RESEARCH ARTICLE

Feasibility of Passive ECG Bio-sensing and EMA Emotion Reporting Technologies and Acceptability of Just-in-Time Content in a Well-being Intervention, Considerations for Scalability and Improved Uptake

P. Cummings[1] · A. Petitclerc[2] · J. Moskowitz[3] · D. Tandon[3] · Y. Zhang[4] · L. A. MacNeill[4] · N. Alshurafa[5] · S. Krogh-Jespersen[4] · J. L. Hamil[3] · A. Nili[4] · J. Berken[6] · W. Grobman[7] · A. Rangarajan[6] · L. Wakschlag[4]

Abstract

Researchers increasingly use passive sensing data and frequent self-report to implement personalized mobile health (mHealth) interventions. Yet, we know that certain populations may find these technical protocols burdensome and intervention uptake as well as treatment efficacy may be affected as a result. In the present study, we predicted feasibility (participant adherence to protocol) and acceptability (participant engagement with intervention content) as a function of baseline sociodemographic, mental health, and well-being characteristics of 99 women randomized in the personalized preventive intervention Wellness-for-Two (W-4-2), a randomized trial evaluating stress-related alterations during pregnancy and their effect on infant neurodevelopmental trajectories. The W-4-2 study used ecological momentary assessment (EMA) and wearable electrocardio-graph (ECG) sensors to detect physiological stress and personalize the intervention. Participant adherence to protocols was 67% for EMAs and 52% for ECG bio-sensors. Higher baseline negative affect significantly predicted lower adherence to both protocols. Women assigned to the intervention group engaged on average with 42% of content they received. Women with higher annual household income were more likely to engage with more of the intervention content. Researchers should carefully consider tailoring of the intensity of technical intervention protocols to reduce fatigue, especially among participants with higher baseline negative affect, which may improve intervention uptake and efficacy findings at scale.

Keywords EMA · ECG bio-sensor · Feasibility · Acceptability

✉ P. Cummings
peter.cummings@northwestern.edu

[1] Department of Psychiatry and Behavioral Sciences, Northwestern Feinberg School of Medicine, Chicago, IL, USA

[2] Laval University School of Psychology, 2325 Rue des Bibliothèques, QC, Québec G1V 0A6, Canada

[3] Department of Medical Social Sciences, Northwestern Feinberg School of Medicine, Chicago, IL, USA

[4] Department of Medical Social Sciences, Northwestern Feinberg School of Medicine, Institute for Innovations in Developmental Sciences, Chicago, IL, USA

[5] Department of Preventive Medicine, Northwestern Feinberg School of Medicine, Chicago, IL, USA

[6] Department of Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

[7] Department of Obstetrics & Gynecology, Northwestern Feinberg School of Medicine, Chicago, IL, USA

🌀 Springer

With technological innovation, researchers increasingly evaluate personalized mHealth interventions to improve well-being (Yim et al., 2020). By harnessing smartphone capabilities, we obtain more precise measurement in real time with ecological momentary assessments (EMAs; Smyth & Stone, 2003). Investigators employ novel wearable passive sensing technology, such as electrocardiograph (ECG) bio-sensors, to capture physiological indicators of health. While pointing to high adherence rates across wearable device studies, Chandrasekaran et al. (2020) also suggest usage may be more common among younger and wealthier individuals (Chandrasekaran et al., 2020). Regarding EMA adherence, van Genugten et al. (2020) find these protocols similarly feasible but note that participants with greater baseline negative affect are more likely to report perceived burden (van Genugten et al., 2020).

We examined the feasibility of EMA and ECG bio-sensor wear-time protocols and the acceptability of the personalized intervention component of the Wellness-for-Two (W-4-2) randomized trial. Wakschlag et al. (2021) designed the W-4-2 trial to assess the efficacy of a personalized intervention tailored to deliver timely content based on maternal stress, for improving infant neurodevelopment Wakschlag et al. (2021). The W-4-2 research team adapted the *Mothers and Babies* (MB) intervention that includes 12 individual sessions targeted to women at 10–22 weeks gestation to reduce prenatal distress (Muñoz et al., 2007). By adding just-in-time (JIT) content to supplement the intervention, the W-4-2 team reinforced material covered during the 12 sessions to further improve efficacy (Barrera et al., 2021). The W-4-2 team designed the JIT component with daily signals provided by ECG data collected by a wearable biosensor device and self-reported perceived stress from EMAs (Wakschlag et al., 2021).

Recent meta-analyses investigating predictors of adherence to mHealth interventions in various clinical or healthy samples determined an average study duration period of between 7 and 12 days with an average of 5 prompts per day. None of the studies was able to determine a relationship between protocol duration or sample characteristics and compliance (de Vries et al., 2021; Williams et al., 2021; Wrzus & Neubauer, 2022). Compared to those studies that did not use incentives for the data collection protocols, those that did had significantly higher rates of compliance (Wrzus & Neubauer, 2022). Two of these studies recommended that objective measurements such as wearable devices be used in parallel with EMAs, so as to avoid reliance on self-report data (de Vries et al., 2021; Wrzus & Neubauer, 2022). However, Wrzus and Neubauer (2022) noted that more research is needed to address whether incorporating such objective measurements with wearable technology affects compliance rates to interventions using EMA technology (Wrzus & Neubauer, 2022).

In the present study, we sought to explore predictors of feasibility and acceptability of the technical components of the W-4-2 intervention for pregnant women. We operationalized feasibility as adherence to the EMA and bio-sensor protocol wear-time requirements necessary for optimal intervention functionality. We operationalized acceptability as engagement with the supplementary JIT intervention content. We sought to identify which baseline sociodemographic, mental health, or well-being characteristics predicted better adherence to and engagement with the technical components of the W-4-2 JIT intervention. We hypothesized that participants' age, income, education, and baseline affect would predict adherence and engagement.

As a result of our analyses, we hope to refine the delivery of interventions for pregnant women, develop JIT intervention better suited for those less likely to adhere to measurement protocols, and subsequently improve efficacy at scale. Research involving pregnant women is a priority to improve maternal and child health outcomes and address related disparities. According to the 2022 National Academies of Sciences, Engineering, and Medicine report, pregnant women continue to constitute an underrepresented population in clinical research (National Academies of Sciences, Engineering, and Medicine, 2022). Our findings will contribute to the development of mHealth interventions designed to address the adverse effects of maternal stress, so they are more convenient and accessible (Brown et al., in press).

Method

Participants

We recruited participants through 6 university-affiliated obstetric care clinics in the metropolitan Chicago area. Participants were (1) at least 18 years of age; (2) English language proficient; (3) currently pregnant and between 18 and 22 weeks gestation; and (4) had a smartphone device. We recruited participants from 6/20/2019 to 8/11/2021, initially by approaching eligible patients after their prenatal care appointments in the clinic. After the onset of the COVID-19 pandemic, we shifted to remote recruitment by contacting potentially eligible women via email who were scheduled for prenatal visits in the clinic. In total, we approached 3,156 women to participate in the study. We screened 344 participants, of whom 316 were eligible. In total, we randomized 100 pregnant women receiving care across these 6 clinics during their first to second trimester to either the W-4-2 intervention (49) or usual care (51). Of those 100 randomized pregnant women, 1 participant withdrew their consent for participation and asked that their data not be used in future publications. A total of 99 women were included in this analysis. Research staff collected and managed study survey data and

randomization procedures using REDCap electronic data capture tools hosted at Northwestern University (Harris et al., 2009).

The W-4-2 trial is currently still underway, and we are actively following participants through the second year of their child's life. As such, the present analysis focuses only on data captured at baseline prior to availability of study effects. The datasets generated during and analyzed during the current study, a copy of the SAS v 9.3 coding file, and an output reflecting a condensed codebook are available in the OSF repository, DOI 10.17605/OSF.IO/RUQ3M.

Intervention

The W-4-2 trial adapted a well-validated, prenatal wellness intervention by adding a personalized approach (Muñoz et al., 2007). We designed this personalized strategy using two measurement methods to monitor stress based on the hypothesis that altering the gestational environment via stress reduction would have the greatest effects when based on women's real-time biopsychosocial state rather than a one-size-fits-all approach (Brown et al., in press).

Participants randomized to the W-4-2 intervention group received the *Mothers and Babies Course* (Muñoz et al., 2007)—a manualized intervention based on cognitive-behavioral therapy (CBT) and attachment theory. Participants completed the 12 sessions on a weekly basis. Participants were given the option to double-up on sessions in a given week for ease of scheduling purposes. Interventionists delivered content from each of the manualized 12 *Mothers and Babies* sessions in one-on-one format. Within each session, interventionists focused on content aligning with 3 CBT modules—pleasant activities, thoughts, and contact with others—and content related mostly to promoting parent-child attachment (Muñoz et al., 2007). We designed each session to be delivered in approximately 20 min, with sessions conducted in-person, by phone, or video conferencing depending on participant preference and availability.

For each of the 12 intervention sessions, a participant could receive up to 5 JIT intervention content messages. These text messages included either a brief, static message; a link to a brief 1–2 item survey; or a link to external content such as a worksheet, video, a guided meditation, or a mindfulness practice. Of the 5 inter-session JIT messages within a session-specific bank, 1 inter-session JIT message may have been included more than once for emphasis. We designed these JIT messages to reinforce the content from the most recent of the 12 sessions they received and provide an opportunity to practice the skills covered. On a given day, a participant who was determined to be stressed would only receive 1 JIT message from a particular session. If a participant's most recent session was doubled up due to scheduling, they would receive 1 JIT message from each of the prior sessions (Barrera et al., 2021).

Participants received a JIT message if more than 50% of their prior day ECG wear-time minutes were determined to be stress-positive, or if their EMA perceived stress scale 4-item (PSS-4) score was above the population average for women of 4.7 (Cohen & Williamson, 1988). We used the PSS-4 solely for this stress detection system, and not in this study's analyses, where we used the 10-item version of the PSS (PSS-10). If a participant did not complete any EMAs, nor provide any ECG data on a given day, the JIT intervention stress detection algorithm would deliver JIT content the next day based on a 50-50 chance determination. To reduce burden, we would not deliver inter-session JIT intervention content on a given day if a participant received content the day before. We made these modifications to our stress detection system based on our pilot study findings that, for pregnant women, 50% of their gestational experience may constitute a stressful day (Wakschlag et al., 2021).

We decided to use compensation to enhance compliance with the protocols after a review of pilot study data. In the pilot study, we sent participants 5 EMAs per day. Based on exit interview data, and data from a usability survey administered after the intervention, participants from the pilot study indicated that receiving 4 EMAs per day was more feasible. We also used the pilot study data to further develop our stress detection algorithm. From those analyses, we found that we needed at least 2 EMA responses for the self-report component of the machine-learning system to function properly. In the pilot study, participants responded to 47% of EMAs received. We developed automated weekly compliance SMS updates to encourage or reinforce compliance (e.g., "Well done! You have reached your goal every day this week for responding to EMAs."; Wakschlag et al., 2021).

In the present study, we determined each participant to be compliant with the EMA protocol based on whether they provided responses to at least 3 of the 4 EMAs sent per day. The cumulative percentage of EMA compliance was therefore reflective of the number of days in the 14-week intervention period that a participant met the threshold of completing at least 3 of the 4 EMAs received on a given day. Given that there does not currently exist consensus on the duration necessary to determine a stressful event, we evaluated the ECG data based on all available data provided (King et al., 2019). As a result, we determined each participant to be compliant with the bio-sensor on a given day if they had any valid stress positive or stress negative data after applying all noise-filtering processes.

During the consent process, we informed participants that they would receive compensation for, among other study-related procedures, their compliance rates for both EMAs and ECG bio-sensor wear-time over the 14-week intervention period, based on a tiered system (tier 1=0–35%, tier 2=35–

70%, tier 3=70%+). For EMA compliance, participants earned the following amounts for each tier: 1=$20; 2=$35; 3=$50. For the bio-sensor, participants earned the following amounts for each tier: 1=$30; 2=$52; 3=$75. We provided compensation at the end of the 14-week intervention period.

EMAs

For 7 days leading up to a participant's scheduled baseline, tech training, and randomization appointment, we invited all participants to respond to EMAs via text message. Participants received 4 EMAs per day based on their indicated sleep-wake schedule with accommodations made if they reported working night shifts. Within each EMA, we asked participants to report on their perceived stress level (Perceived Stress Scale [PSS-4]; Cohen et al., 1983), and on the extent to which they felt two positive (happy, excited) and three negative (worried, irritable/angry, sad) emotions on a scale of *0 = not at all* to *4 = very much*, in the past hour. From this week-long run-in period, we calculated a baseline, pre-intervention level of average daily positive and negative affect.

Once randomized, all participants were invited to complete the 4 times daily EMA protocol for a 14-week period. At the end of each week, participants received automated SMS feedback about their adherence to the EMA protocol for that week.

ECG Bio-sensor and Machine-Learned Model

Participants randomized to the study were asked to wear the BioStampRC ECG bio-sensor (mc10, Lexington, MA, USA), a flexible wearable patch for gathering raw electrocardiograph (ECG) data at 250Hz. The BioStampRC device included a patch-like adhesive to be affixed on the left side of a participant's chest. From this device, we were able to capture data on heartrate variability. King et al. (2019) demonstrated the BioStampRC device's effectiveness in using heartrate variability to predict physiologic and perceived stress among a sample of pregnant women King et al. (2019).

After completing the baseline assessment and prior to randomization, we invited participants to complete an in-person or remote tech training appointment. We designed the protocol for this visit to review how to use the bio-sensor prior to randomization. After the onset of the COVID-19 pandemic, we shifted to remote tech training appointments. In such cases, we shipped the BioStampRC device, adhesives, and the study tablet for syncing data on the remote, secure cloud storage directly to participants. During these tech training sessions, we reviewed with participants how to properly affix the adhesive and discussed the wear-time protocol, including when to wear the device and what to do while showering or exercising. Specifically, we discussed with participants that the BioStampRC adhesives were waterproof and technically able to withstand moisture exposure from prolonged use during exercise and bathing. However, we also informed participants that women in the pilot study reported such exposure tended to make the adhesives stickier and harder to remove, and thus we recommended they remove and reapply during exercise and bathing. We asked participants to wear the device during the day for up to 12 hours and to sync their data on the study tablet at each day's end in order for that day's data to be used. Over the course of the 14-week intervention period, we asked participants to wear the BioStampRC device on a 2-week on, 1-week off cycle (which amounted to 65 days in total) to minimize discomfort associated with the adhesive (Liu et al., 2018).

We designed a novel physiologic stress detection model using machine-learning to detect stress from ECG data on heartrate variability from inter-beat intervals assigned to each 1-min segment of data. We developed an algorithm to process the resulting ECG signaling as either physiologic stress-positive or stress-negative for each minute of wear time (Ng et al., in press). We adapted a process model to filter out invalid wear-time data (e.g., from skin stretching around the adhesive or from changes in movement and posture; Zhang et al., 2018). Furthermore, we employed a process previously used to evaluate heartrate variability data and filter out any observed findings outside of a normal human range (Manikandan & Soman, 2012). As a result, we were able to determine whether a participant experienced any physiologic stress-positive minutes during the day if they returned valid wear-time after the sync process. This machine-learning model combined with self-reported perceived stress from EMA data composed the stress detection system that dictated whether to send JIT intervention content to participants.

We provided ongoing support from interventionists and the study team for any participants who experienced technical difficulties. Data managers from the study team reviewed the wear-time data daily in an interactive dashboard. In the event of any missing data, study staff contacted participants up to 3 times per week offering to troubleshoot any issues with the sensor or discuss any other related issues. As a result of this troubleshooting, we documented any instances of technical issues or skin irritation problems experienced by participants throughout the 14-week intervention period.

Feasibility

We assessed feasibility in terms of a participants' compliance with the EMA and ECG bio-sensor wear-time protocols. First, we calculated the percentage of any EMAs completed of those that were sent to a given participant. Then, we evaluated participants' EMA responses over the 14-week intervention period and considered a participant compliant at the daily level if they responded to at least 3 of the 4 EMAs received that day. Regarding the ECG wear-time, we considered a participant compliant at the daily level if they synced their device as

instructed at the end of each day and had valid wear-time data. We calculated each participant's compliance with the bio-sensor protocol as the percentage of days they provided valid wear-time data at the end of the day out of the 65 days we asked them to wear the device. If a participant reported technological issues or skin irritation from the BioStampRC adhesive that prevented them from wearing the device, a "credit" was applied toward their compliance for that day for the purposes of compensation. However, for the purposes of this analysis we used the raw, uncredited compliance percentages. We reported the number and percentage of participants endorsing any technical issues (e.g., issues with study tablet, problems syncing data) or skin irritation in our analyses.

Acceptability

We analyzed intervention participants' (*n*=48) acceptability of W-4-2 in terms of their engagement with the JIT content received. We operationalized engagement as the percentage of unique, interactive, inter-session JIT intervention content that they clicked on via their smartphone, over the 14-week period.

Survey Data

All participants completed surveys at baseline, post-intervention (after 14 weeks), approximately 1 month before delivery, and at 5 postnatal time-points (1, 3, 7.5, 12, and 24 months after delivery).

For the present study, we focused on sociodemographic and mental health data from the baseline survey battery. As part of the baseline survey, participants provided sociodemographic characteristics including age, race (White, Black, Other), ethnicity (Hispanic vs non-Hispanic), education (greater than a college degree, less than or equal to a college degree), combined annual household income (greater than or equal to $100,000/year, less than $100,000/year), depressive symptomatology (PROMIS-8b Depression battery; Pilkonis et al., 2011), and perceived stress (Perceived Stress Scale [PSS-10]; Cohen et al., 1983). Importantly, while we reported racial and ethnic identity descriptive data for the entire sample of randomized women, we did not include these predictors for any analyses in light of recent recommendations around conducting psychological research using these variables as predictors (Morris et al., 2020).

Statistical Analyses

Feasibility of the EMA and Bio-sensor Protocols

We reported descriptive data on our sample using counts and percentages for each categorical baseline predictor variable

and means and standard deviations for each continuous baseline predictor variable.

For feasibility analyses using EMA and ECG wear-time compliance, we first looked at the distributions of the continuous variables (% of EMAs completed of those received, % of days a participant provided sensor data of those requested). After reviewing these distributions with Q-Q plots and histograms, we identified that both outcomes were bimodal, with participants tending to comply near both ends of the spectrum. In light of these findings, we decided to use tiers of compliance for EMAs and the ECG bio-sensor as feasibility outcome variables. In this way, we could provide results from statistical analyses that predicted the association between sample characteristics and higher tiers of compensation. While analytically, researchers currently recommend using continuous measures of compliance, our method best approaches the question of what participants find the intervention more or less feasible and acceptable given the nature of our data (de Vries et al., 2021; Williams et al., 2021; Wrzus & Neubauer, 2022). We provided descriptive data including counts and percentages for each tiered compensation variable.

We conducted bivariate analyses to test the associations between each baseline predictor variable and the EMA compliance tiers, as well as between each baseline predictor and the sensor compliance tiers. We reported counts and percentages for categorical variables and medians and interquartile ranges for continuous variables by each compliance tier. Given that our dependent variable for these analyses was a 3-level ordinal variable, we conducted non-parametric tests. For categorical predictor variables, we conducted Wilcoxon-Mann-Whitney tests. For continuous variables, we conducted Spearman's correlation tests.

Then, we tested multivariable prediction models of both compliance outcomes. For these analyses, we used ordinal logistic regression to test the association between baseline predictor variables and the compliance tiers. We reported proportional odds ratios and 95% confidence intervals adjusting for all baseline predictors. We interpreted estimates in terms of the odds of each predictor variable's association with the highest tier of compliance compared to the low and middle tiers.

Acceptability of the JIT Intervention Content

Among participants randomized to the intervention, we examined the distribution of the continuous engagement variable (% of unique, non-static inter-session JIT intervention content clicked on). We confirmed that this distribution met the assumption of normality using Q-Q plots and histograms. Then, we conducted bivariate analyses to test the associations between each baseline predictor variable and the continuous engagement outcome. We used independent *t*-tests for

dichotomous predictor variables and Pearson's correlation tests for continuous predictor variables.

Lastly, we conducted a multivariable prediction model to test the associations between baseline predictor variables and engagement. We used ordinary least squares (OLS) regression adjusting for each baseline predictor variable to predict engagement. We reported unstandardized beta coefficients and 95% confidence intervals.

Results

Participants

Within Table 1, we provided descriptive data on the sample of 99 women randomized in the study. Participants were on average 33 years old; most of them identified as White (71%) and non-Hispanic (87%); were highly educated (59% had more than a college degree); and reported high levels of socioeconomic status (74% reported combined household annual incomes greater than $100,000). The participants' average baseline depressive symptom *t*-scores were just below the population mean (mean *t*-score 48.42); their average baseline positive affect was 2.17 and their average baseline negative affect was 0.66 both on a scale of 0–4; and their average perceived stress on the PSS-10 was 14.69, near the mean score among a sample of pregnant women (Solivan, et al., 2015).

On average, participants responded to 67% of the EMAs they were sent and wore and synced their sensors according to protocol on average 52% of days over the 14-week intervention period. Regarding technical issues with the bio-sensor, 42% of participants reported ever having technical issues related to their sensor, and 24% endorsed ever having skin irritation associated with the adhesives. On average, 56% of participants complied with the EMA protocol (i.e., responding to at least 3 of the 4 EMAs in a day) at least 70% of the time, which was the criterion for top-tiered EMA compliance. Regarding the bio-sensor, 41% of participants complied with the BioStampRC protocol (i.e., providing non-missing, valid wear-time data and syncing to the cloud server at the day's end) at least 70% of the time, which was the criterion for top-tiered sensor compliance. Participants randomized to the W-4-2 intervention engaged with the unique, non-static, inter-session JIT intervention content they received 42% of the time, on average, over the 14-week period.

Bivariate Analyses

In Tables 2 and 3, we presented bivariate associations between participants' baseline characteristics and feasibility of the technical components of the intervention as measured by EMA and bio-sensor compliance. Participants' baseline negative affect scores were associated with both their EMA and

Table 1 Descriptive data from the randomized sample in W-4-2 (*n*=99)

Variable	Total
Age *n*, % with valid data	99 (100%)
Mean (sd)	33.2 (4.7)
Race *n*, % with valid data	97 (98%)
White	69 (71%)
Black	12 (12%)
Other	16 (17%)
Ethnicity *n*, % with valid data	99 (100%)
Non-Hispanic	86 (87%)
Education	98 (99%)
> College degree	58 (59%)
Combined annual income	99 (100%)
≥ $100,000	74 (75%)
Arm *n*, % with valid data	99 (100%)
Intervention	48 (48%)
Positive affect *n*, % with valid data	91 (92%)
Mean (sd)	2.17 (0.67)
Negative affect *n*, % with valid data	91 (92%)
Mean (sd)	0.66 (0.48)
PROMIS Depression *t*-score *n*, % with valid data	97 (98%)
Mean (sd)	48.42 (7.19)
PSS-10 *n*, % with valid data	99 (100%)
Mean (sd)	14.69 (6.66)
Sensor protocol	99 (100%)
Compliance %, mean (sd)	52% (34%)
Tech issues, *n* (%)	42 (42%)
Skin irritation, *n* (%)	24 (24%)
Compliance compensation tiers	
0–40%	39 (39%)
40–70%	20 (20%)
70+%	40 (41%)
EMA protocol	99 (100%)
% of EMAs completed mean (sd)	67% (32%)
Compliance compensation tiers	
1=0–40%	30 (30%)
2=40–70%	14 (14%)
3=70+%	55 (56%)
Engagement (intervention only)	45 (94%)
% of JIT content* accessed, mean (sd)	42% (23%)

Abbreviations: *sd* standard deviation, *PROMIS* Patient-Reported Outcomes Measurement Information System, *PSS* Perceived Stress Scale, *EMA* ecological momentary assessment, *JIT* just-in-time

* JIT content assessed only includes the non-static, interactive, and unique messages participants were sent that they clicked on

sensor compliance tier. Participants in the lowest EMA compliance tier were more likely to report higher baseline negative affect than those in the higher tiers median (inter-quartile range) negative affect: tier 1=0.85 (0.52), tier 2=0.72 (0.62),

tier 3=0.43 (0.48), Spearman's $\rho = -0.43$. $p < .0001$. Similarly, participants in the lowest sensor compliance tier were more likely to report higher baseline negative affect than those at the higher tiers median (IQR) negative affect: tier 1=0.72 (0.54), tier 2=0.41 (0.86), tier 3=0.50 (0.45), Spearman's $\rho = -0.27$ $p=.01$. We found similar results for both EMA and bio-sensor compliance tiers for baseline PSS-scores in Tables 2 and 3, such that participants who reported higher stress were less likely to comply with protocols at higher tiers. No other baseline predictors were associated with either the EMA or sensor compliance tiers.

In Table 4, we presented bivariate associations between participants' baseline characteristics and acceptability of the intervention as measured by engagement with the JIT content. Participants reporting combined annual incomes greater than or equal to $100,000 engaged on average with 47% of intervention content they received, while those reporting less than $100,000 engaged on average with 23% ($p<.0001$). No other baseline predictors were associated with engagement with the intervention content received.

In our sample, baseline PSS-10 scores were highly correlated with baseline positive affect ($r= -0.43$, $p<.0001$), baseline negative affect ($r=0.52$, $p<.0001$), and depressive symptoms ($r=0.62$, $p<.0001$). Furthermore, we considered PSS-10 scores as a main outcome of the forthcoming W-4-2 efficacy results to be presented in future analyses. To avoid collinearity

Table 2 Bivariate feasibility analyses describing associations between each baseline predictor variable and EMA compliance tier (1=0–40%, 2=40–70%, 3=70%+)

Predictor	Total	EMA compliance tiers			p
		1	2	3	
Age n, % with valid data[1]	99 (100%)				
Median (IQR)	33 (6.0)	33.0 (6.0)	34.0 (4.0)	32.0 (7.0)	.47
Race n, % with valid data	97 (98%)				
White, n (%)	69 (71%)	16 (23%)	12 (17%)	41 (60%)	
Black, n (%)	12 (12%)	6 (50%)	2 (17%)	4 (33%)	
Other, n (%)	16 (17%)	6 (38%)	0 (0%)	10 (62%)	--
Ethnicity n, % with valid data	99 (100%)				
Non-Hispanic, n (%)	86 (87%)	24 (28%)	12 (14%)	50 (58%)	
Hispanic, n (%)	13 (13%)	6 (46%)	2 (15%)	5 (39%)	--
Education[2]	98 (99%)				
≤ College degree, n (%)	40 (41%)	15 (38%)	4 (10%)	21 (52%)	
> College degree, n (%)	58 (59%)	15 (26%)	10 (17%)	33 (57%)	.44
Combined annual income[2]	99 (100%)				
<$100,000, n (%)	25 (25%)	11 (44%)	2 (8%)	12 (48%)	
≥ $100,000, n (%)	74 (75%)	19 (26%)	12 (16%)	43 (58%)	.21
Arm n, % with valid data[2]	99 (100%)				
Intervention, n (%)	48 (48%)	15 (31%)	10 (21%)	23 (48%)	
Control, n (%)	51 (52%)	15 (29%)	4 (8%)	32 (63%)	.29
Positive affect n, % with valid data[1]	91 (92%)				
Median (IQR)	2.20 (0.88)	2.29 (0.73)	2.08 (0.59)	2.17 (1.00)	.49
Negative affect n, % with valid data[1]	91 (92%)				
Median (IQR)	0.58 (0.60)	0.85 (0.52)	0.72 (0.62)	0.43 (0.48)	<.0001[*]
PROMIS Depression t-score[1] n, % with valid data	97 (98%)				
Median (IQR)	48.4 (8.9)	50.1 (11.3)	49.8 (11.0)	47.9 (8.0)	.13
PSS-10 n, % with valid data[1]	99 (100%)				
Median (IQR)	14.0 (10.0)	18.5 (9.0)	14.0 (14.0)	13.0 (9.0)	.0003[*]

Abbreviations: *IQR* interquartile range, *PROMIS* Patient-Reported Outcomes Measurement Information System, *PSS* Perceived Stress Scale, *EMA* ecological momentary assessment, JIT=just-in-time

[1] *P*-values are derived from the non-parametric Spearman's correlation test

[2] *P*-values derived from the non-parametric Wilcoxon-Mann-Whitney test

[*] Statistically significant at alpha of .05

Table 3 Bivariate feasibility analyses describing associations between each baseline predictor variable and the ordinal bio-sensor compliance tier (1=0–40%, 2=40–70%, 3=70%+)

Predictor	Total	Sensor compliance tiers			p
		1	2	3	
Age n, % with valid data[1]	99 (100%)				
Median (IQR)	33.0 (6.0)	34.0 (5.0)	33.0 (8.5)	32.0 (6.5)	.19
Race n, % with valid data	97 (98%)				
White, n (%)	69 (71%)	22 (32%)	14 (20%)	33 (48%)	
Black, n (%)	12 (12%)	9 (75%)	3 (25%)	0 (0%)	
Other, n (%)	16 (17%)	6 (37%)	3 (19%)	7 (44%)	--
Ethnicity n, % with valid data	99 (100%)				
Non-Hispanic, n (%)	86 (87%)	31 (36%)	18 (21%)	37 (43%)	
Hispanic, n (%)	13 (13%)	8 (62%)	2 (15%)	3 (23%)	--
Education[2]	98 (99%)				
≤ College degree, n (%)	40 (41%)	16 (40%)	8 (20%)	16 (40%)	
> College degree, n (%)	58 (59%)	23 (40%)	12 (20%)	23 (40%)	1.00
Combined annual income[2]	99 (100%)				
<$100,000, n (%)	25 (25%)	13 (52%)	4 (16%)	8 (32%)	
≥ $100,000, n (%)	74 (75%)	26 (35%)	16 (22%)	32 (43%)	.18
Arm n, % with valid data[2]	99 (100%)				
Intervention, n (%)	48 (48%)	19 (40%)	13 (27%)	16 (33%)	
Control, n (%)	51 (52%)	20 (39%)	7 (14%)	24 (47%)	.44
Positive affect n, % with valid data[1]	91 (92%)				
Median (IQR)	2.20 (0.88)	2.27 (0.73)	1.95 (1.00)	2.09 (0.92)	.96
Negative affect n, % with valid data[1]	91 (92%)				
Median (IQR)	0.58 (0.60)	0.71 (0.54)	0.41 (0.86)	0.50 (0.45)	.01[*]
PROMIS Depression t-score[1] n, % with valid data	97 (98%)				
Median (IQR)	48.4 (8.9)	49.8 (9.2)	49.0 (10.3)	48.4 (6.9)	.32
PSS-10 n, % with valid data[1]	99 (100%)				
Median (IQR)	14.0 (10.0)	17.0 (10.0)	16.0 (11.0)	12.0 (9.0)	.01[*]

Abbreviations: *IQR* interquartile range, *PROMIS* Patient-Reported Outcomes Measurement Information System, *PSS* Perceived Stress Scale

[1] *P*-values are derived from the non-parametric Spearman's correlation test

[2] *P*-values derived from the non-parametric Wilcoxon-Mann-Whitney test

[*] Statistically significant at alpha of .05

in our multivariable models, we did not include baseline PSS-10 scores in the multivariable prediction analyses.

Multivariable Models

In Table 5, we presented multivariable models testing the relationship between each baseline characteristic and the feasibility outcomes (compliance with EMA and bio-sensor protocols by tier) and the acceptability outcome (engagement with the unique, non-static JIT intervention content). We found that for every 1-unit increase in average baseline negative affect, participants were on average 86% less likely to comply at the top tier of EMA compliance versus the combined middle and low tiers (*p*<.05). Similarly, for every 1 unit

increase in average baseline negative affect, participants were on average 66% less likely to comply at the top tier of sensor compliance versus the combined middle and low tiers (*p*<.05). We found these associations while holding the effects of all other predictor variables constant. No other baseline predictors were associated with either feasibility outcome in the multivariable analyses.

Regarding engagement, as shown in Table 5, we found that compared to participants reporting annual household incomes less than $100,000, those reporting incomes at or above $100,000 engaged on average with 22% more of the unique, non-static inter-session JIT intervention content received (*p*<.05). We found this association while holding the effects of all other baseline predictor variables constant. No other

Table 4 Bivariate acceptability analyses among intervention participants (n=48) describing the associations between each baseline predictor variable and engagement with W-4-2 intervention measured as cumulative % of non-static inter-session JIT content received and clicked on. Note: Of the 48 intervention participants, 3 participants did not engage with any JIT content they received

Predictor (categorical)	Total	Engagement Mean (sd)	p
Race n, % with valid data	44 (92%)		
White, n (%)	29 (66%)	43.51 (23.31)	
Black, n (%)	8 (18%)	38.07 (20.55)	
Other, n (%)	7 (16%)	43.09 (25.55)	--
Ethnicity n, % with valid data	45 (94%)		
Non-Hispanic, n (%)	43 (96%)	42.66 (22.99)	
Hispanic, n (%)	2 (4%)	23.19 (14.35)	--
Education[1]	45 (94%)		
≤ College degree, n (%)	17 (38%)	46.76 (22.72)	
> College degree, n (%)	28 (62%)	38.79 (22.92)	.26
Combined annual income[1]	45 (94%)		
<$100,000, n (%)	10 (22%)	22.53 (11.47)	
≥ $100,000, n (%)	35 (78%)	47.31 (22.47)	<.0001[*]
Predictor (continuous)[2]		Engagement r	p
Age n, % with valid data	45 (94%)		
Mean (sd)	34.4 (4.5)	0.10	.52
Positive affect n, % with valid data	42 (88%)		
Mean (sd)	2.08 (0.71)	−0.04	.82
Negative affect n, % with valid data	42 (88%)		
Mean (sd)	0.68 (0.48)	−0.30	.05
PROMIS Depression t-score n, % with valid data	44 (92%)		
Mean (sd)	49.4 (7.8)	0.05	.76
PSS-10 n, % with valid data	45 (94%)		
Mean (sd)	14.8 (7.3)	−0.16	.30

Abbreviations: *sd* standard deviation, *PROMIS* Patient-Reported Outcomes Measurement Information System, *PSS* Perceived Stress Scale, *JIT* just-in-time

[1] P-values are derived from independent t-tests

[2] r and p-values derived from Pearson's correlation tests

[*] Statistically significant at alpha of .05

Discussion

baseline predictors were associated with our engagement outcome in the multivariable analysis.

Eysenbach (2005) expressed concern regarding the presentation of efficacy results, emphasizing the need for a more nuanced discuss regarding what worked and did not work in mHealth interventions Eysenbach (2005). Dagher et al. (2022) suggested bolstering efficacy results by striving recruit and retain those historically left behind in mHealth research including women with lower incomes Dagher et al. (2022). We found pregnant women with higher baseline negative affect were less likely to comply with monitoring technology protocols in a personalized intervention to reduce prenatal distress. Furthermore, those with lower incomes were less likely to engage with JIT intervention content. Going forward, interventionists should consider diverse perspectives, priorities, and values of pregnant women during implementation strategy development for optimal tailoring and personalization of content. Findings presented here, even within this well-resourced sample, suggest that a deeper understanding of how best to optimize uptake for women from varied socioeconomic contexts will be key for scalability. This should include community-engaged partnerships and/or focus groups to ensure that lived experience, priorities, and values of women from varied backgrounds are heard and incorporated.

Table 5 Multivariable ordinal logistic regression model results with adjusted odds ratios (ORs) for measures of feasibility (EMA and ECG-biosensor compliance tier: 1=0–40%, 2=40–70%, 3=70%+) and multivariable OLS regression model results with unstandardized beta (β) coefficients for measure of acceptability (engagement: cumulative % of non-static inter-session JIT content received and clicked on)

Predictor	Feasibility						Acceptability	
	EMA compliance tiers (n=88)			Sensor compliance tiers (n=88)			Engagement (n=41)	
	OR	95% CI		OR	95% CI		β	95% CI
Age	0.97	(0.87, 1.07)		0.96	(0.88, 1.05)		−0.13	(−1.66, 1.40)
Combined annual income (< $100K as ref)	1.65	(0.60, 4.52)		1.65	(0.65, 4.20)		22.37	(5.79, 38.94)[*]
Education (≤ college degree as ref)	1.84	(0.73, 4.63)		1.04	(0.45, 2.40)		−9.69	(−23.95, 4.58)
Positive affect	0.74	(0.32, 1.67)		0.67	(0.32, 1.42)		−7.17	(−19.25, 4.91)
Negative affect	0.14	(0.04, 0.45)[*]		0.34	(0.12, 0.97)[*]		−14.97	(−33.34, 3.41)
PROMIS Depression t-score	1.03	(0.96, 1.11)		1.01	(0.95, 1.08)		0.41	(−0.56, 1.38)
Arm (control as ref)	0.54	(0.21, 1.37)		0.80	(0.35, 1.84)		--	--

Abbreviations: *OR* odds ratio, *CI* confidence interval, *ref* reference category, *PROMIS* Patient-Reported Outcomes Measurement Information System
[*] Statistically significant at alpha of .05

Participants in our study responded on average to 67% of EMA prompts sent, a lower rate compared to a recent meta-analysis (average adherence = 79%, sd = 14%; Wrzus & Neubauer, 2022). Williams et al. (2021) suggest that inconsistent reporting across studies may bias results toward studies with higher compliance. Furthermore, they commented that such inconsistencies may result in inadequacies of such meta-analyses to accurately determine protocol features associated with better or worse compliance (Williams et al., 2021). Interventionists interested in improving well-being among pregnant women will require protocols much longer than the average of between 7 and 12 days reported in recent meta-analyses (de Vries et al., 2021; Williams et al., 2021; Wrzus & Neubauer, 2022). We hypothesize that such researchers may improve EMA adherence by adopting less intensive protocols still capturing a typical week. For instance, while reporting on affective changes throughout gestation, Lazarides et al. (2021) utilized a weekly, 4-day (2 weekdays, 2 weekends) EMA protocol, and reported higher (86%) adherence Lazarides et al. (2021). In W-4-2, participants likely perceived the combination of the EMA and bio-sensor protocols together as more burdensome, adversely affecting compliance.

Participants in our study wore the bio-sensor, synced their device, and provided sufficient, valid wear-time data, on average, 52% of the 14-week intervention period. Throughout the W-4-2 study, we provided daily consultation for participants who did not provide wear-time data. Our study staff provided services for those 42% of participants who endorsed at any time a technical issue with the bio-sensor. While we resolved these issues, they do represent an obstacle for researchers considering adding objective wearable technology for their interventions. Throughout the study, 24% of our sample endorsed skin irritation with the adhesive preventing them from using the bio-sensor. Future interventionists interested in combining objective and self-report data could benefit from enhanced staff training protocols, fidelity analyses, and a "warm hand-off" to reduce the prevalence of such issues related to the wearables.

In the future, researchers may consider using sensor data and smartphone behavior to predict likelihood of response to an EMA (Liao et al., 2018). With such predictive, machine-learning based analyses, we can limit the number of EMAs to which participants need to respond for a given protocol, further personalizing interventions. Such research may reduce the necessary number of consecutive days participants need to wear devices and respond to EMA prompts. While such enhancements are possible, they do require a fully sustainable and usable system of subjective and objective data (Ng et al., in press). We hope that the present analysis can be informative for future interventions to ensure optimal system usability.

We found negative affect, depressive symptomatology, and perceived stress were negatively associated with our feasibility outcomes, aligning with other research suggesting these factors may be associated with decreased motivation and lower intervention uptake (Sokolovsky et al., 2014). Walsh et al. (2015) noted that, among pregnant adolescents, compared to participants with lower baseline negative affect scores, those participants with higher baseline negative affect were less likely to complete their EMA protocol after the first session Walsh et al. (2015). We hypothesize that participants with greater baseline negative affect in our study may negatively appraise the measurement process. This could then

translate to a lower inclination to participate in subsequent data collection activities.

Intervention participants on average engaged with 42% of the non-static JIT intervention content they received. We found that greater household annual income was associated with increased engagement with the intervention. Previous research suggests that socioeconomically advantaged participants are more likely to report favorable usability ratings of smartphones and apps (Rahmati et al., 2012). Future investigators may pursue factor-level analyses exploring whether socioeconomic status predicts uptake of specific intervention content to determine if modifications to intervention content may be necessary for broader reach.

We used SMS delivery for EMA dissemination and JIT intervention content delivery. While SMS-based delivery is the most prevalent modality for EMA-based interventions, it is limited in its capacity to differentiate willful from accidental non-compliance (Wrzus & Neubauer, 2022). Recent mHealth interventions focusing on pregnant women have moved toward app-based development to enhance usability. In a sample of studies, EMA compliance (74–84%) was higher than what we saw in our study (Allen et al., 2018; Faherty et al., 2017; Sanjuan et al., 2019). Participants reported greater engagement with app-based platforms when user-friendly and directly useful features were present that enabled them to interact with and observe their results over time (Hartmann et al., 2019). Our study team chose not to implement a dashboard feature visualizing change in stress over time to avoid providing participants in the control group access to features that could produce intervention effects, potentially complicating efficacy findings. Researchers may benefit from an implementation science informed approach with qualitative feedback from focus groups of participants to inform enhancements to intervention end-user focused design.

Given concerns about fatigue, Mishra et al. (2021) evaluated adaptive, machine learning models to assess an individual's receptivity when determining timing of JIT content dissemination Mishra et al. (2021). We recommend researchers prioritize reducing fatigue to improve uptake among diverse populations with greater negative affect. These participants may be more sensitive to perceived burden and could stand to gain the most from well-being interventions. Future researchers may benefit from adaptive trial designs, testing incremental utility of intervention components, and optimizing intervention feasibility and acceptability to the broadest possible range of women.

We conducted the first of its kind personalized prenatal stress and depression prevention intervention using biosensing to tailor information input. Our sample was non-representative and well-resourced. These preliminary findings speak to both the promise of such technologies for maternal-child health research and the need for population-based examination of these issues. Partnerships with community members will be vital to delineating the optimized approach for women in particular contexts. We believe this approach has potential to advance personalized and scalable interventions to improve well-being at this critical life-stage.

Given the high socioeconomic status of our sample, participants may have been more likely to engage with technological components of study protocols with little reinforcement (Wrzus & Neubauer, 2022). Our results cannot speak to feasibility and acceptability of technological enhancements among a more socioeconomically diverse sample of pregnant women. Future researchers can modify their experimental designs with stratified or quota sampling to specifically address differences in feasibility and acceptability by socioeconomic status as well as by racial and ethnic identity. Despite these limitations, we believe that the present study represents an opportunity to inform decision-making among future mHealth interventionists to reduce burden, enhance usability, and improve uptake at scale.

References

Allen, A., Tosun, N., Carlson, S., & Allen, S. (2018). Postpartum changes in mood and smoking-related symptomatology: An ecological momentary assessment investigation. *Nicotine & Tobacco Research, 20*(6), 681–689. https://doi.org/10.1093/ntr/ntx118

Barrera, A. Z., Hamil, J., & Tandon, D. (2021). Integrating SMS text messages into a preventive intervention for postpartum depression

delivered via in-home visitation programs: Feasibility and acceptability study. *JMIR Formative Research, 5*(11), e3099 PMID:34792478.

Brown, H., Krogh-Jespersen, S., Tandon, T., Graham, A., Seghete, K., & Wakschlag, L. (in press). Looking ahead: Pre- and perinatal interventions for maternal distress to prevent neurodevelopmental vulnerability. In A. Wazana, T. Oberlander, & E. Szekely (Eds.), *Prenatal stress and child development*. Springer.

Chandrasekaran, R., Katthula, V., & Moustakas, E. (2020). Patterns of use and key predictors for the use of wearable health care devices by US adults: Insights from a national survey. *Journal of Medical Internet Research, 22*(10), e22443.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*, 386–396.

Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health: Claremont Symposium on Applied Social Psychology* (pp. 31–67). Sage.

Dagher, L., Nedunchezhian, S., El Hajjar, A. H., Zhang, Y., Deffer Jr., O., Russell, A., et al. (2022). A cardiovascular clinic patients' survey to assess challenges and opportunities of digital health adoption during the COVID-19 pandemic. *Cardiovascular Digital Health Journal, 3*(1), 31–39. https://doi.org/10.1016/j.cvdhj.2021.10.007

de Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies, 22*(5), 2361–2408. https://doi.org/10.1007/s10902-020-00324-7

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1), e11.

Faherty, L. J., Hantsoo, L., Appleby, D., Sammel, M. D., Bennett, I. M., & Wiebe, D. J. (2017). Movement patterns in women at risk for perinatal depression: Use of a mood-monitoring mobile application in pregnancy. *Journal of the American Medical Informatics Association, 24*(4), 746–753. https://doi.org/10.1093/jamia/ocx005

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research Electronic Data Capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381.

Hartmann, R., Sander, C., Lorenz, N., Böttger, D., & Hegerl, U. (2019). Utilization of patient-generated data collected through mobile devices: Insights from a survey on attitudes toward mobile self-monitoring and self-management apps for depression. *JMIR Mental Health., 6*(4), e1167 PMID: 30942693.

King, Z. D., Moskowitz, J., Egilmez, B., Zhang, S., Zhang, L., Bass, M., et al. (2019). Micro-stress EMA: A passive sensing framework for detecting in-the-wild stress in pregnant mothers. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies., 3*(3), 1–22. https://doi.org/10.1145/3351249

Lazarides, C., Moog, N. K., Verner, G., Voelkle, M. C., Henrich, W., Heim, C. M., et al. (2021). The association between history of prenatal loss and maternal psychological state in a subsequent pregnancy: An ecological momentary assessment (EMA) study. *Psychological Medicine*, 1–11. https://doi.org/10.1017/S0033291721002221

Liao, P., Dempsey, W., Sarker, H., Hossain, S. M., al'Absi, M., Klasnja, P., & Murphy, S. (2018). Just-in-time but not too much. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(4), 1–21.

Liu, Y., Tian, L., Raj, M. S., Cotton, M., Ma, Y., Ma, S., et al. (2018). Intraoperative monitoring of neuromuscular function with soft, skin-mounted wireless devices. *NPJ Digital Medicine, 1*(1), 19.

Manikandan, M. S., & Soman, K. (2012). A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomedical Signal Processing and Control, 7*(2), 118–128. https://doi.org/10.1016/j.bspc.2011.03.004

Mishra, V., Künzler, F., Kramer, J.-N., Fleisch, E., Kowatsch, T., & Kotz, D. (2021). Detecting receptivity for mHealth interventions in the natural environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(2), 1–24.

Morris, A. S., Wakschlag, L., Krogh-Jespersen, S., Fox, N., Planalp, B., Perlman, S. B., et al. (2020). Principles for guiding the selection of early childhood neurodevelopmental risk and resilience measures: HEALthy Brain and Child Development Study as an exemplar. *Adversity and Resilience Science, 1*(4), 247–267.

Muñoz, R. F., Le, H.-N., Ippen, C. G., Diaz, M. A., Urizar Jr., G. G., Soto, J., Mendelson, T., Delucchi, K., & Lieberman, A. F. (2007). Prevention of postpartum depression in low-income women: Development of the Mamás y Bebés/ Mothers and Babies Course. *Cognitive and Behavioral Practice, 14*(1), 70–83.

National Academies of Sciences, Engineering, and Medicine. (2022). *Improving representation in clinical trials and research: Building research equity for women and underrepresented groups*. The National Academies Press. https://doi.org/10.17226/26479

Ng, A., Wei, B., Jain, J., Ward, E., Tandon, D., Moskowitz, J., Krogh-Jespersen, S., Wakschlag, L. S., & Alshurafa, N. (in press). Predicting next-day perceived and physiological stress of pregnant women using machine learning and explainability: Algorithm development and validation. *JMIR mHealth and uHealth* 13/05/2022, *10*(8), 33850.

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., & Cella, D. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): Depression, anxiety, and anger. *Assessment, 18*(3), 263–283.

Rahmati, A., Tossell, C., Shepard, C., Kortum, P., & Zhong, L. (2012). Exploring iPhone usage: The influence of socioeconomic differences on smartphone adoption, usage and usability. In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (pp. 11-20).

Sanjuan, P. M., Pearson, M. R., Poremba, C., Amaro, H. d. L. A., & Leeman, L. (2019). An ecological momentary assessment study examining posttraumatic stress disorder symptoms, prenatal bonding, and substance use among pregnant women. *Drug and Alcohol Dependence, 195*, 33–39. https://doi.org/10.1016/j.drugalcdep.2018.11.019

Smyth, J. M., & Stone, A. A. (2003). Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being, 4*(1), 35–52.

Sokolovsky, A. W., Mermelstein, R. J., & Hedeker, D. (2014). Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine & Tobacco Research, 16*(3), 351–358.

Solivan, A. E., Xiong, X., Harville, E. W., & Buekens, P. (2015). Measurement of perceived stress among pregnant women: A comparison of two different instruments. Maternal and Child Health Journal, 19(9), 1910–1915. https://doi:10.1007/s10995-015-1710-5

van Genugten, C. R., Schuurmans, J., Lamers, F., Riese, H., Penninx, B. W., Schoevers, R. A., Riper, H. M., & Smit, J. H. (2020). Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *Journal of Clinical Medicine, 9*(2), 322.

Wakschlag, L. S., Tandon, D., Krogh-Jespersen, S., Petitclerc, A., Nielsen, A., Ghaffari, R., Mithal, L., Bass, M., Ward, E., Berken, J., Fareedi, E., Cummings, P., Mestan, K., Norton, E. S., Grobman, W., Rogers, J., Moskowitz, J., & Alshurafa, N. (2021). Moving the dial on prenatal stress mechanisms of neurodevelopmental vulnerability to mental health problems: A personalized prevention proof of concept. *Developmental Psychobiology, 63*(4), 622–640.

Walsh, K., Basu, A., & Monk, C. (2015). The role of sexual abuse and dysfunctional attitudes in perceived stress and negative mood in pregnant adolescents: An ecological momentary assessment study. *Journal of Pediatric and Adolescent Gynecology, 28*(5), 327–332. https://doi.org/10.1016/j.jpag.2014.09.012

Williams, M. T., Lewthwaite, H., Fraysse, F., Gajewska, A., Ignatavicius, J., & Ferrar, K. (2021). Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: Systematic review and meta-analysis. *Journal of Medical Internet Research, 23*(3), e17023. https://doi.org/10.2196/17023

Wrzus, C., & Neubauer, A. B. (2022). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment.* https://doi.org/10.1177/10731911211067538

Yim, S. J., Lui, L., Lee, Y., Rosenblat, J. D., Ragguett, R. M., Park, C., Subramaniapillai, M., Cao, B., Zhou, A., Rong, C., Lin, K., Ho, R. C., Coles, A. S., Majeed, A., Wong, E. R., Phan, L., Nasri, F., & McIntyre, R. S. (2020). The utility of smartphone-based, ecological momentary assessment for depressive symptoms. *Journal of Affective Disorders, 274*, 602–609.

Zhang L, King Z, Egilmez B, Reeder J, Ghaffari R, Rogers J, et al. (2018). Measuring finegrained heart-rate using a flexible wearable sensor in the presence of noise. IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN); 2018: IEEE. https://doi.org/10.1109/BSN.2018.8329683