Missing Data in Sense2Stop MRT

August 3, 2020

This supplementary appendix describes the extent of the missing data in the Sense2Stop MRT and the considerations we make to address this issue for the primary analysis.

1 Notation

- Each minute within a user's 12-hour day is a decision time. Over the course of 12 days $t = 1, \ldots, T$ where $T = 720 \times 10 = 7200$.

- An available decision time is a time corresponding to the minute after the peak of either a detected stressed episode or a not detected stressed episode in which the participant is avaialable.

- A participant is deemed available if they are not currently driving, not physically active, their phone battery is greater than 10%, they did not receive an EMA in the last 10 minutes and they did not receive an EMI in the last hour.

- Recall that the primary outcome $Y_{t+1} = [Y_{t+1,1}, \ldots Y_{t+1,120}]$ is a response vector of length 120 where each entry corresponds to a minute following an available decision time. $Y_{t+1,m} = 1$ if at minute m the user is within a *detected stressed* episode; $Y_{t+1,m} = 2$ if at minute m the user is physically active and $Y_{t+1,m} = 3$ if at minute m the user is within a *not detected stressed* episode, $1 \leq i \leq m$.

2 Missing Data in the Primary Outcome

We refer to Figure 1 to understand the extent of the missing data within the primary outcome. For each cell in Figure 1, we provide the average fraction of classifiable minutes of the 120 minutes following an available decision time. A classifiable minute is one which belongs to either a detected stressed episode, a not detected stressed episode or an unknown episode that is primarily due to physical activity.

2.1 Available Decision Time Stress Missclassifications

In order to track the available decision times that should or should not be used for the primary analysis we look into two data sources for validation. Namely,

1. The phone log data source. This provides information on whether an available decision time is at the peak of either a detected stressed or not detected stressed episode. This data source contains the available decision times (i.e., the times at which an individual was randomized to receive an intervention prompt).

2. The online episode level stress classification data source. This data source provides the data from online stress episode classifications at all times throughout an individual's day (i.e., not just at available decision times).

In theory, these two files should be matching in terms of the stress episode classifications at an available decision time. We investigate whether this is so.

There are 5047 available decision times in total. Their distribution across day in MRT and participant ID is shown in Figure 2.

Figure 3 shows that there are 143 of the 5047 available decision times that do not agree in terms of their stress episode classification amongst the two data sources. The reason for this is still under investigation.

Figure 1: *Average fraction of classifiable minutes out of the 120 minutes following an available decision time. A classifiable minute is one which belongs to either a detected stressed episode, a not detected stressed episode or an unknown episode that is primarily due to physical activity.*

Average fraction of classifiable minutes in 120 minutes following an available decision time (using stress episode data stream)

ID	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Average Within ID
202	0.82	0.89	0.68	0.9	0.73	0.64	0.6	0.6	0.44	0.76	0.71
203	nan	nan	nan	nan	nan	nan	0.86	0.38	nan	nan	0.62
205	0.72	0.92	0.76	0.9	0.76	0.8	0.97	0.85	0.82	0.76	0.83
207	0.78	0.74	0.77	0.51	0.9	0.7	0.74	0.79	0.57	nan	0.72
208	0.72	0.78	0.65	0.54	nan	0.59	0.42	0.62	nan	nan	0.62
211	0.92	0.72	0.87	0.67	0.9	0.96	0.72	0.87	0.75	0.78	0.82
212	0.85	nan	0.98	0.2	0.6	nan	nan	nan	nan	nan	0.66
213	0.64	0.33	nan	nan	nan	nan	nan	nan	nan	nan	0.48
214	0.73	0.9	0.76	0	0.88	0.48	0.54	nan	nan	nan	0.61
216	0.71	0.88	0.63	0.74	0.79	0.8	nan	1.27	0.71	0.45	0.78
219	0.86	0.91	0.74	0.91	0.73	0.97	0.83	0.86	0.83	0.9	0.85
221	0.68	nan	nan	nan	nan	nan	nan	nan	nan	nan	0.68
222	0.98	0.83	nan	nan	nan	nan	nan	nan	nan	nan	0.9
223	1	1	0.92	0.91	0.98	0.9	0.2	nan	nan	1	0.86
224	0.2	nan	nan	nan	nan	nan	0.78	0.4	0.1	nan	0.37
225	0.68	0.96	0.96	0.57	1	0.83	0.82	0.64	0.54	0.97	0.8
226	0.95	0.89	0.63	0.83	nan	0.9	0.9	0.89	0.89	0.99	0.87
227	0.78	0.86	0.87	0.88	0.98	0.99	0.94	0.83	0.89	0.78	0.88
228	0.97	0.81	0.73	0.71	0.93	0.84	0.72	0.73	0.73	0.87	0.8
229	0.99	nan	nan	nan	nan	nan	nan	nan	nan	nan	0.99
231	nan	0.88	0.52	nan	nan	nan	nan	nan	nan	nan	0.7
233	0.85	0.84	0.79	0.48	0.82	nan	0.66	nan	nan	nan	0.74
234	nan	0.54	nan	nan	nan	nan	nan	nan	nan	nan	0.54
235	0.75	0.79	0.82	0.54	0.62	nan	nan	nan	nan	nan	0.7
238	0.9	nan	0.83	0.73	0.83	0.74	nan	nan	0.62	0.89	0.79
240	0.85	nan	nan	nan	0.72	0.93	0.93	0.57	0.9	1	0.84
242	0.55	nan	nan	nan	nan	nan	nan	nan	0.68	0.4	0.54
243	0.66	0.33	nan	nan	nan	0.13	nan	nan	nan	nan	0.37
244	0.51	0.46	0.74	0.66	0.48	0.9	0.47	0.31	0.38	0.68	0.56
245	0.35	0	nan	0	0.52	0.24	nan	0.76	nan	0.45	0.33
248	0.43	0	0	0.35	0.1	nan	nan	0.53	nan	nan	0.24
249	0.76	0.92	0.75	0.45	0.52	0.54	0.8	nan	0.99	0.46	0.69
250	0.8	0.62	0.48	0.24	0.82	0.6	0.88	0.87	0.78	0.93	0.7
251	0.78	nan	0.89	nan	0.63	nan	0.41	0.83	0.55	nan	0.68
252	0.88	1	0.99	0.97	0.98	0.87	0.91	0.95	0.82	1	0.94
253	0.94	0.98	1	0.98	0.98	0.95	0.91	0.88	0.81	0.84	0.93
255	0.92	nan	nan	nan	nan	nan	nan	nan	nan	nan	0.92
256	0.66	0	0.79	nan	0.35	0.27	0.73	0.33	0.2	0.32	0.41
258	nan	0.05	0	0.8	1	nan	nan	nan	nan	nan	0.46
259	nan	0	nan	nan	nan	nan	nan	nan	nan	nan	0
260	0.82	0.9	0.92	0.87	0.94	0.87	0.68	0.8	0.47	0.12	0.74
261	0.84	1	0.97	0.71	0.79	0.73	nan	0.86	0.83	0.9	0.85
262	0.73	0.94	0.84	0.77	0.99	0	0.84	0.85	0.81	0.76	0.75
264	nan	0.9	0.78	0.64	0.76	0.45	0.46	0.83	0.88	0.79	0.72
265	0.8	0.65	0.77	0.86	0.66	0.3	0.64	nan	0.91	0.57	0.68
266	nan	1	0.25	nan	nan	nan	nan	nan	nan	nan	0.62
267	nan	0.96	0.99	0.89	0.93	nan	0.9	0	nan	nan	0.78
268	0.52	nan	nan	nan	nan	nan	nan	nan	nan	nan	0.52
269	0.76	0.43	0.68	0.68	0.59	0.39	0.48	0.79	0.73	0.38	0.59

Figure 2: *Number of available decision times during the Sense2Stop MRT.*

ID	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
202	4	15	6	2	7	5	2	4	8	20
203	0	0	0	0	0	0	10	6	0	0
205	21	22	14	41	23	18	28	18	19	19
207	24	28	30	27	34	33	12	25	17	0
208	6	19	16	10	0	10	18	5	0	0
211	9	31	32	15	13	12	30	15	10	21
212	25	0	9	2	8	0	0	0	0	0
213	7	4	0	0	0	0	0	0	0	0
214	17	20	5	1	12	12	8	0	0	0
216	17	36	24	17	17	12	0	14	7	6
219	9	35	26	37	19	24	21	27	21	28
221	4	0	0	0	0	0	0	0	0	0
222	12	19	0	0	0	0	0	0	0	0
223	5	22	17	16	21	19	2	0	0	1
224	4	0	0	0	0	0	19	3	1	0
225	8	20	8	8	1	18	40	16	11	20
226	6	37	27	35	0	12	7	15	26	10
227	14	31	26	41	31	28	45	30	30	19
228	19	28	15	27	24	10	27	11	9	15
229	8	0	0	0	0	0	0	0	0	0
231	0	4	4	0	0	0	0	0	0	0
233	11	28	43	24	11	0	25	0	0	0
234	0	12	0	0	0	0	0	0	0	0
235	17	24	22	14	6	0	0	0	0	0
238	2	0	23	10	17	16	0	0	6	36
240	4	0	0	0	4	13	13	4	11	3
242	2	0	0	0	0	0	0	0	6	2
243	5	3	0	0	0	3	0	0	0	0
244	11	18	27	19	13	20	18	8	10	18
245	8	1	0	1	11	11	0	5	0	2
248	7	1	1	4	2	0	0	22	0	0
249	8	18	11	6	27	13	31	0	38	15
250	20	32	16	9	21	22	22	47	36	10
251	22	0	14	0	13	0	7	7	2	0
252	30	27	25	30	43	22	9	31	6	31
253	14	19	27	18	21	25	13	8	18	18
255	4	0	0	0	0	0	0	0	0	0
256	5	1	13	0	8	6	7	9	4	4
258	0	2	1	1	1	0	0	0	0	0
259	0	1	0	0	0	0	0	0	0	0
260	24	25	20	28	27	11	5	4	22	4
261	29	22	11	14	29	23	0	24	22	26
262	18	18	24	15	11	1	9	20	15	18
264	0	28	11	16	16	17	15	23	8	28
265	4	18	24	9	16	14	10	0	18	13
266	0	17	2	0	0	0	0	0	0	0
267	0	35	13	16	11	0	33	1	0	0
268	4	0	0	0	0	0	0	0	0	0
269	11	10	19	20	13	14	13	23	18	16

Figure 3: *Stress episode miss-classifications.*

Episode classification discrepencies at available decision times. Using two data sources: (1) Phone log file; (2) stress episode classification file.		
(1) Phone log file: Stress Classification used by randomization algorithm	**(2) Episode Level Stress Classification**	**Count**
detected stressed	detected stressed	266
detected stressed	not detected stressed	1
detected stressed	unknown	2
detected stressed	no data available	1
not detected stressed	detected stressed	20
not detected stressed	not detected stressed	4638
not detected stressed	unknown	113
not detected stressed	no data available	6
Total number of available decision times		5047