Sense2Stop Data Curation Documentation

Contents

As the intended design and planned analyses of the Sense2Stop study have already been described (Battalio et al., 2021), a major focus of this documentation will be on articulating decisions made when constructing the dataset used to test the primary aim hypotheses. Scientific and practical considerations underpinning these decisions are also discussed in this documentation.

1. Primary aim

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

Primary Aim Hypothesis
Delivering (vs. not delivering) a prompt recommending a stress regulation exercise will reduce the likelihood of being stressed in the subsequent two hours; this effect will be stronger if the prompt was delivered when the individual was stressed (vs. not stressed).

The ***primary proximal outcome*** is trichotomous, a classification of each minute during the 120-minute period immediately following micro-randomization as 'probably stressed', 'probably not stressed', or 'physically active'.

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, focusing on information not described in the main Sese2Stop protocol paper (Battalio et al., 2021). Participants were asked to complete a total of 3 lab visits during the study. Since the 3rd lab visit was scheduled to occur *after the participant completes the MRT*, we direct readers to the main Sense2Stop protocol paper for details surrounding the 3rd lab visit.

1st lab visit: A smartphone and wearable devices were loaned to participants. The smartphone contained pre-installed apps, including a data-collection app and three additional self-regulation apps, which included a collection of guided self-regulation exercises. Wearable devices consisted of a chest band and straps to be worn on the left wrist and right wrist. Participants were oriented on study procedures, including that:

- They may access any of the self-regulation apps as often as they wanted.
- Pressing a 'start of day' button each day within the data-collection app was required before the app could send any ***Ecological Momentary Assessment (EMA)*** for that day. The main protocol paper for the Sense2Stop study (Battalio et al., 2021) describes three possible types of EMAs that could have been triggered by the data-collection app.
- Wearable devices should be worn at all times, except when sleeping or bathing (wearable devices were not waterproof). We note that the wearable devices continued to collect data regardless of whether the participant pressed the 'start of day' button, as long as neither the smartphone nor wearable devices were switched off.
- Wearable devices and smartphone must be returned during the 3rd lab visit.

During the 1st lab visit, study staff also (manually) pre-set 'wake time' and 'sleep time' on the smartphone in consultation with the participant. 'Wake time' refers to the time of day at which the data-collection app will automatically notify the participant to press the 'start of day' button within the app. 'Sleep time' refers to the time of day at which the data-collection app will automatically pause all notifications for that day until the next time the 'start of day' button is pressed. Incorporating 'wake time' and 'sleep time' into the study design ensured that the data-collection app did not trigger EMAs when a participant was asleep.

Finally, participants were instructed by study staff to stop smoking the morning of their 2nd lab visit. In effect, participants were asked by study staff to view the date of their 2nd lab visit as their ***Quit Day***.

2nd lab visit: Study staff (manually) activated the micro-randomization capabilities in the data-collection app. 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 micro-randomization capabilities activated, this caused the functionality of the 'start of day' button to change slightly: pressing the 'start of day' button each day within the data-collection app was now required before the app could send prompts to perform a self-regulation exercise **and** EMAs for that day. Notably, study staff did not convey this change in functionality to participants.

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

75 adult smokers between the ages of 18 and 65 years were enrolled into 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 reported 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 inform study staff that they wish to withdraw but did not complete their 2nd lab visit.

Note that both C1 & C2 are prior to the start of the micro-randomized portion of the study. Additionally, study staff beta tested 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 beta testing.

Finally, participants who were never micro-randomized 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.

Table 1. Participants excluded entirely from all analyses.

Criterion violated	No. of participants	Participant IDs
C1	9	204, 209, 210, 220, 232, 236, 237, 239, 246
C2	3	201, 257, 263
C3	5	101, 102, 103, 104, 105
C4	9	206, 215, 217, 218, 230, 241, 247, 254, 270
Total	26	

From here onward, we will only focus on the remaining N = 75-26 = 49 participants. Figures and summary statistics we display from here onward will not include any information from the 26 participants.

Definition of 'First Day' of the MRT, 'Last Day' of the MRT, and 'Quit 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
Quit Day: the date when a participant completed their 2nd lab visit, i.e., Quit Day = First Day

Exception: One participant (Participant 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 on Day 4.

Delayed completion of 2nd lab visit
Only one participant (Participant 251) delayed completing their 2nd lab visit. This participant completed their 2nd lab visit one day later than scheduled due to bad weather, i.e., four (instead of three) days after completing their 1st lab visit. For this specific participant, the 'First Day' and 'Quit Day' will still be the date when they completed their 2nd lab visit and 'Last Day' will still be 10 days after their First Day.

Subsequent implications of 'First Day' and 'Last Day'
Since data collection performed by wearable devices cannot be terminated by study staff 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 the primary analyses. Specifically, the primary analyses will exclude any data collected after 11:59PM on the participant's 'Last Day'.

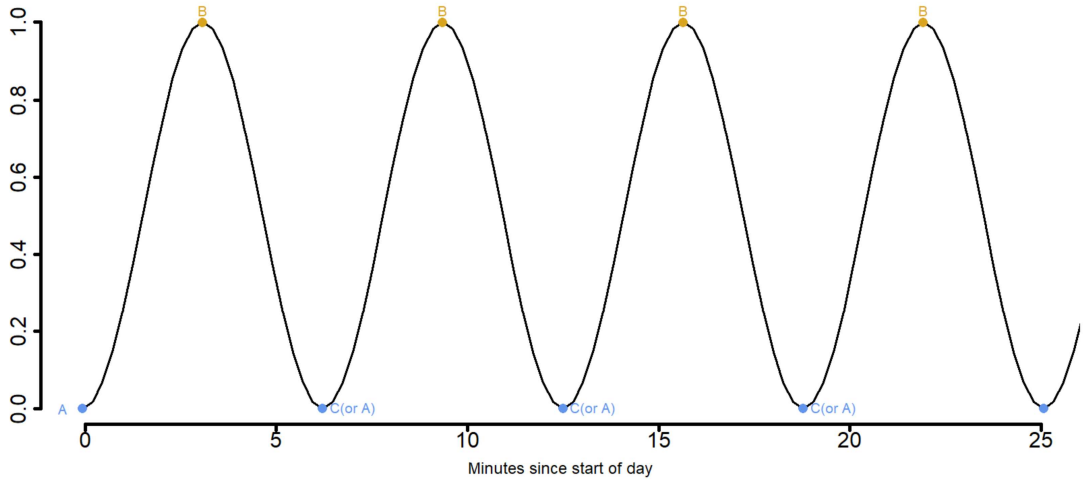
4. Working with sensor-derived assessments to construct the primary proximal outcome

Background

A detection algorithm developed by Sarker and colleagues (Sarker et al., 2016; Sarker et al., 2017) was used to detect when an individual is likely to experience 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 ranging from 0 to 1; a value closer to 1 (closer to 0) signifies that stress is more likely (less likely) to have occurred at that specific moment. We caution the reader that higher (or lower) values in the stress likelihood time series should not necessarily be interpreted as more (or less) intense experience of stress.

Subsequently, ***episodes***, defined as time intervals in which there was 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 between them (see Figure 1 below). We refer to the beginning, peak, and end of an episode as 'A', 'B', and 'C', respectively.

Figure 1. A conceptual stress likelihood time series. In this figure, 'C (or A)' is used to describe instances when 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 was then used to label the episode more specifically as one of three types:

- a ***probably stressed*** episode, if the area under the curve is greater than or equal to a pre-specified threshold *c* and good quality data was available for more than 50% of the time between A and B
- a ***probably not stressed*** episode, if the area under the curve was less than a pre-specified threshold *c* and good quality data was available for more than 50% of the time between A and B
- an ***unknown*** episode, either if good quality data was not available for more than 50% of the time between A to B or if the individual was physically active between A and B

In the Sense2Stop study, c=0.36. 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

Data on episode type and specific time-stamps for A, B, C serve as the starting point for constructing the primary proximal outcome; we refer to this data as the ***episode classification data stream***.

The primary proximal outcome is based on the trichotomous variable $Y_{i,d,t}$, defined to have a value of 'yes', 'no', or 'active', if participant i is within a probably stressed episode, within a probably not stressed episode, or was physically active, respectively, at minute t of day d. Here, since the subscript t indexes each minute within a day, t ranges between 0 and 1439 (=24×60−1); since the subscript d indexes each day between 'First Day' and 'Last Day', d ranges between 0 and 4 (for Participant 213) or between 0 and 10 (for the remaining 48 participants). Next, we will describe the procedure used for constructing the trichotomous variable $Y_{i,d,t}$.

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

Overview: This procedure was motivated by the following:

Observation 1. the episode classification data stream did not differentiate between episodes that were unknown due to the individual being physically active from those that were unknown due to poor data quality

Hence, we needed to use information from other data sources to help distinguish between these two cases.

Observation 2. the duration of time between B and C exceeded 5 minutes in a sizeable number of probably stressed or not-stressed episodes

We believe that the actual duration of time between B and C during stress/no-stress episodes may be shorter than what the data (i.e., the episode classification data stream) suggests. Hence, we devised rules to censor episodes and applied this rule to all types of episodes.

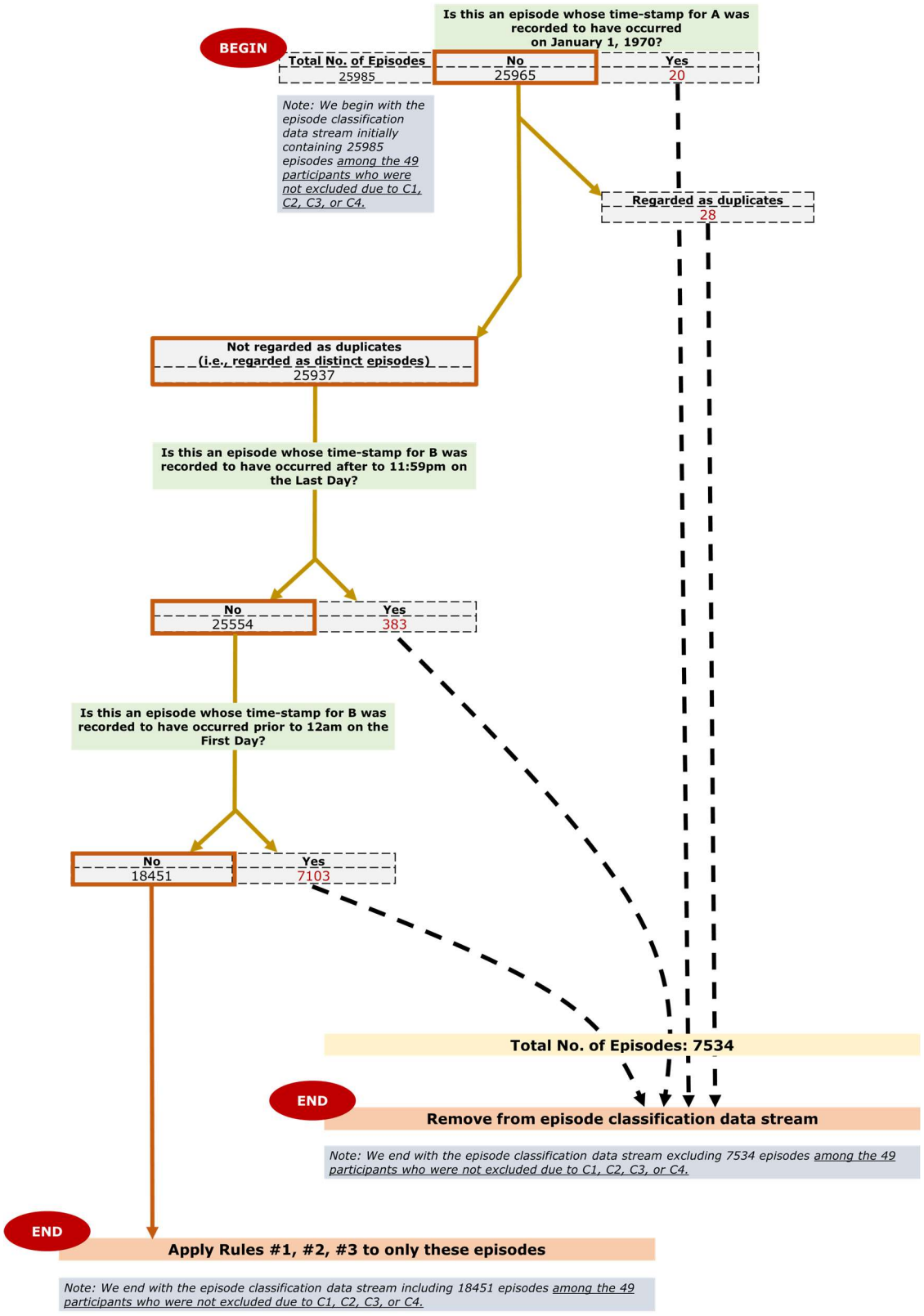
Preliminary data cleaning steps. The following episodes were removed from the episode classification data stream:

1. Episodes whose time-stamp for A was recorded to have occurred on January 1, 1970[1].
2. Episodes regarded as ***duplicates***. Specifically, there were episodes with identical time-stamps for A and B, but differ in their time-stamp for C. Only the episode with the shortest duration of time between A and C was retained while all other episodes were regarded as duplicates.
3. Episodes whose time-stamp for B was recorded to have occurred either prior to 12am on the First Day or after 11:59pm on the Last Day.

Removing the above episodes from the episode classification data stream, left us with 18451 episodes. Of these 18451 episodes, 1074 were labelled as probably stressed, 11606 were labelled as probably not stressed, and 5771 were labelled as unknown. From here onward, we will only focus on these 18451 episodes.

[1] Specifically, these were episodes whose time-stamp for A was recorded to have occurred on 12:00am, January 1, 1970 in Universal Coordinated Time (UTC).

Figure 2. No. of episodes removed from the episode classification data stream after performing the preliminary data cleaning steps.

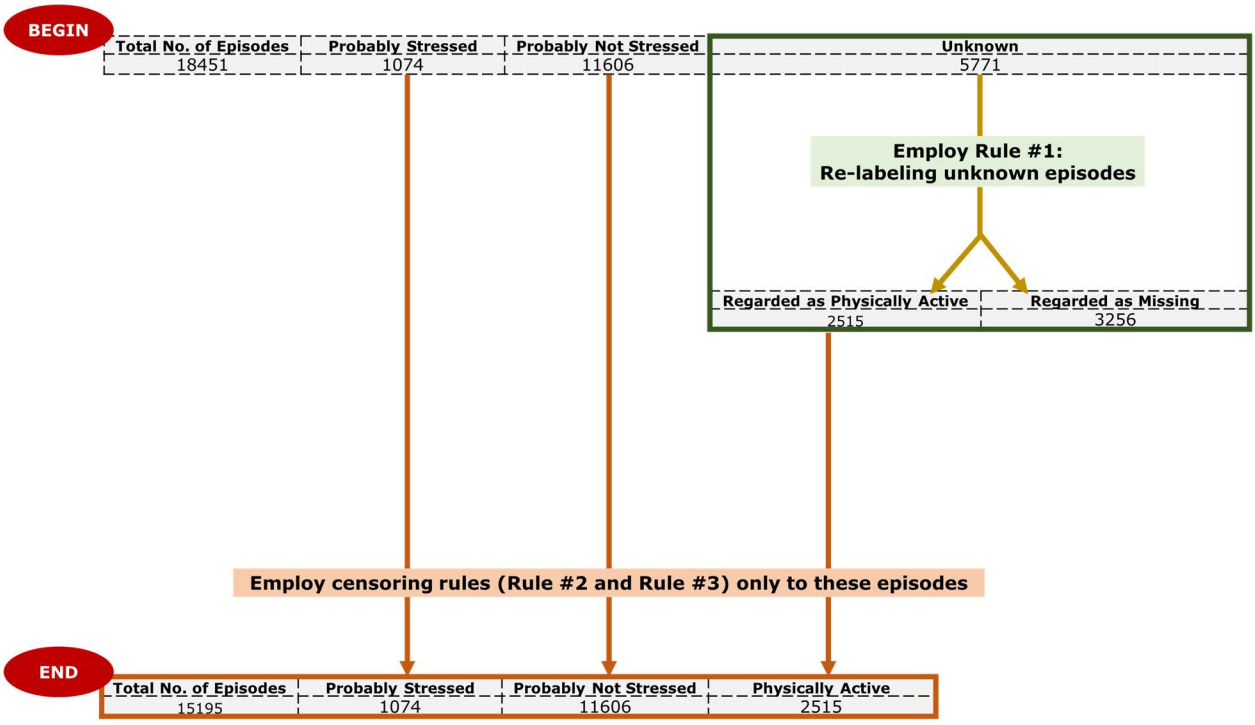


Re-labeling unknown episodes. We use Rule #1, which leverages physical activity data, to differentiate between episodes that were labeled as unknown because the individual was physically active from those labeled as unknown due to poor data quality. In brief, physical activity data was represented as a minute-level indicator for whether physical activity was detected (=1) or not (=0) by an activity detection algorithm. After employing Rule #1, we found that 2515 of the 5771 unknown episodes could be regarded as physically active episodes.

Rule #1: Re-labeling unknown episodes

1 IF physical activity was detected in more than 50% of the minutes between A and B
2 THEN regard the time between A and C as a physically active episode.
3 ELSE regard the information on the trichotomous variable $Y_{i,d,t}$ between A and C as missing.

Figure 3. No. of episodes of each type after employing Rule #1.



Censoring episodes. From here onward, we will focus on the 15195 episodes in the lower left-hand corner of Figure 2. We used two rules (described below) to censor episodes (all types). These rules (described below) were developed with guidance from subject matter experts. We use the term ***censoring*** in this setting to mean that we identified a specific time point in an episode beyond which there is uncertainty as to whether an individual continued to be stressed, not stressed, or physically active.

First, we found that in 7070 (about 47%) out of the 15195 episodes, the length of time between B and C exceeded 5 minutes. Specifically, the maximum length of time between B and C was 664.99 minutes (about 11 hours) among probably stressed episodes, 904 minutes (about 15 hours) among probably not stressed episodes, and 3308 minutes (about 55 hours) among physically active episodes.

We expected the actual duration of time between B and C to be much shorter than observed. We suspect that these long durations (in all episode types) may result from the episode-labeling algorithm (i.e., as implemented in the Sense2Stop study here https://github.com/MD2Korg/stream-processor/blob/master/src/main/java/md2k/mcerebrum/cstress/features/StressEpisodeClassification.java) not using information about data quality after B to determine the specific time-stamp for C. Hence, we used Rule #2 for censoring episodes (any type) for which the time between B and C exceeded 5 minutes. Episodes which were 5 minutes or shorter were not censored. This rule capitalizes on heart rate data, represented as the average number of beats over a 1-minute-long window; heart rate data was one of several sensor-derived assessments used in constructing the stress likelihood time series (Figure 1) for this study. Employing the Rule # 2 required constructing a minute-by-minute binary indicator (i.e., for each minute t of day d) for whether heart rate data was observed.

Rule #2: Censoring episodes (any type) when time between B and C exceeded 5 minutes

1 IF no heart rate data was observed between B and C
2 THEN censor the episode at B AND regard the information on the trichotomous variable $Y_{i,d,t}$ between B and C as missing.
3 ELSE IF some heart rate data was observed between B and C, BUT there is a period of at least five consecutive minutes within B and C having no observed heart rate data
4 THEN censor the episode at the beginning of this no heart rate period (which we denote by ν) AND regard the information on the trichotomous variable $Y_{i,d,t}$ between ν and C as missing.
5 ELSE do not censor the episode

We use C* to denote either the end of an episode if the episode was not censored or the specific time-point at which an episode was censored after Rule #2 was employed. Specifically:

- If an episode was not censored, then C* would be C.
- If an episode was censored, then C* would be either B or ν.

.

Figure 4. No. of episodes impacted by censoring rules in Rule #2. Among episodes having more than 5 minutes elapsed between B and C, we display percentiles on minutes elapsed between B and C.

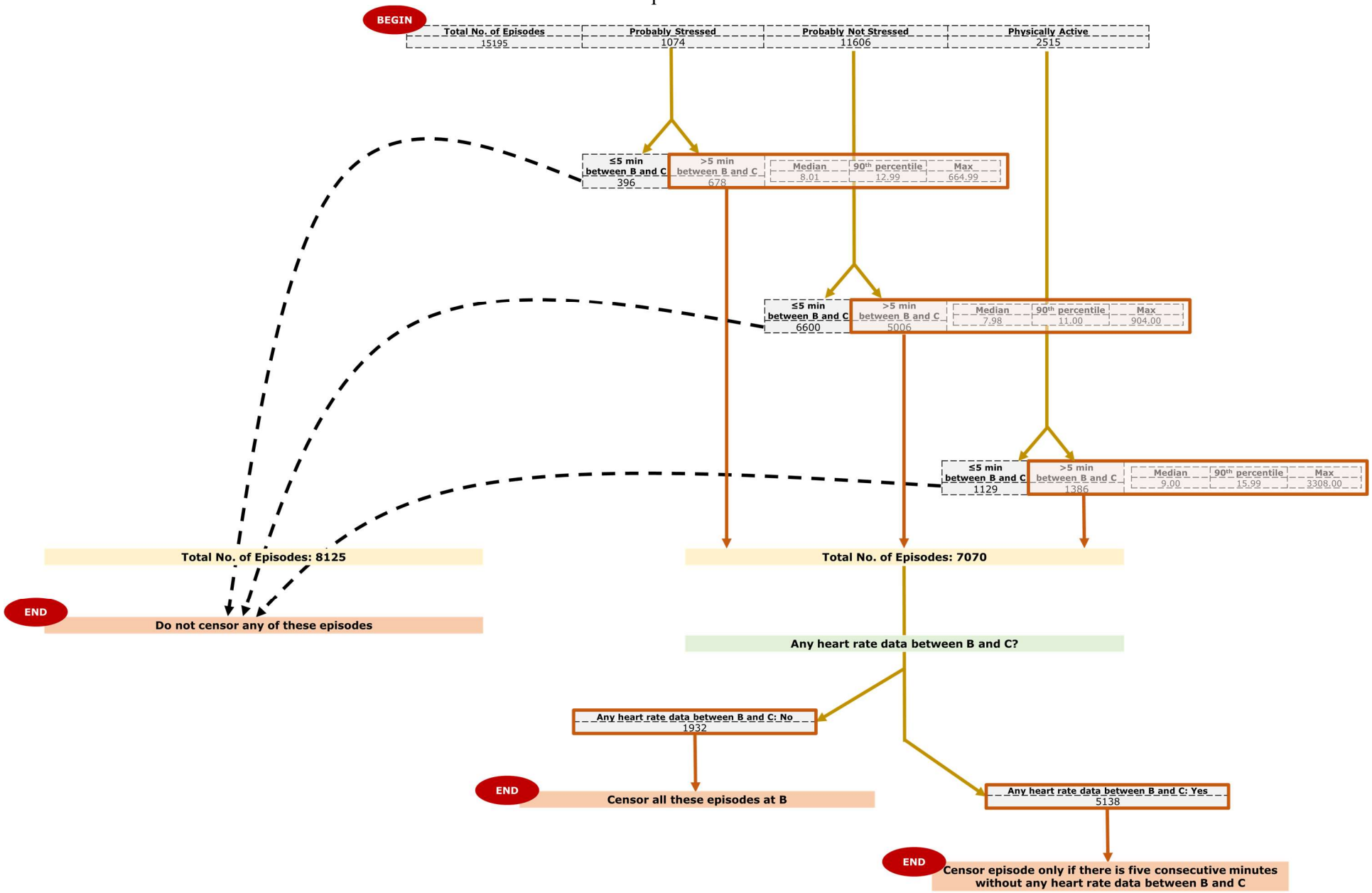


Figure 5. Of the episodes whose time between B and C exceeded 5 minutes (the 7070 episodes in Figure 4), we display the percentage which were censored after applying the Rule #2.

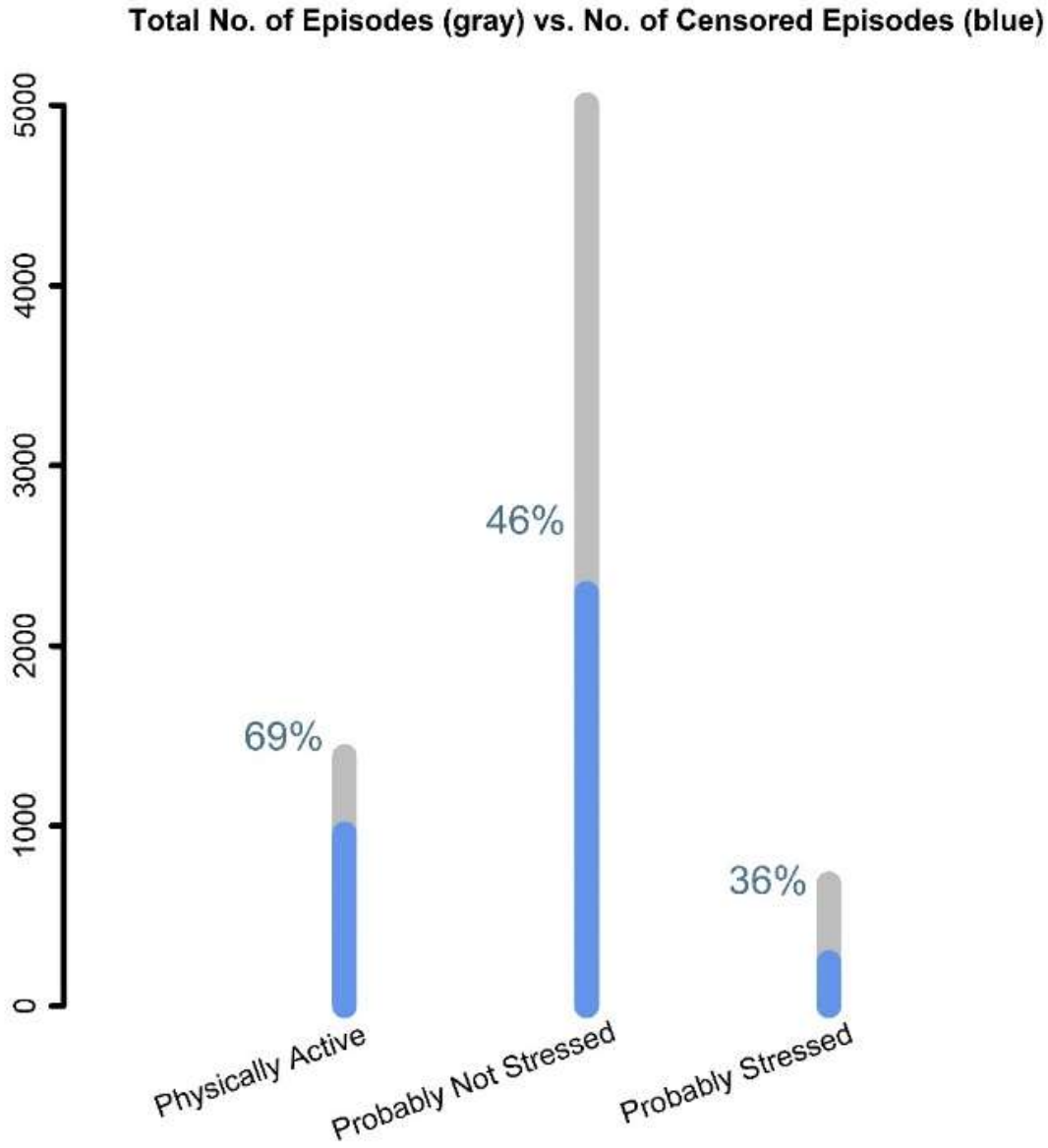


Table 2. Among the 7070 episodes
in which time between B and C exceeded 5 minutes: Distributions of time between B and C* after applying Rule #2

Episode Type	No. of episodes	Median	90th percentile	Max
Probably Stressed	678	6.01	11.00	18.02
Probably Not Stressed	5006	5.01	9.99	18.01
Physically Active	1386	0.00	10.00	20.04
Total	7070			

Second, after employing Rule #2, we observed (see Table 3) that the length of episodes is still higher than expected: the maximum time between A and C* was 4116 minutes (about 69 hours) among probably stressed episodes, 809 minutes (about 13 hours) among probably not stressed episodes, and 75 minutes (about 1.25 hours) among physically active episodes.

Table 3. Distributions of time between A and C* after employing Rule #2 among the 15195 episodes

Episode Type	No. of episodes	Median	90th percentile	Max
Probably Stressed	1074	8.01	15.99	4116.01
Probably Not Stressed	11606	7.99	17.02	809.00
Physically Active	2515	9.00	18.86	75.00
Total	15195			

We employ Rule #3 to further censor episodes based on a specified threshold of 17 minutes—this threshold is the 90th percentile of the time between A and C*.

Rule #3: Re-defining C* after employing Rule #2

```
1       IF time between A and C* > 17 minutes
2               THEN re-define C* to be 17 minutes after A
3       ELSE do not re-define C*
```

After employing Rule #3:

- If an episode was not censored, then C* would be C.
- If an episode was censored, then C* would be either B, ν, or 17 minutes after A.

Table 4 displays the median, 90th percentile and max time between A and C* (in minutes) after applying Rule #3.

Table 4. Distributions of time between A and C* after employing Rule #3 among the 15195 episodes

Episode Type	No. of episodes	Median	90th percentile	Max
Probably Stressed	1074	8.01	15.99	17.00
Probably Not Stressed	11606	7.99	17.00	17.00
Physically Active	2515	9.00	17.00	17.00
Total	15195			

Figure 6. The percentage of episodes which were censored after employing Rule #3 (of the 15195 episodes; see Figure 3).

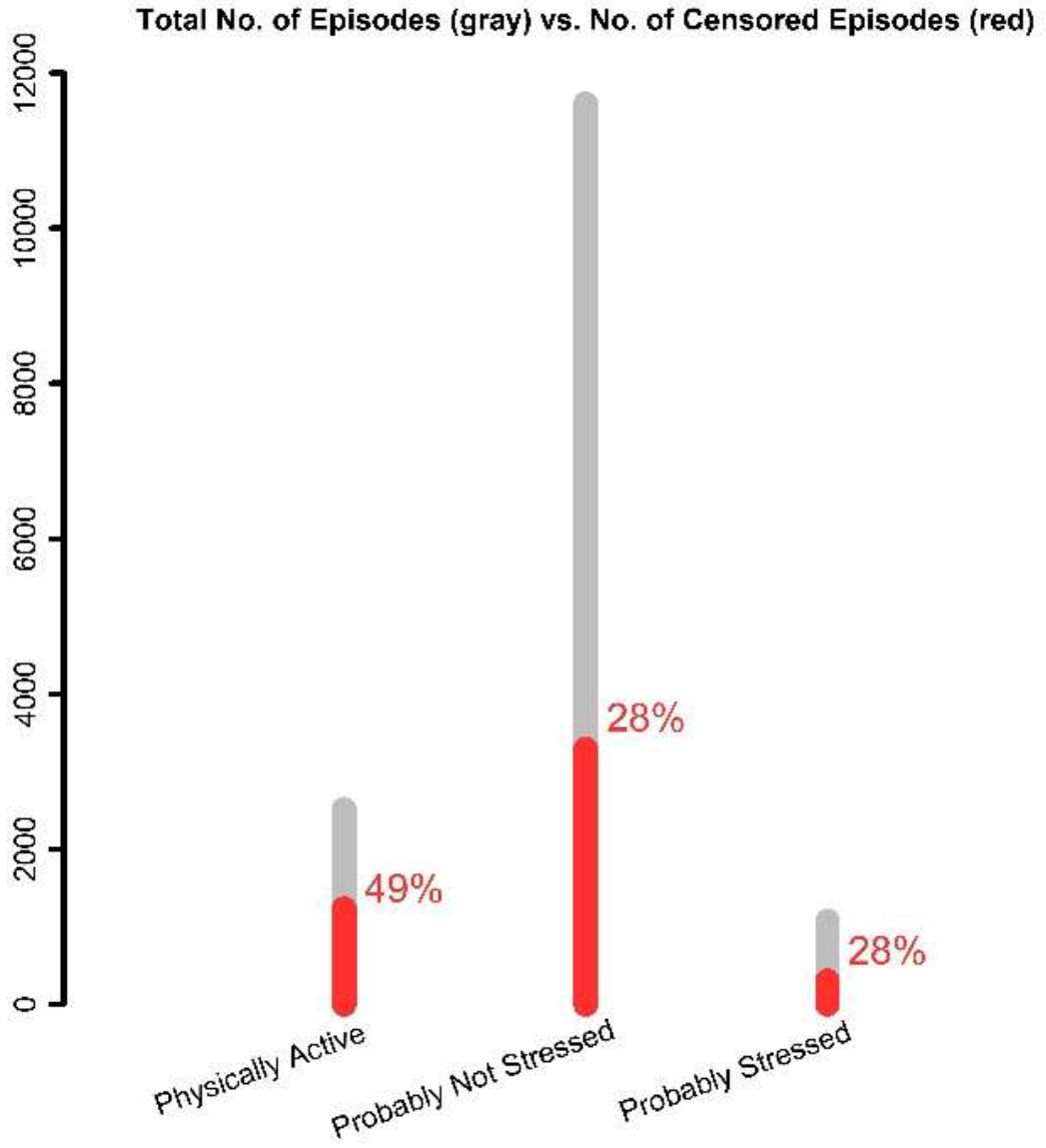
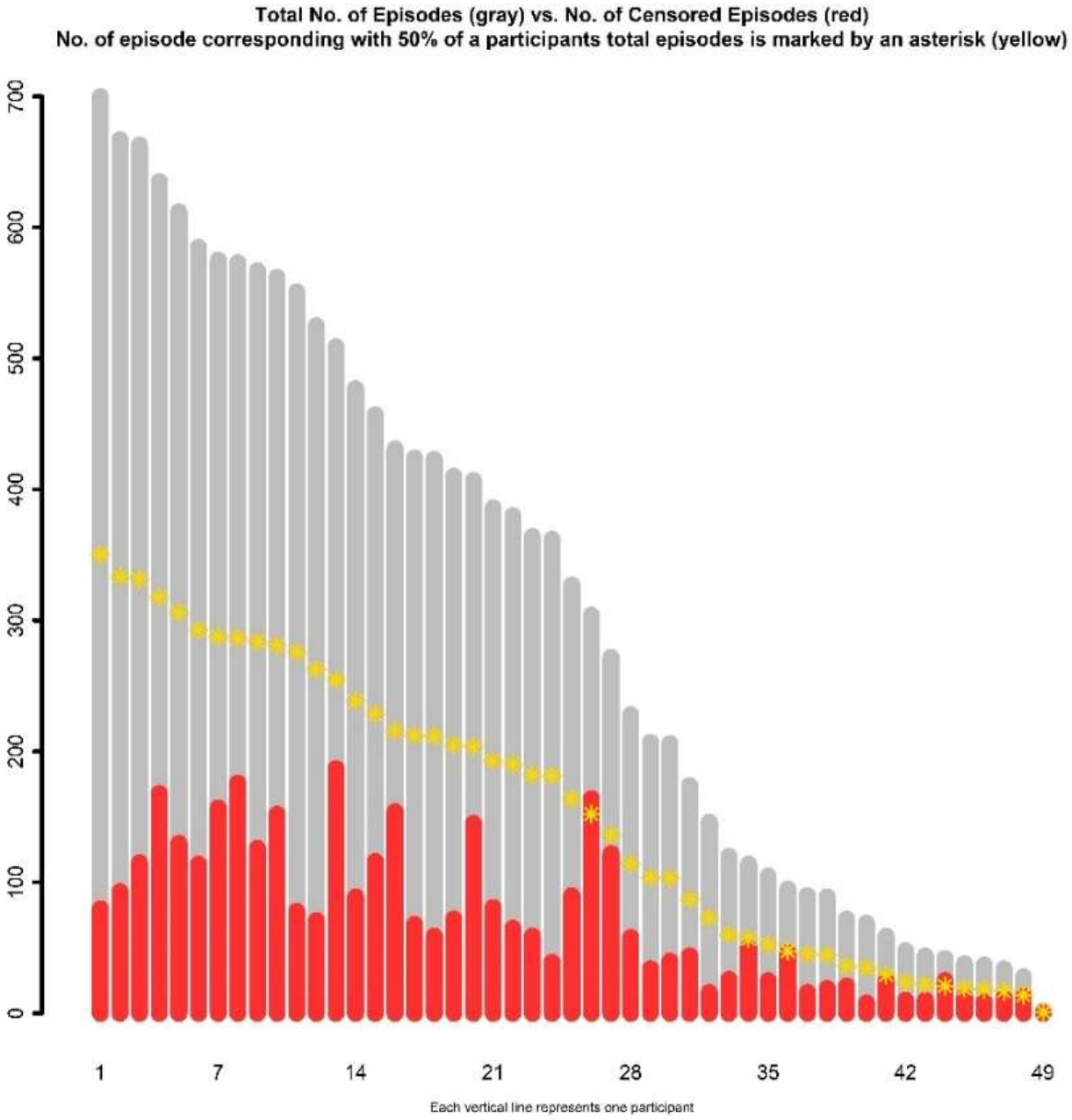


Figure 7. The no. of episodes which were censored, by participant, after employing Rule #3 (of the 15195 episodes; see Figure 3). The figure shows that for the vast majority of participants, less than 50% of their episodes were censored (depicted by red bars lying below the yellow asterisks).



Operational definition of $Y_{i,d,t}$. We regard participant *i* at minute *t* of day *d* as:

- ***probably stressed*** (i.e., assign a value of ***'yes'*** to the trichotomous variable $Y_{i,d,t}$) if participant *i* was within a probably stressed episode at minute *t* of day *d*
- ***probably not stressed*** (i.e., assign a value of ***'no'*** to the trichotomous variable $Y_{i,d,t}$) if participant *i* was within a probably not stressed episode at minute *t* of day *d*
- ***physically active*** (i.e., assign a value of ***'active'*** to the trichotomous variable $Y_{i,d,t}$) if participant *i* was within a physically active episode at minute *t* of day *d*

Missing values in $Y_{i,d,t}$. The trichotomous variable $Y_{i,d,t}$ is considered ***missing*** if we are not able to assign a value of 'yes', 'no', or 'active' to $Y_{i,d,t}$ based on the operational definition of $Y_{i,d,t}$ above.

5. Eligibility for micro-randomization

Background

A participant was regarded as ***eligible for micro-randomization*** at minute *t* of day *d* (and thus micro-randomization occur) if they met a pre-specified set of criteria at that minute. A subset of the criteria were described in the main protocol paper for Sense2Stop (Battalio, et al., 2021). In addition, micro-randomization did not occur at minute *t* of day *d* if (a) the participant did not press the 'start of day' button on that day; or (b) *t* was after 'sleep time' on that day; or (c) *t* was before activation of the 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' were regarded as ineligible for micro-randomization.

Micro-randomization at each eligible minute was ***stratified*** – that is, the probability of an eligible minute *t* of day *d* being micro-randomized to a prompt versus no prompt was balanced according to two factors: (1) whether or not the participant has experienced a post-quit smoking lapse by eligible minute *t* of day *d*; and (2) whether an eligible minute *t* of day *d* was within a probably stressed episode or within a probably not stressed episode. Specifically, the randomization probabilities were constructed with the goal that if the participant has not yet lapsed, an average of 3 prompts will be delivered per day (a daily average of 1.5 during probably stressed episodes and a daily average of 1.5 during probably not stressed episodes); and if the participant has already lapsed, an average of 2.5 prompts will be delivered per day (a daily average of 1 prompt during probably stressed episodes and a daily average of 1.5 during probably not stressed episodes). Additional details are provided in the Sense2Stop main protocol paper (Battalio et al., 2021).

The code for calculating the randomization probabilities can be found here: https://github.com/MD2Korg/mCerebrum-EMAScheduler/blob/master/ema_scheduler/src/main/java/org/md2k/ema_scheduler/scheduler/emi/ProbabilityEMI.java

We use $I_{i,d,t}$ to denote whether ($I_{i,d,t}$ =1) or not ($I_{i,d,t}$ =0) participant *i* was eligible at minute *t* on day *d* for micro-randomization. From here onward, we use *t** to denote a particular eligible minute on day *d* for participant *i*.

Procedure for constructing the variable $I_{i,d,t}$

Data collection in the study provided the following information about each micro-randomization at *t**:

D1. The specific timestamp[2] at which minute *t** was deemed eligible for micro-randomization

[2] This timestamp was represented as milliseconds elapsed since 12am of January 1, 1970 in Universal Coordinated Time (UTC)

D2. The randomization probability associated with *t**, i.e., the probability that a prompt would be delivered
D3. The stress stratum associated with *t**, i.e., whether *t** was labeled within a probably stressed or within a probably not-stressed episode for the purpose of calculating the randomization probability at *t**.
D4. The assigned intervention option at *t**, i.e., whether the randomization resulted in a decision to deliver a prompt or not to deliver a prompt

The data included 5506 timestamps of eligible minutes between the 'First Day' and 'Last Day' across all 49 participants. In other words, there were 5506 micro-randomizations between the 'First Day' and the 'Last Day' across the 49 participants.

We re-labeled some of these eligible minutes as ineligible due to the following issues:

Observation 1. The data contained eligible minutes *t** for which the most recent evidence of episode classification was more than 5 minutes prior to *t**. This is a concern because the study was designed such that micro-randomization should occur immediately following the classification of an episode.

Observation 2. The data contained eligible minutes *t** for which the most recent classification was unknown or missing. This is a concern because the study was designed such that micro-randomizations should occur only when *t** is within either a probably stressed or a probably not-stressed episode.

Observation 3. The data contained eligible minutes *t** for which the randomization probabilities were either exactly zero or exactly one; this happened primarily in eligible minutes within probably not stressed episodes. Although the study was designed to allow these probabilities, scenarios for which the probabilities would be exactly one or zero were not expected to occur during the trial. Further, these probabilities violate the positivity assumption inherent in many intent-to-treat estimators (this violation occurs when certain subgroups in a sample rarely or never receive some treatments of interest), including the estimator to be utilized in estimating the treatment effect for this trial's primary aim.

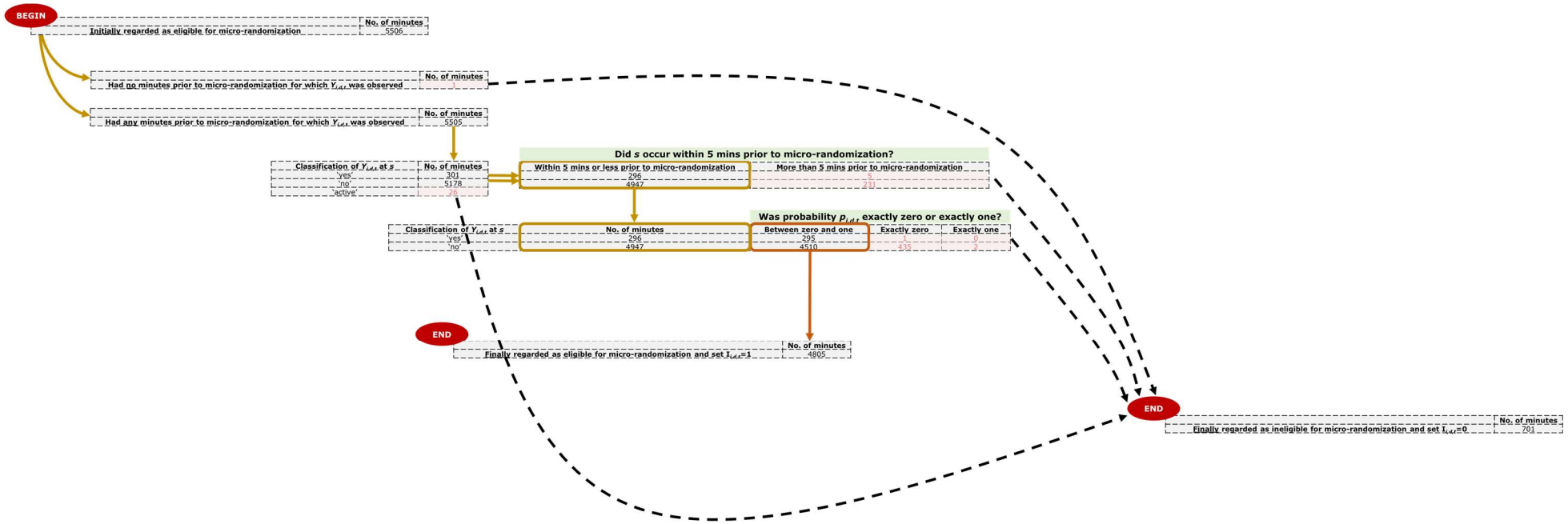
Hence, we employed Rule #4 to re-label *t**'s for which observations 1-3 hold as ineligible for micro-randomization. Figure 8 displays the number of minutes *t** we regarded as ineligible for micro-randomization after applying Rule 4. We ended up with 4805 eligible minutes after Rule 4.

Rule #4: Re-labeling t^*'s as ineligible for micro-randomizations. Below, $p_{i,d,t}$ denotes the randomization probability associated with participant i at minute t of day d.

```
1   Rule 4a:
2   For each t* belonging to participant i on day d:
3       Determine the closest minute s prior to t* for which Y_{i,d,t} was observed.
4       IF s exists, THEN
5           IF t*−s ≤ 5 minutes, THEN
6               IF Y_{i,d,t} was physically active or missing
7                   THEN set I_{i,d,t} = 0
8               ELSE keep I_{i,d,t} = 1
9           ELSE (IF t*−s > 5, THEN)
10              Set I_{i,d,t} = 0
11      ELSE (if s does not exist, THEN)
12          Set I_{i,d,t} = 0
```

```
13  Rule 4b:
14  For each t* belonging to participant i on day d:
15  IF p_{i,d,t} =0 OR p_{i,d,t} =1 , THEN
16      Set I_{i,d,t} = 0
17  ELSE (if p_{i,d,t} >0 AND p_{i,d,t} < 1)
18      keep I_{i,d,t} = 1
```

Figure 8. Number of eligible minutes t^* re-labeled as ineligible after employing Rule #4.



Additional Information

Classification of $Y_{i,d,t}$ at s	No. of minutes	No. of seconds between s and when micro-randomization occurred		
		Median	90th percentile	Max
'yes'	301	2.59	62.43	51592.68
'no'	5178	2.50	122.13	81482.65

6. Intervention decision for each minute between 'First Day' and 'Last Day'

Notation:
Recall that in Section 5, we introduced the notation t^* to denote a particular eligible minute on day d for participant i. In the current (Section 6) and following (Section 7) section, the notation t will be used to denote any minute of participant i on day d. Hence, the set of eligible minutes t^* is a subset of the set of all minutes t of participant i on day d.

Procedure for constructing the variable $A_{i,d,t}$:
Let $A_{i,d,t}$ be a dichotomous variable denoting whether the decision for participant i at minute t on day d was to deliver any prompt ($A_{i,d,t}=1$) or no prompt ($A_{i,d,t}=0$). Since it is possible for micro-randomization to occur at ineligible minutes (i.e., among the 701 minutes in Figure 8), the value of $A_{i,d,t}$ will be set according to two factors: whether participant i at minute t on day d was eligible for micro-randomization and whether participant i at minute t on day d was actually micro-randomized. Specifically, for participant i on day d:

- For minutes t with no micro-randomization and were ineligible *(i.e., $t \neq t^*$)*, $A_{i,d,t}=0$.
- For minutes t with micro-randomization and were ineligible *(i.e., $t \neq t^*$)*, set $A_{i,d,t}=1$ if the assigned intervention option was to deliver any prompt; set $A_{i,d,t}=0$ if the assigned intervention option was to deliver no prompt.
- For minutes t which were eligible (i.e., $t = t^*$), set $A_{i,d,t}=1$ if the assigned intervention option was to deliver any prompt; set $A_{i,d,t}=0$ if the assigned intervention option was to deliver no prompt.

The above is summarized in Rule #5; this rule is equivalent to the above and applicable to any minute t of participant i on day d regardless of whether $t \neq t^*$ or $t = t^*$.

Rule #5: Constructing $A_{i,d,t}$ for each minute t on day d of participant i

```
IF micro-randomization did not occur at t
        THEN set A_{i,d,t} =0
ELSE (micro-randomization occurred at t)
        set A_{i,d,t} =1 if the assigned intervention option was to deliver a prompt at t and
        set A_{i,d,t} =0 if the assigned intervention option was to not deliver a prompt at t
```

7. Stratification of micro-randomizations

Background:
The treatment effects of interest for the primary aim hypothesis are the contrasts:

- the effect of delivering a prompt at t^* (vs. not delivering a prompt at t^*), given that no prompts were delivered in the next 120 minutes if t^* is **within a probably stressed episode**
- the effect of delivering a prompt at t^* (vs. not delivering a prompt at t^*), given that no prompts were delivered in the next 120 minutes if t^* is **within a probably not stressed episode**

Hence, to facilitate analysis, we needed to construct an indicator $X_{i,d,t}$ for whether micro-randomization at an eligible minute (i.e., $t = t^*$) was within a probably stressed episode or a probably not stressed episode.

Constructing $X_{i,d,t}$ for minutes t which were ineligible (i.e., $t \neq t^*$):
The value of $X_{i,d,t}$ at minutes t which were ineligible (i.e., $t \neq t^*$) does not matter for the primary aim analysis. Without loss of generality, we regard the value of $X_{i,d,t}$ at ineligible minutes to be missing.

Constructing $X_{i,d,t}$ for minutes t which were eligible (i.e., $t = t^*$):

The data collected from the study permits the following two approaches to constructing X_{i,d,t^*}. The two approaches do not necessarily yield identical values of X_{i,d,t^*} due to the possibility of deviations in the software implementation from the intended study design.

Approach 1. When participant *i* was eligible for micro-randomization at minute t^* of day *d*, **define the variable X_{i,d,t^*} to be the classification of our constructed trichotomous variable *Y* at time *s*, where *s* denotes the nearest minute prior to t^* for which $Y_{i,d,t}$ was observed.**

Approach 2. When participant *i* was eligible for micro-randomization at minute t^* of day *d*, **define the variable X_{i,d,t^*} to be the stratum associated with t^* as recorded in the data (D3 in Section 5).**

Using the 4805 micro-randomizations we regarded as eligible for micro-randomization after applying Rule #4, we compared both approaches to constructing X_{i,d,t^*} and found that the resulting value of X_{i,d,t^*} only differ in 14 (about 0.3%) out of the 4805 micro-randomizations. In other words, in the vast majority of cases (4791 out of the 4805, or about 99.7% of the 4805 micro-randomizations), both approaches yielded identical values for X_{i,d,t^*}. **Decision: We utilized Approach 1 in constructing X_{i,d,t^*}.**

8. Summary of variables that will be used in primary aim analysis

- $Y_{i,d,t}$ (Section 4) – a trichotomous variable taking on values, 'yes' (i.e., probably stressed), 'no' (i.e., probably not stressed), or 'active' (i.e., physically active) when observed for participant *i* at minute *t* of day *d*
- $I_{i,d,t}$ (Section 5) – a dichotomous variable taking on values '1' (i.e., eligible for micro-randomization), or '0' (i.e., ineligible for micro-randomization) for participant *i* at minute *t* of day *d*
- $A_{i,d,t}$ (Section 6) – a dichotomous variable taking on values '1' (i.e., intervention decision was to deliver a prompt), or '0' (i.e., intervention decision was to not deliver a prompt) for participant *i* at minute *t* of day *d*
- $X_{i,d,t}$ (Section 7) – for minutes *t* which were eligible (i.e., $t = t^*$): a dichotomous variable taking on values '1' (i.e., micro-randomization occurred within a probably stressed episode), or '0' (i.e., micro-randomization occurred within a probably not stressed episode) for participant *i* at minute *t* of day *d*; for minutes *t* which were ineligible (i.e., $t \neq t^*$): 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.