SigmaScheduling: Uncertainty-Informed Scheduling of Decision Points for Intelligent Mobile Health Interventions

Asim H. Gazi*, *Member, IEEE*, Bhanu Teja Gullapalli*,
Daiqi Gao, Benjamin M. Marlin, Vivek Shetty, and Susan A. Murphy

***Abstract*—Timely decision making is critical to the effectiveness of mobile health (mHealth) interventions. At predefined timepoints called "decision points," intelligent mHealth systems such as just-in-time adaptive interventions (JITAIs) estimate an individual's biobehavioral context from sensor or survey data and determine whether and how to intervene. For interventions targeting habitual behavior (e.g., oral hygiene), effectiveness often hinges on delivering support shortly before the target behavior is likely to occur. Current practice schedules decision points at a fixed interval (e.g., one hour) before user-provided behavior times, and the fixed interval is kept the same for all individuals. However, this one-size-fits-all approach performs poorly for individuals with irregular routines, often scheduling decision points after the target behavior has already occurred, rendering interventions ineffective. In this paper, we propose SigmaScheduling, a method to dynamically schedule decision points based on uncertainty in predicted behavior times. When behavior timing is more predictable, SigmaScheduling schedules decision points closer to the predicted behavior time; when timing is less certain, SigmaScheduling schedules decision points earlier, increasing the likelihood of timely intervention. We evaluated SigmaScheduling using real-world data from 68 participants in a 10-week trial of Oralytics, a JITAI designed to improve daily toothbrushing. SigmaScheduling increased the likelihood that decision points preceded brushing events in at least 70% of cases, preserving opportunities to intervene and impact behavior. Our results indicate that SigmaScheduling can advance precision mHealth, particularly for JITAIs targeting time-sensitive, habitual behaviors such as oral hygiene or dietary habits.**

***Index Terms*—mobile health, closed-loop system, uncertainty quantification, micro-randomized trial, behavior prediction**

I. Introduction

At predefined timepoints called "decision points," just-in-time adaptive interventions (JITAIs) intelligently decide whether and how to provide mobile health support based on biobehavioral sensor or survey data [1]. A challenge for many JITAIs is that their effectiveness can depend on delivering interventions shortly before a target behavior, but identifying such pre-behavior time windows in real-time is often infeasible. For instance, while wrist-worn sensors can be used to detect eating [2], a reminder to take medication prior to a meal would be rendered ineffective if triggered only after the meal has already begun [3]. More commonly, the challenge arises due to practical constraints in implementation. For the "Oralytics" JITAI, which leverages a commercially available smart toothbrush and a companion app to promote oral self-care [4], data from the brush becomes available only once each night. As a result, decision points for a given day must be scheduled in advance based on prior brushing patterns to anticipate – rather than respond to – forthcoming behavior.

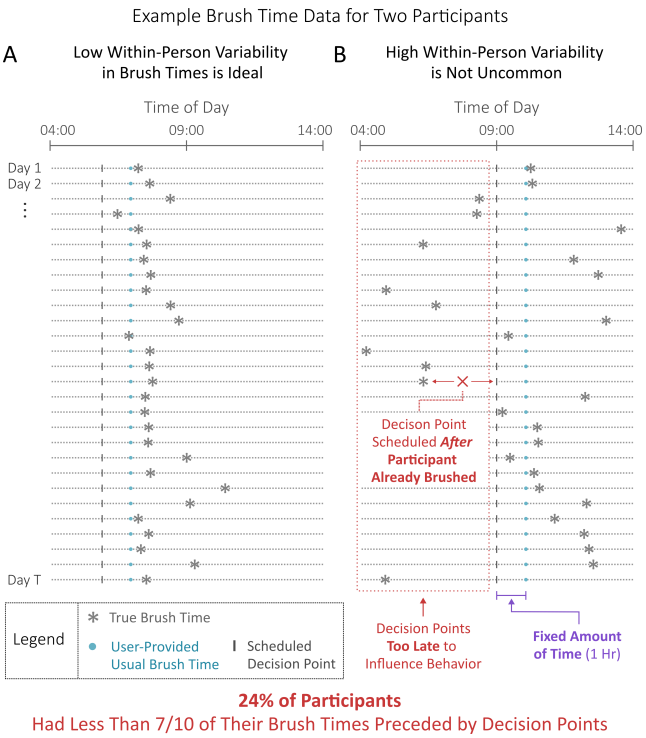

Fig. 1. Example data illustrating heterogeneity in the predictability of routines for two participants in a trial of "Oralytics." For each participant, 28 days of morning brush time data are shown. Days are stacked vertically, with each day's times plotted horizontally. (A) Data from a participant that exhibited less variability in brush times with respect to their user-provided times. (B) Data from a participant that exhibited more variability.

The status quo approach to scheduling decision points is to ask individuals to provide expected behavior times prior to JITAI deployment and then schedule decision points the same fixed duration – for all users – before expected behavior times [5]–[7]. While this status quo may perform satisfactorily for individuals with relatively predictable routines, the status quo

*Equal contribution. This research was supported by the National Institutes of Health (K99EB037411, P50DA054039, P41EB028242, UH3DE028723, R01HL125440, P30AG073107, and 5P30AG073107-03 GY3 Pilots).

A. H. Gazi (agazi@seas.harvard.edu), B. T. Gullapalli, D. Gao, and S. A. Murphy are with the Department of Statistics and the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. S. A. Murphy holds concurrent appointments at Harvard University and as an Amazon Scholar. This paper describes work performed at Harvard University and is not associated with Amazon.

B. M. Marlin is with the Manning College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA.

V. Shetty is with the School of Dentistry, University of California, Los Angeles, Los Angeles, CA, USA.

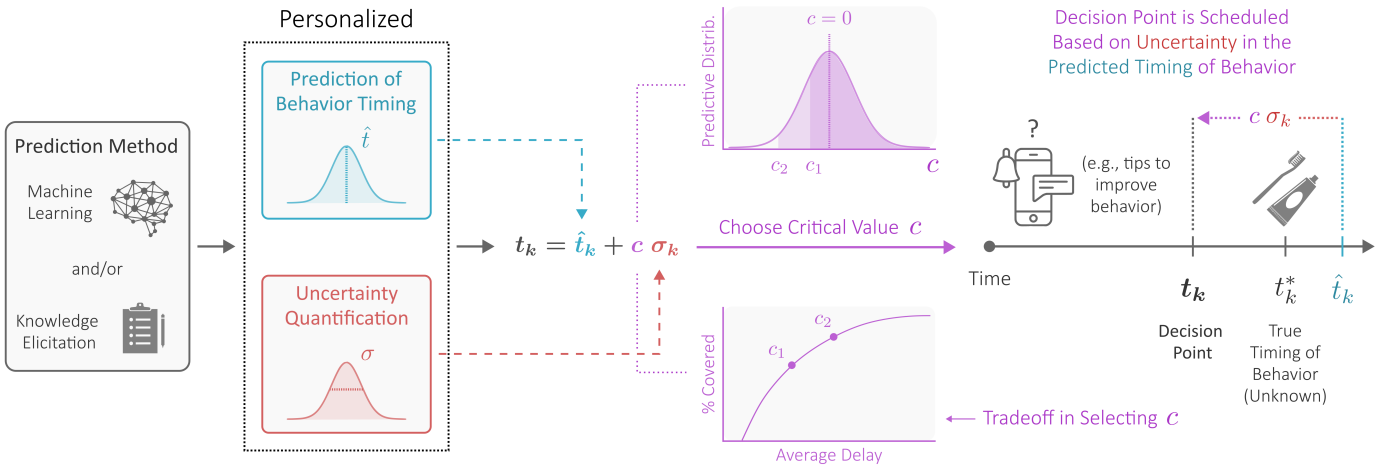

Fig. 2. SigmaScheduling approach. SigmaScheduling uses personalized predictions of behavior times – along with quantifications of uncertainty – to schedule decision points according to the probability of preceding the target behavior. The critical value, c, dictates the number of uncertainty measures (e.g., standard deviations) away from the mean that decision points are scheduled. Choosing c involves a tradeoff between the proportion of true behavior times preceded or "covered" by decision points and the average delay from decision points to behavior times. Based on the choice of c, SigmaScheduling takes predicted behavior times and the uncertainty quantified at each timestep, k, and schedules the kth decision point accordingly.

often falters for more variable behavior patterns, as shown in Fig. 1. Increasing the fixed duration from one to four hours could address the issue in Fig. 1B – but would unnecessarily increase the delay from decision points to brush times for more predictable routines (e.g., Fig. 1A). In settings where behavior predictability varies widely across the population, the one-size-fits-all status quo can squander valuable intervention opportunities. For example, *nearly a quarter of participants in the first deployment of Oralytics had 70% or fewer of their brushing events preceded by a decision point.*

This paper proposes and evaluates SigmaScheduling: a method to dynamically schedule decision points based on uncertainty in predicted behavior times. Our hypothesis is that SigmaScheduling will more effectively precede target behaviors with decision points by scheduling decision points earlier when routines are less predictable and closer beforehand otherwise. To test this hypothesis, we apply SigmaScheduling to real-world data from 68 participants of a 10-week trial of Oralytics [4], using (1) user-provided times and (2) online Bayesian machine learning (ML) for brush time prediction and uncertainty quantification (UQ). Our contribution is to demonstrate that for both approaches, SigmaScheduling increases the proportion of participants whose brushing events are preceded by a decision point at least 70% of the time.

II. Methods

A. SigmaScheduling

Fig. 2 illustrates SigmaScheduling. At each opportunity to schedule a decision point targeting a predicted behavior, indexed by $k \in \{1, 2, ..., K\}$, SigmaScheduling needs a prediction of the behavior time, $\hat{t}_k$, and UQ with an uncertainty measure, σ_k. The decision point, t_k, is scheduled via (1).

$$t_k = \hat{t}_k + c\,\sigma_k \tag{1}$$

The true behavior time, t_k^*, is forecasted via $\hat{t}_k$, and c is a design parameter. The status quo approach instead schedules via $t_k = \hat{t}_k + F$, where $F \leq 0$ is a fixed duration. Effectively, SigmaScheduling replaces 1 multiplied by a constant (i.e., $\hat{t}_k + 1 * F$) with σ_k multiplied by a constant (i.e., $\hat{t}_k + \sigma_k * c$).

The key design choice for SigmaScheduling is the parameter c. Theoretically, c represents a *critical value* and can be selected based on the desired probability that $t_k^* > t_k$. For example, if the posterior distribution $P(t_k^*|\hat{t}_k)$ is modeled as a Gaussian distribution, then c represents a z-score. In practice, selecting c involves a tradeoff between "coverage," the proportion of behavior times preceded by decision points (i.e., $\frac{1}{K} \cdot P$, where $P = \sum_{k=1}^{K} \mathbb{I}[t_k < t_k^*]$), and the mean delay from decision points to behavior times (i.e., $\frac{1}{P} \sum_{k=1}^{K} \mathbb{I}[t_k < t_k^*] \cdot (t_k^* - t_k)$). In general, c can vary across time (i.e., $\{c_k\}_{k=1}^{K}$) or individuals, but in this work, c is fixed.

B. Prediction and UQ with User-Provided Times

User-provided times are one avenue of predicting behavior times. A simple approach for UQ is to compute σ_k via the standard error for a prediction interval based on past errors in user-provided times [8]. In particular, σ_k in (1) can be computed via $\sigma_k = s_\epsilon \sqrt{1 + 1/k}$, where $s_\epsilon = \sqrt{\frac{1}{k-1} \sum_{i=1}^{k} (\epsilon_i - \bar{\epsilon}_k)^2}$ is the standard deviation (SD) of past errors accumulated until timestep k, with mean $\bar{\epsilon}_k = \frac{1}{k} \sum_{i=1}^{k} \epsilon_i$, $\epsilon_i = t_i^* - \hat{t}_i$. In this work, we set $\sigma_1 = 0$. The approach described forms a (biased) prediction interval based on the sample SD of residuals, which likens c in (1) to critical values of the t-distribution.

C. Prediction and UQ via Online Bayesian ML

Another avenue of predicting behavior times (i.e., $\hat{t}_k$) is via ML. In this work, we focus on online Bayesian ML to match the approach planned for the next iteration of Oralytics. Online learning enables a trained ML model to personalize to the target individual as more of the individual's data is observed. Bayesian algorithms produce posterior predictive distributions, $P(t_k^*|\hat{t}_k)$, from which the SD can be used for σ_k.

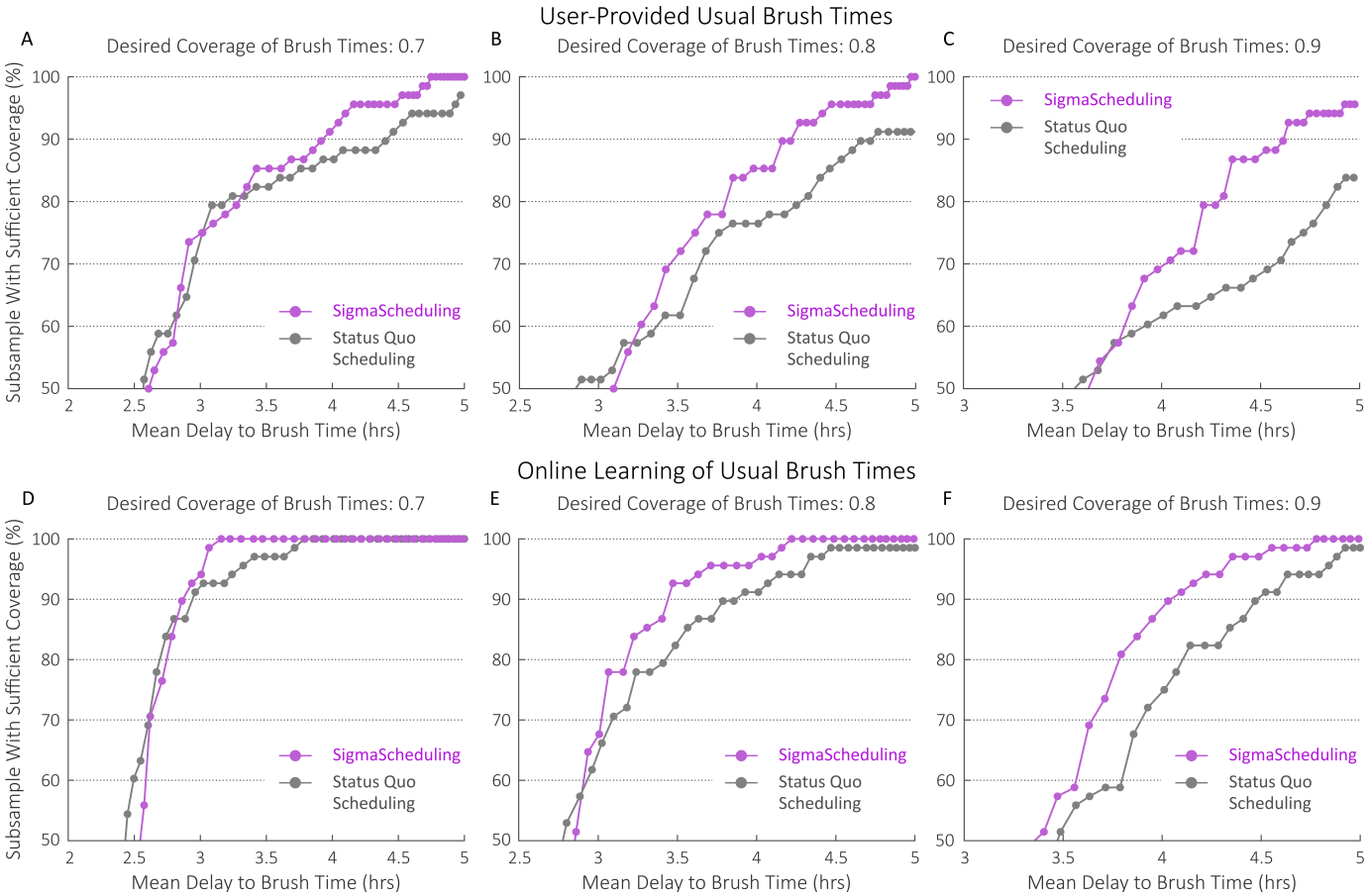

Fig. 3. Comparison of SigmaScheduling with Status Quo Scheduling for select coverage cutoffs. Note that the x-axes differ between columns of the subplot grid (i.e., A and D have the same x-axis, but B does not). The delays quantified along the x-axis refer to means across all participants of each participant-specific mean delay from decision points to the brush times that were successfully covered. Subsamples shown on the y-axis refer to the proportion of 68 participants that met the desired coverage of brush times specified for each subplot. For example, a y-value of 89.7% indicates that 61 participants' decision points were scheduled to satisfy the desired coverage. Each point on the SigmaScheduling curves corresponds to a different critical value, c, used in decision point scheduling, while each point on the status quo scheduling curves represents a possible fixed duration, F. Curves illustrate the tradeoff between coverage and delay and compare the two methods across c and F. Similar to receiver operating characteristic (ROC) curves, the optimal curve would be a horizontal line at a y-value of 100%. (A) Comparison between scheduling approaches when using user-provided usual brush times with a desired coverage of brush times per participant of 0.7. (B) Comparison using user-provided usual brush times with a desired coverage of 0.8. (C) Comparison using user-provided usual brush times with a desired coverage of 0.9. (D) Comparison using machine learning (ML) model predictions of usual brush times with a desired coverage of 0.7 for each participant. (E) Comparison using ML with a desired coverage of 0.8. (F) Comparison using ML with a desired coverage of 0.9.

D. Oralytics Data Used for Evaluation

This paper evaluates SigmaScheduling using brush time data from 68 participants in the first trial of Oralytics [4]. Of the 72 participants analyzed by Trella et al., we excluded one with less than a week of analyzable brush times and three with corrupted data due to user-provided brush times between 0:00 and 04:00. Participants were adults at risk for dental disease. The 70-day micro-randomized trial had two daily decision points (morning and evening); was approved by the University of California, Los Angeles Institutional Review Board (IRB#21–001471) and registered on ClinicalTrials.gov (NCT05624489) [5]; and lasted Sept. 2023 to July 2024 [4].

Participants provided typical morning and evening brush times on weekdays and weekends. Ground truth brush times, $\{t_k^*\}_{k=1}^K$, were sensed via smartbrush (Oral-B 8000). Only brush times, t_k^*, without prior intervention for the same k were analyzed to focus on self-determined behavior times, which decision points should be scheduled to precede. A mean of 53 $\pm$ 23 ($\pm$ SD) brush times were analyzed per participant.

E. Online Bayesian Linear Regression for Oralytics

Leave-one-participant-out cross validation was used to produce N participant-specific Bayesian linear regression (BLR) models for brush time prediction and UQ. BLR was used to match the upcoming deployment of Oralytics. Gaussian priors for regression weights were set as in prior work [4]. Prior means were set via a generalized estimating equation (GEE) fit to the $N-1$ participants' training data: the GEE point estimate was used for statistically significant regressors; zero otherwise. Prior SDs were set by fitting a GEE model to each training participant's data: the SD of point estimates was used for statistically significant regressors; half the SD otherwise. An inverse-gamma prior was used for the observation noise variance. The regressors were user-provided brush times; day of the week; morning or evening; the minimum, maximum, and coefficient of variation of past week brush times, most re-

TABLE I
AREA UNDER ROC-LIKE CURVES AS FUNCTION OF DESIRED COVERAGE

Desired Coverage	User-Provided *Status Quo*	User-Provided Σ-*Scheduling*	Model-Predicted *Status Quo*	Model-Predicted Σ-*Scheduling*
0.66	86.13	**92.77**	**121.18**	119.74
0.70	77.94	**85.51**	112.21	**113.07**
0.75	64.61	**75.24**	95.31	**100.93**
0.80	49.15	**62.94**	80.38	**89.38**
0.85	38.53	**54.00**	66.00	**77.67**
0.90	22.71	**39.67**	45.17	**59.48**
0.95	6.16	**18.17**	15.44	**34.84**
0.99	0.00	0.00	0.02	**3.84**

Boldface indicates superior performance. Rows are not comparable.

cent brush time; and number of no-brush days. For the held-out participant, priors were updated online with $\{t_i^*\}_{i=1}^{k-1}$ before predicting $\hat{t}_k$ and producing Gaussian posterior, $P(t_k^*|\hat{t}_k)$.

F. Comparing SigmaScheduling Versus Status Quo

We compared SigmaScheduling and status quo scheduling using tradeoff curves analogous to receiver operating characteristic (ROC) curves. The ROC-like curves plotted the proportion of participants with the desired coverage of brush times versus the mean delay across participants from decision points to brush times, varying c from 0 to –3 for SigmaScheduling and F from 0 to –6 hours for status quo. To align with Oralytics constraints, decision points scheduled before 04:00 for morning brushing and before 16:00 for evening brushing were clipped. Scheduling performance is improved if the proportion of participants with desired coverage increases for a given mean delay, analogous to increasing the true positive rate for a given false positive rate on a ROC curve. Similarly, the areas under the the ROC-like curves can be compared, analogous to comparing ML classifiers using areas under ROC curves. Areas were bounded below by 50% (i.e., at least a majority achieve desired coverage) and to the right by 5 hours.

The threshold for desired coverage per participant was varied from 0.66 to 0.99. The lower limit of 0.66 aligns with the threshold of 2/3 used for feasibility analyses [9]. We evaluated the percentage of participants with desired coverage, rather than the mean coverage, because we aim to achieve the desired coverage for a majority of participants, rather than to sacrifice coverage for some participants in exchange for others (e.g., 0.7 coverage for two participants versus 0.9 and 0.5).

III. RESULTS

Fig. 3 compares status quo scheduling and SigmaScheduling via ROC-like curves described in II-F. Interpreting Fig. 3B as an example, if the goal is for participants to have at least 8/10 of their brush times preceded by decision points, status quo scheduling achieves this at a mean delay of ∼4.5 hours for ∼85% of participants. For the same mean delay, SigmaScheduling achieves the desired coverage for ∼95% of participants. Table I details the areas under the curves.

IV. DISCUSSION

Using 68 participants' real-world data from the first deployment of the Oralytics JITAI, we find that SigmaScheduling improves the likelihood that interventions are delivered at opportune moments, before the target behavior of tooth brushing, in over 7/10 of instances. The improvement stems from SigmaScheduling's core principle of personalization: decision points are scheduled a longer duration before predicted brush times for participants with less predictable routines, while using shorter durations otherwise. In comparison, status quo scheduling uses the same duration for all participants.

SigmaScheduling's advantage becomes more pronounced as desired coverage increases. However, under more lenient criteria such as when covering only 2/3 of brushing events per participant, status quo scheduling performs comparably or better. One of the primary reasons for this is that for participants with predictable routines, the majority of true brush times fall fairly close to predicted brush times. In this dataset, participants with irregular routines were the minority. If it is acceptable to exclude the minority of participants with less predictable routines, (i.e., the threshold for desired coverage is lower), status quo scheduling suffices.

This work has limitations. Our evaluation of SigmaScheduling is limited to retrospective analysis within the oral self-care domain. Future studies are needed to deploy SigmaScheduling, extend its application to domains such as dietary behavior [6], [7], and evaluate the behavioral impact of improved coverage.

V. CONCLUSION

This paper introduces SigmaScheduling, a novel method for scheduling JITAI decision points that dynamically adapts the lead time between decision points and predicted behavior times to the uncertainty in predicted times. Applied to Oralytics data, SigmaScheduling more consistently schedules decision points prior to toothbrushing (i.e., during windows of receptivity [1]). Future work will evaluate SigmaScheduling in the upcoming clinical trial deployment of Oralytics and apply SigmaScheduling to other domains such as dietary behavior.

REFERENCES

[1] I. Nahum-Shani *et al.*, "Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support," *Annals of Behavioral Medicine*, 2018.

[2] B. M. Bell *et al.*, "Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review," *npj Digital Medicine*, vol. 3, 2020.

[3] A. Ahmed *et al.*, "Proton Pump Inhibitors (PPI)," *StatPearls*, 2023.

[4] A. L. Trella *et al.*, "A deployed online reinforcement learning algorithm in an oral health clinical trial," *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 28792–28800, 2025.

[5] I. Nahum-Shani *et al.*, "Optimizing an adaptive digital oral health intervention for promoting oral self-care behaviors: Micro-randomized trial protocol," *Contemporary Clinical Trials*, vol. 139, 2024.

[6] L. Zhao *et al.*, "A micro-randomized pilot study to examine the impact of just-in-time nudging on after-dinner snacking in adults with type 2 diabetes: A study protocol," *Diabetes, Obesity and Metabolism*, 2023.

[7] M. Fuchs *et al.*, "Exploring message receptivity and protocol adherence in a clinical study: A micro-randomized trial protocol," in *Proceedings of the CHI 2025 Workshop on Envisioning the Future of Interactive Health*, (Yokohama, Japan), ACM, 2025.

[8] S. Geisser, *Predictive Inference: An Introduction*. CRC Press, 1993.

[9] L. N. Coughlin *et al.*, "A mobile health intervention for emerging adults with regular cannabis use: A micro-randomized pilot trial design protocol," *Contemporary Clinical Trials*, vol. 145, 2024.