JOOL MRT Analysis

Susan A. Murphy and Daniel Almirall

February 23, 2018

This supplementary appendix describes the analysis method used to address the primary and secondary aims of the JOOL micro-randomized quality improvement trial. The method is based on a generalization of Robins' multiplicative structural nested log-linear model (Robins, 1997) for use with data arising from a micro-randomized trial. It is an extension of the approaches described in Boruvka, Almirall, Witkiewitz, and Murphy (2017) and Dempsey, Liao, Kumar, and Murphy (2017) with the use of a log-link function to accommodate the primary outcome in this study, which is binary.

1 Notation and Study Set-up

- Decision time is denoted by the pair (d, t): $d = 1, \ldots, D$ corresponding to the $D = 89$ decision days in study; $t = 1, \ldots, T$ corresponding to the $T = 6$ times per day. There are a total of $89 * 6 = 534$ decision times in the study.
- Availability: $I_{d,t} = 1$ if user is available on day d, time t and is 0 otherwise. Availability is defined in the main body of the manuscript.
- Time of day randomization: $B_{d,t} = 1$ if the time of day randomization indicates that treatment will be randomized at time t and is 0 otherwise.
- Note that because users can only have one time per day at which they might be randomized, at any times $j > t$ after $B_{d,t} = 1$, we have $I_{d,j} = 0$.
- Push notification (treatment) randomization: $A_{d,t} = 1$ if randomized to a push and is 0 otherwise. We use the treatment randomization only if $B_{d,t} = 1$
- History (covariate data): $H_{d,t}$ is all of the data we have observed on the user, up to and including day d, time t, including $I_{d,t}$ but excluding randomizations, $B_{d,t}, A_{d,t}$.
- Note that a user is provided a push on (d, t) only if $I_{d,t} = 1$ and $B_{d,t}A_{d,t} = 1$.
- Response (Outcome): $Y_{d+1,t}$ this is whether $(= 1)$ or not $(= 0)$ the user engaged with the JOOL app by charting in the next 24 hours; that is, whether the user engaged after time t on day d but prior to time t on day $d + 1$. Note that $Y_{d+1,t}$ is binary.
- Covariates: The analysis for each aim will include two sets of covariates—$X_{d,t}$ and $Z_{d,t}$—both of which are features of the history $H_{d,t}$. Note that $X_{d,t} \subseteq Z_{d,t}$.
- Moderators: The covariates in $X_{d,t}$ denote the candidate moderators of interest. These are covariates thought to influence the effect of a push. These are described in the main body of the manuscript (see Table 1) and are repeated here for convenience:
  - The Primary Aim analysis uses an intercept only model $X_{d,t} = (1)$, i.e., no candidate moderators are examined in this analysis.

- The Secondary Aim 1 analysis uses $X_{d,t} = (1, \texttt{weekinstudy})$, where `weekinstudy` is a covariate that ranges from 0 to 12, corresponding to the week in the study (which ranges from 1 to 13, respectively).
- The Secondary Aim 2 analysis uses $X_{d,t} = (1, \texttt{weekend})$, where `weekend` is a binary covariate denoting weekend days (= 1 on Saturdays or Sundays) vs weekdays (= 0 on Mondays, Tuesdays, Wednesdays, Thursdays or Fridays).
- The Exploratory Aim analysis has two parts:
  * Part(a)
    $X_{d,t} = (\texttt{8:30AM, 12:30PM, 5:30PM, 6:30PM, 7:30PM, 8:30PM})$,
    where each of these is a binary covariate denoting the corresponding time of day. Note that $\texttt{8:30AM}+\texttt{12:30PM}+\texttt{5:30PM}+\texttt{6:30PM}+\texttt{7:30PM}+\texttt{8:30PM} = 1$, so this model does not include an intercept.
  * Part(b)
    $X_{d,t} = (\texttt{weekend} \times (\texttt{12:30PM, 5:30PM, 6:30PM, 7:30PM, 8:30PM}),$
    $\texttt{weekday} \times (\texttt{8:30AM, 12:30PM, 5:30PM, 6:30PM, 7:30PM, 8:30PM}))$,
    where `weekday` is binary covariate denoting weekdays (= 1 on Mondays, Tuesdays, Wednesdays, Thursdays or Fridays) vs weekend days (= 0 on Saturdays or Sundays). Note $\texttt{weekday} + \texttt{weekend} = 1$, so this model also does not include an intercept.

- $Z_{d,t}$ is a vector of control covariates used to reduce noise. All analyses (all aims) used $Z_{d,t} = (1, \texttt{weekinstudy, dayssincechart, pushedindicator, pushsuccessratio, hascharted10})$ where `dayssincechart` is the number of days since the user last charted, `pushedindicator` is an indicator for whether a push has been sent in the past (once a push is sent this indicator has a value of 1.0 for all remaining time points, otherwise it has a value of 0), `pushsuccessratio` is the total number of charts within 24 hours of receiving a push notification any time in the past divided by the total number of push notification sent any time in the past (note that `pushsuccessratio` is nested within `pushedindicator`; i.e., if no push notifications are sent in the past, the value is zero), and `hascharted10` is an indicator whether the user has charted at least 10 times. Note that `hascharted10` is equivalent to whether the user has unlocked the "What makes you tick feature?" in the JOOL app. In order to ensure that $X_{d,t} \subseteq Z_{d,t}$: In the Secondary Aim 2 analysis, $Z_{d,t}$ also includes `weekend`; and in the two Exploratory Aim analyses, $Z_{d,t}$ also includes the indicators for the time of day and the time of day indicators crossed with `weekend`, respectively.

- There were $n = 1255$ users in the JOOL micro-randomized quality-improvement trial.

2 The Moderated Causal Effects

At every push vs no push decision point (d, t), there are two potential outcomes for the response: The first one is whether the user charted in the next 24 hours given a push now and no other pushes over the next 24 hours. The second one is whether the user charted in the next 24 hours given no push now and no push over the next 24 hours. These potential outcomes are denoted as follows. At 8:30AM, the two potential outcomes $Y_{d+1,1}$ are $Y_{d+1,1}(\overline{B_{d-1,6}A_{d-1,6}}, 1, \bar{0}_5)$ and $Y_{d+1,1}(\overline{B_{d-1,6}A_{d-1,6}}, \bar{0}_6)$. For the remainder of the day (i.e., $j \neq 1$) the two potential outcomes $Y_{d+1,j}$ are $Y_{d+1,j}(\overline{B_{d,j-1}A_{d,j-1}}, 1, \bar{0}_5)$ and $Y_{d+1,j}(\overline{B_{d,j-1}A_{d,j-1}}, 0, \bar{0}_5)$. Note that $\bar{0}_5 = (0,0,0,0,0)$. We are interested in

$$\frac{E[Y_{d+1,t}(\overline{B_{d,t-1}A_{d,t-1}}, 1, \bar{0}_5)|I_{d,t} = 1, X_{d,t}]}{E[Y_{d+1,t}(\overline{B_{d,t-1}A_{d,t-1}}, \bar{0}_6)|I_{d,t} = 1, X_{d,t}]} \tag{1}$$

which is the relative probability of charting given a push now and no other pushes over the next 24 hours versus the probability of charting given no push now nor over the next 24 hours, conditional on $X_{d,t}$ and being available ($I_{d,t} = 1$). This causal effect is on the "risk" ratio scale; a value greater than 1 indicates the push is effective given $X_{d,t}$.

In the manuscript, we model the causal risk ratio in display (1) using $e^{X_{d,t}^T \beta}$, that is, using a log-linear model, where β is unknown.

3 Randomization Probabilities

The JOOL MRT push vs no push randomization probabilities (if available, i.e., $I_{(d,t)} = 1$) are given as follows:

$$\begin{aligned}
\pi_t(H_{d,t}) &= P[B_{d,t}A_{d,t} = 1|H_{d,t}]I_{d,t} \\
&= \left(\prod_{j=1}^{t-1}(1 - I_{d,j} + I_{d,j}(1 - B_{d,j}))\right)\left(\frac{1}{6-t+1}\right)\left(\frac{1}{2}\right) I_{d,t} \\
&= \left(\frac{1}{6-t+1}\right)\left(\frac{1}{2}\right) I_{d,t}.
\end{aligned}$$

The last equality holds because a necessary (not sufficient) condition for $I_{d,t} = 1$ is if at all times $j < t$ on the same day d, $B_{d,j} = 0$.

4 Estimation

Here, we describe our approach to estimating the unknown parameters β.
The method uses a weighted regression approach. The weights $W_{d,t}$, which are a function of the randomization probabilities, are defined as follows:

$$W_{d,t} = \underbrace{\left(\prod_{j=t+1}^{6}\frac{1 - B_{d,j}A_{d,j}}{1 - \pi_j(H_{d,j})}\right)\left(\prod_{j=1}^{t-1}\frac{1 - B_{d+1,j}A_{d+1,j}}{1 - \pi_j(H_{d+1,j})}\right)}_{W_{1,d,t}}\underbrace{\left(\frac{p_{d,t}}{\pi_t(H_{d,t})}\right)^{B_{d,t}A_{d,t}}\left(\frac{1 - p_{d,t}}{1 - \pi_t(H_{d,t})}\right)^{1-B_{d,t}A_{d,t}}}_{W_{2,d,t}}.$$

By design, all users who receive a push at decision point (d,t) will not receive another push in the next 24 hours. However, this is not true of users who do not receive a push at (d,t); these users may or may not receive a push over the next 24 hours. Intuitively, the numerator of the weights $W_{1,d,t}$ subset the data so that only users who do not receive a push in the next 24 hours are used in estimation (which is consistent with the causal estimand in display (1)). The denominator of $W_{1,d,t}$ is used to account for the fact that such users may be different from users who do receive a push over the next 24 hours. The weights $W_{2,d,t}$ allow us to examine causal effects that are marginal over measures in $H_{d,t}$ that are not in $X_{d,t}$. The numerator probabilities in $W_{2,d,t}$, $p_{d,t} = p_t(X_{d,t})$ can only depend on $H_{d,t}$ via $X_{d,t}$; in this analysis, we selected $p_{d,t} = .2$ as this is the average of the product of the time of day and treatment randomization probabilities (ignoring availability).

The analysis method permits the use of auxiliary covariates $Z_{d,t}$ to reduce noise. In particular, we use the working model $e^{Z_{d,t}^T\alpha}$ for $E(e^{-(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}^T\beta}Y_{d+1,t} \mid Z_{d,t}, I_{d,t}=1)$. This working model need not be correct for the estimator of β to be consistent; rather, this model is used to reduce estimation variance.

Let O denote the complete, observed study data; and let $\theta = (\alpha, \beta)$. We obtain $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ as the solution to θ in $0 = \mathbb{P}_n U(O;\theta)$ where the estimating function $U(O;\theta)$ is defined as

$$U(O;\theta) = \sum_{d=1}^{D}\sum_{t=1}^{6} I_{d,t}W_{d,t}\left(e^{-(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}^T\beta}Y_{d+1,t} - e^{Z_{d,t}^T\alpha}\right)\begin{pmatrix}(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}\\ Z_{d,t}\end{pmatrix}.$$

5 Variance-Covariance Estimation

Define Σ_n by

$$\mathbb{P}\sum_{d=1}^{D}\sum_{t=1}^{6} I_{d,t}W_{d,t}e^{-(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}^T\hat{\beta}}\left(Y_{d+1,t} - e^{Z_{d,t}^T\hat{\alpha}+(B_{d,t}A_{d,t}-p_{d,t})\hat{\beta}}\right)\begin{pmatrix}(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}\\ Z_{d,t}\end{pmatrix}$$
$$\times\sum_{d=1}^{D}\sum_{t=1}^{6} I_{d,t}W_{d,t}e^{-(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}^T\hat{\beta}}\left(Y_{d+1,t} - e^{Z_{d,t}^T\hat{\alpha}+(B_{d,t}A_{d,t}-p_{d,t})\hat{\beta}}\right)\begin{pmatrix}(B_{d,t}A_{d,t}-p_{d,t})X_{d,t},\\ Z_{d,t}\end{pmatrix}^T$$

Define M_n by

$$\mathbb{P}\sum_{d=1}^{D}\sum_{t=1}^{6} I_{d,t}W_{d,t}\begin{pmatrix}(B_{d,t}A_{d,t}-p_{d,t})^2e^{-(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}^T\hat{\beta}}Y_{d+1,t}X_{d,t}X_{d,t}^T, & (B_{d,t}A_{d,t}-p_{d,t})e^{Z_{d,t}^T\hat{\alpha}}X_{d,t}Z_{d,t}^T\\ (B_{d,t}A_{d,t}-p_{d,t})e^{-(B_{d,t}A_{d,t}-p_{d,t})X_{d,t}^T\hat{\beta}}Y_{d+1,t}Z_{d,t}X_{d,t}^T, & e^{Z_{d,t}^T\hat{\alpha}}Z_{d,t}Z_{d,t}^T\end{pmatrix}$$

Note that M_n is not symmetric. The limiting asymptotic variance of $\begin{pmatrix}\sqrt{n}(\hat{\beta}-\beta_0)\\ \sqrt{n}(\hat{\alpha}-\alpha_0)\end{pmatrix}$ is given by $V_n = M_n^{-1}\Sigma_n\left(M_n^{-1}\right)^T$.

6 Inference

For the hypothesis tests reported in the manuscript, all of which were of the form H_0 : $c^T\theta = 0$ for some linear combination of θ based on the column vector c (no multivariate tests were conducted), we used standard Wald test statistics $c^T\hat{\theta}/\sqrt{cV_nc^T}$ compared to a standard normal distribution.

Similarly, all $(1-\alpha)\times 100$ percent confidence intervals reported in the manuscript were based on the formulae $c^T\hat{\theta} \pm z_{\alpha/2} \times \sqrt{cV_nc^T}$, where $z_{alpha/2}$ is the $(1-\alpha)\times 100$ percentile of the standard normal distribution.

References

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

Dempsey, W., Liao, P., Kumar, S., & Murphy, S. A. (2017). The stratified micro-randomized trial design: sample size considerations for testing nested causal effects of time-varying treatments. *arXiv preprint arXiv:1711.03587*.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality* (pp. 69–117). Springer.