

Power Calculation for the Mobile-Assistance for Regulating Smoking Micro-Randomized Trial

Jamie Yap

{email: jamiyap@umich.edu; website: <https://github.com/jamiyap>}

June 15, 2021

This documentation is a supplement to the manuscript titled ‘The Mobile-Assistance for Regulating Smoking (MARS) Micro-Randomized Trial Design Protocol’ (Nahum-Shani, et al., 2021), submitted for consideration to the Journal of Contemporary Clinical Trials. In this documentation, we describe the set-up and results of the power calculation for the Primary Aim and Secondary Aim of the MARS Micro-Randomized Trial. Where relevant, we direct readers to R code files.

The work in this documentation was funded by NIH grant U01CA229437 and developed at the d3lab: <https://d3lab.isr.umich.edu/>

Keywords: Causal Inference, Intensive Longitudinal Data, Adaptive Interventions, Mobile Health (mHealth), Micro-Randomized Trial, Power Calculation

1 Study Set-Up

- **Decision Point:** Each day of the 10-day Micro-Randomized Trial (MRT) is divided into six consecutive non-overlapping blocks of time; each block is approximately 2 hours and 20 minutes long. For the purpose of calculating power, we assume that micro-randomization occurs at the beginning of each block. Further, we assume that the end of the sixth block of the current day will be followed immediately (i.e., no time gap) by the first block of the next day. Together, these assumptions imply that the decision points, which we index by $t = 1, 2, 3, \dots, 60$, will be spaced approximately 2 hours and 20 minutes apart.
- **Occurrence of Micro-Randomizations:** Due to practical and ethical considerations, micro-randomizations will not occur if driving is detected or if the participant activates ‘sleep-mode’ on the smartphone. $I_t = 1$ if micro-randomization at t occurs, i.e., $I_t = 1$ if the participant is not driving and does not have ‘sleep-mode’ activated their mobile phone at t ; $I_t = 0$, otherwise.
- **Primary Outcome:** The primary proximal outcome of the MRT is engagement in self-regulatory strategies (self-reported via a survey delivered on the participant’s mobile phone) approximately 1 hour following micro-randomization. The primary proximal outcome is binary where $Y_{t, 60 \text{ mins}} = 1$ denotes that the participant engaged in self-regulatory strategies within the next hour following decision point t ; $Y_{t, 60 \text{ mins}} = 0$, otherwise.
- **Interventions:** At each decision point t , a participant is randomized to be offered **no prompt** with probability 0.5, a prompt to engage in a relatively low effort self-regulatory strategy (which we refer to here as a **low effort prompt**) with probability 0.25, and a prompt to engage in a relatively high effort self-regulatory strategy (which we refer to here as a **high effort prompt**) with probability 0.25.
 - $A_t = 1$ if participant is randomized to **any kind of prompt** at t ; $A_t = 0$ otherwise (i.e., participant is randomized to be offered **no prompt** at t)

- $L_t = 1$ if participant is randomized to a **low effort prompt** at t ; $L_t = 0$ otherwise (i.e., participant is randomized to either **no prompt** or a **high effort prompt**)
- $H_t = 1$ if participant is randomized to a **high effort prompt** at t ; $H_t = 0$ otherwise (i.e., participant is randomized to either **no prompt** or a **low effort prompt**)

In Figure 1, we illustrate the MRT data for the 1^{st} , 2^{nd} , 3^{rd} , ..., 60^{th} decision points.

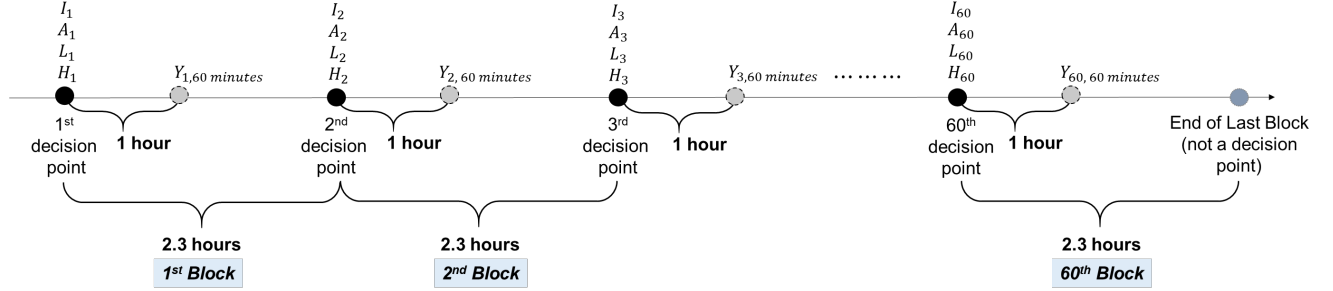


Figure 1: Illustration of the MRT Data.

2 The Causal Effects

The causal effect of interest for the Primary Aim is

$$\frac{Pr \{Y_{t, 60 \text{ mins}} = 1 | A_t = 1, I_t = 1\}}{Pr \{Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1\}}$$

The above risk ratio is the probability (i.e., ‘risk’) of engaging in self-regulatory strategies in the next hour following decision point t , when the participant was randomized to **any kind of prompt** at t , relative to (i.e., divided by) the probability of engaging in self-regulatory strategies in the next hour following decision point t , when the participant was randomized to **no prompt** at decision point t

The causal effect of interest for the Secondary Aim is

$$\frac{Pr \{Y_{t, 60 \text{ mins}} = 1 | L_t = 1, H_t = 0, I_t = 1\}}{Pr \{Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 1, I_t = 1\}}$$

The above risk ratio is the probability (i.e., ‘risk’) of engaging in self-regulatory strategies in the next hour following decision point t , when the participant was randomized to **low effort** prompt at decision point t , relative to (i.e., divided by) the probability of engaging in self-regulatory strategies in the next hour following decision point t , when the participant was randomized to **high effort** prompt at decision point t

3 Models for Causal Relative Risk

We model the causal relative risk for the Primary Aim as

$$\log (Pr \{Y_{t, 60 \text{ mins}} = 1 | A_t, I_t = 1\}) = \beta_0 + \beta_1 A_t$$

and for the Secondary Aim as

$$\log (Pr \{Y_{t, 60 \text{ mins}} = 1|L_t, H_t, I_t = 1\}) = \theta_0 + \theta_1 L_t + \theta_2 H_t$$

In other words, the Primary Aim risk ratio based on the model can be expressed as:

$$\frac{Pr \{Y_{t, 60 \text{ mins}} = 1|A_t = 1, I_t = 1\}}{Pr \{Y_{t, 60 \text{ mins}} = 1|A_t = 0, I_t = 1\}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Similarly, the Secondary Aim risk ratio based on the model can be expressed as:

$$\frac{Pr \{Y_{t, 60 \text{ mins}} = 1|L_t = 1, H_t = 0, I_t = 1\}}{Pr \{Y_{t, 60 \text{ mins}} = 1|L_t = 0, H_t = 1, I_t = 1\}} = \frac{e^{\theta_0 + \theta_1}}{e^{\theta_0 + \theta_2}} = e^{\theta_1 - \theta_2}$$

4 Relationship between Parameters across Models

Now, we have that

$$\begin{aligned} Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 0, I_t = 1) &= e^{\beta_0} \\ Pr(Y_{t, 60 \text{ mins}} = 1|L_t = 0, H_t = 0, I_t = 1) &= e^{\theta_0} \\ Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 1, I_t = 1) &= e^{\beta_0 + \beta_1} \\ Pr(Y_{t, 60 \text{ mins}} = 1|L_t = 1, H_t = 0, I_t = 1) &= e^{\theta_0 + \theta_1} \\ Pr(Y_{t, 60 \text{ mins}} = 1|L_t = 0, H_t = 1, I_t = 1) &= e^{\theta_0 + \theta_2} \end{aligned}$$

Since $Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 0, I_t = 1)$ and $Pr(Y_{t, 60 \text{ mins}} = 1|L_t = 0, H_t = 0, I_t = 1)$ are equivalent, $\beta_0 = \theta_0$. Additionally, observe that since $Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 1, I_t = 1)$ is a weighted average of $Pr(Y_{t, 60 \text{ mins}} = 1|L_t = 1, H_t = 0, I_t = 1)$ and $Pr(Y_{t, 60 \text{ mins}} = 1|L_t = 0, H_t = 1, I_t = 1)$,

$$\begin{aligned} e^{\beta_0 + \beta_1} &= Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 1, I_t = 1) \\ &= Pr(Y_{t, 60 \text{ mins}}, L_t = 1, H_t = 0|A_t = 1, I_t = 1) + Pr(Y_{t, 60 \text{ mins}}, L_t = 0, H_t = 1|A_t = 1, I_t = 1) \\ &= Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 1, I_t = 1, L_t = 1, H_t = 0) \cdot Pr(A_t = 1, I_t = 1, L_t = 1, H_t = 0) \\ &\quad + Pr(Y_{t, 60 \text{ mins}} = 1|A_t = 1, I_t = 1, L_t = 0, H_t = 1) \cdot Pr(A_t = 1, I_t = 1, L_t = 0, H_t = 1) \\ &= (e^{\theta_0 + \theta_1}) \cdot \frac{1}{2} + (e^{\theta_0 + \theta_2}) \cdot \frac{1}{2} \end{aligned}$$

Since $e^{\beta_0} = e^{\theta_0}$, the above expression can be simplified to

$$e^{\beta_1} = (e^{\theta_1}) \cdot \frac{1}{2} + (e^{\theta_2}) \cdot \frac{1}{2}$$

Therefore,

$$\begin{cases} \beta_0 &= \theta_0 \\ \beta_1 &= \log \left(\frac{1}{2}e^{\theta_1} + \frac{1}{2}e^{\theta_2} \right) \end{cases}$$

5 Null and Alternative Hypotheses

The null and alternative hypotheses for the Primary Aim and the Secondary Aim can be expressed in terms of the parameters in the models for the causal relative risks.

For the Primary Aim, the null and alternative hypotheses are

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0$$

For the Secondary Aim, the null and alternative hypotheses are

$$H_0 : \theta_1 - \theta_2 = 0 \quad \text{vs.} \quad H_1 : \theta_1 - \theta_2 \neq 0$$

6 Hypothesis Test and Power

Let type-I error rate be denoted by α and $z_{1-\frac{\alpha}{2}}$ the $1 - \frac{\alpha}{2}$ percentile of the standard normal distribution. Further, let $\hat{\beta}$ and $\hat{\theta}$ denote estimates of $\beta = (\beta_0, \beta_1)^T$ and $\theta = (\theta_0, \theta_1, \theta_2)^T$.

For the Primary Aim, $Z = \frac{\mathbf{L}\hat{\beta}}{\sqrt{\mathbf{L}\widehat{Var}(\hat{\beta})\mathbf{L}^T}}$ and $\mathbf{L} = (0, 1)^T$; the null hypothesis will be rejected if

$$|Z| > z_{1-\frac{\alpha}{2}}$$

For the Secondary Aim, $Z = \frac{\mathbf{L}\hat{\theta}}{\sqrt{\mathbf{L}\widehat{Var}(\hat{\theta})\mathbf{L}^T}}$ and $\mathbf{L} = (0, 1, -1)^T$; the null hypothesis will be rejected if

$$|Z| > z_{1-\frac{\alpha}{2}}$$

Power for the Primary Aim and the Secondary Aim are then estimated using $\mathbb{P}_M(\mathbb{1}(|Z| > z_{1-\frac{\alpha}{2}}))$ where \mathbb{P}_M denotes an empirical average across M Monte Carlo simulations and $\mathbb{1}$ is the indicator function. In `calc-power.R`, each of the M synthetic datasets were generated using the function `GenerateData()`. Type 1 error rates for the Primary Aim and Secondary Aim are each set to $\alpha = 0.05$.

7 Quantities Elicited from Domain Scientists

The following quantities serve as inputs to the simulation-based approach to power calculation and are to be elicited from domain scientists:

1. The probability that randomization will occur at decision point t , i.e., the probability that the participant is not driving and ‘sleep-mode’ is not activated at decision point t
2. The average within-person correlation, averaged only across decision points for which the participant was not driving and ‘sleep-mode’ was not activated.
3. The probability of engaging in self-regulatory strategies (self-reported) in the next hour following decision point t , when the participant was randomized to **no prompt** at decision point t :

$$Pr\{Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1\}$$

4. The probability of engaging in self-regulatory strategies (self-reported) in the next hour following decision point t , when the participant was randomized to **low effort** prompt at decision point t :

$$Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 1, H_t = 0, I_t = 1\}$$

5. The probability of engaging in self-regulatory strategies (self-reported) in the next hour following decision point t , when the participant was randomized to **high effort** prompt at decision point t :

$$Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 1, I_t = 1\}$$

In `calc-power.R`, the value of the latter three quantities are specified in the following lines of code. These lines specify that $Pr\{Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1\} = 0.15$, $Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 1, H_t = 0, I_t = 1\} = 0.21$, and $Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 1, I_t = 1\} = 0.16$.

```
prob_engaged_given_none <- .15
prob_engaged_given_low_effort <- .21
prob_engaged_given_high_effort <- .16
```

The relationship of the above quantities elicited from domain scientists with parameters in the models for the causal relative risk is

$$\begin{aligned}\theta_0 &= \log(Pr\{Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1\}) \\ \theta_1 &= \log\left(\frac{Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 1, H_t = 0, I_t = 1\}}{Pr\{Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1\}}\right) \\ \theta_2 &= \log\left(\frac{Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 1, I_t = 1\}}{Pr\{Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1\}}\right)\end{aligned}$$

Since $Pr(Y_{t, 60 \text{ mins}} = 1 | A_t = 0, I_t = 1)$ and $Pr(Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 0, I_t = 1)$ are equivalent, the equations can be equivalently expressed in a manner that does not involve A_t :

$$\begin{aligned}\theta_0 &= \log(Pr(Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 0, I_t = 1)) \\ \theta_1 &= \log\left(\frac{Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 1, H_t = 0, I_t = 1\}}{Pr(Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 0, I_t = 1)}\right) \\ \theta_2 &= \log\left(\frac{Pr\{Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 1, I_t = 1\}}{Pr(Y_{t, 60 \text{ mins}} = 1 | L_t = 0, H_t = 0, I_t = 1)}\right)\end{aligned}$$

8 Results

Let N denote the total number of participants within the study. Table 1 provides power for the Primary Aim (*Column 9*) and the Secondary Aim (*Column 10*) at $N = 100$ when

- the minimum simulated within-person correlation is about 0.65 (*Column 1*)
- the maximum simulated within-person correlation is about 0.67 (*Column 2*)

- the pre-specified probability under no prompt is 0.15 (*Column 3*)
- the pre-specified probability under low effort prompt is 0.21 (*Column 4*)
- the pre-specified probability under high effort prompt is 0.16 (*Column 5*)
- the pre-specified probability that the participant is not driving or in ‘sleep mode’ at a decision point is 0.80, 0.85, 0.90, 0.95, or 1 (*Column 8*)

The risk ratios implied by the values of the above quantities are also displayed for the Primary Aim (*Column 6*) and the Secondary Aim (*Column 7*).

Table 1: Power Calculation and Assumptions for Simulation Study 1

Min. Simulated Within-Person Correlation (Column 1)	Max. Simulated Within-Person Correlation (Column 2)	Prob. Engaged at Next DP given No Prompt at Current DP (Column 3)	Prob. Engaged at Next DP given Low Effort Prompt at Current DP (Column 4)	Prob. Engaged at Next DP given High Effort Prompt at Current DP (Column 5)	Risk Ratio Primary Aim (Column 6)	Risk Ratio Secondary Aim (Column 7)	Prob. not driving or in sleep mode at Current DP (Column 8)	Power Primary Aim (Column 9)	Power Secondary Aim (Column 10)
0.650	0.670	0.15	0.21	0.16	1.233	1.312	0.80	0.842	0.820
0.652	0.673	0.15	0.21	0.16	1.233	1.312	0.85	0.846	0.853
0.651	0.674	0.15	0.21	0.16	1.233	1.312	0.90	0.863	0.869
0.655	0.674	0.15	0.21	0.16	1.233	1.312	0.95	0.886	0.853
0.656	0.671	0.15	0.21	0.16	1.233	1.312	1.00	0.887	0.895

Table 2 provides the effects that can be detected for the Primary Aim and the Secondary Aim with at least 80% power if the pre-specified probability under no prompt is higher (0.20) and lower (0.10) than the value specified before (0.15) in Table 1. As in Table 1, the calculations were performed assuming $N = 100$.

Table 2: Power Calculation and Assumptions for Simulation Study 2

Min. Simulated Within-Person Correlation (Column 1)	Max. Simulated Within-Person Correlation (Column 2)	Prob. Engaged at Next DP given No Prompt at Current DP (Column 3)	Prob. Engaged at Next DP given Low Effort Prompt at Current DP (Column 4)	Prob. Engaged at Next DP given High Effort Prompt at Current DP (Column 5)	Risk Ratio Primary Aim (Column 6)	Risk Ratio Secondary Aim (Column 7)	Prob. not driving or in sleep mode at Current DP (Column 8)	Power Primary Aim (Column 9)	Power Secondary Aim (Column 10)
0.639	0.659	0.1	0.155	0.107	1.310	1.449	0.8	0.842	0.845
0.652	0.673	0.2	0.267	0.210	1.192	1.271	0.8	0.856	0.857

9 References

1. Friedrich Leisch, Andreas Weingessel and Kurt Hornik (2021). bindata: Generation of Artificial Binary Data. R package version 0.9-20. <https://CRAN.R-project.org/package=bindata>
2. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.6. <https://CRAN.R-project.org/package=dplyr>
3. McDaniel, L. S., Henderson, N. C., and Rathouz, P. J. (2013) Fast pure R implementation of GEE: application of the Matrix package The R Journal, 5/1:181–187
4. Nahum-Shani, I., Potter, L., Lam, C., Yap, J., Moreno, A., Stoffel, R., Wu, Z., Dempsey, W., Kumar, S., Murphy, S., Rehg, J., Wetter, D. (2021). The Mobile-Assistance for Regulating Smoking (MARS) Micro-Randomized Trial Design Protocol. Submitted to the Journal of Contemporary Clinical Trials