y_{t+1}|I_t = 1, H_t)$. Note that the consistency of the estimator for β does not require

$Z_t^{\mathrm{T}}\alpha$ to be a correct model for $E(Y_{t+1}|I_t = 1, H_t)$. We use $p_t(H_t)$ to denote the randomization probability at decision point t, which may possibly depend on H_t.

The WCLS estimator for β is calculated as follows. Suppose $(\hat{\alpha}, \hat{\beta})$ is the (α, β) value that solves the following estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} I_{it}W_{it}\left[Y_{i,t+1} - Z_{it}^{\mathrm{T}}\alpha - \{A_{it} - \tilde{p}_t(S_{it})\}S_{it}^{\mathrm{T}}\beta\right]\begin{bmatrix} Z_{it} \\ \{A_{it} - \tilde{p}_t(S_{it})\}S_{it} \end{bmatrix} = 0; \qquad (6)$$

then $\hat{\beta}$ is the WCLS estimator for β. $\tilde{p}_t(S_{it})$ is an arbitrary probability as long as it depends on H_{it} through at most S_{it} and it is bounded away from 0 and 1; i is the index for the i^{th} individual, and W_{it} is defined as

$$W_{it} = \left\{\frac{\tilde{p}_t(S_{it})}{p_t(H_{it})}\right\}^{A_{it}}\left\{\frac{1 - \tilde{p}_t(S_{it})}{1 - p_t(H_{it})}\right\}^{1-A_{it}}. \qquad (7)$$

W_{it}, the ratio of two probabilities, serves as a change of probability: It makes it as if the treatment A_{it} is randomized with probability $\tilde{p}_t(S_{it})$. It is used to marginalize the causal excursion effect over variables in H_{it} but not in S_{it}. As long as $\tilde{p}_t(S_t)$ depends on H_{it} through at most S_{it} and it is bounded away from 0 and 1, the particular choice of $\tilde{p}_t(S_t)$ doesn't affect the consistency of $\hat{\beta}$. For instance, one can set it to be 0.5 (or any constant between 0 and 1) for all individuals and all decision points, or set it to be the predicted value from a logistic regression fit of $A_t \sim S_t$. If the true randomization probability $p_t(H_t)$ depends at most on S_t, then one can also set $\tilde{p}_t(S_t)$ to be equal to the true randomization probability, in which case (12) is mathematically equivalent to (5). $\tilde{p}_t(S_t)$ can impact the standard error of $\hat{\beta}$. In addition, when the causal excursion effect model $\beta(t, s) = s^{\mathrm{T}}\beta$ is misspecified, $\tilde{p}_t(S_t)$ impacts the limit of $\hat{\beta}$. See the Appendix of Boruvka et al. (2018) for more technical details on how the limit of $\hat{\beta}$ is impacted by $\tilde{p}_t(S_t)$ in this case.

Now we present a way to obtain the general WCLS estimator for time-varying randomization probability through standard statistical software that implements GEE. Suppose the assumed causal excursion effect model is (6) and the working model for $E(Y_{t+1}|I_t = 1, H_t)$ is $Z_t^{\mathrm{T}}\alpha$; then the WCLS estimator $\hat{\beta}$ and its standard error can be obtained by (i) incorporating $I_t W_t$ as the "prior weights", (ii) choosing a working independence correlation structure, and (iii) fitting GEE with dependent variable Y_{t+1} and independent variables Z_t and $\big(A_t - \tilde{p}_t(S_t)\big)S_t$. Then the estimated coefficient for $\big(A_t - \tilde{p}_t(S_t)\big)S_t$ is the WCLS estimate $\hat{\beta}$.

Standard Error Formula for WCLS.

Below we provide the formula for the standard error of the WCLS estimator $\hat{\beta}$. For $(\hat{\alpha}, \hat{\beta})$ that solves estimating equation (12), variance can be estimated by

$$\widehat{\mathrm{Var}}\left(\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}\right) = \frac{1}{n} M_n^{-1} \Sigma_n (M_n^{-1})^{\mathrm{T}},$$

where

$$M_n = -\mathbb{P}_n \sum_{t=1}^{T} I_t W_t \begin{bmatrix} Z_t Z_t^{\mathrm{T}} & \{A_t - \tilde{p}_t(S_t)\} Z_t S_t^{\mathrm{T}} \\ \{A_t - \tilde{p}_t(S_t)\} S_t Z_t^{\mathrm{T}} & \{A_t - \tilde{p}_t(S_t)\}^2 S_t S_t^{\mathrm{T}} \end{bmatrix}$$

and

$$\Sigma_n = \mathbb{P}_n \sum_{t=1}^{T} \Big\{ Y_{t+1} - Z_t^{\mathrm{T}}\alpha - \big(A_t - \tilde{p}_t(S_t)\big) S_t^{\mathrm{T}}\beta \Big\}^2 I_t W_t \begin{bmatrix} Z_t Z_t^{\mathrm{T}} & \{A_t - \tilde{p}_t(S_t)\} Z_t S_t^{\mathrm{T}} \\ \{A_t - \tilde{p}_t(S_t)\} S_t Z_t^{\mathrm{T}} & \{A_t - \tilde{p}_t(S_t)\}^2 S_t S_t^{\mathrm{T}} \end{bmatrix}.$$

Here, $\mathbb{P}_n$ denotes sample average over *n* individuals. The standard error formula can be modified for the setting in which the randomization probability is constant over time (i.e., the setting in the main paper) by letting $\tilde{p}_t(S_t) = p$ and $W_t = 1$.

Appendix C

We conduct a simulation study to illustrate the claim that including variables that are correlated with Y_{t+1} in Z_t may reduce the variance of the WCLS estimator. The generative model mimics features of the HeartSteps data and is set up as follows. For simplicity we assume that the feasible component options are always not restricted. At decision point t, the covariate X_t is drawn from the empirical distribution of the log-transformed 30-minute step count preceding a decision point in the HeartSteps data. For simplicity X_t is generated independently of previous outcomes and treatments. The treatment A_t is generated from a Bernoulli distribution with .6 success probability; this mimics the .6 randomization probability of activity suggestions in HeartSteps. The proximal outcome Y_{t+1} is generated from a Gaussian distribution with mean

$$1.6085 + 0.4037 \times X_t + 0.0655 \times Y_t + 0.1229 \times (A_t - 0.6)$$

and standard deviation 2.716. The coefficients in the above display are the estimated coefficients from a WCLS fit on the HeartSteps data with the same control variables $(1, X_t, Y_t)$ and constant treatment effect model. The standard deviation is the empirical standard deviation of the residual in Y_{t+1} from the above WCLS fit. As in the HeartSteps data set, for each simulated trial we generate 37 individuals, each with 210 decision points.

For each data set generated from the above generative model, we consider four WCLS fits for the true treatment effect 0.1229 and compare their performance. All four WCLS assume the constant treatment effect model, and they differ in the choice of the working model. The first WCLS fit (WCLS-1) includes control variables $(1, X_t, Y_t)$; the second WCLS fit (WCLS-2) includes control variables $(1, X_t)$; the third WCLS fit (WCLS-3) includes control variables $(1, Y_t)$; and the fourth WCLS fit (WCLS-4) includes only the intercept. The bias, standard deviation (SD), and coverage probability (CP) of 95% confidence interval are listed in Supplementary Table 1. All four WCLS estimators are consistent with nominal confidence

interval coverage because their assumed constant treatment effect model holds under this generative model. (This again illustrates that the consistency of the WCLS estimator does not require the control part of the model to be correct.) On the other hand, the choice of working model affects the efficiency of the estimator. In particular, WCLS-1 and WCLS-2 have smaller standard errors than WCLS-3 and WCLS-4 because the former two include X_t, a covariate that is highly correlated with the proximal outcome Y_{t+1}.

Appendix D

To assess sensitivity of the result to potential non-linearity in Question 2 of Section "Analysis Using Data from HeartSteps MRT," we fit a local 2-degree polynomial regression with smoothing span 2/3 and tricubic weighting to estimate the causal excursion effect over time (the default setting for many local regression software, such as the lowess function in R (R Core Team, 2019)). The estimated effect from local regression is presented in Supplementary Figure 1 (black curve). Comparing this estimated effect with the estimated effect based on the linear model (blue curve in Figure 1, with blue shaded area being the pointwise 95% confidence interval), we see that the two estimates are relatively close to each other, indicating that the linear model fits well.

Appendix E

A Synthetic Data Set Mimicking HeartSteps

The HeartSteps data set that was used in the illustrative data analysis is not publicly available. To allow readers to try out the R code which implements the WCLS estimator, we included a synthetic data set which was generated by mimicking some features of the HeartSteps

data. These features include the number of individuals, the number of decision points, the fact that treatments are randomized 5 times a day, and the quantitative relationship among the proximal outcome variable (30-minute step count following each decision point), the treatment indicator, and some time-varying covariates including 30-minute step count preceding each decision point and location of the individual.

The synthetic data set and the example R code to analyze it can be downloaded at https://github.com/tqian/paper_mrt_PsychMethods/tree/main/synthetic_data. Below is a brief description of the files in the github repository folder:

- synthetic_data_37subject_210time.csv: A synthetic data set in csv format.
- analysis_synthetic_data.R: R code to analyze the synthetic data, which is similar to the R code used to analyze the original HeartSteps data set (https://github.com/tqian/paper_mrt_PsychMethods/blob/main/analysis.R)
- analysis_synthetic_data_with_result.pdf: A PDF file that describes the variables in the synthetic data set, walks through the steps in the analysis code, and shows the R output.
- xgeepack.R and estimate.R: Code that implements the WCLS estimator. (They will be loaded by analysis_synthetic_data.R)

Appendix F

Below we provide a SAS code example that utilizes PROC GEE to obtain the WCLS estimator for the synthetic data set (description of which is provided in Appendix E).

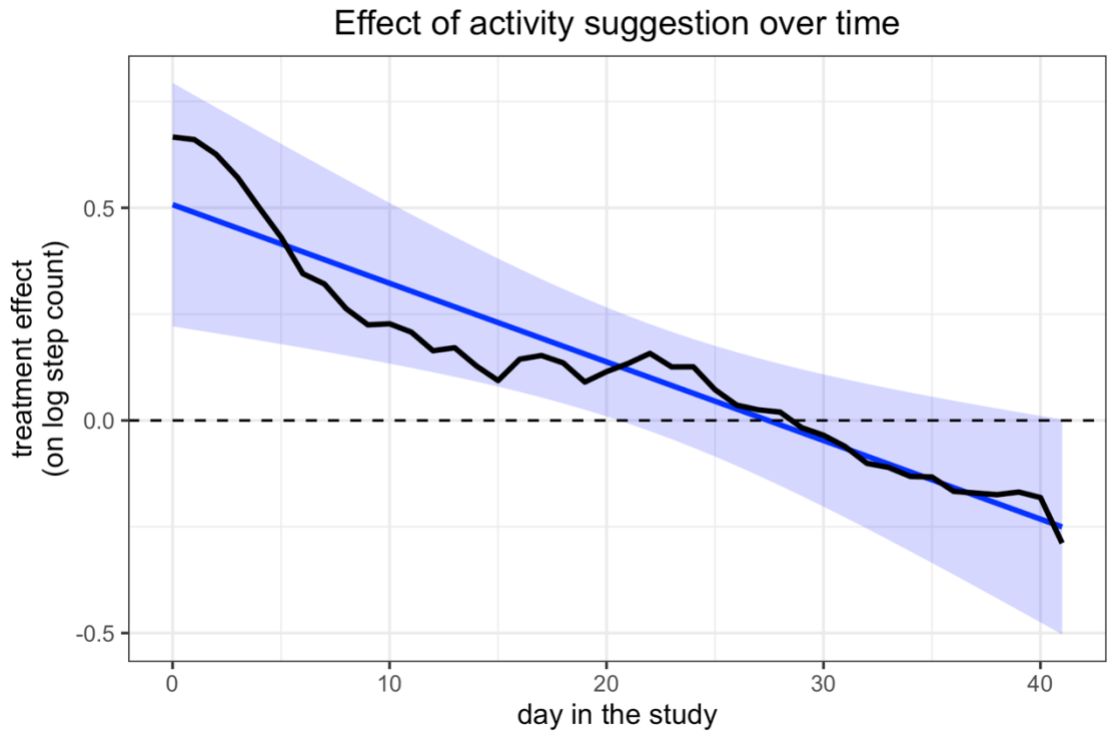
```
* import the synthetic_data_37subject_210time_SAS.csv data set;
* you may need to modify the directory;

FILENAME REFFILE '/home/synthetic_data_37subject_210time_SAS.csv';
```

```
PROC IMPORT DATAFILE=REFFILE
     DBMS=CSV
     OUT=example_data;
     GETNAMES=YES;
RUN;

* add variable: centered treatment indicator;
* (suppose the randomization probability is 0.6 in this case);
data example_data;
     set example_data;
     send_ctr = send - 0.6;
run;

* Marginal effect;
proc gee data = example_data;
     class userid;
     model jbsteps30_log = jbsteps30pre_log send_ctr / dist = normal;
     repeated subject = userid / corr = ind;
     weight avail;
run;

* Marginal effect with more control variables in WCLS fit;
proc gee data = example_data;
     class userid;
     model jbsteps30_log = jbsteps30pre_log jbsteps30_log_lag1 send_ctr / dist = normal;
     repeated subject = userid / corr = ind;
     weight avail;
run;

* Effect change over time;
proc gee data = example_data;
     class userid;
     model jbsteps30_log = jbsteps30pre_log study_day_nogap send_ctr
               send_ctr * study_day_nogap / dist = normal;
     repeated subject = userid / corr = ind;
```

```
     weight avail;
run;

* Effect moderation by outcome at previous time point;
proc gee data = example_data;
     class userid;
     model jbsteps30_log = jbsteps30pre_log jbsteps30_log_lag1
           location_homework send_ctr send_ctr * jbsteps30_log_lag1 / dist = normal;
     repeated subject = userid / corr = ind;
     weight avail;
run;
```

Supplementary Figure 1.

Estimated effect of activity suggestion on proximal outcome as a linear function of days in study, and corresponding 95% pointwise confidence intervals



Note. Figure for the sensitivity analysis in Appendix D regarding "Question 2: Does the effect of the activity suggestions change with each additional day in the study?" in section "Analysis Using Data from HeartSteps MRT." The black curve is the estimated effect using local 2-degree polynomial regression with smoothing span 2/3 and tricubic weighting. The blue line represents the estimated causal excursion effect across the 42 study days, assuming a linear time trend, and the shaded blue area is the pointwise 95% confidence interval.

Supplementary Table 1.

Simulation results for Appendix C: Efficiency gain from including prognostic variable in the working model

	bias	standard deviation	95% coverage probability
WCLS-1	-0.001	0.067	96.7%
WCLS-2	-0.001	0.067	96.9%
WCLS-3	-0.001	0.074	95.8%
WCLS-4	-0.001	0.074	95.7%

Note. All four WCLS assumes the constant treatment effect model, and they differ in the choice of the working model. WCLS-1 includes control variables $(1, X_t, Y_t)$; WCLS-2 includes control variables $(1, X_t)$; WCLS-3 includes control variables $(1, Y_t)$; WCLS-4 includes only the intercept.