

# **Towards Tailoring Just-in-time Adaptive Intervention Systems for Workplace Stress Reduction: Exploratory Analysis of Intervention Implementation**

Jina Suh, Esther Howe, Robert Lewis, Javier Hernandez, Koustuv Saha, Tim Althoff, Mary Czerwinski

Submitted to: JMIR Mental Health  
on: May 17, 2023

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 36

    Figures ..... 37

        Figure 1..... 38

        Figure 2..... 39

    Multimedia Appendixes ..... 40

        Multimedia Appendix 1..... 41

        Multimedia Appendix 2..... 41

# Towards Tailoring Just-in-time Adaptive Intervention Systems for Workplace Stress Reduction: Exploratory Analysis of Intervention Implementation

Jina Suh<sup>1</sup> PhD; Esther Howe<sup>2</sup>; Robert Lewis<sup>3</sup>; Javier Hernandez<sup>1</sup> PhD; Koustuv Saha<sup>1</sup> PhD; Tim Althoff<sup>4</sup> PhD; Mary Czerwinski<sup>1</sup> PhD

<sup>1</sup>Microsoft Research Redmond US

<sup>2</sup>University of California, Berkeley Berkeley US

<sup>3</sup>Massachusetts Institute of Technology Cambridge US

<sup>4</sup>University of Washington Seattle US

## Corresponding Author:

Jina Suh PhD

Microsoft Research

1 Microsoft Way

Redmond

US

## Abstract

**Background:** Integrating stress-reduction interventions into the workplace may improve the health and well-being of employees, and there is an opportunity to leverage ubiquitous everyday work technologies to understand dynamic work contexts and facilitate stress-reduction wherever work happens. Sensing-powered just-in-time adaptive intervention (JITAI) systems have the potential to adapt and deliver tailored interventions, but such adaptation requires a comprehensive analysis of contextual and individual-level variables that may influence intervention outcomes and be leveraged to drive the system's decision-making.

**Objective:** This study aims to identify key tailoring variables that influence momentary engagement in digital stress-reduction micro-interventions to inform the design of similar JITAI systems.

**Methods:** To inform the design of such dynamic adaptation, we analyzed data from 43 participants during a four-week deployment of a system that incorporates passively sensed data across everyday work devices to send just-in-time stress-reduction micro-interventions in the workplace. We evaluated 27 trait-based factors (i.e., individual characteristics), state-based factors (i.e., workplace contextual and behavioral signals, momentary stress), and intervention-related factors (i.e., location, function) across 1,585 system-initiated interventions. We built logistical regression models to identify which factors contribute to the momentary engagement, the choice of interventions, the engagement given an intervention choice, the user rating of interventions engaged, and the stress reduction from the engagement.

**Results:** We found that women (OR=0.41, 95% CI [0.21, 0.77]), those with higher Neuroticism (OR=0.57, 95% CI [0.39, 0.81]), those with higher Cognitive Reappraisal skills (OR=0.69, 95% CI [0.52, 0.91]), and those that chose Calm (OR=0.43, 95% CI [0.23, 0.78]) and Address (OR=0.40, 95% CI [0.16, 0.97]) interventions were significantly less likely to experience stress reduction, while those with higher Agreeableness (OR=1.73, 95% CI [1.10, 2.76]) and those that chose prompt-based (OR=6.65, 95% CI [1.53, 36.45]) or video-based (OR=5.62, 95% CI [1.12, 34.10]) interventions were significantly more likely to experience stress reduction. We also found that work-related contextual signals such as lower Meeting Counts (OR=0.62, 95% CI [0.49, 0.78]) and lower Engagement Skewness (OR=0.64, 95% CI [0.51, 0.79]) were associated with a higher likelihood of engagement, indicating that state-based contextual factors such as being in a meeting or time of the day may matter more for engagement than efficacy. In addition, a JIT intervention that was explicitly rescheduled to a later time was more likely to be engaged (OR=1.77, 95% CI [1.32, 2.38]).

**Conclusions:** JITAI systems have the potential to integrate timely support into the workplace. Based on our findings, we recommend that individual, contextual, and content-based factors be incorporated into the system for tailoring as well as for monitoring unhelpful usage behaviors across subgroups and contexts.

(JMIR Preprints 17/05/2023:48974)

DOI: <https://doi.org/10.2196/preprints.48974>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org>, my preprint will be published as a full article.

## Original Manuscript

## Original Paper

# Towards Tailoring Just-in-time Adaptive Intervention Systems for Workplace Stress Reduction: Exploratory Analysis of Intervention Implementation

## Abstract

**Background:** Integrating stress-reduction interventions into the workplace may improve the health and well-being of employees, and there is an opportunity to leverage ubiquitous everyday work technologies to understand dynamic work contexts and facilitate stress-reduction wherever work happens. Sensing-powered just-in-time adaptive intervention (JITAI) systems have the potential to adapt and deliver tailored interventions, but such adaptation requires a comprehensive analysis of contextual and individual-level variables that may influence intervention outcomes and be leveraged to drive the system's decision-making.

**Objective:** This study aims to identify key tailoring variables that influence momentary engagement in digital stress-reduction micro-interventions to inform the design of similar JITAI systems.

**Methods:** To inform the design of such dynamic adaptation, we analyzed data from the implementation and deployment of a system that incorporates passively sensed data across everyday work devices to send just-in-time (JIT) stress-reduction micro-interventions in the workplace to 43 participants during a four-week deployment. We evaluated 27 trait-based factors (i.e., individual characteristics), state-based factors (i.e., workplace contextual and behavioral signals, momentary stress), and intervention-related factors (i.e., location, function) across 1,585 system-initiated interventions. We built logistical regression models to identify which factors contribute to the momentary engagement, the choice of interventions, the engagement given an intervention choice, the user rating of interventions engaged, and the stress reduction from the engagement.

**Results:** We found that women (OR=0.41, 95% CI [0.21, 0.77]), those with higher Neuroticism (OR=0.57, 95% CI [0.39, 0.81]), those with higher Cognitive Reappraisal skills (OR=0.69, 95% CI [0.52, 0.91]), and those that chose Calm (OR=0.43, 95% CI [0.23, 0.78]) and Address (OR=0.40, 95% CI [0.16, 0.97]) interventions were significantly less likely to experience stress reduction, while those with higher Agreeableness (OR=1.73, 95% CI [1.10, 2.76]) and those that chose prompt-based (OR=6.65, 95% CI [1.53, 36.45]) or video-based (OR=5.62, 95% CI [1.12, 34.10]) interventions were significantly more likely to experience stress reduction. We also found that work-related contextual signals such as higher Meeting Counts (OR=0.62, 95% CI [0.49, 0.78]) and higher Engagement Skewness (OR=0.64, 95% CI [0.51, 0.79]) were associated with a lower likelihood of engagement, indicating that state-based contextual factors such as being in a meeting or time of the day may matter more for engagement than efficacy. In addition, a JIT intervention that was explicitly rescheduled to a later time was more likely to be engaged (OR=1.77, 95% CI [1.32, 2.38]).

**Conclusions:** JITAI systems have the potential to integrate timely support into the workplace. Based on our findings, we recommend that individual, contextual, and content-based factors be incorporated into the system for tailoring as well as for monitoring unhelpful usage behaviors across subgroups and contexts.

**Keywords:** workplace stress; just-in-time; JITAI; engagement; micro-intervention; stress reduction; psychotherapy

## Introduction

Work is a major source of stress in the United States, affecting over half of Americans throughout most of the day [1]. Workplace stress leads to increased risk of mental and physical health disorders, decreased productivity and job satisfaction, and higher rates of accidents and employee costs [2–5]. Integrating stress reduction strategies directly into the workplace has shown to be effective and recommended [6, 7]. However, incorporating these individual-based techniques (e.g., cognitive-behavioral skills training, meditation, exercise) [7, 8] into the workday can be challenging due to work culture [9, 10] or psychological barriers [9–11].

Information workers who frequently use computing technology face challenges, such as prolonged desk-bound and sedentary behaviors, that contribute to chronic physical and mental health issues [12, 13]. High computer use without adequate breaks, high levels of multitasking, and constant connectivity demanded by information and communication technology (ICT) has been found to be associated with increased stress and burnout [14–17]. Despite ICTs being associated with increased stress, workplace computing tools can also be leveraged to understand and reduce stress [18, 19].

Passive sensing capabilities via ubiquitous devices have shown potential in health and wellbeing domains through monitoring and assessing individuals over time [20], and such data may be harnessed to provide precision mental health support. However, introducing new devices (e.g., wearables) to an organization can be costly (e.g., \$100 or more per worker) and could be impractical for real-world, daily functional use with compliance and quality issues [21, 22]. On the other hand, everyday technologies commonly used at work (e.g., webcam, keyboard, software telemetry) can be harnessed for passive sensing and behavioral analysis, offering a more feasible approach to infer affect [23], physiology [24], attention [25], or stress [18, 26].

Recently, technology-mediated support for mental health has generated interest for its ability to provide flexible and always-available access. However, these systems often provide generalized support that does not account for individual variabilities or contexts [27]. Despite an abundance of mental health apps [28, 29], digital interventions that fit the specific workplace context are still highly sought after [27, 30].

Digital mental health interventions, known as “digital micro-interventions,” leverage technology to adapt existing evidence-based psychosocial interventions (EBPIs) and leverage technology affordances to provide individual components of traditional psychotherapy focused on managing proximal symptoms (e.g., relaxation for stress) in the hopes of achieving broad, distal objectives (e.g., overcoming depression) [31]. Just-in-time adaptive interventions (JITAI) have been introduced as a way to deliver personalized, contextualized, and adaptable interventions using dynamic human behavior data captured through ubiquitous sensing technologies [27, 32–34]. This concept has been explored in various health contexts, such as promoting physical activity, stress management, and weight management, with recent interest in applying them to positive coping skill use [35].

Despite their promises, JITAI systems are not yet pervasive with many applications still relying on ecological momentary assessments (EMAs [36]). Recent developments in algorithmic and machine-learning approaches to dynamic adaptation have begun to show improvements in timing, receptivity, and engagement [37–40]. Operationalizing the adaptation of the system requires choosing appropriate tailoring variables and intervention options to drive the system’s decision-making regarding intervention timing and content. Prior research has examined factors associated with engagement in digital mental health interventions (DMHIs) [41, 42], including personal-related (e.g., demographics, personalities), content-related (e.g., perceived fit, usefulness), and technology-related

factors (e.g., technical issues, privacy). However, most prior work investigates study-long engagement, rather than examining *in-the-moment* engagement factors, which are crucial for improving the usability of interventions in real-world contexts [31], especially since the integration of intervention usage into life is a core facilitator for engagement [42].

A recent meta-analysis [36] has also shown that tailoring is a significant aspect of JITAI design associated with greater efficacy. Despite JITAI systems' potential to provide precision support [27], few studies demonstrate the value of just-in-time support in improving user engagement [43, 44]. Although the use of passively sensed data for contextual understanding and system adaptation is often recommended for the design of JITAI systems [27], many still rely mostly on EMAs, app usage, or simple temporal features primarily from a single modality (i.e., mobile devices) or without considering individual or intervention-related factors [36, 44]. Passive sensing technologies offer numerous sets of contextual variables (e.g., location, calendar, movement, activity), presenting a challenge in designing sensing-capable JITAI systems: identifying a core set of tailoring variables among all possible variables that the system should consider for optimizing effective engagement [45]. Therefore, our primary goal and key contribution is to identify crucial tailoring variables that influence *momentary* engagement in digital stress-reduction micro-interventions to inform the design of similar JITAI systems.

In this paper, we analyze engagement data from a four-week longitudinal deployment of a workplace stress-reduction intervention system [19]. We leverage everyday workplace devices as unobtrusive and passive sensors to gain a glimpse into participants' daily work activities (e.g., emails, meetings, computer activity). This system leverages the cloud to integrate passively collected data and EMAs to deliver JIT nudges to engage in digital micro-interventions across devices. Unlike laboratory experiments or controlled studies, our study allows observing users' moment-by-moment interactions with the system in naturalistic work environments. We combine passively sensed work contexts, the system usage, including which intervention participants chose and liked, individual demographics, personality traits, and coping styles to understand how the work context influences engagement patterns and to understand the appropriate conditions that lead to momentary intervention engagement and positive outcomes. We leverage such data from 43 participants to contextualize 1,585 system-initiated interventions. From statistical modeling of the impact of individual, contextual, and intervention-related factors on engagement outcomes, we confirmed that individual factors (e.g., age, gender, personality traits, coping skills), as well as contextual and content-related factors (e.g., availability, intervention modality), significantly influenced momentary intervention engagement, intervention choice, user ratings, and stress reduction outcomes. These findings suggest tailoring guidelines for JITAI systems whereby contextual and personalized factors can be used to find a positive balance between user preferences and maximal intervention efficacy.

In preparing this paper, we referred to the Guidelines and Checklist for the Reporting on Digital Health Implementations (iCHECK-DH) [46].

## Methods

### Objectives

There are multiple considerations that influence JITAI systems' decision points (i.e., "a time at which an intervention decision is made" [32]), such as the right timing for a prompt, the right intervention for the moment, or the intervention likely to be engaged in. Therefore, this study aimed to identify factors that may contribute to improving participant engagement for JITAI systems. The study builds on a pilot implementation and deployment of a just-in-time (JIT) stress-reduction micro-



interventions in a real-world workplace setting, targeting information workers who spend most of their working hours at processing information with computing devices [47]. The deployed system did not adapt the intervention content based on changing state or user context and, therefore, is not a full JITAI system. Instead, the study conducted a retrospective analysis of the deployment data to understand factors that influence the engagement and efficacy of workplace stress-reduction JITAI systems for future development.

## Participants

Information workers were recruited from a large technology organization via randomly distributed email advertisements. We recruited primarily US-based workers for the ease of system troubleshooting and to minimize any country-specific organizational factors. We enrolled participants on a first-come, first-served basis as long as we could satisfy system compatibility. To ensure system compatibility, interested participants completed a brief screener survey about their work set up (e.g., primary device specification and OS, web camera availability). Eligible participants, whose primary device specification met our sensing software requirements, were asked to install and run the study system on their primary desktop for 30 minutes. Only 43 participants that could run the sensing software on their desktop for 30 minutes or more were selected to participate in the study. The sample size of roughly 40 was determined for statistical power within a larger experimental study [19] that compared the JIT condition with a baseline condition for intervention effectiveness.

Our intake survey included several demographics measures such as age, gender, and role. 67.4% of the 43 participants identified as male. This distribution of gender closely aligns with the current industry demographics for large technology companies [48], and therefore, we consider this gender representation acceptable for our analysis. Three participants self-reported to be 18-25 years old, 11 to be 26-35 years old, 18 to be 36-45 years old, eight to be 46-55 years old, two to be 56-65 years old, and one to be 66+ years old. 55.8% of the participants reported as being in engineering/development role, and 32.6% reported as being in sales/business strategy.

Other intake measures included the Depression, Anxiety, and Stress Scale 21 (DASS-21) [49], the brief Big Five Personality Inventory (BFI-10) [50], the Emotional Regulation Questionnaire (ERQ) [51] designed to measure the tendency to regulate emotions through cognitive reappraisal and expressive suppression, and the 6-item Brief Resilience scale [52] that measures the ability to bounce back from stress. The average stress-level of the stress subscale of DASS-21 reported by participants was 5 out of 21 ( $\sigma = 3.8$ ) which is within normal ranges. Our participants scored an average of 3.8 for agreeableness, 4.1 for conscientiousness, 2.6 for extraversion, 2.8 for neuroticism, and 3.5 for openness out of 5. Cognitive reappraisal was scored at 4.7 out of 7 ( $\sigma = 1.1$ ), expressive suppression was at 3.7 out of 7 ( $\sigma=1.3$ ), and resilience was at 3.5 out of 5 ( $\sigma=0.9$ ) on average.

## Study implementation and procedure

We deployed our system to 43 consented participants for four weeks. During every workday of the four-week study period, the system asked users to complete five EMAs per day during their reported work hours to capture their subjective stress ratings from the past 30 minutes (i.e., “How would you rate your level of stress during the last 30 minutes?”). When the system determined that the user’s stress level may be high, the system sent JIT nudges via a chatbot, asking users to engage in a stress-reduction micro-intervention. In the background, the system captured usage data as well as passively sensed contextual data. For detailed description of the system architecture, please see Section 1 of Multimedia Appendix 1.

## *JIT heuristics*

The system determined higher-than-baseline stress level based on our JIT heuristics informed by computed stress scores and self-reported stress levels. Stress scores were computed in real time per individual as an average of 5 components ranging between 0 and 1, each representing five components that previous work has identified as sources of stress: (1) the number of email received [15, 53], (2) the total number of meetings in a given day [54], (3) the percentage of time into the day [55], (4) the amount of facial expressions (via the Facial Action Coding System [56]) from corrugator (i.e., brow furrowing) and lip depressor (i.e., frowning) minus zygomatic major (i.e. smiling) [55, 57, 58], and (5) heart rate [59, 60]. Self-reported stress levels were obtained from EMAs.

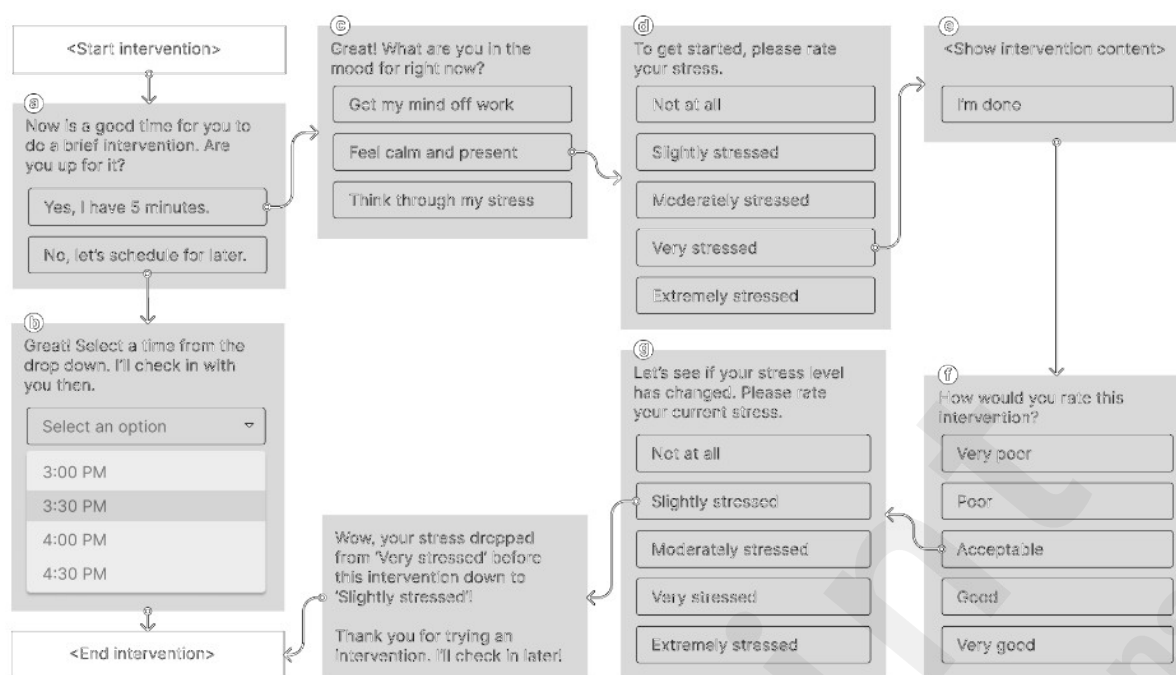
In our JIT heuristics, first, we compute each user's baselines as the average of the computed stress scores and self-reported stress levels based on the data from the first week of using the system. These individualized baselines (captured at week one of the four-week study) are used as thresholds for delineating high-stress from low-stress during subsequent weeks (weeks two to four). During the first week, we use the default baseline at the middle of the score range. Then, we send intervention nudges only if it is during the working hours that the users have stated at intake, if they have not explicitly scheduled an intervention at a later time that day, if they have not completed an intervention in the past hour, if there has not been an intervention nudge in the past two hours, and if there have not been four or more intervention nudges that day. For detailed description of our stress inference and heuristics, please see Section 1 of Multimedia Appendix 1.

## *Micro-interventions*

Micro-interventions [31] used in our study were translated from components of Cognitive Behavioral Therapy (CBT) and Dialectical Behavioral Therapy (DBT), two empirically supported and widely used psychotherapy modalities [61, 62]). These were under five-minute interventions that were either a short video, a single-turn text prompt, or a brief therapeutic conversation with the chatbot. The micro-interventions used in our study can be categorized by (a) the function served for users, (b) the modality in which the intervention was delivered, and (c) the intended location to perform the intervention. Table 1 lists all categories of micro-interventions and an example for each category.

Micro-interventions are primarily categorized into three functional categories which align approximately with the amount of effort required. 'Get my mind off work' interventions are low-effort interventions designed to help users take their mind away from work with positive activities that promote emotion regulation [63]. 'Feel calm and present' interventions are medium-effort interventions that help users feel calm and present by drawing inspiration from the mindfulness practices. 'Think through my stress' interventions are high-effort interventions that help users think through their stress and directly address and problem-solve stress-inducing components of their lives. For simplicity, we refer to these three intervention categories as 'Distract', 'Calm', and 'Address', respectively, in the rest of the paper. Overall, there were 18 interventions per functional category. Please see Section 2 of Multimedia Appendix 1 for more details about the micro-interventions.

Figure 1. (a) System sends a nudge to users to perform an intervention. (b) Users can opt to postpone the intervention at a later time. (c) If users choose to do an intervention, they can select from one of three intervention categories. (d) Users first self-report their current stress level. (e) The system shows the intervention content for users to interact with. (f) User rates the intervention. (g) Users self-report their stress levels after the intervention.



## User engagement flow

Figure 1 illustrates a series of user engagement steps when a system nudge is sent. When the system sends a nudge to perform an intervention (Figure 1a), users can choose to delay it to a later time that day (Figure 1b). The system does not send another nudge until that time. Users may ignore the nudge, and the system expires the nudge after 30 minutes of inactivity. If the user decides to engage in an intervention, they can choose from three intervention categories (Figure 1c), and then the system randomly selects a new intervention within the category. Before beginning the intervention, the users are asked to subjectively rate their current stress level (Figure 1d). Then, the intervention content is shown to the user (Figure 1e). Once the intervention is done, the users are asked to rate the intervention (Figure 1f) and rate their stress level again (Figure 1g) before concluding the intervention flow. An example screenshot of the full user engagement flow can be found in Figure S4 of Multimedia Appendix 1.

## Other study tasks

Beyond JIT-based interactions with the system, participants could also access interventions on demand, where they could perform the intervention at that moment or schedule it to a later time that day. Participants were also asked to complete morning surveys to log their sleep quality and evening surveys to log their food and beverage intake. We additionally asked participants to rate their sleep quality the night before via a morning survey and log their food and drink consumptions throughout the day via an evening survey. They also completed weekly surveys including the DASS-21 and the Brief Resilience scales. The exit survey solicited feedback about the usability of the system and the perceived helpfulness and impact of the interventions.

## Post-study analysis

This section describes how we processed and analyzed the data collected from the above study to understand which factors contribute to improving participant engagement in a system-initiated intervention and the effectiveness of interventions at a given moment. Table 2 describes the full set of variables selected for our analysis and their descriptive statistics. Section 3 of Multimedia Appendix 1 includes additional detail, including descriptive statistics, inclusion/exclusion of data

points, and correlational analyses conducted between variables.

Table 2. List of the per-participant, per-half hour, per-nudge, and per-intervention variables and their descriptive statistics.

Variable	Distribution
<b>Per-participant</b>	
Age	18-35 (14), 36-45 (18), 46+ (11)
Gender	Man (29), Woman (14)
Cognitive Reappraisal	$\bar{x}=4.69$ , $\sigma=1.09$ , min=2, max=7
Expressive Suppression	$\bar{x}=3.74$ , $\sigma=1.26$ , min=2, max=6
Resilience	$\bar{x}=3.51$ , $\sigma=0.89$ , min=2, max=5
Agreeableness	$\bar{x}=3.79$ , $\sigma=0.74$ , min=2, max=5
Conscientiousness	$\bar{x}=4.12$ , $\sigma=0.83$ , min=2, max=5
Extraversion	$\bar{x}=2.62$ , $\sigma=0.86$ , min=1, max=4
Neuroticism	$\bar{x}=2.85$ , $\sigma=1.04$ , min=1, max=5
Openness	$\bar{x}=3.48$ , $\sigma=0.79$ , min=2, max=5
Engagement Skewness	$\bar{x}=-0.09$ , $\sigma=0.58$ , min=-1.41, max=1.41
<b>Per-half hour</b>	
Nudge Probability	$\bar{x}=0.06$ , $\sigma=0.04$ , min=0.02, max=0.30
<b>Per-nudge</b>	
Meeting Counts	$\bar{x}=0.30$ , $\sigma=0.51$ , min=0, max=3
No Meeting Minutes	$\bar{x}=5.43$ , $\sigma=11.79$ , min=0, max=50
Self Event Counts	$\bar{x}=0.12$ , $\sigma=0.37$ , min=0, max=3
Email Messages Sent	$\bar{x}=0.39$ , $\sigma=0.88$ , min=0, max=8
Email Messages Read	$\bar{x}=3.22$ , $\sigma=5.09$ , min=0, max=54
Chat Messages Count	$\bar{x}=3.88$ , $\sigma=5.90$ , min=0, max=64
Adhoc Call Count	$\bar{x}=0.04$ , $\sigma=0.21$ , min=0, max=2
Number of Attention Signals	$\bar{x}=1442.64$ , $\sigma=1316.68$ , min=0, max=5705
Nudge Source	JIT algorithm (1337), Rescheduled (248)
Engaged	TRUE (563), FALSE (1022)
<b>Per-intervention (chosen)</b>	
Category	Address (112), Calm (338), Distract (200)
Engaged	TRUE (563), FALSE (87)
Modality	Video (113), Prompt (384), Conversation (24)
Location	At desk (422), Inside (84), Outside (15)
<b>Per-intervention (completed)</b>	
Stress Reduction	$\bar{x}=0.29$ , $\sigma=0.53$ , min=-1, max=3
Rating	$\bar{x}=3.61$ , $\sigma=1.02$ , min=1, max=5
Improved	TRUE (150), FALSE (371)
Liked	TRUE (289), FALSE (232)
Stress Before	$\bar{x}=2.14$ , $\sigma=0.97$ , min=1, max=5
Category	Address (77), Calm (275), Distract (72)
Modality	Video (60), Prompt (340), Conversation (24)

Location	At desk (325), Inside (84), Outside (15)
----------	------------------------------------------

## Engagement outcome variables

From participants' interactions with the system, we reconstructed each participant's step-by-step interaction with the system as represented in Figure 1 to extract several outcome variables, such as whether a system-initiated intervention was completed and the effectiveness and rating of interventions once engaged. We note that the number of data points varies based on where in the engagement flow the participant exits.

### Engagement label

We labeled each system-initiated intervention as “engaged” in an intervention (i.e., *Engaged* = TRUE) if the participant explicitly marked the intervention as done (i.e., clicking on “I’m done” button in Figure 1e), regardless of whether they completed any subsequent prompts (i.e., Figure 1f-g). Any ignored, incomplete, or timed-out nudges were considered not engaged. Each system-initiated intervention that was triggered at a postponed time (Figure 1b) was categorized as “rescheduled” (i.e., *Nudge Source* = Rescheduled). Although these rescheduled nudges looked identical to JIT nudges, we hypothesized that the participants would be more likely to engage in a system-initiated intervention if they postponed the intervention to a time that is more suitable for engagement.

### Intervention choice

The characteristics of interventions, such as location, modality, and effort level, are important to consider because certain interventions might not be feasible in certain situations (e.g., the participant cannot go outside) [64, 65]. If a participant chose an intervention category (Figure 1c), we marked that nudge with binary labels of *Distract Chosen*, *Calm Chosen*, or *Address Chosen*. Each of these interventions was further labeled with *Modality* and *Location* based on the specific intervention that the system chose within the category.

### Intervention effectiveness and rating

We looked at two outcome metrics – momentary *Stress Reduction* and intervention *Rating*. We further binarized these outcome metrics to determine if a certain intervention engagement *Improved* stress (i.e., self-reported stress rating was lowered after the intervention use) and if the participant *Liked* the intervention (i.e., rated as ‘Good’ or ‘Very good’).

## Tailoring variables for post-deployment analysis.

We used per-participant demographics and validated scale responses (coping skills and personality traits) to explain individual differences. We harnessed passively sensed data streams to explain the context surrounding the system initiation and intervention engagement. Based on participants' interactions with the system, we also account for the probability distribution of system nudges throughout the day and the participant's likelihood of engaging in interventions in the morning or the afternoon.

### Participant Characteristics

We included demographic variables, such as age and gender, and personality traits that have been shown to impact engagement [42]. Because of the small sample sizes on either end of the age groups, we combined the lower two and upper three age groups to create more balanced age groups (i.e., 18-35, 36-45, and 46+). We included emotion regulation [51] and resilience [52] as measures of a



person's ability to cope with stressors. We also included the Big Five personality traits [66–68] because they are known to impact stress [69, 70] and engagement in mental health treatment [42].

### Passively sensed context

To understand the context surrounding a system-initiated intervention, we leveraged two sources of passively sensed data: custom sensing software and Viva Insights.

The custom sensing software ran on participants' desktops and can capture activities that may not be associated with work, such as browsing the internet or using non-work-related software. It captured data from the 5 components used to compute stress scores in real time, as described earlier. The sensing software also captured general user computer activity events such as mouse and keyboard interactions into a single metric, *Number of Attention Signals*, which could be an important indicator for presence. We hypothesized that presence at the computer could lead to higher engagement in interventions as the nudges were designed to grab the attention of participants at work. The range of values for *Number of Attention Signals* was fairly large (max=5705) compared to other variables, so we divided the values by 2000 to estimate a comparable coefficient and confidence intervals during modeling (i.e., to have odds ratios within 2 decimal points). When interpreting the effect sizes, we corrected for this factor of 2000. We also hypothesized that the likelihood of engaging in a stress-reduction intervention during active participation in a meeting is low. Therefore, we included *No Meeting Minutes* to represent the total number of minutes without a scheduled meeting with others and *Self Event Count* as the total number of calendar events with only the participant as the attendee.

Viva Insights captures de-identified activity aggregates in 30-minute windows for Microsoft tools across all devices associated with an individual's work account. From Viva Insights, we included *Meeting Counts*, *Adhoc Call Count*, and *Email Messages Sent/Read*. We excluded *Chat Messages Count* from our analysis because the nudges were delivered through Teams, and our data source cannot be used to discern if the messages were coming from the bot. Because Viva Insights data is limited to half-hourly windows, we associated the contextual metrics with each system-initiated intervention by taking the half-hour window that holds the nudge timestamp.

Both data sources had several overlaps or similar metrics, such as ones related to meetings or emails. Correlational analysis between two data sources is described in Section 3 of Multimedia Appendix 1. It is important to note that contextual data at work can be noisy due to individual differences in the usage of work tools for nonwork purposes (e.g., personal use, subscriptions). We do not differentiate work vs nonwork data because it is challenging, for example, to isolate work-related emails as the only source of stress.

### System-initiated intervention probability

Although each of the system nudges could be considered as an independent, repeated observation, the timing of the system-initiated interventions was sometimes dependent on when EMAs were administered. Because the system's JIT heuristic runs every 5 minutes to check if a nudge needs to be sent based on the stress score and the EMA stress ratings, the most likely hours for receiving a system nudge is shortly after the EMA, leading to each participant receiving more nudges during certain hours of the day than others. To account for such variability in receiving system-initiated interventions, we incorporate the momentary nudge probability in our analysis. Because the range of nudge probabilities is small ( $\mu=0.06$ ,  $\sigma=0.04$ ), we multiply the measure by 100 to represent it in percentages.

### Temporal engagement skewness

Prior research has found that different hours of the day were seen as good or not-so-good timing for stress interventions [33]. To examine if a certain participant has a temporal tendency to engage, we computed the Fisher-Pearson coefficient of skewness, or *Engagement Skewness*, on the hourly intervention engagements per participant. A positive *Engagement Skewness* means that the participants tend to engage at the beginning of their workday, and a negative *Engagement Skewness* means that the participants tend to engage towards the end of their workday. We used the skewness metric instead of simply looking at the engagement during the morning and the afternoon to account for individual differences in working hours. We incorporated this skewness per participant in our analysis.

## Analysis

Taking all the tailoring and outcome variables into account, our analysis focused on estimating the effects of contextual, individual, and intervention characteristics on binary outcome variables (e.g., *Engaged*, *Distract Chosen*, *Liked*). Thus, we built a logistic regression model predicting each outcome based on a combination of per-participant characteristics, per-half-hour nudge probability, per-nudge contextual metrics, or per-intervention metrics as fixed effects. The outputs of the logistical regression models are presented as odds ratios in Table 3-Table 7, representing a ratio of odds (e.g., probability of engaging vs probability of not engaging) under two different conditions (e.g., being a woman vs not being a woman). The data processing was conducted using Python packages (e.g., numpy, pandas, scipy, seaborn) and the models were tested using R libraries (e.g., lme4, car, performance).

We determined the significance of the fitted model against the null hypothesis model using the analysis of deviance. We conducted analyses of variance to estimate the significance of fixed effects. Because the data is unbalanced (i.e., unequal number of observations for each level of a factor), we obtain ANOVA Type II sums of squares [71]. For categorical variables of 3 levels or more (e.g., *Age Group*, *Category*), we estimated pairwise differences using Tukey's HSD procedure. Multicollinearity in fixed effects was tested using the variance of inflation factor (VIF), and none of our models exhibited a multicollinearity issue. Results for ANOVA, multicollinearity, and Tukey analyses are found in Section 4 of Multimedia Supplementary 1.

## Ethical considerations

The study was reviewed by the Microsoft Research Institutional Review Board (OHRP IORG #0008066, IRB #IRB00009672) prior to the research activities and was formally approved. In addition to the ethics review, our study obtained approvals from Microsoft's privacy, security, and legal review officers prior to conducting the study. All participants provided consent as part of the onboarding process and agreed for their de-identified data to be used for research purposes. The consent described the installation of our system, interaction with interventions, surveys, and joining the study data with cloud-based, device-independent telemetry data. All data, collected anonymously or otherwise, were collected and stored in a secured and access-controlled location. All data were joined and de-identified prior to analysis by the research team. For their participation and data, each participant was compensated with a \$400 Amazon gift card.

## Implementation considerations

The implementation of the system used in the deployment study has dependencies that restrict interoperability and sustainability. The system requires that the participants utilize Microsoft platforms (e.g., M365, Windows, Viva Insights, and Teams) and own a decent desktop that can perform vision-based data processing on device. Our current implementation focused on the

understanding of individual, contextual, and intervention-related factors to drive the tailoring of JITAI systems and does not allow generalizability beyond the supported architecture. However, we believe that recent advancements in generative AI technologies and the implementation of interoperability layers across different technology ecosystems can enable more accessible implementation of the system. As this study was a pilot implementation, any budget planning, sustainability model, or interoperability for sustained deployment were not in scope for this paper.

## Results

To see which factors influenced engagement, we modeled *Engaged* as a function of per-participant characteristics, per-half hour nudge probability, and per-nudge contextual metrics. The logistic regression model of *Engaged* with all fixed effects was significantly different from the null hypothesis model (i.e.,  $\text{Engage} \sim 1$ ;  $\chi^2(21)=115.5$ ,  $p<0.001$ ). Table 3 outlines the odds ratios and confidence intervals for each predictor. Predictors *Engagement Skewness*, *Meeting Count*, *Number of Attention Signals*, and *Nudge Source* remained significant after applying the Benjamini-Hochberg correction for multiple comparisons. Reviewing the coefficients of our fixed effects, we found that being 46+ years old, being a woman, higher *Cognitive Reappraisal*, higher *Number of Attention Signals*, and receiving a rescheduled nudge were associated with a higher likelihood of intervention engagement. On the other hand, higher *Engagement Skewness* and *Meeting Count* were associated with to a lower likelihood of intervention engagement.

Table 3. Odds ratios (OR) and confidence intervals (CI) for each predictor of *Engaged* for all system-initiated nudges.

Predictors	Engaged	
	OR	CI
(Intercept)	0.1 <sup>a</sup>	0.03 – 0.30
Age Reference (18-35)		
Age Group (36-45)	1.17	0.88 – 1.56
Age Group (46+)	1.47 <sup>b</sup>	1.07 – 2.02
Gender Reference (Man)		
Gender (Woman)	1.37 <sup>b</sup>	1.05 – 1.79
Cognitive Reappraisal	1.14 <sup>b</sup>	1.01 – 1.29
Expressive Suppression	0.92	0.83 – 1.02
Resilience	1.05	0.88 – 1.25
Agreeableness	0.87	0.72 – 1.05
Conscientiousness	1.13	0.96 – 1.33
Extraversion	1.03	0.89 – 1.21
Neuroticism	0.97	0.83 – 1.12
Openness	1.15	0.99 – 1.34
Engagement Skewness	0.64 <sup>a</sup>	0.51 – 0.79
Nudge Probability	1.02	1.00 – 1.05
Meeting Counts	0.62 <sup>a</sup>	0.49 – 0.78
No Meeting Minutes	1	0.99 – 1.01
Self Event Counts	1.15	0.85 – 1.54
Email Messages Sent	1.05	0.92 – 1.20
Email Messages Read	1.01	0.99 – 1.03
Adhoc Call Count	0.89	0.52 – 1.49
Number of Attention Signals	1.39 <sup>a</sup>	1.17 – 1.66
Trigger Source Reference (System)		
Trigger Source (Rescheduled)	1.77 <sup>a</sup>	1.32 – 2.38
Observations = 1585; Tjur's $R^2=0.073$		



<sup>a</sup> statistically significant ( $p < 0.05$ ) after Benjamini-Hochberg correction

<sup>b</sup> statistically significant ( $p < 0.05$ ) before any correction

Next, we modeled each of the three choice outcome measures – *Distract Chosen*, *Calm Chosen*, and *Address Chosen* – as a function of per-participant characteristics, per-half hour nudge probability, and per-nudge contextual metrics. Table 4 outlines the odds ratios and confidence intervals for all models.

For choosing Distract interventions, we found that higher *Openness* was associated with a higher likelihood. On the other hand, being a woman and higher *Number of Attention Signals* were associated with a lower likelihood of choosing Distract interventions. We found that higher *Cognitive Reappraisal* was associated with a higher likelihood and lower *Openness* was associated with a lower likelihood of choosing Calm interventions. We also found that higher *Resilience* was associated with a higher likelihood and lower *Cognitive Reappraisal* was associated with a lower likelihood of choosing Address interventions.

To understand the effect of intervention choice on engagement, we modeled *Engaged* as a function of per-participant characteristics, per-half hour nudge probability, per-nudge contextual metrics, and per-intervention characteristics. The logistic regression model of *Engaged* was significant ( $\chi^2(27)=69.595$ ,  $p \ll 0.001$ ). Table 5 outlines the odds ratios and confidence intervals for each predictor. We found that a higher *Resilience*, choosing a prompted-based intervention, and choosing a video-based intervention were associated with a higher likelihood of engagement. On the other hand, choosing an intervention that could be performed inside was associated with a lower likelihood of engagement.

Table 4. Odds ratios (OR) and confidence intervals (CI) for each predictor of *Distract Chosen*, *Calm Chosen*, *Address Chosen* for all participant-chosen interventions.

	Distract Chosen		Calm Chosen		Address Chosen	
Predictors	OR	CI	OR	CI	OR	CI
(Intercept)	0.11 <sup>b</sup>	0.01 – 0.73	1.84	0.32 – 10.62	0.44	0.04 – 4.52
Age Ref. (18-35)						
Age Group (36-45)	1.65 <sup>b</sup>	1.03 – 2.67	0.62 <sup>b</sup>	0.40 – 0.96	0.88	0.46 – 1.65
Age Group (46+)	0.71	0.40 – 1.26	0.95	0.57 – 1.57	1.6	0.85 – 3.07
Gender Ref. (Man)						
Gender (Woman)	0.62 <sup>b</sup>	0.39 – 0.98	1.31	0.87 – 1.96	1.29	0.73 – 2.26
Cognitive Reappraisal	1.01	0.82 – 1.25	1.27 <sup>b</sup>	1.06 – 1.53	0.67 <sup>a</sup>	0.53 – 0.84
Expressive Suppression	0.99	0.83 – 1.19	1.06	0.90 – 1.25	0.95	0.75 – 1.18
Resilience	0.93	0.69 – 1.25	0.87	0.67 – 1.13	1.42 <sup>b</sup>	1.00 – 2.02
Agreeableness	0.75	0.55 – 1.03	1.02	0.77 – 1.36	1.46	0.99 – 2.20
Conscientiousness	1.30	0.98 – 1.75	0.92	0.71 – 1.18	0.74	0.51 – 1.05
Extraversion	1.18	0.91 – 1.53	1.00	0.79 – 1.27	0.79	0.57 – 1.11
Neuroticism	0.88	0.68 – 1.14	1.00	0.79 – 1.26	1.36	0.99 – 1.89
Openness	1.67 <sup>a</sup>	1.30 – 2.16	0.71 <sup>b</sup>	0.57 – 0.88	0.83	0.60 – 1.13
Engagement Skewness	1.12	0.74 – 1.68	1.11	0.78 – 1.59	0.86	0.53 – 1.38
Nudge Probability	1.00	0.96 – 1.05	1.01	0.97 – 1.05	0.97	0.92 – 1.03
Meeting Counts	0.84	0.55 – 1.25	1.07	0.76 – 1.52	1.14	0.71 – 1.77
No Meeting Minutes	1.00	0.99 – 1.02	1.00	0.99 – 1.01	1.00	0.98 – 1.02

Self Event Counts	1.08	0.63 – 1.79	0.83	0.53 – 1.30	1.19	0.65 – 2.05
Email Messages Sent	1.05	0.84 – 1.30	1.00	0.82 – 1.22	0.99	0.74 – 1.30
Email Messages Read	1.03	0.99 – 1.07	0.99	0.96 – 1.03	0.96	0.90 – 1.01
Adhoc Call Count	0.88	0.36 – 1.94	0.93	0.45 – 1.93	1.26	0.45 – 3.05
Number of Attention Signals	0.65 <sup>a</sup>	0.48 – 0.87	1.18	0.92 – 1.53	1.33	0.94 – 1.87
Trigger Source Ref. (System)						
Trigger Source (Rescheduled)	1.01	0.61 – 1.63	0.76	0.49 – 1.16	1.39	0.81 – 2.35
Observations = 650	<i>Tjur's R<sup>2</sup></i> =0.115		<i>Tjur's R<sup>2</sup></i> =0.059		<i>Tjur's R<sup>2</sup></i> =0.089	
<sup>a</sup> statistically significant (p < 0.05) after Benjamini-Hochberg correction						
<sup>b</sup> statistically significant (p < 0.05) before any correction						

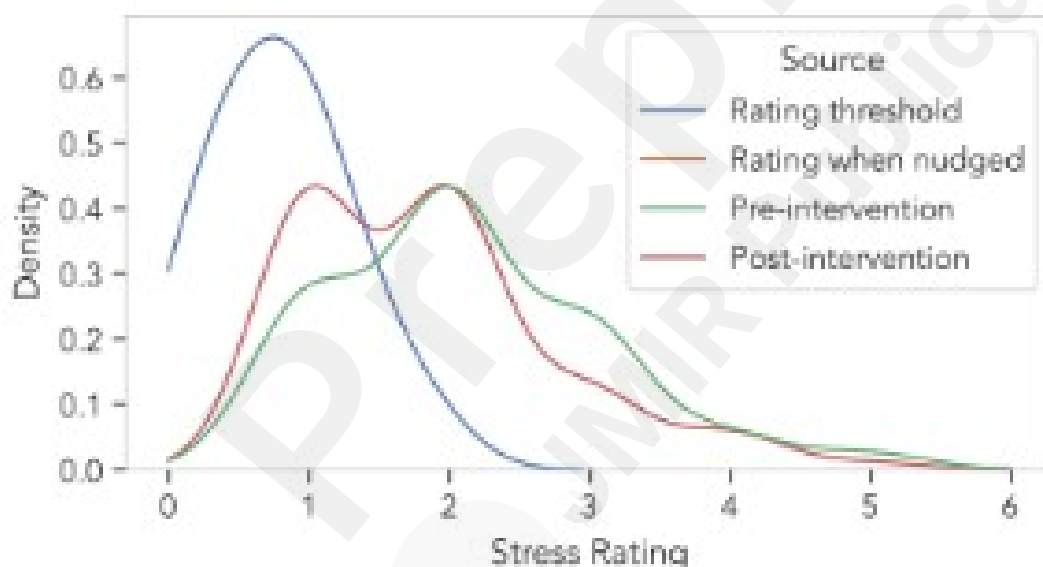
Table 5. Odds ratios (OR), confidence intervals (CI), and p-values for each predictor of *Engaged after chosen*, i.e., Engaged for all interventions after participants chose a category.

Predictors	Engaged after chosen	
	OR	CI
(Intercept)	0.05 <sup>b</sup>	0.00 – 0.94
Age Reference (18-35)		
Age Group (36-45)	1.06	0.54 – 2.05
Age Group (46+)	1.65	0.71 – 4.02
Gender (Woman)	1.65	0.88 – 3.16
Cognitive Reappraisal	1.11	0.83 – 1.48
Expressive Suppression	1.25	0.95 – 1.57
Resilience	1.63 <sup>b</sup>	1.05 – 2.59
Agreeableness	1.24	0.77 – 2.00
Conscientiousness	0.81	0.55 – 1.17
Extraversion	0.72	0.48 – 1.08
Neuroticism	1.39	0.96 – 2.07
Openness	1.24	0.88 – 1.76
Engagement Skewness	1.01	0.60 – 1.73
Nudge Probability	1.01	0.95 – 1.07
Meeting Counts	0.65	0.41 – 1.06
No Meeting Minutes	1	0.98 – 1.02
Self Event Counts	1.94	0.86 – 5.61
Email Messages Sent	1.16	0.84 – 1.67
Email Messages Read	0.97	0.93 – 1.03
Adhoc Call Count	0.54	0.23 – 1.33
Number of Attention Signals	0.9	0.61 – 1.33
Trigger Source Reference (System)		
Trigger Source (Rescheduled)	1.79	0.85 – 4.10
Engagement Skewness	1.01	0.60 – 1.73
Category Reference (Distract)		
Category (Calm)	1.14	0.60 – 2.14
Category (Address)	0.55	0.22 – 1.43
Modality Reference (conversation)		
Modality (prompt)	3.53 <sup>b</sup>	1.35 – 9.60
Modality (video)	5.86 <sup>b</sup>	1.69 – 21.68

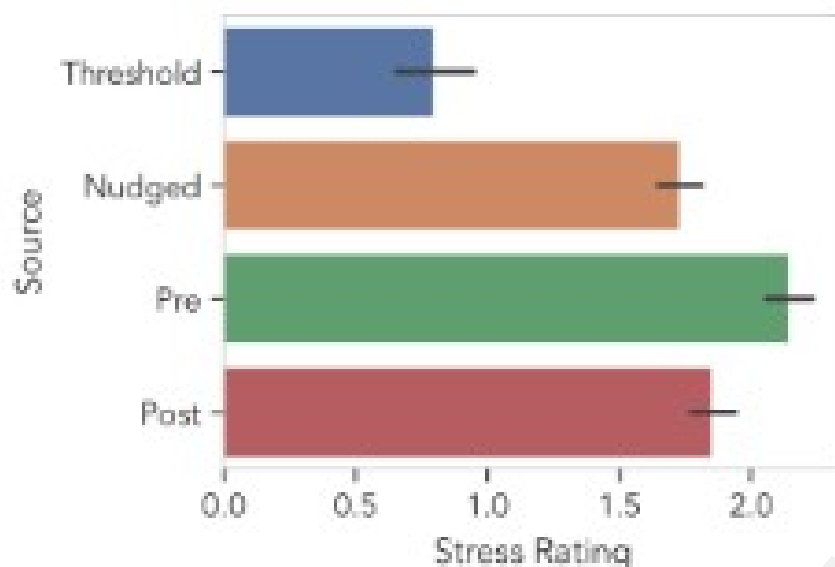
Location Reference (at desk)		
Location (Inside)	0.43 <sup>b</sup>	0.23 – 0.83
Location (Outside)	0.34	0.11 – 1.18
Observations = 650; Tjur's $R^2=0.12$		
<sup>a</sup> statistically significant ( $p < 0.05$ ) after Benjamini-Hochberg correction		
<sup>b</sup> statistically significant ( $p < 0.05$ ) before any correction		

Finally, we analyzed the factors associated with higher intervention rating and intervention effectiveness. The logistic regression model of *Liked* as a function of per-participant characteristics, per-half hour nudge probability, per-nudge contextual metrics, and per-intervention metrics was significant ( $\chi^2(28)=92.385$ ,  $p \ll 0.001$ ). The odds ratios and confidence intervals for each predictor can be seen in Table 6. We found that being a woman, higher *Cognitive Reappraisal*, higher *Extraversion*, and higher *Stress Reduction* were associated with a higher likelihood of liking the intervention. On the other hand, being 36-45 years old, higher *Expressive Suppression* and higher *Nudge Probability* were associated with a lower likelihood of liking the intervention.

Figure 2. (a) A kernel density estimate plot of subjective stress ratings shows that the distribution of stress ratings when the system sent intervention nudges is higher than the individual thresholds. The pre-intervention stress ratings are distributed at a higher rating than post-intervention stress ratings and the ratings when the nudges were sent. (b) Average momentary stress ratings at different points in time with 95% confidence intervals.



(a) Distribution of momentary stress ratings



(b) Comparison of momentary stress ratings

When we examine the momentary stress ratings surrounding the system nudges and the intervention use (Figure 2a), we see that the distributions of EMA, pre-intervention, and post-intervention stress ratings sit higher than the individual thresholds used for sending the nudges. The average stress rating measured just before the intervention use ( $\bar{x}=2.14$ ) is higher than the average stress rating used for system nudges ( $\bar{x}=1.73$ ), indicating that the participants were more likely to engage in interventions when their stress ratings were higher than average (Figure 2b).

We then examined the factors associated with stress reduction. The logistic regression model of *Improved* was significant ( $\chi^2(30)=186.66$ ,  $p \ll 0.001$ ), and the resulting odds ratios and confidence intervals per predictor can be found in Table 7. We found that higher *Agreeableness*, higher *Nudge Probability*, higher *Stress Before*, higher *Rating*, and getting prompt-based or video interventions were associated with a higher likelihood of improvement on their stress rating. On the other hand, being a woman, higher *Cognitive Reappraisal*, higher *Neuroticism*, and choosing Calm or Address interventions were associated with a lower likelihood of improvement on their stress rating.

Table 6. Odds ratios (OR), confidence intervals (CI), and p-values for each predictor of *Liked* for all interventions that participants engaged in.

Predictors	Liked	
	OR	CI
(Intercept)	5.95	0.51 – 70.91
Age Reference (18-35)		
Age Group (36-45)	0.51 <sup>b</sup>	0.30 – 0.86
Age Group (46+)	1.01	0.55 – 1.83
Gender (Man)		
Gender (Woman)	2.51 <sup>a</sup>	1.51 – 4.20
Cognitive Reappraisal	1.35 <sup>b</sup>	1.08 – 1.69
Expressive Suppression	0.82 <sup>b</sup>	0.68 – 0.99
Resilience	1.10	0.80 – 1.51
Agreeableness	0.72	0.50 – 1.02
Conscientiousness	0.80	0.58 – 1.09
Extraversion	1.46 <sup>b</sup>	1.10 – 1.95

Neuroticism	0.84	0.64 – 1.11
Openness	0.99	0.75 – 1.30
Engagement Skewness	0.73	0.47 – 1.12
Nudge Probability	0.94 <sup>b</sup>	0.90 – 0.99
Meeting Counts	0.93	0.61 – 1.42
No Meeting Minutes	1.00	0.99 – 1.02
Self Event Counts	0.66	0.40 – 1.10
Email Messages Sent	1.01	0.79 – 1.28
Email Messages Read	1.00	0.96 – 1.04
Adhoc Call Count	0.66	0.24 – 1.69
Number of Attention Signals	1.16	0.84 – 1.59
Trigger Source Reference (System)		
Trigger Source (Rescheduled)	0.78	0.47 – 1.30
Category Reference (Distract)		
Category (Calm)	0.8	0.49 – 1.29
Category (Address)	0.94	0.44 – 2.01
Stress Reduction	2.36 <sup>a</sup>	1.60 – 3.54
Modality Reference (conversation)		
Modality (prompt)	0.52	0.17 – 1.53
Modality (video)	0.75	0.22 – 2.47
Location Reference (at desk)		
Location (Inside)	0.69	0.40 – 1.19
Location (Outside)	2.56	0.79 – 9.27
Observations =521; Tjur's $R^2=0.166$		
<sup>a</sup> statistically significant ( $p < 0.05$ ) after Benjamini-Hochberg correction		
<sup>b</sup> statistically significant ( $p < 0.05$ ) before any correction		

Table 7. Odds ratios (OR), confidence intervals (CI), and p-values for each predictor of *Improved* for all interventions that participants engaged in.

Predictors	Improved	
	OR	CI
(Intercept)	0.00 <sup>a</sup>	0.00 – 0.00
Age Reference (18-35)		
Age Group (36-45)	0.58	0.30 – 1.13
Age Group (46+)	0.98	0.49 – 1.98
Gender Reference (Man)		
Gender (Woman)	0.41 <sup>a</sup>	0.21 – 0.77
Cognitive Reappraisal	0.69 <sup>a</sup>	0.52 – 0.91
Expressive Suppression	1.01	0.79 – 1.28
Resilience	0.84	0.55 – 1.25
Agreeableness	1.73 <sup>b</sup>	1.10 – 2.76
Conscientiousness	1.27	0.86 – 1.89
Extraversion	0.89	0.62 – 1.29
Neuroticism	0.57 <sup>a</sup>	0.39 – 0.81
Openness	1.09	0.78 – 1.55
Engagement Skewness	1.18	0.68 – 2.03
Nudge Probability	1.09 <sup>a</sup>	1.03 – 1.16
Meeting Counts	0.85	0.48 – 1.42
No Meeting Minutes	1.01	0.99 – 1.03
Self Event Counts	1.29	0.69 – 2.36
Email Messages Sent	1.04	0.78 – 1.39

Email Messages Read	0.96	0.90 – 1.01
Adhoc Call Count	0.94	0.28 – 2.63
Number of Attention Signals	1.4	0.95 – 2.05
Trigger Source Reference (System)		
Trigger Source (Rescheduled)	1.08	0.57 – 2.02
Category Reference (Distract)		
Category (Calm)	0.43 <sup>a</sup>	0.23 – 0.78
Category (Address)	0.40 <sup>b</sup>	0.16 – 0.97
Stress Before	5.76 <sup>a</sup>	3.98 – 8.64
Rating	2.47 <sup>a</sup>	1.84 – 3.39
Modality Reference (conversation)		
Modality (prompt)	6.65 <sup>b</sup>	1.53 – 36.45
Modality (video)	5.62 <sup>b</sup>	1.12 – 34.10
Location Reference (at desk)		
Location (Inside)	1.17	0.60 – 2.25
Location (Outside)	0.23	0.03 – 1.10
Observations = 521; Tjur's $R^2$ = 0.338		
<sup>a</sup> statistically significant ( $p < 0.05$ ) after Benjamini-Hochberg correction		
<sup>b</sup> statistically significant ( $p < 0.05$ ) before any correction		

## Discussion

In this paper, we presented a comprehensive and systematic study that simultaneously encompassed trait-based factors (i.e., individual characteristics), state-based factors (i.e., workplace contextual and behavioral signals, momentary stress), and intervention-related factors (i.e., location, function) to identify what drives JIT intervention engagement and efficacy. This study particularly focuses on momentary outcomes to inform the design of *dynamic tailoring*, which is a key component of JITAI systems. In this study, we leveraged surveys, EMA, and passively-sensed data from the deployment of the JIT stress-reduction intervention system to identify significant factors that influence the momentary engagement, the choice of interventions, the engagement given an intervention choice, the user rating of interventions engaged, and the stress reduction from the engagement.

We found that stress ratings immediately after the JIT interventions were significantly lower than those reported in the moments immediately prior to the interventions (Figure 2b). While keeping in mind the study sample, we found that women, those with higher *Neuroticism*, those with higher *Cognitive Reappraisal* skills, and those that chose *Calm* and *Address* interventions were significantly less likely to experience stress reduction, while those with higher *Agreeableness* and those that chose prompt-based or video-based interventions were significantly more likely to experience stress reduction. Surprisingly, contextual signals, such as meeting or email load, did not lead to a significant increase or decrease in stress ratings, which may indicate that trait-based or intervention-related factors matter more for efficacy or that the contextual signals may need finer granularity.

On the other hand, we found that contextual signals such as lower *Meeting Counts* and lower *Engagement Skewness* were associated with a higher likelihood of engagement, indicating that state-based contextual factors such as being in a meeting or time of the day may matter more for engagement than efficacy. In addition, a JIT intervention that was explicitly rescheduled to a later time was more likely to be engaged. This implies that intervention engagement can be improved by giving some level of control to the users over complete automation.

With respect to the choice of interventions and liking the interventions, factors that significantly

influenced the choice were primarily trait-based. Higher *Openness* was associated with a higher likelihood of choosing *Distract* but a lower likelihood of choosing *Calm*, while higher *Cognitive Reappraisal* was associated with a higher likelihood of choosing *Calm* but lower likelihood of choosing *Address*. Higher *Resilience* was associated with a higher likelihood of choosing *Address* as well as the subsequent engagement after the choice. Interventions were liked more by women, those with higher *Cognitive Reappraisal*, and those with higher *Extraversion*, and when the interventions led to stress reduction. One state-based factor that stood out is the *Number of Attention Signals* which negatively impacted the likelihood of choosing *Distract* interventions. Higher *Number of Attention Signals* indicates higher desk-bound work activity (e.g., keyboard, mouse) where getting the mind off work (i.e., how *Distract* interventions were communicated) during intense work activities may not be feasible or desired.

In summarizing our findings, we first categorize these factors into (1) non-modifiable individual factors, (2) modifiable individual factors, (3) contextual factors, and (4) content factors.

The distinction between modifiable and non-modifiable individual factors is important for intervention design. Once modifiable factors are identified, strategies can be deployed to directly influence those factors if those strategies can lead to a greater impact on the engagement or the efficacy of the interventions. Non-modifiable factors are also important to determine which subset of populations can benefit from additional targeted support through organizational or policy-level changes [72]. In our analysis, non-modifiable individual factors include gender and age, and modifiable individual factors, although debatable [73], include emotion regulation skills and personality traits. Our findings revealed that these individual factors not only influence study-long engagement [42] but also influence momentary engagement and can be useful for tailoring JITAI systems. Although our findings for individual factors corroborate with prior studies, it is important to highlight that our study evaluates instance-level engagement metrics through play-by-play analysis of app usage behaviors rather than study-long engagement metrics. Contextual factors such as workload (e.g., meeting counts, email counts) and availability (e.g., presence or activity at the computer) are helpful in the JITAI system's decision-making process for when to interrupt the user. Content factors include intervention-related information such as the amount of effort required, the modality of intervention delivery, and the ideal location for intervention engagement, and these factors can inform JITAI systems in determining which intervention to present to the user given the understanding of the current context.

In this section, we summarize and discuss the findings organized by these four categories of factors with recommendations for design and future research.

## Non-modifiable individual factors

In our analysis, we found several significant effects of gender and age on our outcome measures. Participants who self-reported as being a woman had over twice the likelihood of liking interventions than being a man, despite having less likelihood of improving from engaging in interventions. Being a woman also had 38% less likelihood of choosing *Distract* interventions than being a man, suggesting that there may be an unobserved motivational factor. For example, although not statistically significant, women reported higher momentary stress on average compared to men (2.15 for women vs 1.84 for men;  $t(20.756) = -1.545$ ,  $p = 0.137$ ) which may contribute to choosing more *Calm* or *Address* interventions to reduce their stress. Although the general findings from DMHI studies that women are more likely to engage in digital interventions than men are corroborated by our analysis [42], the fact that women improve less despite engaging more, liking interventions more, and choosing more *Address* (i.e., high reward) interventions is a concern for systems design that only take engagement metrics into account.



Participants that are 46+ years old were 47% more likely to engage in interventions while being 36-45 years old increased the likelihood of choosing *Distract* interventions by 65% and decreased the likelihood of liking interventions by 49% compared to those that are 18-35 years old. Although the effects of different age groups on engagement have mixed results across prior studies, the higher engagement rate for 46+-year-old participants in the study could be explained by a higher rate of interest in digital interventions for older populations [74]. Participants in the 36-45-year-old age group choosing *Distract* interventions and not liking them highlight an opportunity for finding different types of interventions that they might enjoy.

*Recommendation #1: Monitor the usage by age and gender to identify subgroups that may need additional support.* Tailoring JITAI systems to individuals has been suggested to improve engagement and efficacy of interventions in prior research [36]. Our findings suggest that certain age or gender groups may benefit more than others from our system. The lack of engagement or the lack of improvement despite engagement for some groups highlights opportunities for targeting research and design efforts to further understand unobserved barriers to engagement and effective responses. Although it is generally recommended that intervention content be tailored based on individuals, we also recommend that the efficacy and user rating of interventions be closely monitored to ensure that certain groups enjoy the same level of benefits as other groups. In addition, new intervention content could be co-designed with groups that may not be benefiting as much and added to the system on a regular basis to equalize outcomes across subgroups.

## Modifiable individual factors

Our findings corroborate prior research that personality trait is a strong factor in the engagement of DMHIs [42]. Prior study has found that openness to experience is associated with better adherence and lower odds of attrition [75]. In our analysis, we did not find a significant effect on engagement, but we found that *Openness* significantly increases the likelihood of choosing *Distract* but decreases the likelihood of choosing *Calm* interventions. One possible explanation may be the variety in the intervention content, which people with high openness would prefer [76]. *Distract* interventions tend to offer more variety in content with videos of travel destinations and humor as well as opportunities to explore social connections, whereas *Calm* interventions were mostly introspective activities such as breathing or focused observations.

Although the effects were only moderately significant ( $p < 0.061$ ), we found that an increase by one point in *Neuroticism* and *Agreeableness* scales increases the likelihood of choosing *Address* interventions, which were designed to help users directly address and problem-solve stress-inducing components of their lives. For participants with high scores in *Neuroticism* and *Agreeableness* scales, it is possible that interventions that help them directly address their stress were more appealing than others that were designed to distract from stress and refocus on the present. Prior study has also found neuroticism and agreeableness to associate with a stronger interest in the use of stress management apps [77]. We also found that one point increase in *Agreeableness* scale was associated with an increased likelihood of improvement by 73%, whereas the same point increase in *Neuroticism* scale was associated with a lower likelihood of improvement by 43%, despite both having higher tendency to choose *Address* interventions. Since agreeableness is known to be positively associated with the therapeutic alliance in mental health treatments [78], it is possible that the prosocial and cooperative nature of those with high agreeableness [79] allowed them to fully engage in the *Address* interventions that were more action-oriented, leading to a greater improvement. On the other hand, neuroticism has been known to negatively correlate with adherence to mental health recommendations [80] and to a wide variety of mental health treatment outcomes [78].



Prior work has associated extraversion with lower interest in using web-based mental health services over face- to-face interactions with a provider [74]. Even though our interventions were delivered solely through technology, we found that one point increase in *Extraversion* scale increases the likelihood of liking the intervention by 46%. We found no significant association between the intervention content (category, modality, or location) and liking them to explain this behavior. Future research should explore how extraversion facilitates momentary engagement in interventions.

*Recommendation #2: Incorporate personality traits and coping styles in system adaptation.* Corroborating with prior study-long findings that associate personalities with DMHI engagement, our study found that personality traits also influence momentary engagement and efficacy of interventions. We recommend that JITAI systems carefully monitor potentially unhelpful usage behaviors by incorporating personality traits in the system adaptation algorithm because they may impact the choice of interventions that may lead to negative downstream effects on outcomes. For example, it may be beneficial to offer a variety of more effortful interventions for people who report higher scores in openness. For people who report higher scores in neuroticism, the system could suggest less effortful interventions.

Across the board, emotion regulation styles had significant effects. One point increase in the *Cognitive Reappraisal* scale was associated with a 14% increase in the likelihood of engaging, a 27% increase in the likelihood of choosing *Calm* interventions, a 35% increase in the likelihood of liking the intervention, a 33% decrease in the likelihood of choosing *Address* interventions, and a 31% decrease in the likelihood of improvement after engagement. It is important to note the differences in the range of point scales. An increase of 14% for a 7-point scale is equivalent to an increase of 20% for a 5-point scale. On the other hand, one point increase in the *Expressive Suppression* scale was associated with an 18% decrease in the likelihood of liking the intervention. We also found that one point increase in the *Resilience* scale increases the likelihood of choosing *Address* interventions by 42% and increases the likelihood of subsequently engaging in interventions by 63%.

Prior research has explored the role of emotion regulation in stress coping. For example, emotion regulation skills help assess the stressful situation and determine the appropriate emotional response [81] or act as buffers against the negative effects of stress [82]. Emotion regulation has also been theorized as a moderator for increased resilience after encountering a stressful situation [83]. Although our analysis cannot claim the causal direction between coping skills and engagement, our findings suggest that emotion regulation and resilience may play a role in, not only the stress-coping process but in choosing different interventions or deciding to engage in an intervention. While most research has argued for increasing coping skills as an outcome measure or a treatment target [84], our findings suggest promising new research directions in understanding how coping skills could impact our decisions to engage in therapeutic interventions in the moment.

*Recommendation #3: Adopt a dynamic perspective on personality and coping skills.* The role of emotion regulation strategies in altering our decisions and choices in various contexts has been empirically studied in highly controlled laboratory settings [85]. Prior research has also studied personality traits [73, 86] and coping skills [84] as “states” with intrapersonal variations and targeted their modifications through therapeutic strategies (e.g., CBT). Although our study has assumed personality traits, emotion regulation skills, and resilience to be stable over the course of the study, it remains to be seen how shifting the perspectives of these characteristics to be more dynamic would inform the JITAI systems. Because of the potential mediating role of coping skills in perceived stress [87] and the role of perceived stress in outcome improvement (i.e., one point increase in stress rating before the intervention leads to being over 5 times more likely to improve in our findings), coping

skills should be measured periodically and incorporated into the decision-making process of JITAI systems. Therefore, we recommend further research to explore how taking a dynamic approach to personality and coping skills would inform the design of JITAI systems.

## Contextual factors

We found that the more the participant tends to engage at the end of the day, the more likely they would engage in interventions and that the system-initiated interventions that were rescheduled to a later time increased the likelihood of engagement. These findings may suggest a tendency to defer interventions to later in the day. Prior work that applied the Self Determination Theory [88] to JITAI systems suggested that perceived competence and self-regulation abilities may deplete throughout the day, potentially leading to unhealthy choices (e.g., unhealthy food, alcohol) towards the end of the day [34]. This has a serious consequence for those that tend to procrastinate or postpone healthy behaviors (e.g., exercise, stress intervention) towards the end of the day. On the other hand, allowing people to defer an intervention to a specific time may increase self-efficacy and perception of control, which might lead to an increased chance of behavior change [89]. In fact, Howe et al. reported that many participants liked the ability to perform the interventions when they wanted. We also found that the increase in Nudge Probability decreases the likelihood of liking the intervention but increases the likelihood of improvement, revealing that a JIT intervention might be “a bitter pill to swallow” but a useful pill, nonetheless. However, a relentless reminder could lead to distraction [90] and eventual system abandonment [44].

*Recommendation #4: Balance individual preferences with intervention efficacy.* Our findings show encouragement that intelligent timing based on contextual information could improve the engagement and effectiveness of interventions, but perhaps at the cost of a lowered sense of user agency and control, negative perceptions towards the interventions, or leading to unhealthy choices towards the end of the day. It also highlights that relying solely on what people like may not be helpful for stress reduction and that there needs to be an additional investigation into how momentary factors could influence user ratings. Therefore, future JITAI systems should carefully balance user ratings (e.g., likes) with intervention efficacy (e.g., stress reduction, engagement) and help users discover what works best for them. Considering that higher user ratings may not always reflect the effectiveness of the interventions, such systems should aim to simultaneously explore the rating and the improvement in determining the timing of interventions.

In evaluating the work context, as expected, we found that a nudge sent at a time when the user is less likely to be in a meeting but more active at the desk improves engagement. Contrary to our hypothesis, we found no significant associations with *No Meeting Minutes* or *Self Event Counts*. It is possible that there is high variability in the level of focus and attention needed during times carved out for self. For example, our data sources cannot discern if the times carved out for self were work-related (e.g., focus time for reading, writing) or nonwork-related (e.g., running errands, child pick up, exercise). The current study's data sources cannot achieve automatic detection of activities beyond basic work activities, such as meetings, emails, chats, calls, or computer activities.

*Recommendation #5: Assess the value of accurate activity detection on engagement.* Although tailoring to the activity context is the defining promise of JITAI systems, automatically detecting the activities performed within a time window is not an easy task. In addition, how much tailoring and the level of system intelligence are really necessary to maximize engagement and outcomes is unknown, especially given the cost and risks associated with the invasion of privacy in passive sensing. Further research is necessary to understand the cost and benefit of accurate activity detection and intelligent timing on the engagement and effectiveness of JITAI systems.

## Content factors

In our analysis, intervention categories, modalities, and locations showed pronounced effects on engagement and improvement, suggesting the importance of the intervention content in the design of JITAI systems. We found that having an intervention that could be done at the desk more than doubled the likelihood of engagement compared to an intervention that could be done indoors, but not at the desk. It is possible that leaving the desk at the moment of the nudge was not appropriate given the situation or there was an unobserved motivational barrier. In these scenarios, suggesting a different activity, rescheduling the activity or waiting until the next appropriate transition time might have been beneficial. To improve engagement, intervention designers could provide additional desk-based stress reduction techniques to minimize the burden of leaving the desk.

On the other hand, we found that having a prompt- or video-based intervention increased the likelihood of engagement by over 3-fold and improvement by over 5-fold, compared to a conversation-based intervention. Although prompt- and video-based interventions were typically less effortful than conversation-based interventions that require many turn-taking interactions with the bot, conversation-based interventions were designed to address the sources of the stress with the hope of creating a longer-lasting impact. It is possible that the conversations were not usable for participants to fully engage in the content. This finding suggests that quick, effortless interventions could be useful at the moment, but complex, turn-taking interventions need more thoughtful redesign.

*Recommendation #6: Provide a diverse set of interventions.* Our findings revealed that the effects of intervention content types were significant and large. To investigate the impact of choices on improving outcomes and user ratings, future digital intervention systems that provide a catalog of interventions would benefit from characterizing each intervention on multiple dimensions, such as the level of effort, location, modality, etc. Such characterization would help understand the interplay among personality traits, contextual cues, and intervention types that users are likely to choose and benefit from. Therefore, we recommend that JITAI systems provide a variety of interventions to fine-tune its recommendations based on contexts but also to identify interventions that may need redesign.

## Limitations

Our system design and analysis setup have several limitations. The retrospective analysis of the inferred stress scores indicates that there is room for improvement. Our stress scores were computed from a generalized algorithm, but recent studies have shown that stress can be idiosyncratic and, therefore, needs to be modeled at an individual level [18]. Our user engagement flow cannot differentiate the dismissal of the nudge due to bad timing or low stress or both. Our analysis setup does not allow for determining the causal relationship between the individual, contextual, and content factors with engagement, stress reduction, and intervention ratings. Even though the deployment study collected long-term stress measures via the DASS-21 scale, our analysis was limited to momentary stress ratings because micro-interventions are more appropriate for proximal outcomes over distal outcomes [31]. Micro-randomized trials (MRTs) are a promising research direction for JITAI systems to quantify the impact of tailored interventions on both short- and long-term outcomes [91] with careful considerations for the appropriate sample size [92]. Our data was also limited by a small sample population that exhibited low-stress levels, focusing on US-based information workers from a large technology organization, skewed towards engineers and those that identified as male. Prior work has suggested that, when evaluating engagement (or attrition), the severity of symptoms should also be considered [41]. Therefore, future research should evaluate the

system through MRTs with a sample population exhibiting high severity of stress symptoms, while also expanding into other information worker roles from different sectors.

The deployment of the system used in the study is limited by the system requirements. Our system assumes that a single organization may offer a standardized set of software tools to workers for several reasons (e.g., security, privacy, productivity, compliance, scalability) and does not support interoperability across different software tools used by workers within an organization or across organizations. Beyond this challenge, it is important to note that the data required to describe an individual worker's stress necessarily crosses work-nonwork boundaries, and such boundaries are increasingly blurred by technologies [93] and remote work [94]. Our contextual data captured at work were noisy and were aggregated without peeking into the content (e.g., of email or documents) to preserve privacy but also because it is challenging to separate blurred contexts. Such blurred boundaries raise concerns surrounding boundary preferences, data ownership, values and incentives, wellbeing definitions, and power dynamics that may undermine the successful deployment of personal stress support systems [95]. Perhaps more important than the interoperability of a system or clean data is a human-centered implementation process that allows for flexibility and adaptation of the system based on multi-stakeholder perspectives and usage over time. This paper piloted one aspect of such adaptation. However, much work is needed in realizing the full potential of JITAI systems that take sociotechnical considerations into account and where users can participate in drawing the boundary and providing the right level and fidelity of data that meets their needs.

## Conclusion

JITAI systems have the potential to integrate timely support into the workplace. To identify key factors that should be incorporated into the system's adaptable decision-making process, we analyzed data from a four-week deployment of a JIT workplace stress reduction micro-intervention system. Based on our findings, we recommend that individual, contextual, and content-based factors be incorporated into the system for tailoring as well as for monitoring unhelpful usage behaviors across subgroups and contexts. Future work should explore careful balancing of individual preferences, intervention efficacy, and system accuracy to help users discover what works best for them and to continuously improve system recommendations.

## Acknowledgements

This research was funded by Microsoft Research. T.A was supported in part by the NSF grant CAREER IIS-2142794, NSF grant IIS-1901386, and an NIH grant R01MH125179. We are deeply grateful to our participants for their time and willingness to install our system, engage in our interventions, and provide valuable feedback and data. This study would not have been possible without the guidance and support of Dr Patricia Areán, Dr James Fogarty, Dr Eric Horvitz, and Microsoft Research Human Understanding and Empathy (HUE) group. No generative AI tools were used in preparation of this manuscript.

## Authors' Contributions

JS and EH conceptualized and conducted the study and collected the data. All authors contributed to the development of the methodology. JS conducted the analysis and generated tables and visualizations. JS drafted the manuscript. TA and MC provided supervision of the study. All others reviewed and edited the final manuscript.

## Data Availability

The data sets generated during and/or analyzed during this study are not publicly available due to privacy and legal restrictions but are available from the corresponding author on reasonable request

with a clear justification and a license agreement. The request will be reviewed and approved case by case by Microsoft Research Release and Compliance team, at which point a license agreement will be drafted and shared.

## Conflicts of Interest

The research was conducted when JS, EH, RL, JH, KS, and MC were employed at Microsoft Research. JS, EH, JH, and MC were involved in the design, development, and deployment of the innovations and software evaluated in this study.

## Abbreviations

JITAI: Just-in-time adaptive intervention

EBPI: Evidence-based psychosocial Interventions

EMA: Ecological momentary assessment

RQ: Research question

OR: Odds ratio

CI: Confidence interval

MRT: Micro-randomized trial

## References

1. Julie Ray. Americans' Stress, Worry and Anger Intensified in 2018. <https://news.gallup.com/poll/249098/americans-stress-worry-anger-intensified2018.aspx> (Accessed 2021-08-20), 2019.
2. Edwin Boudreaux, Cris Mandry, and Phillip J Brantley. Stress, job satisfaction, coping, and psychological distress among emergency medical technicians. *Prehospital and Disaster Medicine*, 12(4):9–16, 1997.
3. Elizabeth Brondolo, Kahaema Byer, Peter J. Gianaros, Cindy Liu, Aric A. Prather, Kamala Thomas, and Cheryl L. Woods-Giscombé. Apa working group on stress and health disparities.(2017). *Stress and Health Disparities: Contexts, Mechanisms, and Interventions among Racial/Ethnic Minority and Low-Socioeconomic Status Populations*. Retrieved from <http://www.apa.org/pi/health-disparities/resources/stress-report.aspx>, 2017.
4. Thomas W Colligan and Eileen M Higgins. Workplace stress: Etiology and consequences. *Journal of workplace behavioral health*, 21(2):89–97, 2006.
5. Mina Westman and Dalia Etzion. The impact of vacation and job stress on burnout and absenteeism. *Psychology & Health*, 16(5):595–606, 2001.
6. Cary L Cooper and Sue Cartwright. An intervention strategy for workplace stress. *Journal of psychosomatic research*, 43(1):7–16, 1997.
7. Katherine M Richardson and Hannah R Rothstein. Effects of occupational stress management intervention programs: a meta-analysis. *Journal of occupational health psychology*, 13(1):69, 2008.
8. Gail Kinman and Fiona Jones. Lay representations of workplace stress: What do people really mean when they say they are stressed? *Work & stress*, 19(2):101–120, 2005.
9. Maureen F Dollard. *Psychosocial safety climate: A lead indicator of workplace psychological health and engagement and a precursor to intervention success*. Routledge, 2012.
10. Susana García Herrero, Miguel Ángel Mariscal Saldaña, Javier García Rodríguez, and Dale O Ritzel. Influence of task demands on occupational stress: Gender differences. *Journal of Safety Research*, 43(5-6):365–374, 2012.
11. Sylvia Helbig and Lydia Fehm. Problems with homework in cbt: Rare exception or rather

- frequent? *Behavioural and cognitive psychotherapy*, 32(3):291–301, 2004.
12. Eunmi Lee and Yujeong Kim. Effect of university students' sedentary behavior on stress, anxiety, and depression. *Perspectives in psychiatric care*, 55(2):164, 2019.
  13. Yuhan Luo, Bongshin Lee, Donghee Yvette Wohn, Amanda L Rebar, David E Conroy, and Eun Kyoung Choe. Time for break: Understanding information workers' sedentary behavior through a break prompting system. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.
  14. Sara Thomée, Annika Härenstam, and Mats Hagberg. Computer use and stress, sleep disturbances, and symptoms of depression among young adults—a prospective cohort study. *BMC psychiatry*, 12(1):1–14, 2012.
  15. Gloria Mark, Yiran Wang, and Melissa Niiya. Stress and multitasking in everyday college life: an empirical study of online activity. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 41–50, 2014.
  16. Arla Day, Stephanie Paquet, Natasha Scott, and Laura Hambley. Perceived information and communication technology (ict) demands on employee outcomes: the moderating effect of organizational ict support. *Journal of occupational health psychology*, 17(4):473, 2012.
  17. William C Murray and Adam Rostis. Who's running the machine? a theoretical exploration of work stress and burnout of technologically tethered workers. *Journal of individual employment rights*, 12(3):249–263, 2007.
  18. Mehrab Bin Morshed, Javier Hernandez, Daniel McDuff, Jina Suh, Esther Howe, Kael Rowan, Marah Abdin, Gonzalo Ramos, Tracy Tran, and Mary Czerwinski. Advancing the understanding and measurement of workplace stress in remote information workers from passive sensors and behavioral data. In *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. IEEE, 2022.
  19. Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2022.
  20. Victor P Cornet and Richard J Holden. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of biomedical informatics*, 77:120–132, 2018.
  21. Gonzalo J Martinez, Stephen M Mattingly, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, Jessica Young, Nitesh Chawla, Munmun De Choudhury, Sidney D'Mello, Gloria Mark, et al. Predicting participant compliance with fitness tracker wearing and ecological momentary assessment protocols in information workers: Observational study. *JMIR mHealth and uHealth*, 9(11):e22218, 2021.
  22. Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Subigya K Nepal, Andrew T Campbell, Anind K Dey, and Aaron D Striegel. On the quality of real-world wearable data in a longitudinal study of information workers. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 1–6. IEEE, 2020.
  23. Daniel McDuff, Kael Rowan, Piali Choudhury, Jessica Wolk, ThuVan Pham, and Mary Czerwinski. A multimodal emotion sensing platform for building emotion-aware applications. *arXiv preprint arXiv:1903.12133*, 2019.
  24. Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010.
  25. Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. Bored Mondays and focused afternoons: the rhythm of attention and online activity in the workplace. In *Proceedings of*



- the SIGCHI Conference on Human Factors in Computing Systems*, pages 3025–3034, 2014.
26. Javier Hernandez, Pablo Paredes, Asta Roseway, and Mary Czerwinski. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–60. ACM, 2014.
  27. N Bidargaddi, G Schrader, P Klasnja, J Licinio, and S Murphy. Designing m-health interventions for precision mental health support. *Translational psychiatry*, 10(1):1–8, 2020.
  28. Anna Huguet, Sanjay Rao, Patrick J Mcgrath, Lori Wozney, Mike Wheaton, Jill Conrod, and Sharlene Rozario. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. *PLOS One*, 11(5):E0154248, 2016.
  29. Martha Neary and Stephen M Schueller. State of the field of mental health apps. *Cognitive and Behavioral Practice*, 25(4):531–537, 2018.
  30. Elsbeth Marieke De Korte, Noortje Wiezer, Joris H. Janssen, Peter Vink, and Wessel Kraaij. Evaluating an mHealth app for health and well-being at work: Mixed-method qualitative study. *JMIR mHealth and uHealth*, 6(3):1–17, 2018.
  31. Amit Baumel, Theresa Fleming, and Stephen M Schueller. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. *Journal of medical Internet research*, 22(10):e20631, 2020.
  32. Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462, 2018.
  33. Akane Sano, Paul Johns, and Mary Czerwinski. Designing opportune stress intervention delivery timing using multi-modal data. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 346–353. IEEE, 2017.
  34. Donna Spruijt-Metz and Wendy Nilsen. Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Computing*, 13(3):13–17, 2014.
  35. Adrienne S Juarascio, Megan N Parker, Madeline A Lagacey, and Kathryn M Godfrey. Just-in-time adaptive interventions: A novel approach for enhancing skill utilization and acquisition in cognitive behavioral therapy for eating disorders. *International Journal of Eating Disorders*, 51(8):826–830, 2018.
  36. Liyuan Wang and Lynn Carol Miller. Just-in-the-moment adaptive interventions (jitai): A meta-analytical review. *Health Communication*, 35(12):1531–1544, 2020.
  37. Peng Liao, Walter Dempsey, Hillol Sarker, Syed Monowar Hossain, Mustafa al’Absi, Predrag Klasnja, and Susan Murphy. Just-in-time but not too much: Determining treatment timing in mobile health. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), dec 2018.
  38. Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(1), mar 2020.
  39. Varun Mishra, Florian Künzler, Jan-Niklas Kramer, Elgar Fleisch, Tobias Kowatsch, and David Kotz. Detecting receptivity for mhealth interventions in the natural environment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(2), jun 2021.
  40. Florian Künzler, Varun Mishra, Jan-Niklas Kramer, David Kotz, Elgar Fleisch, and Tobias Kowatsch. Exploring the state-of-receptivity for mhealth interventions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4), sep 2020.
  41. Gavin Doherty, David Coyle, and John Sharry. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1421–1430, 2012.

42. Judith Borghouts, Elizabeth Eikey, Gloria Mark, Cinthia De Leon, Stephen M Schueller, Margaret Schneider, Nicole Stadnick, Kai Zheng, Dana Mukamel, Dara H Sorkin, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. *Journal of Medical Internet Research*, 23(3):e24387, 2021.
43. Wendy Hardeman, Julie Houghton, Kathleen Lane, Andy Jones, and Felix Naughton. A systematic review of just-in-time adaptive interventions (jitais) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):1–21, 2019.
44. Niranjana Bidargaddi, Daniel Almirall, Susan Murphy, Inbal Nahum-Shani, Michael Kovalcik, Timothy Pituch, Haitham Maaieh, Victor Strecher, et al. To prompt or not to prompt? a microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. *JMIR mHealth and uHealth*, 6(11):e10123, 2018.
45. Lucy Yardley, Bonnie J Spring, Heleen Riper, Leanne G Morrison, David H Crane, Kristina Curtis, Gina C Merchant, Felix Naughton, and Ann Blandford. Understanding and promoting effective engagement with digital behavior change interventions. *American journal of preventive medicine*, 51(5):833–842, 2016.
46. Caroline Perrin Franck, Awa Babington-Ashaye, Damien Dietrich, Georges Bediang, Philippe Veltsos, Pramendra Prasad Gupta, Claudia Juech, Rigveda Kadam, Maxime Collin, Lucy Setian, Jordi Serrano Pons, S Yunkap Kwankam, Béatrice Garrette, Solenne Barbe, Cheick Oumar Bagayoko, Garrett Mehl, Christian Lovis, and Antoine Geissbuhler. icheck-dh: Guidelines and checklist for the reporting on digital health implementations. *J Med Internet Res*, 25:e46694, May 2023.
47. Carol Collier Kuhlthau. The role of experience in the information search process of an early career information worker: Perceptions of uncertainty, complexity, construction, and sources. *Journal of the American Society for information Science*, 50(5):399–412, 1999.
48. Susanne Hupfer, Sayantani Mazumder, Ariane Bucaille, and Gillian Crossan. Women in the tech industry: Gaining ground, but facing new headwinds, 2024.
49. Peter F Lovibond and Sydney H Lovibond. The structure of negative emotional states: Comparison of the depression anxiety stress scales (dass) with the beck depression and anxiety inventories. *Behaviour research and therapy*, 33(3):335–343, 1995.
50. Beatrice Rammstedt and Oliver P John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212, 2007.
51. James J Gross and Oliver P John. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348, 2003.
52. Bruce W Smith, Jeanne Dalen, Kathryn Wiggins, Erin Tooley, Paulette Christopher, and Jennifer Bernard. The brief resilience scale: assessing the ability to bounce back. *International journal of behavioral medicine*, 15(3):194–200, 2008.
53. Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1717–1728. ACM, 2016.
54. Gloria Mark, Shamsi Iqbal, and Mary Czerwinski. How blocking distractions affects workplace focus and productivity. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 928–934, 2017.
55. Daniel McDuff, Eunice Jun, Kael Rowan, and Mary Czerwinski. Longitudinal observational



- evidence of the impact of emotion regulation strategies on affective expression. *IEEE Transactions on Affective Computing*, 2019.
56. Rosenberg Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.
  57. Leah M Mayo and Markus Heilig. In the face of stress: Interpreting individual differences in stress-induced facial expressions. *Neurobiology of stress*, 10:100166, 2019.
  58. Karim Sadik Kassam. *Assessment of emotional experience through facial expression*. Harvard University, 2010.
  59. Tanja GM Vrijkotte, Lorenz JP Van Doornen, and Eco JC De Geus. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, 35(4):880–886, 2000.
  60. Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Sogaard. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, 92(1):84–89, 2004.
  61. Andrew C Butler, Jason E Chapman, Evan M Forman, and Aaron T Beck. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical psychology review*, 26(1):17–31, 2006.
  62. Anita Lungu and Marsha M Linehan. Dialectical behavior therapy: A comprehensive multi-and transdiagnostic intervention. *The Oxford handbook of cognitive and behavioral therapies*, pages 200–214, 2016.
  63. Marsha Linehan. *DBT Skills training manual*. Guilford Publications, 2014.
  64. Jessica Schroeder, Jina Suh, Chelsey Wilks, Mary Czerwinski, Sean A Munson, James Fogarty, and Tim Althoff. Data-driven implications for translating evidence-based psychotherapies into technology-delivered interventions. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 274–287, 2020.
  65. Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. Multi-stage receptivity model for mobile just-in-time health intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(2), jun 2019.
  66. Paul T Costa and Robert R McCrae. *Revised NEO Personality Inventory (NEO-PI-R) and Neo Five-Factor inventory (NEO-FFI)*. Psychological Assessment Resources, 1992.
  67. Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.
  68. Oliver P. John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. 1999.
  69. Jennifer K Connor-Smith and Celeste Flachsbart. Relations between personality and coping: a meta-analysis. *Journal of personality and social psychology*, 93(6):1080, 2007.
  70. Niall Bolger and Adam Zuckerman. A framework for studying personality in the stress process. *Journal of personality and social psychology*, 69(5):890, 1995.
  71. Øyvind Langsrud. Anova for unbalanced data: Use type ii instead of type iii sums of squares. *Statistics and Computing*, 13(2):163–167, 2003.
  72. Frederick K Ho, Carlos A Celis-Morales, Stuart R Gray, S Vittal Katikireddi, Claire L Niedzwiedz, Claire Hastie, Lyn D Ferguson, Colin Berry, Daniel F Mackay, Jason MR Gill, et al. Modifiable and non-modifiable risk factors for covid-19, and comparison to risk factors for influenza and pneumonia: results from a uk biobank prospective cohort study. *BMJ open*, 10(11):e040402, 2020.
  73. Brent W Roberts, Jing Luo, Daniel A Briley, Philip I Chow, Rong Su, and Patrick L Hill. A

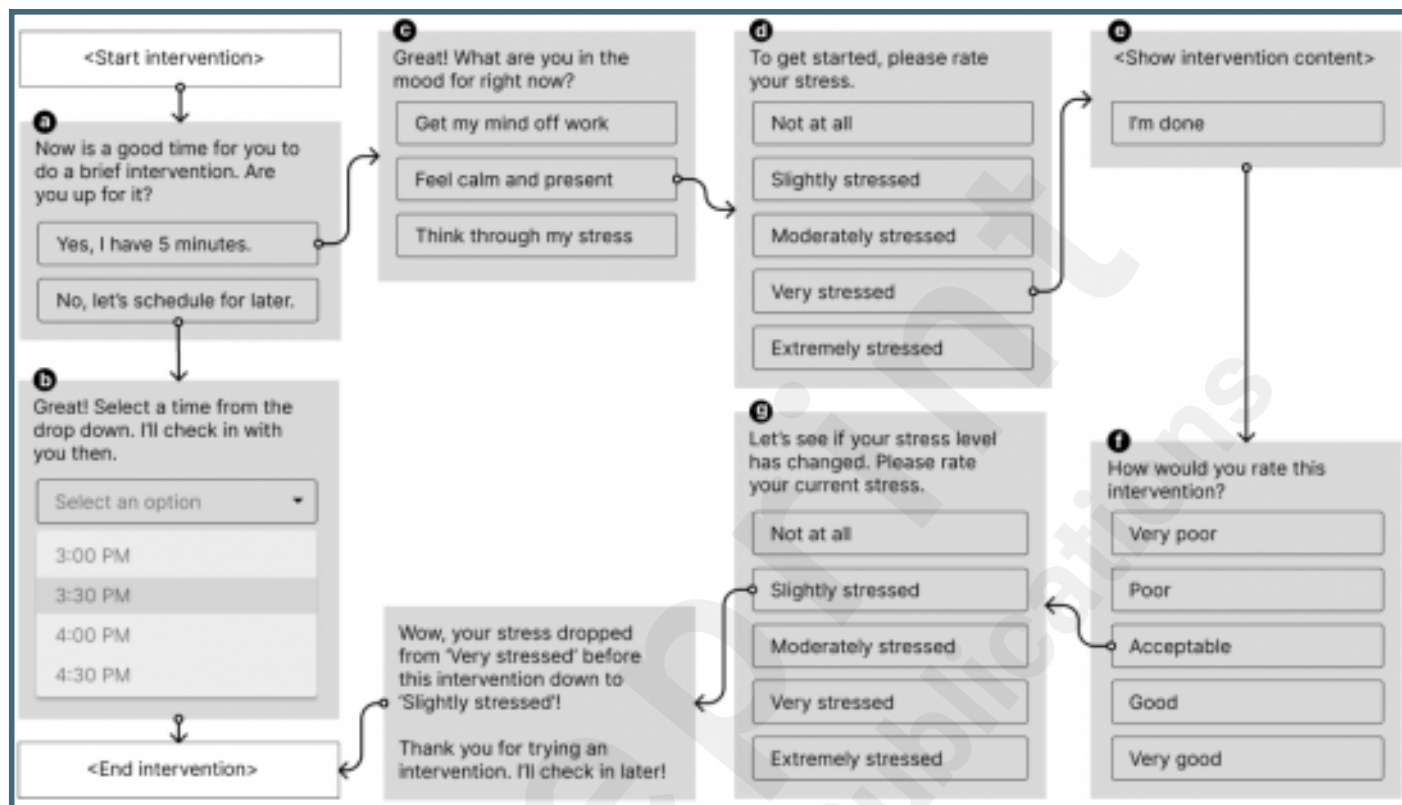
- systematic review of personality trait change through intervention. *Psychological Bulletin*, 143(2):117, 2017.
74. Sonja March, Jamin Day, Gabrielle Ritchie, Arlen Rowe, Jeffrey Gough, Tanya Hall, Chin Yan Jackie Yuen, Caroline Leanne Donovan, Michael Ireland, et al. Attitudes toward e-mental health services in a community sample of adults: online survey. *Journal of medical Internet research*, 20(2):e9109, 2018.
75. Michael Mikolasek, Claudia M Witt, Jürgen Barth, et al. Adherence to a mindfulness and relaxation self-care app for cancer patients: mixed-methods feasibility study. *JMIR mHealth and uHealth*, 6(12):e11271, 2018.
76. Robert R McCrae and Paul T Costa Jr. Conceptions and correlates of openness to experience. In *Handbook of personality psychology*, pages 825–847. Elsevier, 1997.
77. Mari Ervasti, Johanna Kallio, Ilmari Määttänen, Jani Mäntyjärvi, Markus Jokela, et al. Influence of personality and differences in stress processing among finnish students on interest to use a mobile stress management app: survey study. *JMIR mental health*, 6(5):e10039, 2019.
78. Meredith A Bucher, Takakuni Suzuki, and Douglas B Samuel. A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Psychology Review*, 70:51–63, 2019.
79. William G Graziano and Nancy Eisenberg. Agreeableness: A dimension of personality. In *Handbook of personality psychology*, pages 795–824. Elsevier, 1997.
80. AT Pereira, C Cabacos, S Soares, A Araújo, A Manão, AP Amaral, R Sousa, and A Macedo. Relationship between personality, adherence to (mental) health behaviours and psychological distress during the covid-19 pandemic. *European Psychiatry*, 65(S1):S499–S500, 2022.
81. Sarah Myruski, Samantha Denefrio, and Tracy A Dennis-Tiway. Stress and emotion regulation. *The oxford handbook of stress and mental Health*, page 415, 2019.
82. Manjie Wang and Kimberly J Saudino. Emotion regulation and stress. *Journal of Adult Development*, 18(2):95–103, 2011.
83. Allison S Troy and Iris B Mauss. Resilience in the face of stress: Emotion regulation as a protective factor. *Resilience and mental health: Challenges across the lifespan*, 1(2):30–44, 2011.
84. Matthias Berking, Peggilee Wupperman, Alexander Reichardt, Tanja Pejic, Alexandra Dippel, and Hansjörg Znoj. Emotion-regulation skills as a treatment target in psychotherapy. *Behaviour research and therapy*, 46(11):1230–1237, 2008.
85. Alessandro Grecucci and Alan G Sanfey. Emotion regulation and decision making. *Handbook of emotion regulation*, 2, 2014.
86. Krystyna Glinski and Andrew C Page. Modifiability of neuroticism, extraversion, and agreeableness by group cognitive behaviour therapy for social anxiety disorder. *Behaviour Change*, 27(1):42–52, 2010.
87. Chloe C Boyle, Annette L Stanton, Patricia A Ganz, Catherine M Crespi, and Julianne E Bower. Improvements in emotion regulation following mindfulness meditation: Effects on depressive symptoms and perceived stress in younger breast cancer survivors. *Journal of consulting and clinical psychology*, 85(4):397, 2017.
88. Edward L Deci and Richard M Ryan. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 2013.
89. Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision*

- processes*, 50(2):179–211, 1991.
90. Kostadin Kushlev, Jason Proulx, and Elizabeth W Dunn. "silence your phones" smartphone notifications increase inattention and hyperactivity symptoms. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1011–1020, 2016.
  91. Predrag Klasnja, Shawna Smith, Nicholas J Seewald, Andy Lee, Kelly Hall, Brook Luers, Eric B Hekler, and Susan A Murphy. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartsteps. *Annals of Behavioral Medicine*, 53(6):573–582, 2019.
  92. Peng Liao, Predrag Klasnja, Ambuj Tewari, and Susan A Murphy. Sample size calculations for micro-randomized trials in mhealth. *Statistics in medicine*, 35(12):1944–1971, 2016.
  93. Julie B Olson-Buchanan and Wendy R Boswell. Blurring boundaries: Correlates of integration and segmentation between work and nonwork. *Journal of Vocational behavior*, 68(3):432–445, 2006.
  94. Tammy D. Allen, Kelsey L. Merlo, Roxanne C. Lawrence, Jeremiah Slutsky, and Cheryl E. Gray. Boundary management and work-nonwork balance while working from home. *Applied Psychology*, 2021.
  95. Jina Suh, Javier Hernandez, Koustuv Saha, Kathy Dixon, Mehrab Bin Morshed, Esther Howe, Anna Kawakami, and Mary Czerwinski. Towards successful deployment of wellbeing sensing technologies: Identifying misalignments across contextual boundaries. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 1–8. IEEE, 2023.

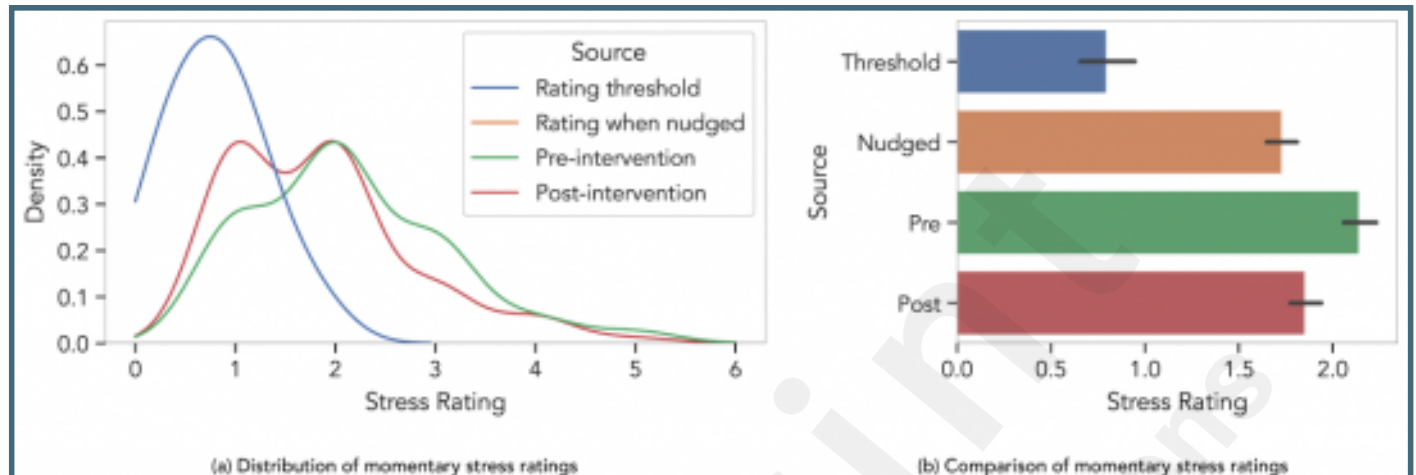
## Supplementary Files

## Figures

(a) System sends a nudge to users to perform an intervention. (b) Users can opt to postpone the intervention at a later time. (c) If users choose to do an intervention, they can select from one of three intervention categories. (d) Users first self-report their current stress level. (e) The system shows the intervention content for users to interact with. (f) User rates the intervention. (g) Users self-report their stress levels after the intervention.



(a) A kernel density estimate plot of subjective stress ratings shows that the distribution of stress ratings when the system sent intervention nudges is higher than the individual thresholds. The pre-intervention stress ratings are distributed at a higher rating than post-intervention stress ratings and the ratings when the nudges were sent. (b) Average momentary stress ratings at different points in time with 95% confidence intervals.



## Multimedia Appendixes



Supplementary information that includes: (1) detailed system architecture and stress inference methods, (2) intervention design and example interaction flows, (3) detailed description of how interaction and context data were extracted, and (4) additional results from ANOVA Type II and multicollinearity analyses conducted for each logistic regression model.

URL: <http://asset.jmir.pub/assets/c32f9c397c1cdfeff34d1715524dfb5.pdf>

Completed checklist of iCHECK-DH guidelines for manuscript “Towards Tailoring Just-in-time Adaptive Intervention Systems for Workplace Stress Reduction: Analysis of Intervention Implementation”.

URL: <http://asset.jmir.pub/assets/998268323ad6b5e0c115479566c2384d.docx>

