Developing an Objective Measure of Mobile Meditation App Habits

Authors: Fowers, Rylan; Stecher, Chad; Chung, Yunro

Keywords: Behavioral habits; Habit formation; Mindfulness meditation; Mental health; mHealth.

Background

Persistent performance of daily health behaviors, such as improved diet, physical activity, and mindfulness meditation can improve physical and mental health outcomes and lower healthcare costs. However, the formation of healthy behavioral habits has proven difficult for many, and most existing behavior interventions fail to assist in the formation process. One barrier to designing more successful interventions is the difficulty of measuring a habit's strength. Currently, behavioral habit strength measures are based on self-reported survey response data, which has known limitations and suffers from potential biases. The goal of this research was to construct and test a novel objective measure of habit strength based on passively observed behavior collected from mobile application data.

Methods

Summary

To compare performance with the survey-based habit strength responses, meditation behavior data were collected from 5572 users of a mobile-health application. The users were surveyed at 6 different times throughout one year, and the habit strength measurements were recorded. Then, from the actual meditation usage of the mobile-health application, metrics were obtained from the 30 days before each survey response. These metrics included basic information on the mindfulness meditation sessions: session duration, and count of sessions by the time of day (morning, afternoon, evening, or night). In addition to these, an engineered daily consistency metric was built (BCT). This consistency metric scores meditation session timing between consecutive days and penalizes days with no meditation. A panel data negative binomial regression model was fit to model the survey-based outcome using these observed behaviors.

Behavior Consistency Score (BCT)

The objective consistency metric was designed to compare two time series of behavior. The time series are assumed to be binary indicators of behavior (i.e., a 1 if the person was acting in the behavior and a 0 if the person was not acting in the behavior). For this paper, the time series were minute-level meditation data from the application Calm. The behavior consistency algorithm can be used to calculate consistency in timing and duration, however, we will focus on the timing consistency since that was the main metric used in this analysis.

The timing consistency algorithm compares the start times of two behaviors. For example, if a person started a behavior at 1:00 PM on day 1 and on the previous day they also started the behavior at 1:00 PM, then they would be perfectly consistent with their timing. This would give them a behavior consistency score of 1 (the max level of consistency). When comparing two days of use, the behavior consistency score will return a value between 0 and 1, with 0 indicating lack of consistency and 1 indicating max consistency. However, sometimes individuals will fail to perform a behavior on a given day. When a day of behavior is being compared to a day of no behavior, then the algorithm will score this as a 0 (no consistency). Lastly, sometimes individuals will fail to perform a behavior on consecutive days, this means the algorithm would be comparing two days of no behavior. In this case, the algorithm returns a -1 as a penalty for missing consecutive days of the behavior. These different cases are displayed in the following formula:

$$\text{Score}(d_1, d_2) = \begin{cases} (0,1], & d_1 > 0 \text{ and } d_2 > 0 \\ -1, & d_1 = 0 \text{ and } d_2 = 0 \\ 0, & \text{otherwise} \end{cases}$$

To allow for flexibility in start times, a window parameter can be used. This window parameter will allow for maximum consistency if the difference in start time is within the given window. For example, using a window of 60 minutes, if a person started a behavior at 1:00 PM on day 1 and on the previous day they started at 1:30 PM, then they would score 1 since the start time was within the 60-minute window. *Figure 1* shows the varying penalties for different windows (a penalty of 1 means a 0 for consistency).

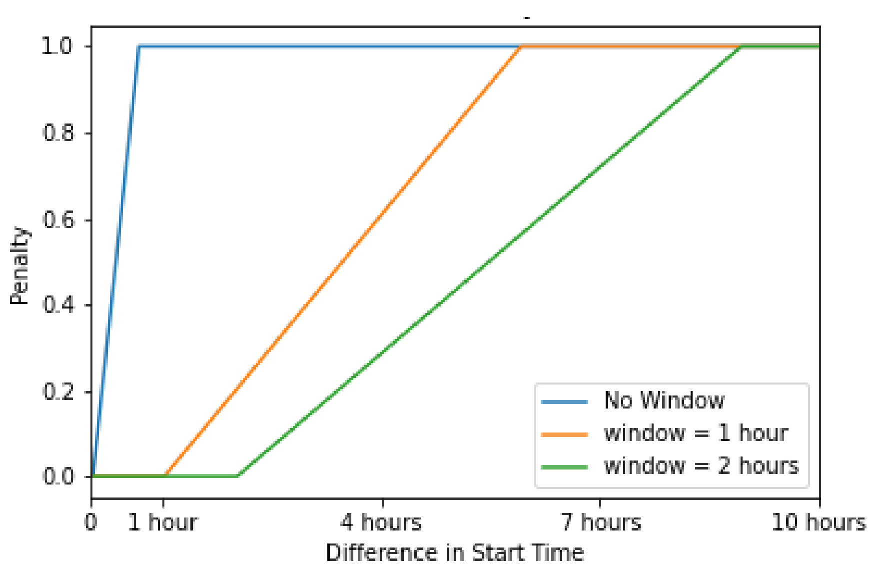

Figure 1: Varying behavior consistency scores for different windows

Now that the underlying piecewise nature of the algorithm and window parameter are understood, we will now discuss how the consistency in time is calculated. Behavior consistency in time is calculated as follows:

$$\tau = 1 - \min(1, \Delta \sqrt{\frac{w}{|s|}})$$

$$\Delta = \left(\frac{|s_1 - s_2|}{w}\right) - 1$$

Where s_i is the start time for each day, Δ is the number of windows different, $|s|$ is the length of the time series, and w is a window size parameter chosen by the user.

Let's walk through two examples using a window size of 60 minutes. Suppose we are comparing two sessions of behavior on two days using a minute-level time series. For the first example, on day *x* the user began their behavior at 1:00 AM, and on day *y* the user started their behavior at 2:00 AM. For the second example, on day *x* the user began their behavior at 12:00 AM, and on day *y* the user started their behavior at 2:00 AM. First, we calculate the difference in windows Δs:

$$\Delta = \left(\frac{|60 - 120|}{60}\right) - 1 = 0$$

$$\Delta = \left(\frac{|0 - 120|}{60}\right) - 1 = 1$$

The first example was within the 60-minute window, so their delta is 0. However, the second example shows the behavior was off by 1 window. Next, we calculate τ, our timing consistency metric for the two examples:

$$\tau = 1 - \min(1, 0 \sqrt{\frac{60}{1440}}) = 1$$

$$\tau = 1 - \min(1, 1 \sqrt{\frac{60}{1440}}) = 1\text{-}.2 = 0.8$$

The length of the time series is 1440 since there are 1440 minutes in a day (60*24). The timing consistency for the first example is a 1 since they were able to start their behavior within the hour window. The timing consistency for the second example is 0.8 since they were outside their window but were still fairly consistent in their start time.

It should be noted that sometimes a person may engage in a behavior multiple times in a single day. This means that the algorithm may need to compare two sessions of behavior with one session of behavior on the previous day. When this is the case, the algorithm will align the sessions of behavior in temporal order, calculate the consistency between all sessions, and take the average as the output. For example, suppose a user meditated at 1:00 PM and 5:00 PM on a given day, but on the previous day, they only meditated once at 1:00 PM. The algorithm would compare the 1:00 PM session with the 1:00 PM session and the 5:00 PM session with no session. This would return scores of 1 (for perfect consistency in timing at 1:00 PM) and 0 (for no consistency in timing at 5:00 PM). The final output would be the average of those two values, which would be 0.5.

To calculate the daily consistency in time for this analysis, each day was compared to the previous day using a window size of 60 minutes. Then, the sum of the scores in the 30 days before the survey responses, were used, along with the other features outlined above, as regressors in a panel data regression to fit the self-reported habit index survey response.

Panel Regression

Each user was surveyed six times throughout the year. This allowed for a panel data setup for the regression analysis. The users were surveyed at 6 different times throughout the year, and the habit strength measurements were recorded. Then, from the actual meditation usage of the mobile-health application, regressors were obtained from the 30-days before each survey response. The regressors used in the model were session duration, count of sessions by the time of day (morning, afternoon, evening, or night), and the time consistency metric. The regression equation was defined as follows:

$$\boldsymbol{SRHI \sim BCT + Duration + Morning + Afternoon + Evening + Night}$$

Before being analyzed in the regression the data was processed. First, rows with missing regressors were dropped. Then, to make a more balanced panel, any individual with 2 or fewer survey responses (observations) was dropped. This left the data with those individuals that had 3 or more survey responses (longitudinal observations). Lastly, the most extreme outliers were dropped (any above the 99^{th} quantile) to avoid unrealistic observations.

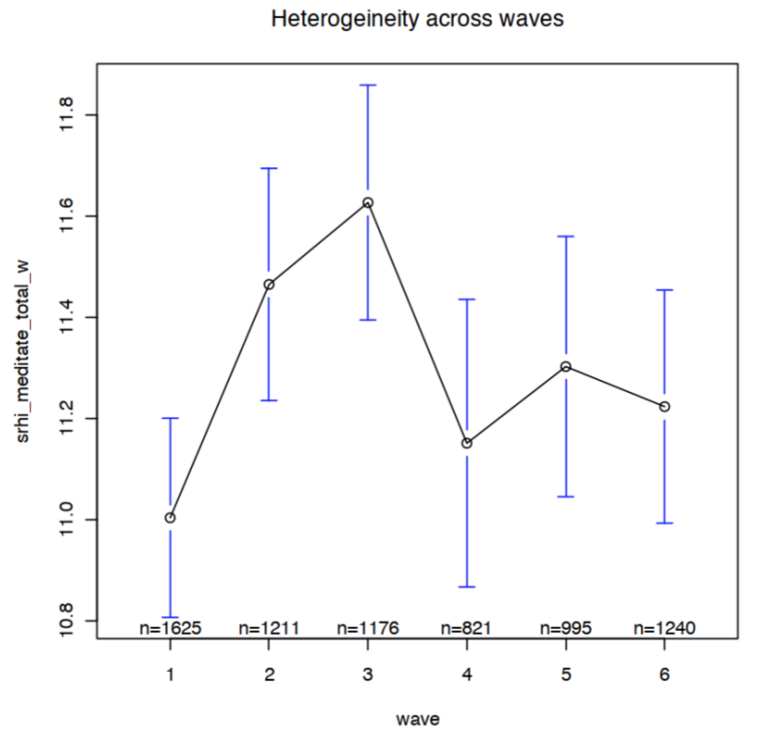

Figure 2: SRHI through the waves

A Breusch-Pagan test for unbalanced panels showed random effects should be used over a pooled regression (p=0.00). Additionally, a Hausman test showed that a fixed-effects model should be used over a random-effects model (p=0.00). However, since fixed effects drop demographics, three models were used in the final output. A random-effects model, a random-effects model with demographics (used to compare the model with and without demographics), and the fixed effects model (drops demographics).

Results

Users were between 18 and 114 years old (mean 41.0 [SD 11.2]), primarily female (68.2%) and white (77.1%). The SRHI ranges from 4 to 20 (strongest habit) with an average of 11.3 (SD=4.10). The behavioral consistency in time (BCT) was on average -8.22 (SD=16.6). The average meditation time in 30-days before the survey was 133.5 minutes (SD=173.8). The remaining feature descriptive statistics can be seen in *Table 1.*

	SRHI	BCT	Duration	Morning	Afternoon	Evening	Night
count	7068	7068	7068	7068	7068	7068	7068
mean	11.28	-8.22	133.49	4.51	1.86	1.65	2.93
std	4.10	16.62	173.80	8.89	4.08	3.70	6.51
min	4.00	-31.00	0.00	0.00	0.00	0.00	0.00
max	20.00	30.00	1011.31	54.00	27.00	25.00	39.00

Table 1: Descriptive statistics of outcome (SRHI) and regressors (duration reported in minutes).

Figure 3 shows the relationship between the objective measurement (BCT) and the reported measurement (SRHI). Those who reported higher habit strength also were more consistent in their timing. In addition, as the SRHI increases so does the average BCT.

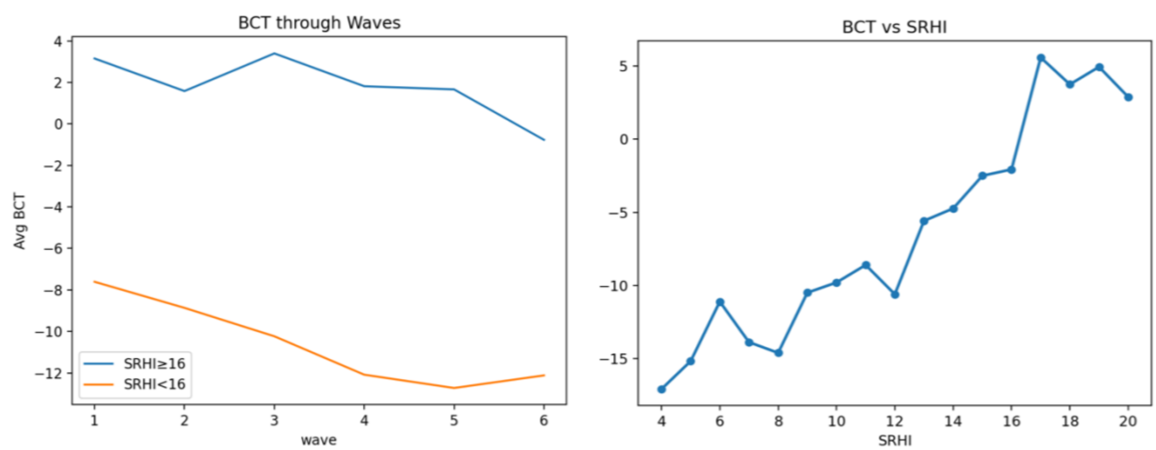

Figure 3: Behavior consistency in time (BCT) A: The mean BCT for high and low SRHI through each wave. B: The mean BCT for each SRHI level.

Figure 4 shows the timing of the day throughout the waves. For the most part, these stay pretty consistent, with morning meditation being the most popular and evening meditation being the least popular.

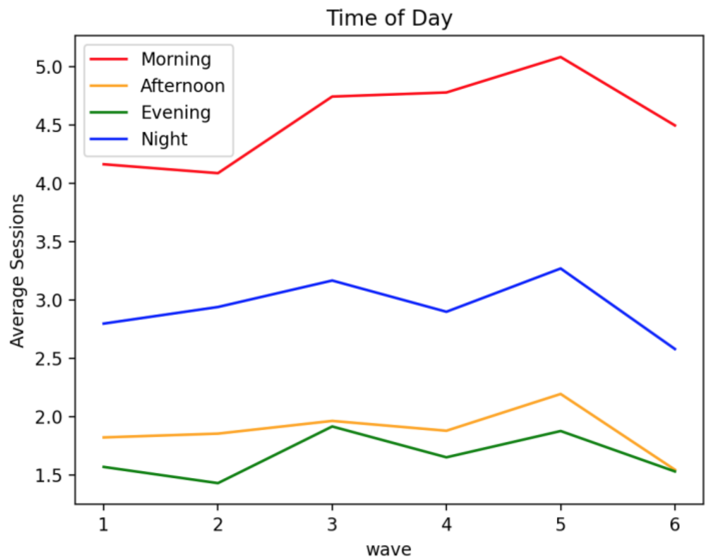

Figure 4: Sessions by the time of day through the waves

The daily consistency metric ($p \leq 0.001$) and the number of meditation sessions in the morning ($p \leq 0.001$) and night ($p \leq 0.001$) were significantly associated with the survey-based habit strength measure. Results suggest that meditation duration ($p=0.62$), number of meditation sessions in the afternoon ($p=0.03$), and the number of meditation sessions in the evening ($p=0.12$), were not significant (at the alpha=0.01 level) in modeling the survey-based habit strength metric. All variables were positively associated with the outcome. The r-squared of the three models were 0.12, 0.12, and 0.04.

	(1)	(2)	(3)
Regressors	SRHI	SRHI	SRHI
BCT	0.0323*** (0.0039)	0.0323*** (0.00391)	0.0202*** (0.0044)
Duration	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Morning	0.0638*** (0.0102)	0.0599*** (0.0103)	0.0473*** (0.0122)
Afternoon	0.0296* (0.0147)	0.0291* (0.0148)	0.0381 (0.0179)
Evening	0.0275 (0.0140)	0.0273 (0.0141)	0.0253* (0.0164)
Night	0.0409** (0.0127)	0.0418** (0.0129)	0.0108 (0.0147)
Demographics Fixed Effects	No No	Yes No	No Yes

*Table 2: Panel regression results. *** for p<0.001, ** for p<0.01 and * for p<0.05.*

Discussion

This strong relationship between SRHI and BCT demonstrates a potential objective alternative to self-reported habit measures. When survey data is not available, this objective measure could potentially be used as a replacement to estimate habit formation and habit strength.

Conclusion

This indicates that individuals are more likely to self-report strong habit measures when meditating in the morning and showing higher daily consistency scores in timing. This demonstrates the potential efficacy of the novel consistency metric; in that, it shows a great significant association with the self-reported survey habit strength measurements. Furthermore, this also provides further insight into the efficacy of the self-reported measure as it can be modeled well by daily consistency measures.

References

1. Jardim TV, Mozaffarian D, Abrahams-Gessel S, et al. Cardiometabolic disease costs associated with suboptimal diet in the United States: A cost analysis based on a microsimulation model. *PLoS Med* 2019; 16: e1002981.

2. Iuga AO, McGuire MJ. Adherence and health care costs. *Risk Manag Healthc Policy* 2014; 7: 35–44.

3. Haith AM, Krakauer JW. The multiple effects of practice: skill, habit, and reduced cognitive load. *Curr Opin Behav Sci* 2018; 20: 196–201.

4. Stecher C, Berardi V, Fowers R, et al. *Using mHealth Data to Identify Mindfulness Meditation Habits and the Associated Mental Health Benefits (Preprint)*. 2021. Epub ahead of print 21 January 2021. DOI: 10.2196/preprints.27282.