Contextualization and Individualization for Just-in-Time Adaptive Interventions to Reduce Sedentary Behavior

Matthew Saponaro
mattsap@aiwhoo.com
A.I. Whoo LLC
Newark, Delaware, USA

Greg Dominick
gdominic@udel.edu
University of Delaware
Newark, Delaware, USA

Ajith Vemuri
kumar@udel.edu
University of Delaware
Newark, Delaware, USA

Keith Decker
decker@udel.edu
University of Delaware
Newark, Delaware, USA

ABSTRACT

Wearable technology opens opportunities to reduce sedentary behavior; however, commercially available devices do not provide tailored coaching strategies. Just-In-Time Adaptive Interventions (JITAI) provide such a framework; however most JITAI are conceptual to date. We conduct a study to evaluate just-in-time nudges in free-living conditions in terms of receptiveness and nudge impact. We first quantify baseline behavioral patterns in context using features such as location and step count, and assess differences in individual responses. We show there is a strong inverse relationship between average daily step counts and time spent being sedentary indicating that steps are steadily taken throughout the day, rather than in large bursts. Interestingly, the effect of nudges delivered at the workplace is larger in terms of step count than those delivered at home. We develop Random Forest models to learn nudge receptiveness using both individualized and contextualized data. We show that step count is the least important identifier in nudge receptiveness, while location is the most important. Furthermore, we compare the developed models with a commercially available smart coach using post-hoc analysis. The results show that using the contextualized and individualized information significantly outperforms non-JITAI approaches to determine nudge receptiveness.

CCS CONCEPTS

• **Applied computing** → **Consumer health**.

KEYWORDS

Just-in-time, Sedentary Behavior, Machine Learning, Contextualization, Individualization, Wearable Devices, Free-living

ACM Reference Format:

Matthew Saponaro, Ajith Vemuri, Greg Dominick, and Keith Decker. 2021. Contextualization and Individualization for Just-in-Time Adaptive Interventions to Reduce Sedentary Behavior. In *CHIL '21: ACM-CHIL, April 08-10, 2021*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3450439.3451874>

1 INTRODUCTION

Sedentary behavior and physical inactivity are among the leading modifiable risk factors for exacerbating the development of cardiovascular and metabolic diseases such as heart disease and type 2 diabetes, and directly contributes to all-cause mortality [3, 6, 29]. Whereas physical inactivity refers to performing insufficient amounts of moderate-to-vigorous intensity physical activity (MVPA) (i.e. meeting < 150 minutes of weekly MVPA), sedentary behavior is characterized as the absence of physical activity, and is defined as any waking behavior performed in a seated, reclined, or supine posture resulting in a very low energy expenditure (i.e. < 1.5 metabolic equivalents of task, or METs) [3]. Increasing the percentage of adults who reduce time in sedentary behavior while increasing MVPA is a global priority for improving cardiovascular health [6, 17]. Population surveillance data indicate nearly 90 percent of US adults are inactive [25, 26], with over half of waking hours spent in sedentary activities [10, 11, 26]. Moreover, recent epidemiological evidence indicates that the prevalence of sedentary behavior in the US is increasing [27]. Thus interventions designed to reduce prolonged periods of sedentary behavior is vital for improving population health outcomes.

Wearables and smartphones have the potential to extend the reach and impact of human coaching to reduce sedentary behavior by delivering actionable feedback in real-time, based on collected data (e.g. heart rate, step count, intensity) [15, 22]. Unfortunately, most commercial smart coaches include questionable behavior change mechanisms, only provide general intervention strategies, and don't make use of the available information specific to their user. For example, Fitbit provides messages (e.g. "it's step 'o clock") at the 50th minute of the hour if the user hasn't walked 250 steps within the hour. Just-in-time adaptive intervention (JITAI) design enables researchers to monitor changes to a person's psychological (e.g. mood), behavioral, and physical states and contexts over time for the purpose of delivering support at the moment it is needed,

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ACM CHIL '21, April 8–10, 2021, Virtual Event, USA

© 2021 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-8359-2/21/04.

<https://doi.org/10.1145/3450439.3451874>

and when the person is most likely to be receptive [18]. This requires the coach-system to adapt to the ever-changing states and contexts of the individual. Such adaptability requires extensive monitoring of the person to decide (a) whether the person is in a state that requires support; (b) what type or amount of support is needed given the current state; and (c) whether the support provided will be positively acted upon or perceived negatively [15]. Few physical activity JITAI are truly adaptive and most studies utilize generic messages that are delivered at pre-set times such as shortly before, or soon after, a desired goal is achieved [5, 9]. The current state-of-the-art remains largely conceptual [1, 4, 14]. Although researchers are striving towards a JITAI, most of the work is in just-in-time interventions which are unable to adapt to the changing contextual environment [5, 28].

The objective of this paper is to move towards a fully autonomous wearable JITAI coach to reduce sedentary behavior. We accomplish the objective by pursuing the following aims: **Aim 1** combines biometric, temporal, and geospatial data from wearable and mobile devices to better understand contextual and individual sedentary behaviors in free-living conditions over a period of two weeks. **Aim 2** defines two measures of nudge success—receptiveness (i.e. belief that the nudged prompted engagement or not) and impact (i.e. steps taken as a result of the nudge)—and examine the contextual influences on these measures in a just-in-time setting. **Aim 3** evaluates the predictors of nudge receptiveness and indicates that location is a key feature while steps (as used by Fitbit) is not important. **Aim 4** demonstrates feasibility to learn nudge receptiveness in a JITAI setting, using post-hoc analysis on human data, and shows learned contextualized and individualized JITAI significantly outperform an existing commercially available wearable smart coach to determine receptive contexts.

The remainder of this paper is outlined as follows: In section 2, we discuss the related work done in contextualizing and individualizing interventions. In section 3, we discuss our human subjects study and data collection protocol. In section 4, we talk about how we pre-process our data. In section 5, we analyze baseline behaviors and nudge responses, and then develop a contextualized and individualized model to determine nudge receptiveness. In section 6, we discuss limitations of our study. Finally, we conclude our work and discuss our next steps in section 7 and 8.

2 RELATED WORK

Mobile applications have been demonstrated to be an effective way to engage users and provide effective behavior changing interventions [16]. Substantial work in individualizing and contextualizing interventions has been done in reducing sedentary behavior [5, 28], increasing physical activity [13, 28], and other behavior change applications [16, 19, 30]; however, most of the just-in-time adaptive work is conceptual [1, 12, 14, 15]. There have been limited just-in-time studies which examine individualization or contextualization in free-living conditions [5, 30, 31]. Machine Learning techniques have been proposed to learn JITAI; however, due to the limited training data, these approaches have been explored using synthetic data [20, 21, 24]. The study in this paper presents key insights in designing JITAI to reduce sedentary behavior using both contextual and individual information through understanding the context for

when people are receptive and how the context effects the impact of nudges to reduce sedentary behavior in free-living conditions. Furthermore, we show a comparison of using both contextual and individualized information (real world data) to learn when to nudge in free-living conditions using Random Forest models.

2.1 Just-in-Time and Synthetic JITAI Analysis

Just-in-time interventions provide data collection challenges due to the need to monitor a person's state, which may comprise multiple tailoring variables, and observe the physical response from nudges in the various contexts. Nudging too often is infeasible because high frequency may burden and annoy the user. Saponaro et al. explored the feasibility to learn contextualized and individualized JITAI to increase physical activity by developing synthetic user models [21]. In order to re-use data across individuals, Saponaro and Decker further evaluate, in a synthetic environment, the trade-off between data availability and relevancy when grouping individuals together based on similar behavior responses to nudges [20]. The work presented in this paper provides real-data analysis of contextualizing and individualizing JITAI and can inform the development of data re-use algorithms.

Thomkins et al. develop a Bayesian pooling algorithm to determine when to nudge to reduce sedentary behavior using others' data as well as analyze the impact of the delivered nudges in terms of steps [24]. In order to provide enough nudge data to the learner, nudge responses are based on a synthetic model which was based in turn on contextualized human data in free-living conditions where on average 2.4 nudges were provided throughout the day [5]. In this paper, we deliver 12 nudges per day which provides a more reliable estimate of free-living behaviors as described in Thomas et al. [23]. We use tree-based models to learn nudge receptiveness due to the natural alignment with human expert-prescribed JITAI that can facilitate jump-starting machine learning algorithms [14]. Though both Thomkins and this paper examine features such as location and time of day, the analysis provided in this paper shows fully real-data contextual impact on nudge receptiveness in terms of receptiveness as well as step count. Additionally, we compare step counts before and after a delivered nudge, which provides a better estimate of impact when only comparing steps during decision points of non-delivered and delivered nudges due to the contextual influence on the nudge outcome. As demonstrated in this paper, geospatial features play an important role on nudge impact, which may differ across nudge decision points. Furthermore, this paper provides an analysis of baseline behaviors in free-living conditions in order to inform JITAI design.

2.2 Individualizing Interventions

Substantially more work has been done in non just-in-time interventions to explore individualization. MyBehavior, a smartphone app, delivers personalized activity suggestions (physical activities and food types) for weight loss using EXP3 (a multi-arm bandit algorithm) to select interventions [19]. The interventions are context free and updated each day according to the user activity history and preferences. Yom-tov et al. uses Reinforcement Learning (RL) to reduce sedentary behavior among type-2 diabetes patients [28]. RL was used to select a personalized message type which helped

Figure 1: Study developed mobile application used to collect contextual data and record nudge response

minimize sedentary behavior. Users were represented to the RL algorithm (free of current context) by their demographics, past activity, expected activity, and message history. Messages were given from once a day to a once a week, corresponding to the physical activity achieved by the users from the time of their previous messages (with no location context). The work presented in this paper examines context-aware and individualized interventions and has shown that context significantly effects intervention outcome.

Zhou et al. uses Reinforcement Learning to personalize goals (daily step count) for individual users [31]. They conducted a 13 participant pilot study to increase physical activity and show that customized goals help users achieve higher daily step counts. Particularly, they show that 10000 steps isn't a good goal (most participants in our study did not achieve this amount). The baseline behavior analysis presented in this paper may be able to jump start learning approaches or guide goal settings as presented by Zhou et al.

3 EXPERIMENT DESIGN

Here, we explain our data collection protocol in order to capture the contextual information of our users in free-living conditions (i.e. day to day living) and their physical responses to the provided nudges in order to better understand sedentary behavior and people's receptiveness to nudges.

3.1 Data Collection Protocol

Participants (N=33) were male (N=20) and female (N=13) who were University of Delaware students (N=20), faculty (N=5), and of-office staff (N=8). This study was approved by the University of Delaware's IRB and all participants consented to participate in this 2 week study. Participants were healthy adults between the ages of 18 and 55. Participants took a short health screen questionnaire to ensure that it was safe to engage in unsupervised physical activity and exercise. During an initial study visit, participant demographics and current physical activity/exercise behaviors were self-reported via questionnaire. participants completed a short questionnaire behavior, providing their typical weekly calendar, using a study-developed mobile application, and wearing a Fitbit.

3.2 Participant Meta-data Questionnaire

Participants completed a short questionnaire that asked about demographics (i.e. age, sex, gender), anthropometrics (i.e. height, weight), average time sleeping and doing physical activity during the week-day and weekends, motivation and energy assessments, and nudge frequency preferences. Self-reported motivation and energy was assessed using a Likert scale of 1-5 where 3 was neutral.

3.3 Calendar Input

Participants provided expected location (i.e. home, a friend's home, work, school, gym, park, driving, other, or I don't know) and expected receptiveness for 15 minute time blocks for each day of a typical week. Expected receptiveness, in this form, reflects a person's belief that they would be more likely to move if a nudge prompted them to do so at that specific moment in time. Expected receptiveness was assessed with a single question using a Likert scale of 1-5 where 3 was neutral. A study administrator validated the calendar form was accurately and fully filled out. This is used to compare the real and expected locations and receptiveness of users in future work.

3.4 Mobile Application Protocol

During the initial visit, participants downloaded a study-based mobile application and let it run in the background during the entire 14-day study period. Participants were instructed to not close the application during the duration of the study. The application could be put into the background, but not closed. Whenever a participant physically traveled to a new meaningful location that was not recorded during the study, participants were requested to tag that location via the app interface. Participants were able to use their phone's touch-screen feature to select a general label of Home, Work, School, Gym, or Other. If participants selected "Other", they were asked to input text regarding the location (e.g. restaurant, department store, grocery store, etc.). Once a participant tagged a new location, they were requested to not tag it again. Additionally, participant GPS data was collected in the background to monitor their location throughout the day. See figure 1. While collecting the tagged and background locations, the phone's "Best Accuracy" setting was used for the GPS. This level of accuracy is intended for use in navigation apps that require precise position information at all times.

Our system provided nudges to participants and were delivered at the 50th minute of each hour from 8:50am to 8:50pm via a push-notification within the mobile application. The nudges provided, mimic the content of the messages provided by Fitbit if the users step count is less than 250 at the 50th minute of each hour. At each nudge time-point, participants were asked to open the app and to evaluate nudges provided by our system regardless if the nudge was successful to prompt anti-sedentary behavior. Nudges were evaluated based on time, energy, interest, nudge receptiveness (i.e. whether the participant believed the nudge prompted them to move), and overall. All evaluation criteria except for nudge receptiveness was evaluated using a Likert scale. Participants were instructed that the nudge was receptive if the participant took non-sedentary actions as a result of receiving the nudge. If a participant was already engaging in some activity upon receiving a nudge (e.g.

currently walking), they were instructed to mark the nudge as receptive only if the nudge prompted the user for an extension or additional non-sedentary activity (e.g. extended their walk).

Participants were requested to always keep their phones with them day and night, for the duration of the study. The total time participants were asked to use the app and evaluate nudges was 14 days.

3.5 Fitbit Protocol

Participants also wore a Fitbit device with heart rate (HR) monitoring capabilities (i.e. Charge HR, Ionic, and Alta). Participants were instructed to wear the Fitbit device on their non-dominant wrist during all waking hours for the 2 week study duration. The option to receive Fitbit notifications was turned off so participants only received notifications from the study-developed mobile application. The devices were only allowed to be taken off while showering, swimming, and sleeping. Because Fitbit does not provide real-time data, data was pulled nightly during quiet-hours (i.e. 3 a.m.) to ensure the previous days data was complete. Fitbit data collected included minute level: heart rate, step count, absolute intensity level predictions (i.e. sedentary, light, moderate, vigorous), and caloric expenditure. It's noted that Fitbit provides estimates of these values and other research[2] shows that, for example, Fitbit underestimates heart rate during first and last portions of physical activity.

3.6 Data Quality and Validation

The day of the initial visit was considered to be "day 0" in order to ensure both the mobile application and Fitbit device were correctly providing data. Furthermore, participant data was monitored throughout the study (particularly on day 0) for issues. Data collected during "day 0" was not included in the analysis. After "day 0", participants provided 14 days of valid data. A day was considered to be valid if there were at least nine nudge evaluations with corresponding Fitbit and GPS data. Fitbit data was manually reviewed for non-wear periods. If the application was closed during the study, participants were requested to re-open the application so it would continue collecting GPS data. GPS data was manually reviewed for significant sequential location differences.

When participants evaluated nudges, the system provided contextual information such as location, whether they were driving or not, and whether they were outside or not. These location-based features were estimated using the user's 10 most recent GPS records (i.e. a small recent sample). Participants were instructed to correct the location-based information if incorrect. A nudge could only be outstanding for 1 hour and 59 minutes before it was considered stale. Stale nudges weren't evaluated. This mechanism helps maintain that the participants' recollection of the context and evaluation of the nudge is accurate.

4 DATA PRE-PROCESSING

We capture several elements of the contextual environment derived from temporal, biometric, and geospatial features. Because nudges were provided every hour, the data was summarized into hourly chunks and enriched using the mapping logic provided below.

4.1 Temporal Features

Temporal features use the date and timestamp of the delivered nudges.

- **TimeOfDay** - Either Morning (8:00AM to 11:59AM), Lunch Time (12:00AM to 1:59PM), Afternoon (2:00PM to 4:59PM), Evening (5:00PM to 8:59PM) based on the timestamp of the nudge. The parts of the day categorizations align with Learner's Dictionary's parts of the day categorizations [7].
- **Weekend?** - Boolean representing whether the nudge was provided on the weekend or not

4.2 Biometric Features

Minute level step counts were aggregated into hourly level features. The step features use the format Steps_{x-y} , where all the minutes of the hour between x and y were aggregated.

- **Steps₀₋₅₀** - Total steps taken in the first 50 minutes of each hour.
- **Steps₄₀₋₅₀** - Total steps taken 10 minutes prior to the nudge up until the nudge was provided.
- **Steps₅₀₋₆₀** - Total steps taken 10 minutes after the nudge was delivered.
- **Steps ≥ 250** - Boolean representing whether or not **Steps₀₋₅₀** ≥ 250 . Fitbit provides its nudges if this feature is False.
- **Steps ≥ 500** - Boolean representing whether **Steps₀₋₅₀** ≥ 500 . If a person consistently walks 500 steps each hour, they would approximately walk about 7,500 steps per day.

4.3 Geospatial Features

Raw GPS data collected consists of a longitude, latitude, and speed. Geospatial features used the 10 latest GPS records prior to nudge delivery to estimate instantaneous location. Particularly, GPS was labeled using the closest tagged location; however, if the closest tagged location is more than 0.5 kilometers away, then the GPS record was labeled as "None of the Above". 0.5 kilometers was used as our threshold criterion due to its ability to correctly encapsulate tagged locations when manually examining building location with respect to building size. Locations were determined to be outside based on pre-specified classifications. For example, Home and work are indoors, while parks are outdoors. However, users specified whether "Other" tagged locations were indoors or outdoors in the mobile application. None of the above locations were defaulted to outside because upon manual review most of the "None of the above" locations were outside.

- **Location** - labeled tagged location (i.e. Home, Work, Gym, Park, Other, None of the Above) at time of nudge.
- **Driving?** - Boolean representing whether driving was being done at time of nudge. A participant was considered driving if their average sampled GPS speed was greater than 12 meters per second. The 12 meters per second threshold was used to mimic Apple's driving estimation criterion.
- **Outdoors?** - Boolean representing whether the participant was outside at time of nudge.

Figure 2: Average daily step counts for each participant over 14 days.

5 EVALUATION AND RESULTS

In order to move towards providing JITAI to reduce sedentary behavior, we pursue the following aims: **Aim 1:** understand baseline behaviors using the integrated step and location data. We assess our participant population at both the individual and group level. In order to understand behaviors in free-living conditions, we pursue the following sub-aims: **Sub-aim 1** examine daily step counts and hours spent sedentary to assess the overall degree of sedentary behavior in free-living conditions. **Sub-aim 2** show a relationship between average daily steps and total hours spent sedentary—a key finding that activity is done progressively throughout the day, rather than in large bouts of activity. **Sub-aim 3** compare times spent sedentary by participant location to understand where is the greatest effect that a nudge can provide. **Aim 2:** examine the contextual influences on two defined measures of nudge success—receptiveness and impact. In order to understand the influences on nudge outcomes, we pursue the following sub-aims: **Sub-aim 1** define nudge success. **Sub-aim 2** evaluate people’s baseline receptiveness to nudges. **Sub-aim 3** examine nudge receptiveness and impact in each location. By understanding the influences on nudge outcomes, JITAI can be optimized by utilizing these influences that facilitate successful nudges. **Aim 3:** determine important features of nudge receptiveness. **Aim 4:** demonstrate the potential to learn individualized and contextualized JITAI through a post-hoc analysis of the collected data; as well as show that individualized and contextualized JITAI significantly outperform an existing commercially available wearable smart coach to determine receptive contexts. By examining these aims, we found interesting behavioral patterns, and impact and indicators of successful nudges that can inform the development of JITAI.

5.1 Understanding Contextualized and Individualized Sedentary Behaviors in Free-living Conditions

Sub-aim 1: We examine the total number of steps as well as whether individuals walked more than 250 steps in an hour or not—Fitbit’s criterion for sedentary behavior to send nudges. In order to achieve health benefits from walking, individuals should take an average of 7500 steps per day [8]. Fitbit establishes a daily step goal of 10000 steps. In figure 2, we see that 18 participants (about 55 percent)

reach enough steps for health benefits. Furthermore, a large majority of our participant population failed to achieve Fitbit’s 10000 daily step goal half of the time. From the data, we see that our participants do not reach the established goals. When considering JITAI design, we need to set attainable goals in order to effectively modify sedentary behavior.

Figure 3: Distribution of the number of days participants achieved the Fitbit step goal over the duration of the study.

Figure 4: Total hours spent sedentary for each participant’s average daily step counts.

Sub-aim 2: We group individuals into 3 equally-sized bins based on the average daily step counts in figure 2 in order to identify relationships between daily physical activity and hourly-level sedentary behavior. Participants with the lowest daily step counts were binned into the least active group and participants with the highest daily step counts were binned into the most active group. We use Fitbit’s sedentary behavior criterion in our analysis (i.e. steps < 250 for the

first 50 minutes of the hour). Figure 4 shows that there is a relationship between daily total step-count activity levels and hourly sedentary behavior; particularly, people that walk more spend less hours being sedentary. We apply a linear regression model to the data and observe an R^2 value of 0.67 with outliers removed ($R^2=0.63$ with outliers) indicating a good fit. One potential explanation for this is that, for this participant population, people who are more active may tend to be more conscientious about limiting their time in sedentary activities.

Sub-aim 3: We examine location data in order to understand how long people spend at each location and how long they are sedentary at each location. Our participants spent most of their times at home, then work, then at others as shown in 5(a) and 5(b). This indicates that JITAI design should focus on home and work strategies in order to have the most effective hours throughout the day. Half of the time at home and work are spent sedentary, while little time is spent sedentary at the gym and park. Knowing that a person is at the gym or at a park, it may not be worthwhile to provide nudges to reduce sedentary behavior; however, we note that Fitbit would send nudges during these times. It should also be noted that nudges to improve sedentary behavior are different than nudges to increase physical activity. For example, targeting people with nudges at the park to run may be effective, but since most time at the park is not sedentary, nudges relating to reducing sedentary behavior at the park are not effective. Providing nudges to improve physical activity (MVPA) is outside of this paper's scope.

When considering the gym, approximately 18 percent of the time was considered sedentary. Because people are expected to be non-sedentary at a gym, there could be 3 reasons for why sedentary behavior is appearing in the graph above: 1) the person was indeed sedentary, 2) the Fitbit definition of sedentary doesn't reflect sedentary behavior (e.g. the person was sitting down lifting weights, rather than walking to increase step count), 3) the way the data was discretized into hourly chunks eliminated the granularity to accurately detect non-sedentary behavior. Further examination into data summarization methods and trade-offs for accuracy and meaningfulness could be examined. Furthermore, additional analysis could be done using other data streams such as heart rate to validate the sedentary behavior definition used by Fitbit to provide nudges and data quality.

5.2 Analyzing Influences on Nudge Outcomes

Sub-aim 1: We define and examine two measures of nudge success: receptiveness and impact. Receptiveness to a nudge is the recipient's belief that they engaged in non-sedentary behavior as a result of the nudge. In the mobile application, after participants received a nudge, they were required to indicate via an in-app questionnaire whether the nudge prompted an action [8]. If participants indicated that it did prompt action, then we consider the participant was receptive to the nudge. Nudge impact is the difference in steps taken after and before the nudge was delivered (i.e. $\text{Steps}_{50-60} - \text{Steps}_{40-50}$). In non-sedentary contexts, nudge impact may be heavily influenced by the subtrahend (i.e. Steps_{40-50}) and thus a high impact may be achieved via vigorous activity. In this paper, we only examine nudge impact in sedentary contexts which makes the subtrahend small (always less than 250 steps). There are other ways to define nudge

Figure 5: 5(a) shows the total time participants spent at each location with respect to sedentary and non-sedentary time. 5(b) shows the percentage of time spent being sedentary and non-sedentary.

success (e.g. recipients belief regarding the goodness of the nudge); however, we only examine these two success measures in this paper. By understanding the different ways people respond to nudges, we are able to understand the effects of a JITAI when learning contextual influences.

Sub-aim 2: When looking at the receptiveness of individuals by activity level groupings, we see that the least active group are not as receptive as the moderate and most active groups, as shown in Figure 6. We conduct a Mann-Whitney U test to determine whether: 1) there is significant difference in receptiveness between any of the groups and 2) whether members belonging to the most active group are more receptive than members belonging to the least active group. The results indicate that there is no statistical difference between the moderate and most active groups ($p=0.4396$); however, the most active group is significantly more receptive than the least active group ($p=0.039$). Not only do the least active individuals spend the most time being sedentary, we show that the least active individuals are the least receptive to nudges. When considering learning an

individual's JITAI, there will be fewer contexts where a non-active person is receptive compared to someone who is more active.

Sub-aim 3: We examine the effects of location on receptiveness by counting the number of receptive and non-receptive nudges at each location, see Figure 7. Nudges sent while individuals are at Home, Work, and other locations yield approximately 40 percent receptive while at each location. Nudges sent while at the park yield about 85 percent receptive nudges. Based on the time spent at each location, even though that nudges sent at parks are most likely to be receptive, not many nudges can be provided at parks. Interestingly, when looking at the individual receptiveness in Figure 8, we see that people differ in receptiveness at each location. When learning a JITAI, we need to individualize the intervention to accommodate individuals' receptiveness in each location.

Figure 6: Total number of successful nudges by each activity level group

In addition to understanding whether people claimed that they were receptive or not, we examined the actual impact of the nudge by comparing the step counts 10 minutes before and after the nudge was delivered. Three Mann-Whitney U tests were conducted to determine: 1) whether or not there was an impact when people were receptive, 2) whether or not there was more of an impact after the nudge when people were receptive compared to non-receptive, 3) whether or not there was an impact from a non-receptive nudge. The results conclude that there is a significantly higher impact (before vs. after and after receptive vs. after non-receptive) due to the nudge ($p=8 \times 10^{-12}$ and $p=3 \times 10^{-103}$ respectively) and there were no statistical differences ($p=0.4177$) in non-receptive before and after step counts indicating that we can use **Steps**₄₀₋₅₀ as a baseline for nudge impact.

We examine nudge impact when individuals were exhibiting sedentary behavior. Using the previously described activity level groupings, we observe that the more active individuals tend to have higher impacts, compared to less active individuals as seen in figure 9. To validate this observation, we conduct a Wilcoxin Signed-Rank Test for the right tails of the distribution (i.e. step difference ≥ 15).

Figure 7: 7(a) shows the total time participants spent at each location with respect to being receptive to the nudge. 7(b) shows the percentage of receptiveness at each location.

This test only considers meaningful difference in steps, rather than when there were no differences. The tests conclude the most active and moderately active groups are impacted more by nudges than the least active group ($p=.0377$ and $p=.0483$ respectively); however, there were no differences between the moderate and most active groups ($p=.5781$). Furthermore, the skewness and kurtosis of the least active group (6.42 and 46.34 respectively) is higher than the skewness and kurtosis of the moderate (5.857 and 39.85 respectively) and most active groups (4.66 and 26.99) which indicates least active individuals have lower nudge impacts than more active individuals. Interestingly, when observing location, receptive nudges at work had higher impact on step counts compared with receptive nudges at home as seen in figure 10(a) and 10(b). To validate this observation, we conduct another Wilcoxin Signed-Rank Test using all participants (not grouped by activity levels) and compare the impacts at work and home. The tests conclude that the impact in steps is larger at work, than at home ($p=0.0001$).

Figure 8: Total time participants spent at Home 8(a), Work 8(b), and other locations 8(c) where they were receptive (and not)

Figure 9: Difference in step count before and after successful nudges when individuals were sedentary.

5.3 Evaluating the Predictors of Nudge Receptiveness

We analyze Fitbit's nudging model (i.e. if steps ≥ 250 is false, then nudge) in order to evaluate whether this model is a good indicator for nudge outcomes (i.e. correctly classifying when and when not to nudge based on receptiveness). The accuracy (.04661), precision (.4824), recall (.4663), and F1-score (.4722) indicate that Fitbit's model does poorly at accurately determine appropriate contexts to nudge and inappropriate contexts to not nudge. Due to the poor performance measures, we conclude that the step data used is not a good indicator for nudge outcome. When examining the receptive

(a) Home

(b) Work

Figure 10: 10(a) shows the total time participants spent at each location with each hour marked as sedentary or non-sedentary. 10(b) shows the percentage of time spent being sedentary and non-sedentary when at each location.

nudges throughout the study, we see a large disparity in receptiveness among participants as seen in figure 11. Furthermore, when looking at the percent of successful nudges where the individual was sedentary, we see that less than half were in states when the individual walked less than 250 steps (i.e. Fitbit would nudge).

In order to examine which features are good indicators for nudge receptiveness, we develop a random forest model using the first week of data of all participants to learn when individuals are receptive to nudges. We split the training and evaluation data sets by the week so that the evaluation and training set were more or less identical to avoid day-of-week bias (e.g. training on both Mondays where Mondays are not receptive days and evaluating on Tuesday which is a very receptive day). Random forests use samples of data to develop a decision tree and use a majority vote from each decision tree within the forest. For the purposes of this paper, we use 250 decision tree estimators in the random forest for sufficient sample size. We define feature importance as the normalized sum of Information Gain across all features across all trees. Interestingly, we see in figure 12 that the step data was the least important feature, while geospatial and temporal features were most important.

Figure 11: Percent of successful nudges by sedentary behavior for participants

This demonstrates that Fitbits, as well as other smart coaches, need to utilize more contextual information when providing nudges in order to be successful.

Figure 12: Feature Importance based on entropy within Random Forest

5.4 Learning Contextualized and Individualized Models of Receptiveness

In figure 13, we compare four nudging strategies to determine nudge receptiveness in a post-hoc fashion. Performance is evaluated using the average accuracy of each strategy to predict the actual nudge receptiveness across all individuals. We use a baseline strategy of always sending a nudge to evaluate the average receptiveness of individuals. We use Fitbit's nudging algorithm of providing a nudge if the wearer doesn't reach 250 steps by the 50th minute of the hour. The Fitbit algorithm provides an insight on the performance of a smart coach that doesn't utilize all contextual information, nor individualizes to their user. For the general algorithm, we develop a random forest trained using the first week's data of all participants. This algorithm provides insight on how well a smart coach may do if it were to only use general contextual information averaged over all users. For the individual algorithm, we develop a

random forest model for each participant individually using their first week's data. We see that the always method performs worst in correctly identifying contexts that result in receptive nudges. The error bars are the largest indicating the large disparity in individuals' receptiveness to nudges. Interestingly, Fitbit's performance is not much better than always sending nudges throughout the day. The results indicate that Fitbits are good for empowering people to monitor their data but struggle at providing relevant nudges. To compare the Fitbit model with Always, we conduct a Wilcoxin Signed-Rank test. This test compares performance paired on the same contexts. The test concludes that there is no statistical difference in performance between Fitbit and Always ($p=0.1357$). We see that the individualized algorithm performs best and only slightly better than the generalized algorithm. Interestingly, the error bars of the individualized algorithm is smaller than the error bars of the generalized algorithm which indicates tailoring improves the lower bound of performance. To validate these performance observations, we conduct two Wilcoxin Signed-Rank tests comparing the generalized model to the Fitbit model and the individualized model to the generalized model. The tests conclude that the generalized model does better than the Fitbit model ($p=0.0002$) and the individualized model does better than the general model ($p=0.0407$). We also examine the pairwise differences between the individualized and generalized algorithms, see figure 14. For approximately 66 percent of participants, the individualized model did better than the generalized model. The results presented above show that when developing a JITAI, not only should we contextualize the intervention, we need to individualize the JITAI as well.

Figure 13: A Post-hoc Comparison of different nudging algorithms to correctly identify nudge receptiveness.

6 LIMITATIONS

In this paper, we captured the temporal, biometric, and geospatial aspects of the contextual environment. Unfortunately, other known influences on physical activity/sedentary behaviors, such as elements of the social environment (e.g. social support), could not be assessed and thus should be further researched.

Behavior change occurs over extended period of time and the study presented primarily analyzes reactive behavior, rather than habits. A longer study would be able to capture how people progress

Figure 14: Pairwise accuracy differences when using the Individual and General models for each participant.

through the stages of behavior change; however, the results presented in this study provide preliminary data to inform analysis on habitual behavior changes.

Receptiveness is a self-reported measure collected via the mobile application questionnaire. It is unknown whether the participant was truly receptive which could potentially influence the impact analysis since our analysis only considered receptive nudges. We attempt to mitigate this data quality concern through statistical tests. Clustering analysis on impact and receptiveness could be conducted in order to estimate incorrectly reported data.

Participants may not have considered nudges independently (even though instructed to do so) due to the time frame. Furthermore, nudges every hour are infeasible to respond to for long periods of time. Additionally, due to the data collection limitations, the post-hoc analysis doesn't allow for the algorithms to actually coach the individual which could get them to avoid prolonged non-receptive contexts or steer towards receptive contexts. Unfortunately, the algorithm analysis doesn't allow for a true independent evaluation. However, by standardizing the features used across the different algorithms, we are able to do a standardized comparison.

7 CONCLUSION

In this paper, we analyze sedentary behavior and just-in-time nudges in free-living conditions and show the effectiveness of learning contextualized and individualized JITAI. Currently, JITAI as discussed in the current literature is often conceptual in nature. We conduct a novel study that collects biometric, geospatial, and temporal information to understand individuals' context and evaluate delivered nudges. The collected data quantitatively agrees with other estimates regarding lack of physical activity and time spent sedentary. Individuals spend the most time overall, and the majority of time being sedentary, at home. Nudges provided at home have less of an impact than nudges provided at work. This suggests that future interventions should be carefully designed to target sedentary behavior at home. It may be possible that the physical structure of the work place creates a psychological bias for individuals to take more steps when nudged or that people just want a break while at work; however, further investigation would be required.

We demonstrate the need to tailor interventions to individuals due to the vastly different behavioral responses in each of the different contexts. There appears to be commonalities between individuals which indicate shared behaviors (e.g. high average daily step counts indicating few sedentary hours); however, tweaking the intervention strategy in content and context should allow for more effective nudging. This grouping is useful to an extent to resolve the data limitation issues when learning JITAI; however, individualization still needs to occur. More sophisticated grouping mechanisms than the provided ad-hoc grouping could be used. Using these grouped behaviors, areas requiring targeted behavior change mechanisms can be identified. Furthermore, this paper demonstrates that machine learning is a feasible tool to develop individualized JITAI. From 1 week of data, the learning algorithms were able to predict nudge receptiveness with 74 percent accuracy, which significantly outperforms an existing commercially available smart coach. The work presented in this paper improves the understanding of JITAI in free-living conditions to create tailored interventions.

8 FUTURE WORK

One of the limitations of the Fitbit is that it does not provide real-time data. We plan to examine the capabilities of the Apple WatchOS frameworks with respect to JITAI, compare the current Series 3 and 5 hardware platforms, and to pilot test the impact of simple JITAI for increasing walking-based MVPA and reducing sedentary activity. We are also currently working on capturing and using user preferences over contexts and timings to construct automated action and coping plans. JITAI would represent activities that the user is supposed to accomplish at different times and days to achieve their daily or weekly goals, and in the case the user fails to accomplish some of these activities, the system would then dynamically re-plan or reschedule to provide modified plans as JITAI.

In order to further address the data limitation issues to learn JITAI, we plan to examine similarity metrics among individuals using the collected real time data, calendar inputs, and meta-data questionnaire. Particularly, we plan to examine the trade-off between data availability and reliability of the estimated similarities. Once individuals are identified as similar, we then plan to examine how to re-use the information that we have collected. We plan to examine in which contexts an individuals' nudge receptiveness vary significantly. In order to further the state-of-the-art, longitudinal data needs to be collected. Unfortunately, GPS collection for location information requires significant battery consumption which may burden individuals too much to be used in longitudinal settings. We plan on exploring relationships between steps, heart rate, and location in order to estimate important features when information is unavailable. Once a longer time frame dataset is created, we plan to examine how people's receptiveness to nudges changes as they progress through the states of behavioral change.

REFERENCES

- [1] Talko B Dijkhuis, Frank J Blaauw, Miriam W Van Ittersum, Hugo Velthuis, and Marco Aiello. 2018. Personalized physical activity coaching: a machine learning approach. *Sensors* 18, 2 (2018), 623.
- [2] Greg Dominick and Matthew Saponaro. 2019. COMPARING RELATIVE AND ABSOLUTE INTENSITY BETWEEN CHEST AND WRIST-WORN HEART RATE SENSORS IN FREE-LIVING CONDITIONS. In *ANNALS OF BEHAVIORAL*

- MEDICINE, Vol. 53. OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA, S651–S651.
- [3] Bethany Barone Gibbs, Andrea L Hergenroeder, Peter T Katzmarzyk, I-Min Lee, and John M Jakicic. 2015. Definition, measurement, and health risks associated with sedentary behavior. *Medicine and science in sports and exercise* 47, 6 (2015), 1295.
 - [4] Kareann Sok-Fun Khow, Joanne Dollard, Kathy Bray, Carla Smyth, Mellick Chehade, Olga Theou, and Renuka Visvanathan. 2018. A randomized controlled feasibility study to evaluate the effects of a goal-setting coaching intervention using feedback from an accelerometer on sedentary time in older people at risk of falls (SMART-MOVE): a study protocol. *Pilot and feasibility studies* 4, 1 (2018), 173.
 - [5] Predrag Klasnja, Shawna Smith, Nicholas J Seewald, Andy Lee, Kelly Hall, Brook Luers, Eric B Hekler, and Susan A Murphy. 2019. Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps. *Annals of Behavioral Medicine* 53, 6 (2019), 573–582.
 - [6] Carl J Lavie, Cemal Ozemek, Salvatore Carbone, Peter T Katzmarzyk, and Steven N Blair. 2019. Sedentary behavior, exercise, and cardiovascular health. *Circulation research* 124, 5 (2019), 799–815.
 - [7] learnersdictionary. [n.d.]. *Learner Dictionary Parts of the Day*. <https://learnersdictionary.com/qa/parts-of-the-day-early-morning-late-morning-etc>
 - [8] I-Min Lee, Eric J Shiroma, Masamitsu Kamada, David R Bassett, Charles E Matthews, and Julie E Buring. 2019. Association of step volume and intensity with all-cause mortality in older women. *JAMA internal medicine* 179, 8 (2019), 1105–1112.
 - [9] Aniek Lentferink, Hilbrand Oldenhuis, Martijn Groot, Louis Polstra, Hugo Velthuisen, and Julia van Gemert-Pijnen. 2017. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. *Journal of medical Internet research* 19 (2017).
 - [10] Charles E Matthews, Kong Y Chen, Patty S Freedson, Maciej S Buchowski, Bettina M Beech, Russell R Pate, and Richard P Troiano. 2008. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *American journal of epidemiology* 167, 7 (2008), 875–881.
 - [11] Charles E Matthews, Stephanie M George, Steven C Moore, Heather R Bowles, Aaron Blair, Yikyung Park, Richard P Troiano, Albert Hollenbeck, and Arthur Schatzkin. 2012. Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. *The American journal of clinical nutrition* 95, 2 (2012), 437–445.
 - [12] Claire McCallum, John Rooksby, and Cindy M Gray. 2018. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. *JMIR mHealth and uHealth* 6, 3 (2018), e58.
 - [13] Susan Michie, Lucy Yardley, Robert West, Kevin Patrick, and Felix Greaves. 2017. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *Journal of medical Internet research* 19, 6 (2017), e232.
 - [14] Andre Matthias Müller, Ann Blandford, and Lucy Yardley. 2017. The conceptualization of a Just-In-Time Adaptive Intervention (JITAI) for the reduction of sedentary behavior in older adults. *Mhealth* 3 (2017).
 - [15] Inbal Nahum-Shani, Shawna Smith, Bonnie Spring, Linda Collins, Katie Witkiewitz, Ambuj Tewari, and Susan Murphy. 2018. Just-in-time adaptive interventions (JITAI) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462.
 - [16] Johanna Nurmi, Keegan Knittle, Todor Ginchev, Fida Khattak, Christopher Helf, Patrick Zwickl, Carmina Castellano-Tejedor, Pilar Lusilla-Palacios, Jose Costa-Requena, Niklas Ravaja, et al. 2020. Engaging Users in the Behavior Change Process With Digitalized Motivational Interviewing and Gamification: Development and Feasibility Testing of the Precious App. *JMIR mHealth and uHealth* 8, 1 (2020), e12884.
 - [17] World Health Organization et al. 2008. Physical inactivity: a global public health problem.
 - [18] Gillian A O'Reilly and Donna Spruijt-Metz. 2013. Current mHealth technologies for physical activity assessment and promotion. *American journal of preventive medicine* 45, 4 (2013), 501–507.
 - [19] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 707–718.
 - [20] Matthew Saponaro and Keith Decker. [n.d.]. Partial Policy Re-use in Connected Health Systems. ([n.d.]).
 - [21] Matthew Saponaro, Haoran Wei, and Keith Decker. 2017. Towards Learning Efficient Intervention Policies for Wearable Devices. In *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017 IEEE/ACM International Conference on*. IEEE, 298–299.
 - [22] Susan Schembre, Yue Liao, Michael Robertson, Genevieve Dunton, Jacqueline Kerr, Meghan Haffey, Taylor Burnett, Karen Basen-Engquist, and Rachel Hicklen. 2018. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. *Journal of medical Internet research* 20, 3 (2018), e106.
 - [23] J Graham Thomas and Dale S Bond. 2015. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization. *Health Psychology* 34, S (2015), 1261.
 - [24] Sabina Tomkins, Peng Liao, Serena Yeung, Predrag Klasnja, and Susan Murphy. 2019. Intelligent Pooling in Thompson Sampling for Rapid Personalization in Mobile Health. (2019).
 - [25] Richard P Troiano, David Berrigan, Kevin W Dodd, Louise C Masse, Timothy Tillet, and Margaret McDowell. 2008. Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise* 40, 1 (2008), 181–188.
 - [26] Jared M Tucker, Gregory J Welk, and Nicholas K Beyler. 2011. Physical activity in US adults: compliance with the physical activity guidelines for Americans. *American journal of preventive medicine* 40, 4 (2011), 454–461.
 - [27] Lin Yang, Chao Cao, Elizabeth D Kantor, Long H Nguyen, Xiaobin Zheng, Yikyung Park, Edward L Giovannucci, Charles E Matthews, Graham A Colditz, and Yin Cao. 2019. Trends in sedentary behavior among the US population, 2001–2016. *Jama* 321, 16 (2019), 1587–1597.
 - [28] Elad Yom-Tov, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, and Irit Hochberg. 2017. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *Journal of medical Internet research* 19, 10 (2017), e338.
 - [29] Deborah Rohm Young, Marie-France Hivert, Sofiya Alhassan, Sarah M Camhi, Jane F Ferguson, Peter T Katzmarzyk, Cora E Lewis, Neville Owen, Cynthia K Perry, Juned Siddique, et al. 2016. Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the American Heart Association. *Circulation* 134, 13 (2016), e262–e279.
 - [30] Chao Zhang, Daniel Lakens, and Wijnand IJsselstein. 2019. Theory Integration for Lifestyle Behavior Change in the Digital Age: An Adaptive Decision-making Framework. (2019).
 - [31] Mo Zhou, Yonatan Mintz, Yoshimi Fukuoka, Ken Goldberg, Elena Flowers, Philip Kaminsky, Alejandro Castillejo, and Anil Aswani. 2018. Personalizing Mobile Fitness Apps using Reinforcement Learning.. In *IUI Workshops*.