Appendix B: Sample Size Simulations

This appendix briefly describes the design and results of a simulation study used to demonstrate that adequate statistical power can be achieved for testing effects of interest in a hybrid experimental design (HED), given a reasonable sample size. Throughout, we assume investigators would like to plan sample size for the HED described in Figure 6. We report power for detecting (a) the main effect of the JITAI options (A_{it}) on the proximal outcome; (b) the interaction between JITAI options and first-stage (Z_{i1}) and second-stage (Z_{i2}) ADI options in terms of the proximal outcome; (c) the main effects of first-stage and second-stage ADI options (Z_{i2}) on the distal outcome; and (d) the interaction between first-stage or second-stage ADI options and the mean of the JITAI options (representing the message rate of delivery) on the distal outcome.

Data-Generating Model

The data-generating model mimicked the HED example in Figure 6. In this example, there are 112 daily JITAI decision points in which individuals are randomized to 2 JITAI options (A_{it}), specifically message ($A_{it} = 1$) vs. no message ($A_{it} = -1$). The average value $\bar{A}$ over a period of time is therefore directly related to the rate of message delivery, with $\bar{A} = 0$ meaning that a participant received the message 50% of the time.

Additionally, there are two ADI decision points. The first occurs at the beginning of the study where participants are randomized to two first-stage ADI options (Z_{i1}), specifically digital intervention with ($Z_{i1} = 1$) or without ($Z_{i1} = -1$) coaching. The second occurs at week 4 (day 28) where participants who show early signs of non-response ($R_i = 0$) are randomized to two second-stage ADI options (Z_{i2}), specifically step-up ($Z_{i2} = 1$) or continue with ($Z_{i2} = -1$) the initial intervention. In contrast, responders ($R_i = 1$) are not randomized to ADI options at week 4 and instead continue with the initial intervention ($Z_{i2} = 0$). The probability of being a responder is set to .40, .50, or .60 (regardless of Z_{i1}) in different scenarios. The sample size is set to 100, 150 or 200.

The data-generating model for the proximal outcome at time $t + \Delta$, conditional on intervention history and response status, for individual i is as follows:

$$E(Y_{it+\Delta}|\boldsymbol{Z_i}, A_{it}, R_i) = \beta_0 + \beta_1 Z_{i1} + \beta_2 C_{it} Z_{i2} + \beta_3 C_{it} Z_{i1} Z_{i2} + \gamma_0 A_{it} + \gamma_1 Z_{i1} A_{it} + \gamma_2 Z_{i2} C_{it} A_{it} + \gamma_3 Z_{i1} Z_{i2} C_{it} A_{it} + \delta R_i^c$$

In this model, δ is the correlational effect of being predisposed to be a responder, $R_i^c = R_i - (r(Z_{i1}) - 1)$ is the centered response status, and $r(z) = P(R_i = 1|Z_{i1} = z_1)$. This effect is included to reflect the realistic assumption that non-responders and responders will differ systematically in terms of the outcome of interest, above and beyond the effects of the intervention options. Setting $\delta = 0$ gives Model 1 of Appendix A, which is the model for the proximal outcome marginal over R_i. The coefficient values are chosen as follows:

- $\beta_0 = 0.30$
- $\beta_1 = \beta_2 = \beta_3 = -0.03$
- $\gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = -0.02$
- $\delta = -0.08$

Negative values are used to reflect an intuition that the response is a quantity which is desired to be reduced (such as the number of drinks on the next day). Although the responses are generated as normally distributed, the values for the means are chosen to be on the same order of magnitude as the probabilities of binary variables would be. Specifically, the fitted means for the proximal outcome are between 0.05 and 0.55, depending on the values of Z_{i1}, Z_{i2}, and A_{it}. The effects resulting from these coefficient values are summarized in Table 4.

The residuals in $Y_{it+\Delta}$ are generated as multivariate normal distribution with AR-1 correlation within participant, with variance $\sigma^2 = 0.20$ and correlation $\rho = 0.5$. The standardized effect sizes in terms of the proximal outcome are all very small, but this is offset by many observations per individual, and also reflects the fact that the overall effectiveness of the intervention relies on more than one

component (so that the combination is much stronger than any of the parts). A distal outcome was also calculated as the sum of proximal outcomes for each individual: $Y_i^* = \sum_{t=1}^{112} Y_{it}$.

Because effect coding is used, scaled differences in means (Cohen's *d*) for main effects on the proximal outcome can be calculated as twice the absolute value of the coefficient, divided by $\sigma = \sqrt{0.20} \approx 0.4472$. Therefore, $\gamma_0 = -0.02$ represents a standardized main effect of approximately $d = (2 \times .02)/(0.4472) \approx .09$.

As a supplemental study, the simulations were also separately repeated in a scenario in which all null hypotheses of interest are true: specifically, $\beta_1 = \beta_2 = \beta_3 = 0$ and $\gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = 0$, but other aspects of the model were unchanged. This was done to get an estimate of Type One error rate for nominal .05 testing.

After each data was simulated, marginal models were fit to the data, specifically Model 1 of Appendix A for the proximal outcome, and Model 2 of Appendix A for the distal outcome. Weighted and replicated estimating equations (Nahum-Shani et al., 2012), with a working independence assumption and robust (sandwich/Taylor linearized) standard errors were used, in order to make it feasible to fit these models with standard software even though responders were not re-randomized to Z_2 (Lu et al., 2016; Nahum-Shani et al., 2020). To provide the best possible estimate of the standard errors, all effects in Model 1 were estimated jointly, and similarly all effects in Model 2 were estimated jointly.

2000 datasets, each of 100 participants with 112 observations each, were simulated from this model. The parameters of interest were tested for each model. The observed Type One error (in the null scenarios) and power (in the non-null scenarios) were obtained as the proportion of rejecting the null hypothesis for each parameter in each scenario, at a nominal .05. The results are shown in Tables 5 and 6 respectively.

Results

Simulated Type One error rates are shown in Table 5. They were as desired for the proximal outcome model (very close to nominal), although they were sometimes slightly inflated for the distal outcome model. However, this inflation was generally small. The inflation may be due to the asymptotic nature of the sandwich formula, which is a limitation with modest sample sizes (recall that the distal model has only one outcome per participants, not many as in the proximal).

Simulated power values for the non-null scenarios are shown in Table 6. Power for proximal effects was generally very satisfactory. For instance, power for testing the average main effect of JITAI options A on the proximal outcome $Y_{it+\Delta}$ is extremely high. Note that although the effect size is modest, the total amount of information used to detect this effect across JITAI decision points and participants (112 occasions × 100 participants) is substantial. Similarly, power for the interaction between the JITAI options and the first stage ADI options $A_{it} \times Z_{i1}$ is high. Interactions of A_{it} with the second-stage ADI options Z_{i2} are sometimes less powerful, because they are informed only by a subsample of the participants and occasions. Still, power for these interactions is still satisfactory for larger sample sizes.

With respect to the distal effects, the power for the main effect of Z_{i1} on the distal outcome is high, whereas power for the main effect of Z_{i2} is lower because contrasts between levels of Z_{i2} are informed only by non-responders. Further, power was very low for testing the interaction between the ADI options (either Z_{i1} or Z_{i2}) and the rate of JITAI options $\bar{A}$. This is reasonable given the way this particular HED was designed. Since participants were randomized to JITAI options with 0.5 probability at each decision points, due to the Central Limit Theorem, most participants have $\bar{A}$ very close to 0. Estimating the effect of a large change in $\bar{A}$ would therefore require a gross extrapolation from an extremely restricted distribution.

Limitations

This simulation was intended mainly as a proof-of-concept illustration; it should be extended to incorporate additional and more realistic scenarios. For example, it was assumed that no data was

missing, while in practice, missingness and dropout would tend to reduce power. Also, it was assumed that the overall mean of the proximal outcome, averaged over the effects, did not change over time. A richer model should be simulated to incorporate time trends (e.g., . by adding a linear effect of t on $Y_{it+\Delta}$, nonlinear trajectories, and/or interactions between the intervention options and time). Furthermore, a richer data-generating model should also allow different rates of response for participants given different values of initial treatments (i.e., at least Z_{i1}, and perhaps early values of A_{it}).

Discussion

In this simulation, power for testing key scientific questions about the effects of the JITAI options on the proximal outcome, or about the main effects of the ADI options on the distal outcome, was found to be quite satisfactory for a reasonable number of participants (i.e., 100-200). Questions about the effects of the second-stage ADI options might require a somewhat larger sample size, because only second stage data (and in some cases only data from non-responders) is useful. Power for questions involving the rate of JITAI options in relation to the distal outcome was much lower. A satisfactory power for such questions would require a different HED that allows the randomization probabilities to systematically vary between participants. Nonetheless, a model containing $\bar{A}$ is straightforward to fit even to a simpler design, and it might provide some useful information on an exploratory basis, with the understanding that power and precision will be relatively low because the values of $\bar{A}$ are tightly clustered around the mean.

Table 4
Effect Sizes in Data-Generating Model for Simulation

Scientific question	Estimand and raw effect size in data-generating model
Main effect of JITAI options on the proximal outcome, averaging over ADI options	$2\gamma_0$= -0.04
Main effect of the first-stage ADI options on the distal outcome, averaging over JITAI options	$2\beta_1 * 112$= -6.72
Main effect of the second-stage ADI options on the distal outcome, averaging over JITAI options	$2\beta_2 * (112\text{-}28)$= -5.04
Interaction between JITAI options and first-stage ADI options in relation to the proximal outcome	$4\gamma_1$= -0.08
Interaction between first-stage ADI options and JITAI options in relation to the distal outcome.	Subtracting the conditional effects given two $\bar{A}$ values of interest can be interpreted as an interaction. For example, consider the following two conditional effects: (1) the conditional effect of first-stage ADI options when the average rate of JITAI message delivery is 0.6 (i.e., $\bar{A}$ =0.2 when using effect coding for A_{it}) is $2(\beta_1 + \gamma_1) * 0.2 * 112$= -2.24; (2) the conditional effect when the average rate of JITAI message delivery is 0.4 (i.e., $\bar{A} = -0.2$) is 2.24. The difference between these two conditional effects is -4.48.
Interaction between JITAI options and second-stage ADI options in relation to the proximal outcome	$4\gamma_2 = -0.08$
Interaction between second-stage ADI options and JITAI options in relation to the distal outcome.	Subtracting the conditional effects given two $\bar{A}_i^{(2)}$ values of interest can be interpreted as an interaction. For example, consider the following two conditional effects: (1) The conditional effect of second-stage ADI options when the average rate of JITAI message delivery after day 28 is 0.6 (i.e., $\bar{A}_i^{(2)} = 0.2$ when using effect coding for A_{it}) is: $2(\beta_2 + \gamma_2) * 0.2 * (112 - 28)$= -1.68; (2) The conditional effect when the average rate of JITAI message delivery after day 28 is 0.4 (i.e., $\bar{A}_i^{(2)} = -0.2$) is: 1.68. The difference between these two conditional effects is -3.36.

Table 5
Type One Error Rate Estimates

Proximal Effects										
Sample size		**100**	**100**	**100**	**150**	**150**	**150**	**200**	**200**	**200**
Response prob.		**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**
Parameter	**Effect**									
β_1	Z_1	0.06	0.06	0.06	0.05	0.05	0.06	0.06	0.05	0.05
β_2	Z_2	0.06	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.05
β_3	$Z_1 \times Z_2$	0.06	0.06	0.05	0.05	0.04	0.05	0.05	0.06	0.06
γ_0	A	0.06	0.06	0.06	0.05	0.06	0.05	0.05	0.05	0.05
γ_1	$Z_1 \times A$	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.05	0.06
γ_2	$Z_2 \times A$	0.07	0.06	0.06	0.05	0.05	0.06	0.06	0.06	0.05
γ_3	$Z_1 \times Z_2 \times A$	0.06	0.06	0.05	0.05	0.06	0.06	0.06	0.06	0.05
Distal Effects										
Sample size		**100**	**100**	**100**	**150**	**150**	**150**	**200**	**200**	**200**
Response prob.		**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**
Parameter	**Effect**									
β_1	Z_1	0.06	0.07	0.06	0.05	0.06	0.07	0.06	0.05	0.06
β_2	Z_2	0.07	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.07
β_3	$Z_1 \times Z_2$	0.07	0.07	0.06	0.07	0.05	0.06	0.06	0.06	0.06
γ_0	A	0.07	0.07	0.08	0.07	0.07	0.06	0.06	0.06	0.06
γ_1	$Z_1 \times A$	0.08	0.07	0.08	0.07	0.07	0.07	0.05	0.06	0.05
γ_2	$Z_2 \times A$	0.07	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.06
γ_3	$Z_1 \times Z_2 \times A$	0.06	0.07	0.07	0.07	0.06	0.06	0.06	0.05	0.06

Table 6
Power Estimates

Proximal Effects										
Sample size		**100**	**100**	**100**	**150**	**150**	**150**	**200**	**200**	**200**
Response prob.		**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**
Parameter	**Effect**									
β_1	Z_1	0.96	0.97	0.97	1.00	1.00	1.00	1.00	1.00	1.00
β_2	Z_2	0.44	0.55	0.65	0.61	0.71	0.83	0.71	0.83	0.91
β_3	$Z_1 \times Z_2$	0.45	0.54	0.64	0.63	0.73	0.82	0.73	0.85	0.91
γ_0	$\bar{A}$	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
γ_1	$Z_1 \times \bar{A}$	0.93	0.93	0.94	0.99	0.99	0.99	1.00	1.00	1.00
γ_2	$Z_2 \times \bar{A}^{(2)}$	0.57	0.66	0.78	0.73	0.84	0.91	0.86	0.91	0.97
γ_3	$Z_1 \times Z_2 \times \bar{A}^{(2)}$	0.56	0.67	0.74	0.74	0.85	0.91	0.86	0.92	0.96
Distal Effects										
Sample size		**100**	**100**	**100**	**150**	**150**	**150**	**200**	**200**	**200**
Response prob.		**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**	**0.6**	**0.5**	**0.4**
Parameter	**Effect**									
β_1	Z_1	0.97	0.96	0.97	1.00	1.00	1.00	1.00	1.00	1.00
β_2	Z_2	0.40	0.51	0.61	0.56	0.67	0.78	0.67	0.81	0.89
β_3	$Z_1 \times Z_2$	0.42	0.49	0.61	0.58	0.69	0.79	0.68	0.80	0.88
γ_0	$\bar{A}$	0.08	0.09	0.09	0.07	0.07	0.07	0.06	0.07	0.08
γ_1	$Z_1 \times \bar{A}$	0.08	0.07	0.08	0.07	0.08	0.08	0.07	0.06	0.07
γ_2	$Z_2 \times \bar{A}^{(2)}$	0.07	0.08	0.07	0.06	0.06	0.06	0.04	0.06	0.06
γ_3	$Z_1 \times Z_2 \times \bar{A}^{(2)}$	0.07	0.06	0.07	0.06	0.06	0.06	0.07	0.06	0.06

References

Lu, X., Nahum-Shani, I., Kasari, C., Lynch, K. G., Oslin, D. W., Pelham, W. E., Fabiano, G., & Almirall, D. (2016). Comparing dynamic treatment regimes using repeated-measures outcomes: Modeling considerations in SMART studies. *Statistics in Medicine*, *35*(10), 1595-1615.

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