

Appendix D: Annotated Analysis Code for Example Hybrid Experimental Design

This code is written in R and uses the freely available `dplyr`, `tidyr` and `geepack` packages. The first two packages are well-known data management packages. The third is a generalized estimating equations library by Søren Højsgaard, Ulrich Halekoh, Jun Yan, and Claus Ekstrøm all are available on CRAN.

The data is assumed to be in wide format in the object `data_wide`. `Z1` and `Z2` are the assigned ADI options at the two ADI decision points, `A1, . . . , A112` are the assigned JITAI options at each of the 112 JITAI decision points, `R` is the week 4 response status indicator, and `Y1, . . . , Y112` are the proximal outcomes measured following each JITAI decision point.

To further prepare the data, we use the weighting and replicating (W&R) method (see Nahum-Shani et al., 2012). Responders are not randomized to Z_2 so they have no specific value for this indicator. Hence, their data is replicated into two copies, one with each possible value of Z_2 , namely $+1$ and -1 . In addition, non-responders get double weight (4 instead of 2) in the final analysis because they are randomized twice whereas responders are randomized once. Detailed justification and description of this approach is available in multiple papers (e.g., Dziak et al., 2019; Nahum-Shani et al., 2020; Nahum-Shani et al., 2012).

```
rows_to_replicate <- data_wide %>% filter(R==1)
rows_to_replicate$known_weight <- 2
positive_pseudodata <- rows_to_replicate
positive_pseudodata$Z2 <- +1
positive_pseudodata$replicant <- 1
negative_pseudodata <- rows_to_replicate
negative_pseudodata$Z2 <- -1
negative_pseudodata$replicant <- 2
rows_not_to_replicate <- data_wide %>% filter(R==0)
rows_not_to_replicate$replicant <- 1
```

```

rows_not_to_replicate$known_weight <- 4

data_wide_w_r <- bind_rows(rows_not_to_replicate,
                           positive_pseudodata,
                           negative_pseudodata)

data_wide_w_r <- data_wide_w_r %>% arrange(id, replicant)

```

The weighted and replicated data is now stored in `data_wide_w_r`. To perform the analysis for the proximal effects of JITAI options (Model 1 in Appendix A) it is convenient to first convert the dataset from wide form into long form (so that we have rows for each observation within each person, instead of just one row per person). This involves rearranging the A and Y columns.

```

data_long_w_r_A <- data_wide_w_r %>%
  select("id","replicant",starts_with("A")) %>%
  pivot_longer(cols=starts_with("A"), names_prefix = "A",
               names_to = "time", values_to = "A")

data_long_w_r_Y <- data_wide_w_r %>%
  select("id","replicant",starts_with("Y")) %>%
  pivot_longer(cols=starts_with("Y"), names_prefix = "Y",
               names_to = "time", values_to = "Y")

data_long_w_r_other <- data_wide_w_r %>%
  select("id","replicant","known_weight", "Z1","R","Z2")

data_long_w_r <- inner_join(data_long_w_r_other,
                           data_long_w_r_A, by=c("id","replicant")) %>%
  inner_join(data_long_w_r_Y,by=c("id","replicant","time")) %>%
  mutate(wave = as.integer(time) + n_obs*(replicant-1)) %>%
  mutate(stage2 = 1*(time>K)) %>%
  mutate(Z2stage2=Z2*stage2)

```

The last two lines of this code create `stage2`, which is an indicator for whether (=1) or not (=0) the second-stage ADI options Z_2 were assigned by time t (this indicator is labeled C_{it} in Appendix A).

The long-form, weighted and replicated data is now ready for analysis. Generalized estimating equations (GEE) are used to estimate Model 1 in Appendix A. The “independence” correlation means that working independence is used but robust standard errors are applied (see Qian et al., 2022).

```
gee_formula_proximal <- Y ~ Z1 + Z2stage2 + Z1*Z2stage2 +
                        A + A*Z1 + A*Z2stage2 + A*Z1*Z2stage2
gee_proximal <- geese(formula = gee_formula_proximal,
                      id=id, weights = known_weight, waves=wave,
                      data=data_long_w_r, corstr = "independence")
print(summary(gee_proximal))
```

To perform the analysis for the distal effects of ADI options (Model 2 in Appendix A), we return to the wide dataset. In our example, we assume that the distal Y is the sum of the proximal Y 's, and that one of its predictors is the average across the A 's. In the following code, we create the distal outcome (`sum_Y`), the average JITAI options (A 's) over all JITAI decision points for individual i (`mean_A`, which is labeled \bar{A}_i in Appendix A) and the average over only those points after week 4 (`mean_A_stage2`, which is labeled $\bar{A}_i^{(2)}$ in Appendix A).

```
all_Y_cols <- which(substr(colnames(data_wide_w_r),1,1)=="Y")
all_A_cols <- which(substr(colnames(data_wide_w_r),1,1)=="A")
stage2_A_cols <- (which(colnames(data_wide_w_r)=="A1")) :
                  (which(colnames(data_wide_w_r)=="
                        paste("A",K+1,sep=""))))
data_wide_w_r$sum_Y <- apply(data_wide_w_r[,all_Y_cols],1,sum)
data_wide_w_r$mean_A <- apply(data_wide_w_r[,all_A_cols],1,mean)
data_wide_w_r$mean_A_stage2 <- apply(data_wide_w_r
```

```

                                [,stage2_A_cols],1,mean)

data_wide_w_r$mean_A_stage2_Z2 <- data_wide_w_r$mean_A_stage2 *

                                data_wide_w_r$Z2

data_wide_w_r$mean_A_stage2_Z1_Z2 <- data_wide_w_r$mean_A_stage2 *

                                data_wide_w_r$Z1 *

                                data_wide_w_r$Z2

```

Again, we use GEE to estimate Model 2 in Appendix A.

```

gee_formula_distal <- sum_Y ~ Z1 + Z2 + Z1*Z2 +

                                mean_A + mean_A*Z1 + mean_A_stage2_Z2 +

                                mean_A_stage2_Z1_Z2

gee_distal <- geese(formula = gee_formula_distal,

                    id=id,

                    weights = known_weight,

                    waves=replicant,

                    data=data_wide_w_r,

                    corstr = "independence")

print(summary(gee_distal))

```

References

- Dziak, J. J., Yap, J. R., Almirall, D., McKay, J. R., Lynch, K. G., & Nahum-Shani, I. (2019). A data analysis method for using longitudinal binary outcome data from a SMART to compare adaptive interventions. *Multivariate behavioral research*, 54(5), 613-636.
- Nahum-Shani, I., Almirall, D., Yap, J. R., McKay, J. R., Lynch, K. G., Freiheit, E. A., & Dziak, J. J. (2020). SMART longitudinal analysis: A tutorial for using repeated outcome measures from SMART studies to compare adaptive interventions. *Psychological methods*, 25(1), 1-29.
- Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., Waxmonsky, J. G., Yu, J., & Murphy, S. A. (2012). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological methods*, 17(4), 457-477.
<https://doi.org/10.1037/a0029372>
- Qian, T., Walton, A. E., Collins, L. M., Klasnja, P., Lanza, S. T., Nahum-Shani, I., Rabbi, M., Russell, M. A., Walton, M. A., & Yoo, H. (2022). The microrandomized trial for developing digital interventions:

Experimental design and data analysis considerations. *Psychological methods*.
<https://doi.org/https://doi.org/10.1037/met0000283>