Sense2Stop Data Curation Documentation

February 4, 2022

As the intended design and planned analyses of the Sense2Stop study has already been described (Battalio et al., 2021), a major focus of this documentation will be on articulating decisions informed by scientific and practical considerations towards analysis of trial aims.

1. Trial aims

The Sense2Stop study (Battalio et al., 2021) sought to investigate the following scientific questions through a micro-randomized trial (MRT) among smokers who have expressed a willingness to quit smoking:

Primary Aim Hypothesis

Administration of a prompt to perform a stress regulation exercise, as compared to no prompt, will reduce the likelihood of being stressed in the subsequent two hours, and this effect will be stronger if the prompt is administered when the individual is stressed.

The ***primary proximal outcome*** in the Sense2Stop study is trichotomous: it reflects the probability that each moment of time during the 120-minute period immediately following micro-randomization is classified as 'probably stressed', 'probably not stressed', or 'physically active'.

Secondary Aim Hypothesis

Administration of a prompt to perform a stress regulation exercise will reduce the odds of smoking in the subsequent two hours, and this effect will be stronger if the prompt is administered when the individual is stressed.

2. Implemented study protocol during lab visits

To provide context for the decisions described in this documentation, we provide details regarding the 1st and 2nd lab visits not covered in existing work on the Sense2Stop study (Battalio et al., 2021). Participants were asked to complete 3 lab visits during the study.

1st lab visit: A smartphone and wearable devices are loaned to participants. The smartphone contains pre-installed apps, including a Sense2Stop app and three additional apps focused on helping participants perform self-regulation exercises while wearable devices consist of a chest band and straps to be worn on the left and right wrist. Participants were oriented on study procedures, including that:

- Wearable devices should be worn at all times, except when sleeping
- Wearable devices and smartphone must be returned during the 3rd lab visit
- Participants may access any of the apps focused on helping participants perform self-regulation exercises as often as they wanted
- Pressing a 'start of day' button each day within the Sense2Stop app is required before the Sense2Stop app could send a Random EMA for that day

During the 1st lab visit, study staff also (manually) pre-set 'sleep time' on the smartphone in consultation with the participant. 'Sleep time' refers to the time of day at which the Sense2Stop app will automatically pause all notifications for that day until the next time the 'start of day' button is pressed.

We note that the wearable devices continued to collect data regardless of whether the participant pressed the 'start of day' button, as long as both smartphone and wearable devices were not switched off.

2nd lab visit: Study staff (manually) activated the Sense2Stop app's micro-randomization capabilities. Since this activation cannot be done remotely, only participants who completed their 2nd lab visit could have been micro-randomized.

For those participants who had the Sense2Stop app's micro-randomization capabilities activated, this caused the functionality of the 'start of day' to change slightly: pressing a 'start of day' button each day within the Sense2Stop app is now required before the Sense2Stop app could send prompts to perform a self-regulation exercise **and** Random EMAs for that day.

Notably, study staff did not convey this change in functionality to participants. As long as participants did not notice that they can control the dose of the intervention by not pressing the start of day button, this results in a situation akin to 'blinding'.

3. Criteria for including vs. not including participants in all analyses

75 adult smokers between the ages of 18 and 65 years were enrolled in the Sense2Stop study. These participants met eligibility criteria for enrolment (see Battalio et al., 2021), including having met the study's definition of an ***active smoker*** – an individual who reports to have smoked one or more tobacco cigarettes per day for the past year.

Once enrolled into the study, (based on Institutional Review Board-approved study protocol; Spring, 2018), participants were subsequently removed from the study if they:

- C1. Informed study staff either (a) before the scheduled date of their 2nd lab visit or (b) during the day they completed their 2nd lab visit that they wish to withdraw from the study.

- C2. Did not complete their 2nd lab visit but did not inform study staff that they wish to withdraw.

Note that both C1 & C2 are prior to the start of the micro-randomized portion of the trial. Additionally, study staff conducted a 'pilot run' of study procedures on 5 of the 75 participants to identify potential barriers to fidelity to the intended study design and, when necessary, adjusted study procedures for all subsequent participants. Hence, participants were not included in all analyses if they

- C3. Were part of the trial's 'pilot run'.

Finally, participants who never received a notification were not included in all analyses. Specifically, participants were not included in all analyses if they:

C4. Had no micro-randomizations between their 'first day' and their 'last day', inclusive; 'first day' and 'last day' are defined below.

C4 can occur if (a) the participant was not eligible for micro-randomization throughout the study (see Section 4 for details); or (b) due to technical/software issues.

Table 1. Participants excluded entirely from all analyses.

Criterion violated	Participant ID's
C1	9 participants (Participant IDs 204 209 210 220 232 236 237 239 246)
C2	3 participants (Participant IDs 201 257 263)
C3	5 participants (Participant IDs 101 102 103 104 105)
C4	*Among those participants who have not yet been counted toward C1- C3:* 9 participants (Participant IDs 206 215 217 218 230 241 247 254 270)

From here onward, we will only focus on the remaining N = 75-(9+3+5+9) = 49 participants.

Definition of 'first day' and 'last day'
First day (***Day 0***): the date when a participant completed their 2nd lab visit
Last day (***Day 10***): 10 days after the first day

Exceptions:

- One participant (Participant ID 213) informed study staff that they wish to withdraw from the study 5 days after they completed their 2nd lab visit. For this specific participant, the 'last day' will be Day 5.
- One participant (Participant ID 251) completed their 2nd lab visit one day later than scheduled due to bad weather. For this specific participant, the 'first day' will still be Day 0 but their 'last day' will be Day 9.
- One participant (Participant ID 266) completed their 2nd lab visit one day later than scheduled, but the reason for the delay was not recorded in study staff notes. For this specific participant, the 'first day' will still be Day 0 but their 'last day' will be Day 9.

Subsequent implications of 'first day' and 'last day'
Since data collection performed by wearable devices cannot be terminated by the study team remotely and participants may neglect to return wearable devices by their 'last day', any data-points providing information on their physiology or behavior during time periods beyond their 'last day' will not be included in analyses. Specifically, we will not include any data collected after 11:59PM on the participant's 'last day'.

4. Eligibility for micro-randomization
A participant was regarded as ***eligible for micro-randomization*** at a particular moment of time t if they met a pre-specified set of criteria, which are described in Battalio, et al. (2021; Figure 2). In addition to these pre-specified criteria, micro-randomization did not occur at t if (a) the participant did not press the 'start of day' button on that day; and/or (b) t is after 'sleep time';

and/or (c) t is before activation of the Sense2Stop app's micro-randomization capabilities on the participant's 'first day', e.g., if activation occurred at 3pm, all minutes prior to 3pm on the participant's 'first day' will be regarded as ineligible for micro-randomization.

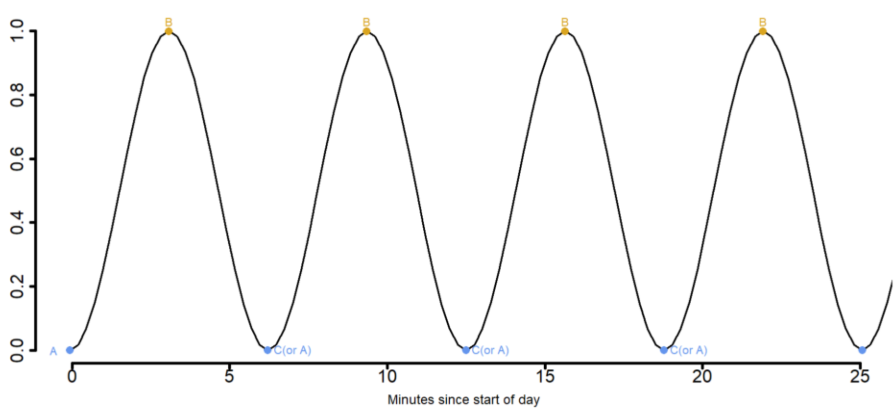
Table 2. Number of micro-randomizations <u>*per participant*</u>

Min	25th-percentile	Median	75th-percentile	Max
1	19	85	180	321

5. Working with sensor-derived assessments to construct the primary proximal outcome
A detection algorithm developed by Sarker and colleagues (Sarker et al., 2016; Sarker et al., 2017) was used to predict when an individual might be experiencing stress, but with some modifications specific to the Sense2Stop study. During the course of the study, data collected by wearable devices were initially transformed into a ***stress likelihood time series***, a continuous-time measure for physiological arousal ranging between 0 to 1. Intuitively, this measure does not directly capture the intensity of physiological arousal, but rather, the likelihood that any physiological arousal occurred at a particular moment in time; a value closer to 1 (closer to 0) signifies that physiological arousal is more likely (less likely) to have occurred at that specific moment.

Subsequently, ***episodes***, defined as a time interval in which there is an increasing trend immediately followed by a decreasing trend in the stress likelihood time series, were constructed from the stress likelihood time series. Within any given episode, the stress likelihood time series thus takes the form of two successive valleys with a peak sandwiched in between (see Figure 1 below). We refer to the beginning, peak, and end of an episode occurred as 'A', 'B', and 'C', respectively.

Figure 1. A conceptual stress likelihood time series. In this figure, the text 'C (or A)' refers to the fact that C also marks the beginning (i.e., A) of a subsequent episode.



Finally, the area under the curve defined by the stress likelihood time series between A and B is then used to label the episode more specifically as one of three types:

- a ***probably stressed*** episode, when the area under the curve exceeds a pre-specified threshold *c* and good quality data is available for more than 50% of the time between A and B; or

- a ***probably not stressed*** episode, when the area under the curve lies below a pre-specified threshold *c* and good quality data is available for more than 50% of the time between A and B; or
- an ***unknown*** episode, either when good quality data is not available for more than 50% of the time between A to B or when a physical activity confound exists between A and B.

The code used for episode-labeling can be found here: https://github.com/MD2Korg/stream-processor/blob/master/src/main/java/md2k/mcerebrum/cstress/features/StressEpisodeClassification.java In the Sense2Stop study, c=0.36 (lines 47-48 in the linked code).

Episode type and specific time-stamps for A, B, C constitute the observed data which we will use as our starting point to construct the primary proximal outcome. For brevity, we will collectively refer to this data as the ***episode classification data stream*** from here onward.

The primary proximal outcome is based on the trichotomous variable $Y_{i,t}$, defined to have a value of 0, 1, or 2 if participant *i* is within a probably not stressed episode, within a probably stressed episode, or is physically active, respectively, at minute t. Here, the subscript t indexes each minute between 'first day' and 'last day', inclusive. In what follows, we will describe a procedure for constructing the trichotomous variable $Y_{i,t}$. The procedure was motivated by noticing that the episode classification data stream does not differentiate between episodes that are unknown due to the existence of a physical activity confound and those that are unknown due to poor data quality. Hence, we needed to bring in information from other data sources to help us distinguish between these cases.

Procedure for constructing the trichotomous variable $Y_{i,t}$

Working with probably stressed and probably not stressed episodes. Initial inspection of the episode classification data stream show that in a sizeable number of probably stressed and probably not stressed episodes, the duration of time between B and C exceeds 15 minutes (see Table 4). Since this observation is not consistent with the duration of time between B and C that we expect to observe, we believe that the actual duration of time between B and C may be shorter than what the data (i.e., the episode classification data stream) suggests. Information on data quality after B was not used when determining the specific time-stamp for C, hence the long duration we sometimes observe between B and C.

Table 3. Number of minutes elapsed between B and C. Statistics were calculated among episodes whose peak lies between 'first day' and 'last day', inclusive

	Min	**25th-percentile**	**Median**	**75th-percentile**	**Max**
Probably not stressed	0.66	2.02	4.01	7.00	904.00
Probably stressed	0.58	0.58	6.01	9.00	664.99
Unknown	0.56	0.56	6.01	11.00	8556.01

Table 4. Number of episodes for which the number of minutes elapsed between B and C is greater than 5 minutes (Column 1) vs. at most 5 minutes (Column 2); greater than 15 minutes (Column 3) vs. at most 15 minutes (Column 3); greater than 60 minutes (Column 5) vs. at most 60 minutes (Column 6). Statistics were calculated among episodes whose peak lies between 'first day' and 'last day', inclusive.

Episode type	**>5 vs. ≤5 min.**		**>15 vs. ≤15 min.**		**>60 vs. ≤60 min.**		**Total no. of episodes**
	(1)	(2)	(3)	(4)	(5)	(6)	
Probably not stressed	5006	6600	105	11501	21	11585	11606
Probably stressed	678	396	28	1046	1	1073	1074
Unknown	3431	2339	1033	4737	322	5448	5770

We use the rule in Box 1 for assigning a new value for C in probably stressed and probably not stressed episodes which exceed 5 minutes. Employing this rule requires constructing a minute-by-minute binary indicator for whether heart rate data is observed. In brief, heart rate data is represented as the average number of beats over a 1-minute-long window. This rule assumes that a generally more precise C may be determined by identifying the first moment of time after B for which a no heart rate data was observed for 5 consecutive minutes.

Box 1. Rule for assigning a new value for C for probably stressed and probably not stressed episodes where the duration of time between B and C exceed 5 minutes

1	IF there is a period of at least five consecutive minutes within B and C having no observed heart rate data
2	THEN regard the beginning of this no heart rate period (which we denote by Č) as the end of an episode in place of C AND regard the information on the trichotomous variable $Y_{i,t}$ between Č and C as missing.
3	ELSE do not assign a new value for C

In effect, probably stressed episodes and probably not stressed episodes are censored when Č is regarded as the end of an episode in place of C.

Working with unknown episodes. We use the rule in Box 2 which leverages physical activity data and heart rate data. In brief, physical activity data is represented as a minute-level indicator for whether physical activity is detected (=1) or not (=0) by an activity detection algorithm.

Box 2. Rule for labeling <u>unknown</u> episodes as physically active episodes

1	IF physical activity is detected in more than 50% of minutes between A to B
2	THEN regard the time between A to C as a physically active episode AND apply the rule in Box 1 to determine whether the physically active episode needs to be censored.
3	ELSE regard the information on the trichotomous variable $Y_{i,t}$ between A and C as missing.

After using the rule in Box 2, we may now regard participant *i* as physically active at minute t (i.e., assign a value of 2 to $Y_{i,t}$) if they were within a physically active episode at minute t.

** ADD LATER **
** TABLE HERE: **
** MIN, MEDIAN, MAX episode duration after applying Box1-2 **

Missing values in the trichotomous variable $Y_{i,t}$. The trichotomous variable $Y_{i,t}$ is said to have a missing value if we are not able to assign a value of 0, 1, or 2 to $Y_{i,t}$ using the rules described above. We will use $M_{i,t}$ to denote an indicator equal to 1 if $Y_{i,t}$ is observed and equal to 0 if $Y_{i,t}$ is missing.

References

[1] Battalio, S. L., Conroy, D. E., Dempsey, W., Liao, P., Menictas, M., Murphy, S., …, and Spring, B. (2021). Sense2Stop: A micro-randomized trial using wearable sensors to optimize a just-in-time adaptive stress management intervention for smoking relapse prevention. Contemporary Clinical Trials, 109, 106534

[2] Spring, B. (2018). Protocol Title: Sense2Stop: Mobile Sensor Data to Knowledge (Version Number 24). Internal report (Northwestern University Institutional Review Board-approved study protocol).

[3] Sarker, H., Tyburski, M., Rahman, M. M., Hovsepian, K., Sharmin, M., Epstein, D. H., ... and Kumar, S. (2016, May). Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 4489-4501).

[4] Sarker, H., Hovsepian, K., Chatterjee, S., Nahum-Shani, I., Murphy, S. A., Spring, B., ..., and Kumar, S. (2017). From markers to interventions: The case of just-in-time stress intervention. In Mobile health (pp. 411-433). Springer, Cham.

[5] Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, and Kumar S. cstress: towards a gold standard for continuous stress assessment in the mobile environment. ACM UbiComp. 2015:493–504.

[6] Hossain, S. M., Hnat, T., Saleheen, N., Nasrin, N. J., Noor, J., Ho, B. J., ..., and Kumar, S. (2017, November). mCerebrum: a mobile sensing software platform for development and validation of digital biomarkers and interventions. In proceedings of the 15th ACM Conference on Embedded Network Sensor Systems (pp. 1-14).

[7] Boruvka, A., Almirall, D., Witkiewitz, K., & Murphy, S. A. (2018). Assessing time-varying causal effect moderation in mobile health. Journal of the American Statistical Association, 113(523), 1112-1121.