Control Covariates for Primary Analysis of Sense2Stop Micro-Randomized Trial

August 4, 2020

This appendix provides a set of covariates and reasons why these should be included in the primary analysis as control variables. Two separate analyses were conducted to arrive at this final set of covariates.

1 Analysis 1: Predicting missing episodes (for justification of MAR)

To justify the MAR assumption for the missing data in our proximal outcome, we investigate features that are predictive of missing episodes (note that a missing episode is either: (i) one which is classified as an unknown episode due to bad quality data or (ii) a completely missing portion of episode classification data within a participant's day) in the 120 minutes following an available decision time.

We consider the following columns of data for good candidate features that may predict missing episodes:

- x_1: ID (Integer with values in $\{1, \ldots, 48\}$)
- x_2: Day in MRT (Integer with values in $\{1, \ldots, 10\}$)
- x_3: Detected stressed episode at available decision time (Binary with values in $\{1, 0\}$)
- x_4: Episode length in minutes (from start to peak of episode). Note that this episode includes an available decision time.
- x_5: Previous Episode is Missing (Binary with values in $\{1, 0\}$). Note that this refers to the episode prior to the episode that contains the available decision time.
- x_6: Previous Episode is Detected Stressed (Binary with values in $\{1, 0\}$). Note that this refers to the episode prior to the episode that contains the available decision time.
- x_7: Previous Episode is Not Detected Stressed (Binary with values in $\{1, 0\}$). Note that this refers to the episode prior to the episode that contains the available decision time.

- x_8: Previous Episode Length in minutes (from start to end of episode). Note that this refers to the episode prior to the episode that contains the available decision time.
- x_9: Previous Day's Proportion of Activity (Proportion of minutes within 12 hour day that correspond to physical activity minutes)
- x_{10}: Previous Day's Proportion of Bad Quality REP Data (Proportion of minutes within 12 hour day that correspond to bad quality REP data)
- x_{11}: Previous Day's Proportion of Bad Quality ECG Data (Proportion of minutes within 12 hour day that correspond to bad quality ECG data)
- x_{12}: Number of Interventions Sent Previous Day
- x_{13}: BMI on Day 1 of study (continuous variable from 18 to 46)
- x_{14}: Gender ($0 =$ Female, $1 =$ Male)
- x_{15}: Age (Integer from 20 to 63)
- x_{16}: Age Started Smoking (Integer from 20 to 63)
- x_{17}: Total Fagerstrom Score (Integer from 0 to 9)
- x_{18}: Weekday ($1 =$ Weekday, $0 =$ Weekend)
- x_{19}: Hour of Day (Integer with values in $\{1, \ldots, 23\}$)
- x_{20}: Is Morning ($1 =$ Morning, $0 =$ Other)
- x_{21}: Is Afternoon ($1 =$ Afternoon, $0 =$ Other)
- x_{22}: Is Night ($1 =$ Night, $0 =$ Other)
- y: Is Current Episode Missing? ($1 =$ Yes, $0 =$ No). The current episode can be either the episode including the available decision time or any episode after this within the 120 minutes following an available decision time.

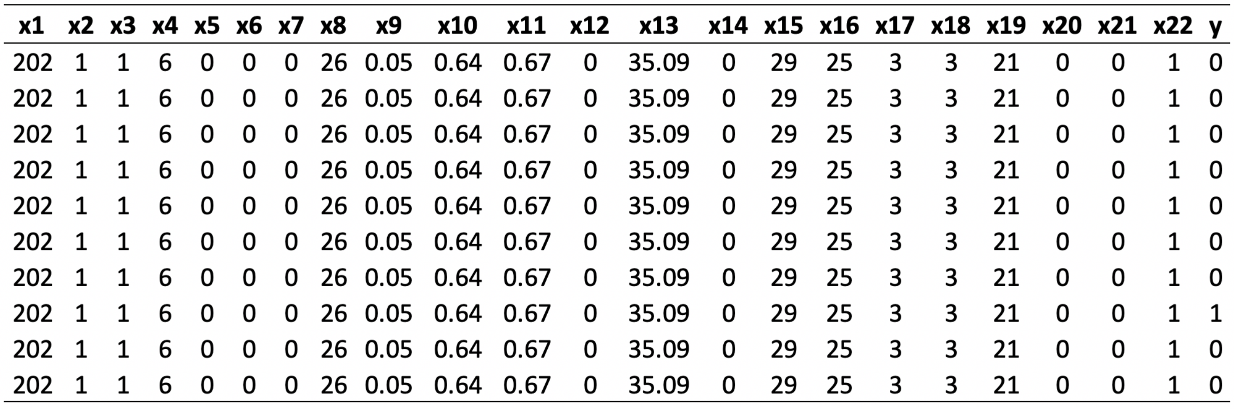
An example of the fist 10 rows of this data set is shown below in figure 1. These 10 rows corresponds to the first available decision time from ID 202 on day 1.

All numerical features were then converted to their standard scores. We note that each row within the data set used for Analysis 1 corresponds to one episode within the 120 minutes following an available decision time (i.e., there are multiple rows corresponding to an available decision time. For example, if there are 5 episodes in the 120 minutes following an available decision time including the episode encompassing the available decision time then this corresponds to 5 rows of data.)

To learn which variables explain missingness we use a logistic regression model for the binary outcome: missing or not missing episode.

We train and test the predictive performance of a logistic regression model with cross validation (i.e., we train on one portion of data and test on another that the model has

Figure 1: *Snapshot of features used in analysis 1.*



x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20	x21	x22	y
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	1
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0
202	1	1	6	0	0	0	26	0.05	0.64	0.67	0	35.09	0	29	25	3	3	21	0	0	1	0

not yet seen and we do this multiple times). For predictive performance, we use the recall score, which is defined as the ratio: (true positives)/(true positives + false negatives). The best model achieved a weighted[1] recall score of 0.59 and a weighted F1 score of 0.65. The five most influential features in descending order are (we use the magnitude and sign of the coefficients to detect the influence of the features given that the data was standardised prior to the model fit):

- (−) Previous Episode Type Is Not Detected Stressed
- (+) Previous Episode Type Is Missing
- (+) BMI on Day 1 of study
- (−) Age
- (+) Age Started Smoking

The sign in parentheses above is the direction of influence these 5 variables had on the missing episodes.

2 Analysis 2: Predicting minute level outcome using pre-decision point measures of the outcome (for reduction of variance)

The other reason we control for variables is to reduce the variance so as to increase the chance that if there is an effect of the notification on the minute level outcome, we will detect this. The natural variables are usual pre-decision point measures of the outcome. Here, this would include:

[1] Calculate metrics for each label, and find their average weighted by support (the number of true instances for each label).

- Number of minutes in a detected stressed episode in the prior 120 minutes
- Number of minutes in a physically active episode in the prior 120 minutes

In addition, for this analysis we include the following features:

- ID (Integer with values in 1, . . . , 48)
- Day in MRT (Integer with values in 1, . . . , 10)
- Detected stressed episode at available decision time (Binary with values in $\{1, 0\}$)
- Minute number after available decision time (Integer with values in $\{1, \ldots, 120\}$)
- Weekday (1 = Weekday, 0 = Weekend)
- Hour of Day (Integer with values in $\{1, \ldots, 23\}$)
- Is Morning (1 = Morning, 0 = Other)
- Is Afternoon (1 = Afternoon, 0 = Other)
- Is Night (1 = Night, 0 = Other)

All numerical features were converted to their standard scores. We note that each row within the data set used for Analysis 2 corresponds to one minute's outcome within the 120 minutes following an available decision time (i.e., there are up to 120 minutes to an available decision time).

To learn which of these variables explain the minute level outcomes we use a multi-class logistic regression model for the multi-class outcome: physically-active-minute, detected-stressed-minute, not-detected-stressed-minute.

We train and test the predictive performance of a multi-class logistic regression model with cross validation (i.e., we train on one portion of data and test on another that the model has not yet seen and we do this multiple times). For predictive performance, we use the recall score, which is defined as the ratio: (true positives)/(true positives + false negatives). The best model achieved an overall weighted recall score of 0.57 and an overall weighted F1 score of 0.63. The two most influential features are (here, we can think of the influence of a feature as the usability of a single feature to distinguish two classes "one-vs-rest").

- Detected stressed episode at available decision time (Binary with values in $\{1, 0\}$)
- Is Night (1 = Night, 0 = Other)

A classification report is shown in Table 1.

h

	precision	recall	f1-score	support
detected stressed	0.14	0.40	0.21	7227
not detected stressed	0.89	0.60	0.71	109325
physically active	0.18	0.44	0.25	15383
weighted avg	0.76	0.57	0.63	131935

Table 1: Classification report for Analysis 2.

3 Recommendation: Final set of control variables

- Previous Episode Type Is Not Detected Stressed
- Previous Episode Type Is Missing
- BMI on Day 1 of study
- Age
- Age Started Smoking
- Detected stressed episode at available decision time
- Is Night