

# The Ava bracelet for collection of fertility and pregnancy data in free-living conditions: An exploratory validity and acceptability study

Alison K. Nulty<sup>1,2</sup> , Elizabeth Chen<sup>3</sup>  and Amanda L. Thompson<sup>1,2,4</sup>

## Abstract

**Objective:** To evaluate the validity and acceptability of the Ava bracelet for collecting heart rate, sleep, mood, and physical activity data among reproductive-aged women (pregnant and nonpregnant) under free-living conditions.

**Methods:** Thirty-three participants wore the Ava bracelet on their non-dominant wrist and reported mood and physical activity in the Ava mobile application for seven nights. Criterion validity was determined by comparing the Ava bracelet heart rate and sleep duration measures to criterion measures from the Polar chest strap and ActiGraph GTX3 + accelerometer. Construct validity was determined by comparing self-report measures and the heart rate variability ratio collected in the Ava mobile application to previously validated measures. Acceptability was evaluated using the modified Acceptability of Health Apps among Adolescents Scale.

**Results:** Mean absolute percentage error was 11.4% for heart rate and 8.5% for sleep duration. There was no meaningful difference between the Ava bracelet, ActiGraph, and construct a measure of sleep quality. Compared to construct measures, Ava bracelet heart rate variability had a significant low negative correlation ( $r = -0.28$ ), mood had a significant low positive correlation ( $r = 0.39$ ), and physical activity level had a significant low ( $r_{\text{level of physical activity}} = 0.56$ ) to moderate positive correlation ( $r_{\text{MET-minutes/week}} = 0.71$ ). The acceptability of the Ava bracelet was high for fertility and low for pregnancy tracking.

**Conclusion:** Preliminary evidence suggests the Ava bracelet and mobile application estimates of sleep and heart rate are not equivalent to criterion measures in free-living conditions. Further research is needed to establish its utility for collecting prospective, subjective data throughout periods of preconception and pregnancy.

## Keywords

Pregnancy, fertility, mobile health, wearables, validity, acceptability

Submission date: 12 June 2021; Acceptance date: 14 February 2022

## Introduction

The use of smartphones and wearable technology have allowed for increased monitoring of daily lifestyle behaviors such as sleep and physical activity. In 2019, approximately 25% of women in the United States reported regularly wearing a smart-watch or fitness tracker.<sup>1</sup> This regular use of wearable technology has started to enable health researchers to collect continuous, real-time data in a free-living environment.<sup>2,3</sup>

Previous studies have determined the use of wearables to be beneficial within health care settings, as well as

<sup>1</sup>Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>2</sup>Carolina Population Center, University of North Carolina, North Carolina, USA

<sup>3</sup>Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>4</sup>Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

### Corresponding author:

Alison K. Nulty, Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516, USA.  
 Email: alkramer@email.unc.edu

maternal and child health research.<sup>4,5</sup> In a study assessing the feasibility of collecting data during pregnancy using Garmin Vívosmart HR activity tracker, Grym et al.<sup>5</sup> found it was feasible for women to wear the device throughout their pregnancy. Other studies have employed the use of wearable devices, with their respective mobile apps, during pregnancy as part of interventions aiming to: (1) promote physical activity and adequate sleep during pregnancy<sup>6,7</sup> or (2) collect longitudinal cohort data on physical activity during several time points throughout preconception and into pregnancy.<sup>8</sup> However, the majority of studies involving wearables during preconception and pregnancy have focused on using the wearable as part of a health behavior intervention and have less often used the wearables for research data collection. Thus, the literature on the ability of wearables to provide adequate and accurate observational data remains scarce.<sup>9–12</sup> Due to the rapid physiological changes that occur during pregnancy, a collection of daily objective and subjective measures of maternal health could be important to understand the short- and long-term influences of pregnancy on both mother and infant. Wearables may provide a low burden tool to gather these data.

The Ava bracelet, a newly developed Food and Drug Administration-registered fertility tracking wearable, has been marketed as a wearable that can accurately predict a woman's fertile window by measuring and analyzing nighttime skin temperature, resting pulse rate, heart rate variability (HRV) ratio, skin perfusion, breathing rate, movement, and sleep.<sup>13</sup> To predict an Ava bracelet wearer's fertile window, women are asked to wear the Ava bracelet only while they sleep and to sync the device with their Ava smartphone application (mobile app). The mobile app also allows women to self-report their mood, bodily symptoms, and a daily custom log to add any other pertinent information. The Ava mobile app also allows women to continue to track data during their pregnancy. This enables women to monitor their own health data and receive weekly updates on their developing fetuses. Additionally, pregnant women have the ability to report their weight, activity level, prenatal measurements (i.e. blood pressure, baby's heart rate, or fundal height), medical diagnoses (i.e. gestational diabetes, gestational hypertension, preeclampsia, etc.), and hydration status on the Ava mobile app. While the Ava company has sponsored and published research on the use and validity of their product, most of their studies have focused on its ability to predict a woman's fertile window, and the bracelet has not yet been assessed by an unaffiliated researcher.<sup>13–15</sup> This study investigated the validity and acceptability of the Ava bracelet for collection of observational data among pregnant and nonpregnant reproductive-aged women.

## Methods

### Study design

The validity and acceptability of the Ava bracelet were assessed through a mixed-methods prospective observational study among reproductive-aged women over a study period of 12 weeks. Given the use of qualitative and quantitative data, we aimed to reach a sample size of 30 participants to satisfy the central limit theorem.<sup>16</sup>

### Recruitment

We chose to recruit both nonpregnant and pregnant women to assess the validity and acceptability among both groups of reproductive-aged women, knowing that health researchers are likely to use wearables, such as the Ava bracelet, to collect data before, during, and after pregnancy. Reproductive-aged (18 to 45 years) women were recruited through advertisements on Instagram, Facebook, and local online listservs (Appendix) by the first author. Women were eligible if they spoke English, had access to a smartphone, did not wear a pacemaker, and regularly slept 4 or more hours per night (specified by the Ava company for accurate data recording). Women were excluded if they had a pacemaker, as some of the sensors on the Ava bracelet could interfere with a pacemaker; or if they did not get at least 4 h of sleep per night.

Recruitment took place between September and November 2020, and data collection ended in December 2020. Women who expressed interest in participating in the study were asked to fill out an online screener survey that included basic information such as race/ethnicity, marital status, current employment status, and pregnancy status. We had 272 women complete the screener survey and 35 were invited to enroll. Participants were chosen on a first-come-first-served basis with an emphasis on recruiting pregnant women. Women who were pregnant throughout the data collection period were considered pregnant, while women who were not pregnant during data collection were considered nonpregnant. All study participants provided informed consent prior to participating in this study. This study protocol was reviewed and approved by the Institutional Review Board of the University of North Carolina at Chapel Hill (IRB No. 20-1732).

### Study procedures

Self-reported sociodemographic data were collected through online questionnaires both before and after the data collection period. After recruitment and consent, each participant scheduled a time to meet the principal investigator (PI) to pick up their devices. During device pick-up, participants also received a paper and electronic copy of study instructions, with separate instructions for

pregnant and nonpregnant women, to address the differences by pregnancy status within the Ava mobile app. All participants were instructed to download the Ava mobile app using the anonymized username and password provided by the research team. Participants were instructed to wear the Ava bracelet (Ava AG, Zurich, Switzerland) on their non-dominant wrist, the ActiGraph GT3X+ (ActiGraph Corporation, Pensacola, FL, USA) monitor on their dominant wrist, and a heart rate chest strap (Polar H10, Polar Electro, Kempele, Finland) for seven consecutive nights during sleep. Participants were also instructed to self-report their mood and physical activity under the designated tabs within the Ava mobile app.

Each morning, participants were sent a secure survey link via a Short Message Service text message where they were asked to record the time they went to bed and the time they woke up. Following the seven nights of data collection, participants returned the devices to the PI and completed a set of online questionnaires to assess their past week's stress, physical activity, sleep quality, and acceptability of the Ava bracelet and mobile app. Lastly, all participants were interviewed via Zoom (Zoom Video Communications, San Jose, CA, USA). Before the interview began, subjects were informed all responses would be de-identified, kept confidential, and recorded if the participant agreed. After completion of the interview, each participant was emailed a \$30 Amazon gift card.

## Measures

### *Actigraphy and Ava mobile application*

Objective measures of sleep and heart rate were obtained using the ActiGraph GT3X+, worn on the dominant wrist, and a heart rate chest strap. These data were collected in 60 s epochs continuously for seven nights during sleep. ActiGraph sleep and heart rate data were extracted using ActiLife software version 6.13.4 (ActiGraph Corporation, Pensacola, FL, USA). A combination of the Cole-Kripke and the Tudor-Locke algorithm was used to detect sleep and wake periods. Nightly sleep duration and nightly sleep efficiency (%) were obtained from the analysis.

Objective measures of heart rate, sleep duration, sleep quality, and HRV ratio were obtained using the Ava bracelet worn on the non-dominant wrist. Subjective measures of daily mood and physical activity were entered into the mobile app by participants (Figures 1 and 2). These data were saved in the mobile app and extracted by the PI after the data collection period ended.

### *Heart rate*

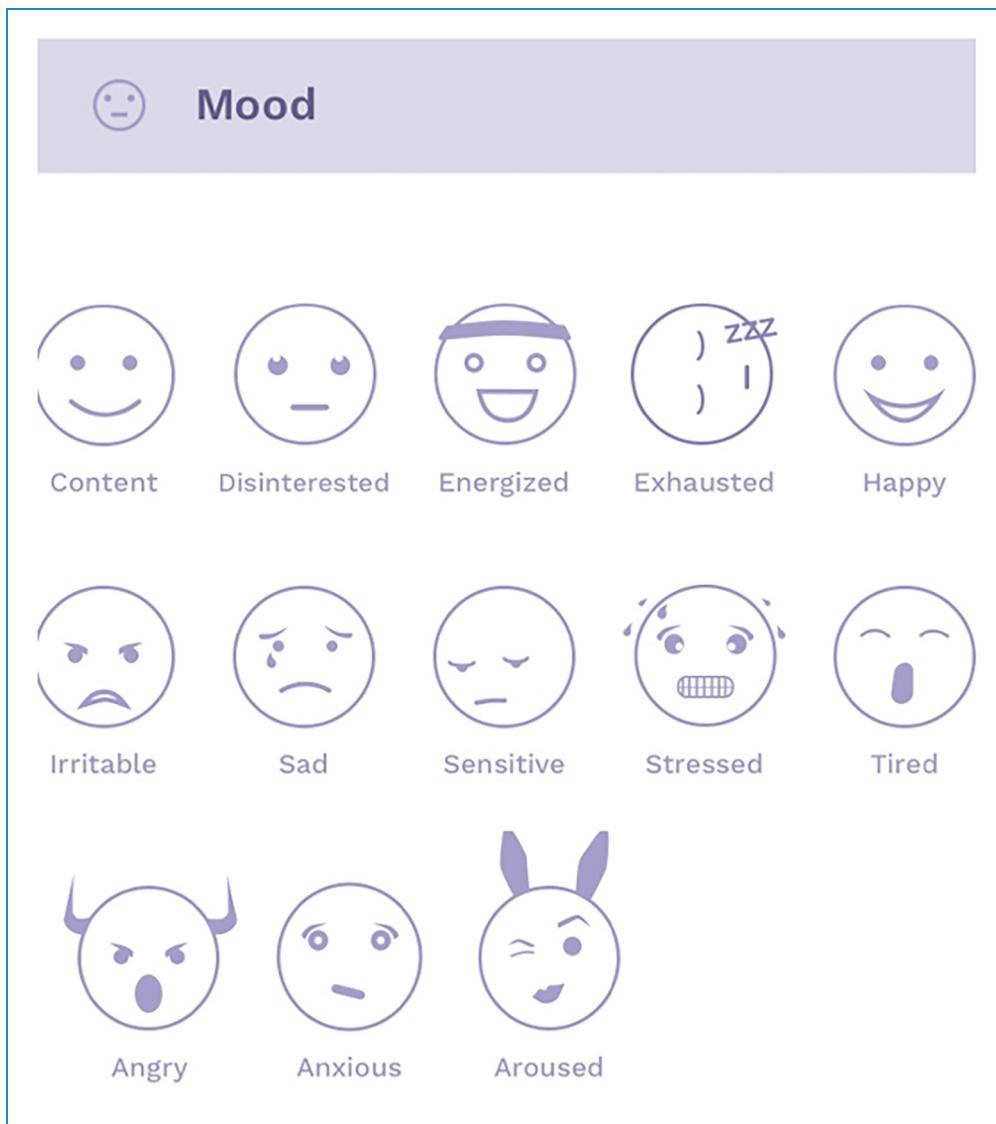
Both the Ava bracelet and the ActiGraph collected mean heart rate (bpm) during sleep. Additionally, the Ava bracelet collected the HRV ratio. HRV refers to the fluctuation in

time between subsequent heartbeats over time,<sup>17</sup> and is considered to be a measure of the body's ability to adapt to physiological and environmental stressors. Adequate HRV is associated with health and self-regulation.<sup>17</sup> The Ava bracelet measures the HRV ratio of the low frequency to high frequency (LF/HF) power.<sup>18</sup> This measure is thought to estimate the ratio between the sympathetic and parasympathetic nervous systems.<sup>17</sup> While the ActiGraph and heart rate chest strap did not collect the HRV ratio, they did collect R-R intervals (ms), the time between two successive R-waves of the QRS complex on an electrocardiogram.<sup>19</sup> The R-R intervals enabled us to calculate SD1/SD2, a measure of the unpredictability of the R-R intervals; where SD1 is equal to the root mean square of successive R-R intervals and SD2 equals SD1 plus the mean R-R interval during the time of data collection.<sup>17</sup> SD1/SD2 is correlated with the LF/HF ratio (HRV ratio). Nightly HRV ratio was measured using the Ava bracelet and nightly SD1/SD2 was calculated for each participant using the R-R interval data collected from the ActiGraph and heart rate chest strap. A lower HRV ratio through the Ava bracelet indicates a woman is less stressed,<sup>18</sup> while a lower SD1/SD2 ratio has been associated with increased stress or anxiety.<sup>20</sup> Therefore, we would expect to see the HRV ratio calculated through the Ava bracelet to be negatively correlated to the SD1/SD2 ratio.

### *Sleep*

The objective sleep measures collected with the Ava bracelet include sleep duration, nighttime proportion of combined deep and rapid eye movement (REM) sleep, and nighttime proportion of light sleep (non-REM). Ava notes that light, deep, and REM sleep are distinguished via accelerometry, specifying that a healthy sleep pattern consists of a night that is 50% to 65% light sleep and 35% to 50% deep + REM sleep.<sup>18</sup> Similarly, to assess sleep quality, ActiGraph measures sleep efficiency which is the proportion of sleep periods that a person spends asleep (normal sleep efficiency >85%).<sup>21</sup> We expect ActiGraph's measure of normal sleep efficiency to be positively correlated to Ava's measure of a healthy sleep pattern.

Subjective sleep quality during the week of data collection was assessed using the 19-item Pittsburgh Sleep Quality Index (PSQI) after all data had been collected.<sup>22</sup> The PSQI has been validated for use among the general population,<sup>22</sup> as well as pregnant women.<sup>23</sup> The 19 items are combined to form 7 component scores (subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction). Each component score ranged from 0 ("no difficulty") to 3 ("severe difficulty"). The sum of the component scores was continuous and ranged from 0 to 16 with higher scores (referred to as global scores) indicating poorer sleep quality. A person with a



**Figure 1.** Example of Ava mobile application collection of daily mood. Participants were asked to select all moods that applied.

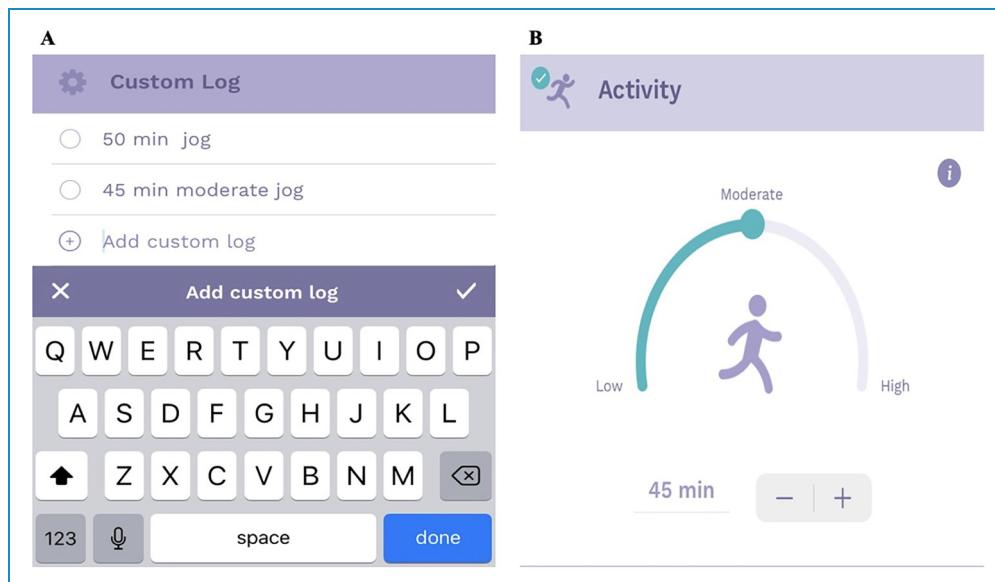
global score  $>5$  is considered to have poor sleep quality. These component measures showed adequate internal consistency (Cronbach's  $\alpha > 0.68$ ).

### Stress

Stress during the week of data collection was assessed using the Cohen 10-item perceived stress scale (PSS) after all data had been collected.<sup>24</sup> Response options for each item ranged from 0 ("never") to 4 ("very often"). Scores ranged from 0 to 40 with higher scores indicating higher stress. These measures showed excellent internal consistency (Cronbach's  $\alpha > 0.87$ ).

Given that the Ava mobile app allowed participants to report several moods other than "stressed" (Figure 1), our goal was to determine if the report of negative moods

(e.g. disinterested, anxious, sad, stressed, irritable) was correlated with a higher score on the PSS scale. Previous literature has shown higher PSS scores are positively correlated with negative emotion, anxiety, and depression.<sup>25</sup> Therefore, we anticipated women who reported more frequent negative moods to have higher PSS scores. An average mood score based on the input in the Ava mobile app was calculated for each participant. For each day, each woman began with a score of 0. Then, based on her input, she would be assigned 1 point per each negative mood listed, -1 point for each positive mood listed (e.g. happy, content), and 0 points for each neutral mood listed (e.g. sensitive, tired, exhausted, or aroused). The participant's daily totals were added together to get a weekly mood score with a higher score reflecting frequent reports of negative moods. To calculate the average mood score,



**Figure 2.** Example of Ava mobile application collection of daily physical activity. Format for physical activity input varied by pregnancy status. A denotes the format for nonpregnant women and B denotes the format for pregnant women.

the total mood score was divided by the number of days of mood data collection (average of 5.24 days per participant). This average mood score was continuous and ranged from -1.5 to 2.75, with a higher score indicating an increased frequency of negative mood reporting.

### Physical activity

Physical activity during the week of data collection was assessed using the International Physical Activity Questionnaire (IPAQ)-Short Form after all data had been collected.<sup>26</sup> This questionnaire has been validated to collect physical activity data among the general population,<sup>26</sup> and recommended for pregnant women when more valid objective measures (i.e. accelerometry) are unavailable.<sup>27</sup> Because the Ava bracelet is only worn during sleep, we were only able to compare the self-report of physical activity within the Ava mobile app (Figure 2) to the results from the IPAQ that each participant completed at the end of data collection. The IPAQ asked participants to report the frequency (times per week), duration (minutes/day), and intensity (vigorous, moderate, or walking) of any physical activity they performed during the past 7 days.

Continuous physical activity scores were calculated in units of metabolic equivalents (MET)-minutes per week and then categorized into physical activity level categories of “high,” “moderate,” or “low.”<sup>28</sup> For vigorous physical activity,  $1 \text{ MET} = 8.0$ ;  $\text{MET minutes per week}_{\text{vigorous}} = 8.0 \times \text{duration} (\text{minutes/day}_{\text{vigorous}}) \times \text{frequency} (\text{times per week}_{\text{vigorous}})$ . For moderate physical activity,  $1 \text{ MET} = 4.0$ ;  $\text{MET minutes per week}_{\text{moderate}} = 4.0 \times \text{duration} (\text{minutes/day}_{\text{moderate}}) \times \text{frequency} (\text{times per week}_{\text{moderate}})$ .

For walking physical activity,  $1 \text{ MET} = 3.3$ ;  $\text{MET minutes per week}_{\text{walking}} = 3.3 \times \text{duration} (\text{minutes/day}_{\text{walking}}) \times \text{frequency} (\text{times per week}_{\text{walking}})$ . Total MET minutes per week refers to the sum of vigorous, moderate, and walking MET minutes per week.<sup>29</sup>

To assess categorical measures of physical activity, participants were scored as “high” physical activity level if they performed vigorous physical activity on 3 or more days, with 1500 total MET minutes in the last week or if they achieved at least 3000 total MET minutes over the last week. Participants were scored as “moderate” physical activity level if they performed vigorous physical activity and/or walking on 3 or more days for at least 30 min per day or if they performed moderate physical activity and/or walking for at least 30 min per day on 5 or more days. They were also scored as “moderate” if they achieved at least 600 total MET minutes over the last week. Participants were scored as “low” physical activity level if they did not meet criteria for moderate or high physical activity.<sup>29</sup>

### Ava bracelet and mobile application acceptability

After all data had been collected, the acceptability of the Ava bracelet and mobile app to monitor reproductive health measures was assessed using a 15-item questionnaire modified from the acceptability of health apps among adolescents (AHAA) scale.<sup>30</sup> AHAA is based on the seven domains within the theoretical framework of acceptability proposed by Sekhon et al.<sup>31</sup> including affective attitude, burden, perceived effectiveness, ethicality, intervention coherence, opportunity costs, and self-efficacy. In our adapted scale, we assessed affective attitude (i.e. how did

participants feel about using the Ava bracelet and mobile app?), burden (i.e. how much time was required to use the Ava bracelet and mobile app), intervention coherence (i.e. did participants understand the use of the Ava bracelet and mobile app and how they worked?), ethicality (i.e. did participants find value in monitoring their reproductive health?), and self-efficacy (i.e. were participants confident they could use the Ava bracelet and mobile app?). Response options for each item ranged from 1 “strongly disagree” to 5 “strongly agree” on a 5-point Likert scale and items included “I am confident that I can use this app to monitor my reproductive health even if I’m not reminded to do it” and “I really enjoy using this app.” Total acceptability scores were calculated by taking the mean of responses within each of the 5 subscales and then totaling the 5 means. The total acceptability scores were continuous ranging from 8.75 to 25 with higher scores indicating

**Table 1.** Measures collected to validate the Ava bracelet.

Measures	Ava bracelet and mobile app	ActiGraph GT3X+ and polar chest strap	Validated questionnaires <sup>a</sup>
<b>Heart rate</b>			
Average heart rate (bpm)	✓	✓	--
Heart rate variability ratio <sup>b</sup>	✓	✓	--
<b>Sleep</b>			
Duration (hours)	✓	✓	--
Quality	✓	✓	✓
Self-reported mood	✓	--	✓
Self-reported physical activity	✓	--	✓

<sup>a</sup>Validated questionnaires used to evaluate sleep quality, mood, and physical activity include the Pittsburgh sleep quality index,<sup>22,23</sup> the perceived stress scale,<sup>24</sup> and the international physical activity questionnaire-short form.<sup>26</sup>

<sup>b</sup>Heart rate variability ratio in the Ava bracelet was calculated by low-frequency waves/high-frequency waves; this ratio should be correlated with calculated SD1/SD2 from the ActiGraph and Polar chest strap measures where SD1 is the Poincaré plot standard deviation perpendicular to the line of identity and SD2 is the Poincaré plot standard deviation along the line of identity.<sup>17</sup>

higher levels of acceptability. These measures showed excellent internal consistency (Cronbach’s  $\alpha > 0.94$ ). In addition to the modified AHAA, participants were also asked an open-ended question: “If given the opportunity, would you use the Ava bracelet and mobile app while trying to conceive? While you were pregnant?”

## Ava bracelet validity

For this analysis, we defined validity as the ability of the Ava bracelet to accurately measure quantitative physiological data.<sup>32</sup> Validity of the Ava bracelet included criterion and construct validity. Criterion validity was determined by comparing the Ava bracelet heart rate and sleep duration measures to the criterion measures collected from the Polar chest strap and ActiGraph GT3X+ accelerometer (Tables 1 and 2).<sup>33,34</sup>

Construct validity was determined by comparing self-reported mood and physical activity, and the HRV ratio collected in the Ava mobile app to previously validated measures of stress, physical activity, and HRV ratio. Valid Ava bracelet measures should be correlated with the construct measures. The definition of acceptability, adapted from the definition by Samkange-Zeeb et al.,<sup>35</sup> was whether women would use the Ava bracelet and mobile app to collect nighttime physiological data, mood data, and physical activity data.

## Statistical analysis

All statistical analyses were performed with Stata software (version 16; StataCorp, College Station, TX, USA). Descriptive statistics were calculated for participant characteristics and acceptability measures.

## Criterion validity

To estimate measurement error, mean absolute percentage error (MAPE) was calculated, with a smaller MAPE representing better accuracy. We interpreted a MAPE of  $< 10\%$  to indicate the Ava bracelet measure was equivalent to the Polar chest strap and ActiGraph measure. Pearson correlation coefficients were calculated and interpreted with the following ratings:  $< 0.60$  low, 0.60 to  $< 0.75$  moderate, 0.75 to  $< 0.90$  good, and  $\geq 0.90$  excellent.<sup>36</sup> Bland–Altman plots were created to visually assess bias in the collected data. Lastly, equivalence testing was used to determine whether the mean values of heart rate and sleep duration from the Ava bracelet fell within the equivalence zone from the criterion measures. The values for the equivalence zone were determined a priori based on prior validity studies assessing heart rate and sleep data collected through wearables.<sup>37,38</sup> Ninety-five percent precision was assumed if the Ava bracelets’ 90% confidence intervals were within an equivalence zone that was between 10% of the criterion (ActiGraph) mean for sleep duration and 5% of the criterion (Polar chest strap) mean for heart rate.

**Table 2.** Ava bracelet validity methods and expectations.

Ava bracelet measure	Criterion measure	Construct measure	Statistical test	Expectation
Heart rate (bpm)	ActiGraph	–	Pearson correlation; equivalence testing	Positive correlation; Ava bracelet values would fall within the 5% equivalence zone
Heart rate variability ratio	–	SD1/SD2	Pearson correlation	Negative correlation
Sleep duration (hours)	ActiGraph	–	Pearson correlation; equivalence testing	Positive correlation; Ava bracelet values would fall within the 10% equivalence zone
Sleep quality	–	ActiGraph and PSQI	Chi-squared test	No significant difference between Ava bracelet, ActiGraph, and PSQI distributions of normal sleep quality
Physical activity	–	IPAQ	Spearman rank-order correlation (activity level); Pearson correlation (MET minutes per week)	Positive correlation
Mood	–	PSS	Pearson correlation	Positive correlation

PSQI: Pittsburgh sleep quality index; IPAQ: International physical activity questionnaire; PSS: perceived stress scale; MET: metabolic equivalents.

### Construct validity

Construct validity was determined by comparing subjective measures of mood and physical activity and the HRV ratio collected in the Ava mobile app to previously validated measures of stress, physical activity, and HRV. Valid measures should be significantly correlated with the results. Correlation coefficients were calculated and interpreted with the following ratings: <0.60 low, 0.60 to <0.75 moderate, 0.75 to <0.90 good, and  $\geq 0.90$  excellent.<sup>36</sup>

## Results

### Sample characteristics

We collected data from 33 women. During the data collection period, 5 women were pregnant and 28 were nonpregnant. Most of the women in this sample identified as non-Hispanic White (70%) and over half of the women were married or in a domestic partnership (58%) (Table 3). Participants provided a total of 229 nights of data collection (mean: 6.94 nights per participant; range: 6–7 nights).

### Criterion validity of the Ava bracelet

#### Heart rate

The Ava bracelet underestimated heart rates compared to the ActiGraph (Table 4 and Figure 3). Equivalence testing

for heart rate is shown in Figure 4. The Ava bracelet did not fall within the 5% equivalence zone and the MAPE was 11.4%. The mean bias (mean difference between ActiGraph and Ava bracelet measures) was  $7.75 \pm 3.93$  bpm. The Ava bracelet heart rate measures were positively correlated with the ActiGraph heart rate measures (Pearson's  $r: 0.92$ ;  $P$ -value: 0.00) as expected.

#### Sleep

The Ava bracelet underestimated sleep duration compared to the ActiGraph (Table 4 and Figure 5). Equivalence testing for sleep duration is shown in Figure 6. The Ava bracelet fell within the 10% equivalence zone and the MAPE was 8.5%. The mean bias (mean difference between ActiGraph and Ava bracelet measures) was  $0.18 \pm 0.88$  h. The Ava bracelet sleep duration measures were positively correlated with the ActiGraph sleep duration measures (Pearson's  $r: 0.73$ ;  $P$ -value: 0.00) as expected.

### Construct validity of the Ava bracelet

#### Heart rate

The Ava bracelet HRV ratio had significant low negative correlation with SD1/SD2 ( $r: -0.28$ ;  $P$ -value: 0.00) as expected (Table 5).

**Table 3.** Ava bracelet usability and feasibility study participant characteristics.

Participant characteristics	Baseline (N = 33)
Age, % (n)	
18–24 years	27.3 (9)
25–34 years	57.6 (19)
35–44 years	15.2 (5)
Race, % (n)	
White	69.7 (23)
Black or African American	3.0 (1)
Asian	21.1 (7)
Other	3.0 (1)
Married/domestic partnership, % yes (n)	57.6 (19)
Employed, % yes (n)	69.7 (23)
Yearly household income level, % yes (n)	
<\$25,000	24.4 (8)
\$25,000 to <\$50,000	3.0 (1)
\$50,000 to <\$75,000	9.1 (3)
\$75,000 to <\$100,000	12.1 (4)
\$100,000 to <\$150,000	24.2 (8)
≥ \$150,000	27.3 (9)
Nonpregnant, % yes (n)	84.8 (28)
Nulliparous	67.9 (19)
Pregnant, % yes (n)	15.2 (5)
Nulliparous	0.2 (1)
Body mass index, mean (SD)	24.7 (5.2)
Currently wears a wearable, % yes (n)	66.7 (22)
Apple Watch	63.6 (14)
Fitbit	36.4 (8)

### Sleep

Using the Ava bracelet definition of a healthy sleep pattern, 49.8% of nights were considered to represent a normal sleep

quality; whereas, using the ActiGraph definition, 81.2% of the nights were considered normal (Table 5). There was no meaningful difference between the Ava bracelet and ActiGraph measures of sleep quality (*P*-value: 0.31) as expected. The PSQI identified 39.7% of participants with good sleep quality. There was no meaningful difference between the Ava bracelet and PSQI measures of sleep quality (*P*-value: 0.37) as expected.

### Stress

The Ava bracelet mood scores had a significantly low positive correlation with PSS scores (*r*: 0.39; *P*-value: 0.03) as expected (Table 5).

### Physical activity

The Ava bracelet physical activity level had a significant low positive correlation with IPAQ physical activity level (*r*: 0.56; *P*-value: 0.00) and a significant moderate positive correlation with IPAQ MET minutes/week (*r*: 0.71; *P*-value: 0.00) (Table 5).

### Acceptability of the Ava bracelet

Using the AHAA, participants reported high total acceptability scores (mean:  $19.3 \pm 4.2$  out of a possible 25) with a range of 8.75 to 25. Among the acceptability subscales assessed (e.g. affective attitude, burden, coherence, ethicality, and self-efficacy), participants had the highest scores under ethicality (mean:  $4.8 \pm 0.5$  out of a possible 5) and the lowest scores under coherence (mean:  $2.8 \pm 1.4$  out of a possible 5) (Table 6). Ninety-four percent (*n* = 31) of participants reported, if given the opportunity, they would use Ava bracelet and mobile app prior to conception as a fertility tool; whereas only 52% (*n* = 17) of participants said they would use it during their pregnancy.

### Discussion

This study explored Ava bracelet validity and acceptability in a cohort of women aged 18 to 44 years. It compared measures of heart rate, sleep, physical activity, and mood from the Ava bracelet and the corresponding mobile app to the criterion measures of the ActiGraph GT3X+ and Polar chest strap and construct measures using validated questionnaires. Results indicated, in free-living environments, the Ava bracelet may estimate sleep duration with a moderate level of accuracy. Whereas, the heart rate estimates (beats per minute) from the Ava bracelet may show large errors, as the Ava bracelet tends to underestimate heart rate. Subjective Ava bracelet measures (HRV ratio, sleep quality, physical activity, and mood) had low to moderate correlation with validated construct measures. These

**Table 4.** Criterion validity of Ava bracelet compared to ActiGraph GT3X+ and polar chest strap (criterion measure).

Measures	n <sup>a</sup>	Ava bracelet	ActiGraph GT3X+ and polar chest strap	Pearson's <i>r</i> ( <i>P</i> -value)	MAPE <sup>b</sup> (%)	Mean bias (SD)	Limits of agreement <sup>c</sup>
Heart rate (bpm) <sup>a</sup>	149	58.9 (8.1)	66.7 (9.9)	0.92 (0.00)	11.4	7.75 (3.93)	0.05–15.46
Sleep duration (hours), mean (SD) <sup>a</sup>	211	6.8 (1.3)	7.0 (1.0)	0.73 (0.00)	8.5	0.18 (0.88)	–1.54–1.89

<sup>a</sup>*n* represents the number of nights of recorded data.

<sup>b</sup>MAPE represents mean absolute percent error: absolute value of (ActiGraph – Ava bracelet)/ActiGraph × 100.

<sup>c</sup>95% limits of agreement.

findings would be valuable to researchers and clinicians aiming to use the Ava bracelet to collect physiological and lifestyle measures among pregnant and nonpregnant women, outside of a clinical setting.

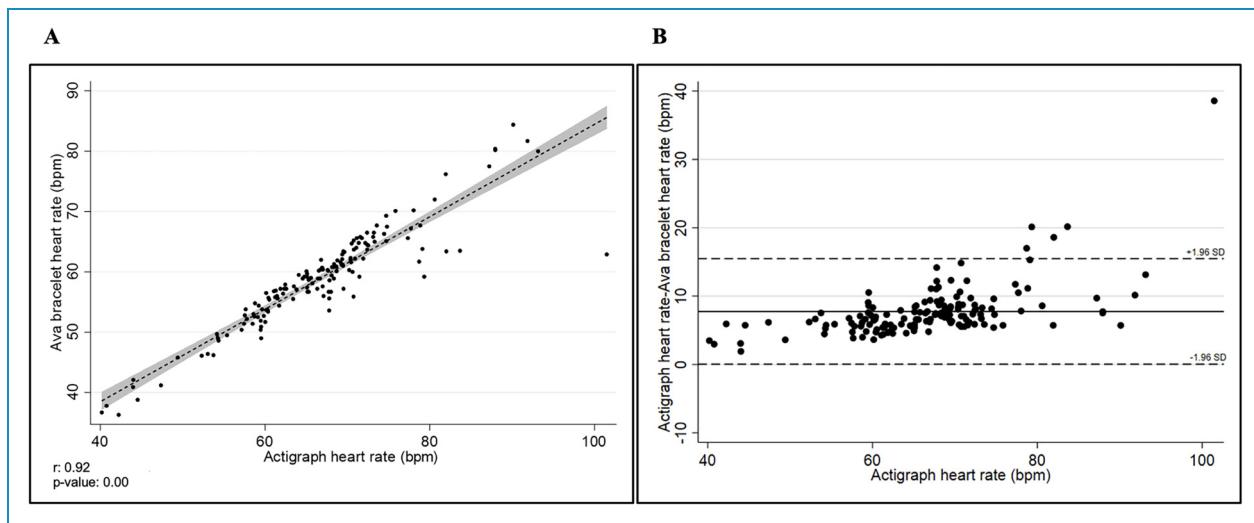
The Ava bracelet heart rate measures (beats per minute) had an excellent correlation to the ActiGraph + Polar chest strap (*r*: 0.92), an unacceptable MAPE value ( $\geq 10\%$ ), and estimates that were outside of the equivalence region. This finding contrasts with a previous study conducted in a laboratory setting that concluded the Ava bracelet, in comparison to polysomnography, provided an accurate measure of heart rate during sleep (MAPE: 0.20%).<sup>39</sup> Other validity studies have also found consumer-grade wearables, including the Apple Watch and Fitbit, were able to provide accurate measures of heart rate during sleep.<sup>40–42</sup> Given these findings, it does not appear that the Ava bracelet is suitable for research aiming to collect heart rate among pregnant or nonpregnant women in free-living conditions. Additionally, Ava bracelet estimates of HRV ratio had a low negative correlation with SD1/SD2. While the finding of a significant negative correlation aligns with our initial hypothesis, the low correlation may be due to the choice to use SD1/SD2 as the construct measure, as typically studies evaluate HRV ratio in comparison to the gold standard electrocardiogram.<sup>43</sup> Thus, while it appears the Ava bracelet's measure of the HRV ratio is correlated with the construct measure, further clinical testing is needed. We are unable to determine its clinical accuracy due to our limited study resources.

The Ava bracelet sleep duration measures had a moderate correlation to the ActiGraph (*r*: 0.73), an acceptable MAPE value ( $< 10\%$ ), and estimates that overlapped the equivalence region. These results are inconsistent with previous validity studies of consumer-grade wearables, as most have found that consumer-grade wearables generally overestimate sleep time for both adults and children.<sup>36,44,45</sup> Despite the acceptable MAPE value, one must also consider the clinical implications of these results. The limit of agreement indicated the Ava bracelet could underestimate sleep duration up to 1.54 h and overestimate up to 1.89 h. The American Academy of Sleep Medicine and Sleep Research Society recommends a healthy adult sleep 7 h

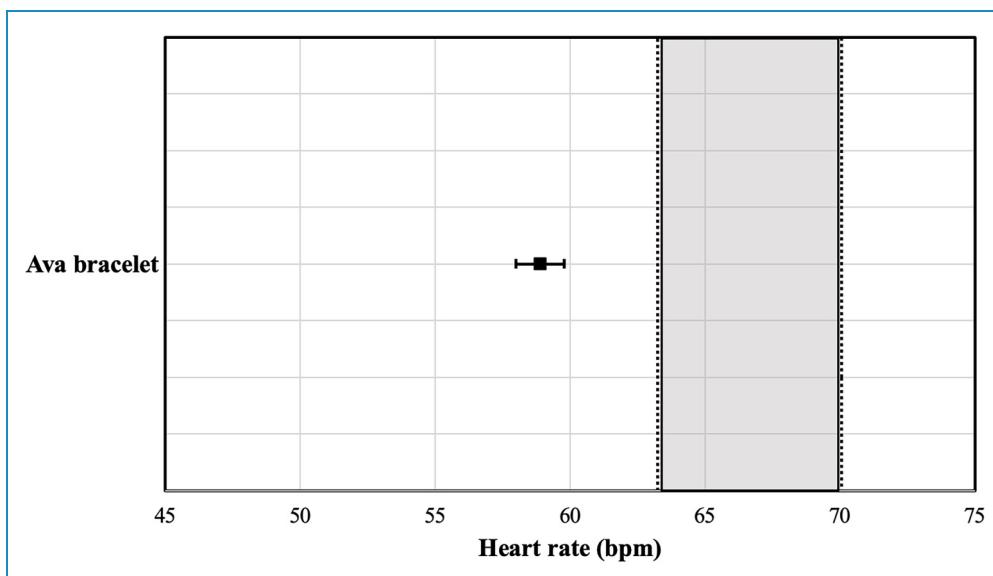
per night for optimal health, noting that  $< 7$  h of sleep is associated with risk of obesity, diabetes, hypertension, heart disease, stroke, depression, and death.<sup>46</sup> Studies analyzing sleep and fertility on pregnancy outcomes have identified poor outcomes (i.e. reduced fecundability, preterm birth, gestational diabetes, and hypertension) with less than 7 h or over 9 h of sleep.<sup>47–50</sup> Therefore, the Ava bracelet's lack of precision by almost 2 h, could lead to misclassification bias, influencing overall results on sleep, fertility, and pregnancy research.

There were no meaningful differences between Ava bracelet estimates of sleep quality compared to the sleep quality estimates from the ActiGraph. Given that both the Ava bracelet and the ActiGraph use accelerometry data to discern sleep quality,<sup>18</sup> it appears that the ability to assess sleep time versus wake time is similar in both devices. The Ava bracelet sleep quality estimates were also similar to data collected using the construct measure, the PSQI. Previous studies have found PSQI measures are correlated with objective sleep data from the gold standard, polysomnography,<sup>51</sup> and from consumer-grade wearables such as the Garmin vívofit 3®.<sup>52</sup> Another study among college-age individuals found data from accelerometer-based wearable trackers to be similar to data collected from a subjective sleep diary.<sup>37</sup> Thus, researchers may be able to substitute the Ava bracelet for the PSQI to collect measures of sleep quality.

The subjective measures of mood input by participants in the Ava mobile app had a significantly low correlation with data obtained from the PSS. The low correlation is likely due to the way we scored mood reported in the Ava mobile app. The calculated mood score was a reflection of all reported moods in the past week—not only stress. The PSS score is a retrospective reflection of the participant's stress level throughout the past week, whereas the Ava mobile app provides the participant's moods on the day each mood was experienced. While results indicate the mood function on the Ava mobile app yields a poor measure of stress, the app may be a useful way for women to report their daily mood, as 94% (*n* = 31) of participants reported at least one day of moods in their app. This could provide researchers with a better way to prospectively evaluate mood during



**Figure 3.** (A) The correlation between the heart rate measured via the ActiGraph and the heart rate measured via the Ava bracelet. (B) Bland Altman plot illustrating the bias between the ActiGraph and the Ava bracelet heart rate measurements, plotted against the ActiGraph (gold standard) heart rate measurements.

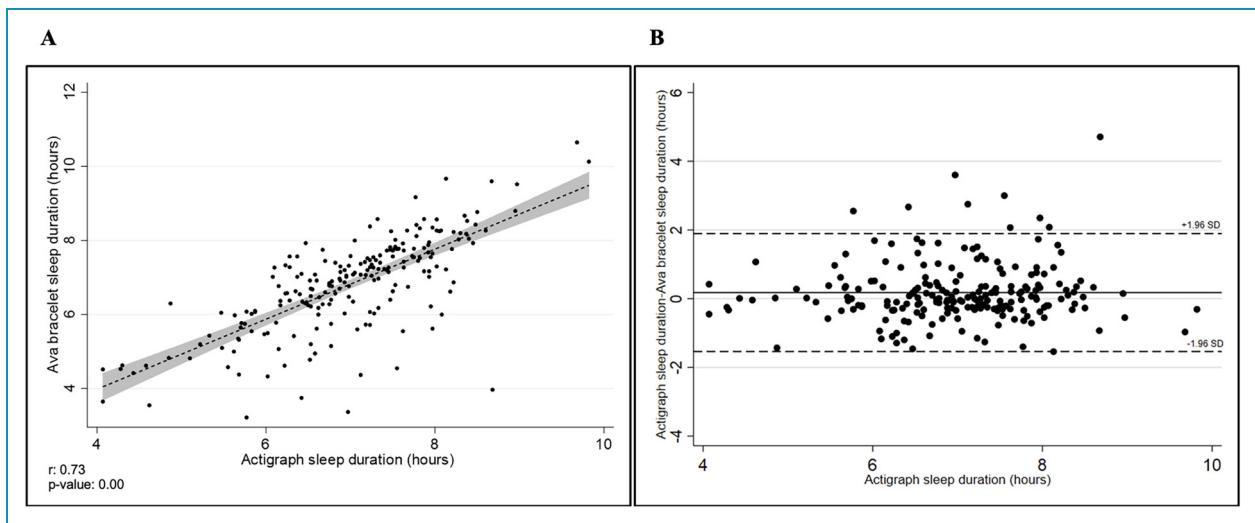


**Figure 4.** Equivalence testing for heart rate. Shaded gray area indicates the proposed equivalence zone ( $\pm 5\%$  of the ActiGraph heart rate mean). Dark bar indicates the 90% confidence interval for the mean heart rate estimated by the Ava bracelet.

the perinatal period, including preconception and postpartum, as women battling infertility during preconception and women in their postpartum period experience higher rates of depression and other mood disorders compared to the general population.<sup>53,54</sup>

The subjective measures of physical activity recorded in the Ava mobile app had a significant low to moderate correlation with data obtained from the IPAQ. The Ava mobile app physical activity data had a stronger correlation with the estimated MET minutes per week versus classification of physical activity level (i.e. low, moderate, or high).

Similar to the PSS, the IPAQ provides a retrospective reflection of the participant's physical activity (i.e. vigorous, moderate, and walking) over the past week. For pregnant women, the Ava mobile app collects similar data as there is a specific area for women to record the duration and intensity of the physical activity they performed each day. However, if a pregnant woman engages in both vigorous and moderate activity on the same day, there is no way for her to denote that in the app, as the app only allows women to enter 1 physical activity intensity each day. Conversely, for nonpregnant women, there is not a



**Figure 5.** (A) The correlation between the sleep duration measured via the ActiGraph and the sleep duration measured via the Ava bracelet. (B) Bland Altman plot illustrating the bias between the ActiGraph and the Ava bracelet sleep duration measurements, plotted against the ActiGraph (gold standard) sleep duration measurements.

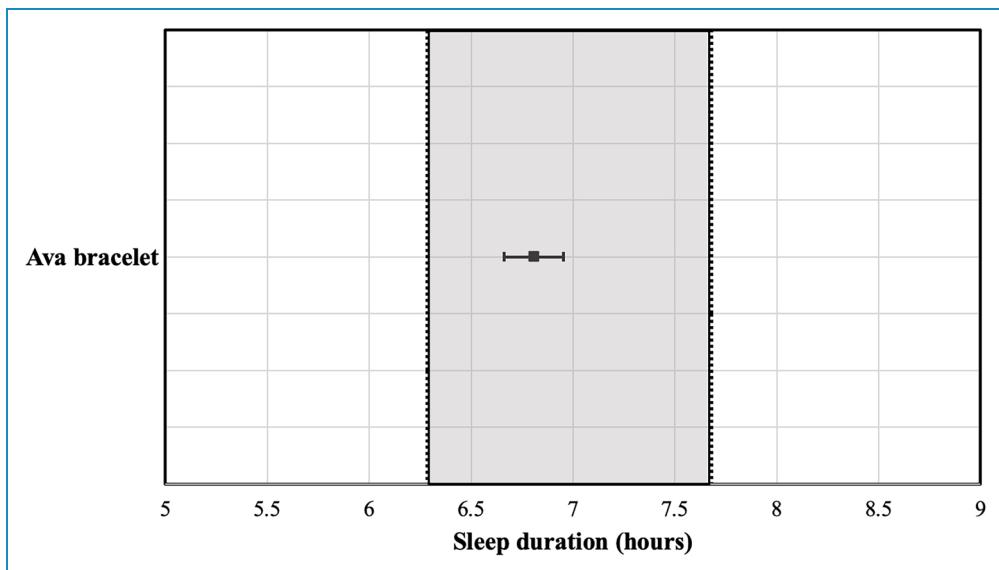
designated area of the Ava mobile app for recording physical activity. To collect this information, participants were instructed to enter it manually in their custom log. While this method yielded higher quality data as participants were able to record several physical activity types, durations, and intensities for each day, it appears that this method of data collection placed a heavy burden on participants as 18% ( $n=6$ ) did not record any physical activity data in their mobile app during the study period. Based on our findings, it would appear that the Ava mobile app would not be a preferred method of collecting physical activity data among nonpregnant women tracking fertility. The Ava mobile app may be a useful way to prospectively collect self-reported physical activity data among pregnant women. However, given the reporting limitations within the app, researchers aiming to use consumer-grade wearables to collect physical activity data should consider how a collection of subjective physical activity data and the inability to report multiple physical activity intensities may impact their results.

Lastly, the acceptability of the Ava bracelet for fertility tracking was high (94%), while acceptability for pregnancy tracking was significantly lower (52%). This difference in acceptability between periods of fertility versus pregnancy tracking may signal that women see utility in monitoring reproductive health data solely as a means to actively prevent or become pregnant, making the need to track reproductive health data obsolete during pregnancy itself. Among the six subscales of acceptability,<sup>30</sup> participants had high affective attitude and ethicality scores and low coherence scores. These results indicate that while participants cared about and supported ways to monitor their reproductive health, there was a lack of understanding

around how to successfully use the Ava bracelet and the mobile app. Therefore, researchers aiming to use the Ava bracelet and mobile app in future studies should implement a thorough training session that introduces participants to the Ava bracelet and mobile app, allowing them to feel confident using the equipment prior to the start of data collection.

### Strengths and limitations

To our knowledge, this is the first consumer-grade wearable study to explore the validity and acceptability of a consumer-grade wearable among a sample of reproductive-aged women. Additionally, this is the first study to evaluate the validity of the sleep, physical activity, and mood measures obtained from the Ava bracelet and mobile app. This study had excellent compliance with a mean number of 6.9 nights of data collection per participant. While this study had many strengths, there were also notable limitations. First, our small sample size of pregnant participants ( $n=5$ ) limited our statistical power to test differences between the pregnant and nonpregnant participants, so results presented should be considered exploratory. While we were able to recruit more than 30 participants, we had hoped to recruit a sample comprised of 50% pregnant and 50% nonpregnant women, however, we were unable to achieve that distribution ( $n=5$  pregnant;  $n=28$  nonpregnant). Given that recruitment took place in the fall of 2020, during the height of the COVID-19 pandemic, it is likely that pregnant women were not as willing to engage in a research study due to concerns over social distancing protocols and potential COVID-19 exposures. Second, the acceptability of the Ava bracelet



**Figure 6.** Equivalence testing for sleep duration. The shaded gray area indicates the proposed equivalence zone ( $\pm 10\%$  of the ActiGraph sleep duration mean). The dark bar indicates the 90% confidence interval for the mean sleep duration estimated by the Ava bracelet.

**Table 5.** Construct validity of Ava bracelet compared to construct measures.<sup>a</sup>

Measures	n <sup>b</sup>	Ava bracelet	Construct measures	Correlation coefficient <sup>c</sup> (P-value)	Expected outcome?
Heart rate variability ratio	149	0.4 (0.1)	SD1/SD2: 0.2 (0.1)	-0.28 (0.00)*	Yes
Healthy sleep quality, % yes (n)	213	49.8 (114)	PSQI: 39.7 (13)	PSQI: - <sup>d</sup> (0.37)	Yes
			ActiGraph: 81.2 (173)	ActiGraph: - <sup>d</sup> (0.31)	Yes
Physical activity level, <sup>e</sup> % (n)			IPAQ-SF		No
High		3.7 (1)	24.2 (8)		
Moderate	148	44.4 (12)	48.4 (16)	0.56 (0.00)*	
Low		51.9 (14)	27.3 (9)		
MET minutes per week, mean (SD)	148	698.57 (569.47)	IPAQ-SF: 1492.1 (1304.7)	0.71 (0.00)*	Yes
Mood score, mean (SD) <sup>f</sup>	173	-0.15 (0.85)	PSS: 13.1 (6.8)	0.39 (0.03)*	Yes

PSQI: Pittsburgh sleep quality index; IPAQ-SF: international physical activity questionnaire-short form; PSS: perceived stress scale; SD: standard deviation; LF/HF: low frequency to high frequency.

<sup>a</sup>Construct measures: SD1/SD2 (a measure of the unpredictability of the R-R intervals correlated with the LF/HF ratio); PSQI; IPAQ-SF; PSS.

<sup>b</sup>n represents the number of nights data was collected using the Ava bracelet. All participants completed the additional questionnaires at the final assessment, therefore the completion rate for the construct measures was 33/33 participants.

<sup>c</sup>Correlation coefficients were determined using spearman rank-order correlation for ordinal variables and Pearson correlation for continuous variables.

<sup>d</sup>Statistical significance was determined using a chi-square test for dichotomous categorical variables.

<sup>e</sup>Six participants did not record any physical activity in the Ava mobile app.

<sup>f</sup>Two participants did not record any mood data in the Ava mobile app.

\*P<0.05.

**Table 6.** Acceptability of Ava bracelet and mobile app.

Measures of acceptability	N = 33
Would use the Ava bracelet, % yes	
Fertility	94.0 (31)
Pregnancy	51.5 (17)
Adapted AHAA scale	
Total acceptability score, <sup>a</sup> mean (SD)	19.3 (4.2)
AHAA subscales, mean (SD)	
Affective attitude	4.1 (1.0)
Burden	3.9 (1.2)
Coherence	2.8 (1.4)
Ethicality	4.8 (0.5)
Self-efficacy	3.7 (1.1)

AHAA: acceptability of health apps among adolescents; SD: standard deviation.

<sup>a</sup>Total acceptability scores were calculated using an adaptation of the AHAA scale by taking the mean of responses within each of the five subscales (i.e. affective attitude, burden, coherence, ethicality, and self-efficacy) and then totaling the five means.<sup>30</sup>

was especially high for tracking fertility. This finding may be influenced by our small sample size and the high prevalence of current wearable use within our study sample ( $n = 22$ , 67%). Therefore, the acceptability of the Ava bracelet within our sample may not be representative of the entire population of reproductive-aged women. Third, to conduct this study in free-living conditions, we were unable to use gold standard measurements such as an electrocardiogram or polysomnography as our criterion and construct measures. For example, heart rate and HRV are most reliably assessed using electrocardiograms; yet due to resource limitations, we used the Polar chest strap and SD1/SD2 (a correlate of HRV). Thus, the conclusions drawn from the data were limited. Additionally, the method of wrist actigraphy to collect sleep data may introduce misclassification bias as it determines sleep duration based on lack of motion.<sup>33</sup> For example, if a participant was awake, but motionless, their sleep duration would be overestimated. However, the limitations of wrist actigraphy are likely non-differential due to the use of accelerometry in both the Ava bracelet and the ActiGraph to measure sleep. Finally, the Ava bracelet can also collect data on skin temperature, respiratory rate, and provide information on ovulation status. However, we did not evaluate any of those

measures in this study and as a result, are unable to comment on the validity of those metrics.

## Conclusion

The Ava bracelet and ActiGraph were comparable to one another in assessing sleep duration and quality, but not heart rate in free-living conditions. However, it should be noted that a lack of precision within the sleep duration estimates may lead to misclassification bias in sleep research. Due to increased potential for sleep duration misclassification and large errors with heart rate measurement when using the Ava bracelet, we recommend that researchers continue using the gold standard ActiGraph, while further clinical testing is needed to validate the HRV measures. Self-reported measures of mood and physical activity collected in the Ava mobile app had a low to moderate correlation with construct measures, yet compliance with mood data input was high. So, we recommend researchers consider using a mobile app, such as Ava, to prospectively evaluate mood during the perinatal period. Acceptability of the Ava bracelet was high for tracking fertility and lower for tracking pregnancy. Given our small sample size and potential for selection bias, future research should aim to address this discrepancy through feasibility and usability studies.

## Contributorship

AKN researched literature and conceived the study. EC and ALT were involved in protocol development and gaining ethical approval. EC was involved in participant recruitment. AKN analyzed the data and wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version.

**Declaration of Conflicting Interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval:** The Institutional Review Board of Our University reviewed and approved this study protocol (IRB No. 20-1732).

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Innovate Carolina, University of North Carolina at Chapel Hill (Dreamers-Who-Do Award and Harriet J. Kupferer Award).

**Guarantor:** AKN

**Informed Consent:** Not applicable, because this article does not contain any studies with human or animal subjects.

**ORCID iDs:** Alison K. Nulty  <https://orcid.org/0000-0002-5191-1336>  
 Elizabeth Chen  <https://orcid.org/0000-0002-7477-694X>

**Trial Registration:** Not applicable, because this article does not contain any clinical trials.

## References

- Vogels E. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center, <https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/> (2020, accessed 15 June 2020)
- de Zambotti M, Cellini N, Menghini L, et al. Sensors capabilities, performance, and use of consumer sleep technology. *Sleep Med Clin* 2020; 15: 1–30.
- Wright SP, Hall Brown TS, Collier SR, et al. How consumer physical activity monitors could transform human physiology research. *Am J Physiol Regul Integr Comp Physiol* 2017; 312: R358–R367.
- Runkle J, Sugg M, Boase D, et al. Use of wearable sensors for pregnancy health and environmental monitoring: Descriptive findings from the perspective of patients and providers. *Digit Health* 2019; 5: 2055207619828220.
- Grym K, Niela-Vilén H, Ekholm E, et al. Feasibility of smart wristbands for continuous monitoring during pregnancy and one month after birth. *BMC Pregnancy Childbirth* 2019; 19:34.
- Kawajiri M, Nakamura Y, Yoshida M, et al. Acceptability and feasibility of a sedentary behavior reduction program during pregnancy: A semi-experimental study. *Healthcare (Basel)* 2020; 8.
- Galea JT, Ramos K, Coit J, et al. The use of wearable technology to objectively measure sleep quality and physical activity among pregnant women in urban Lima, Peru: a pilot feasibility study. *Matern Child Health J* 2020; 24: 823–828.
- Mumford SL, Johnstone E, Kim K, et al. A prospective cohort study to evaluate the impact of diet, exercise, and lifestyle on fertility: design and baseline characteristics. *Am J Epidemiol* 2020; 189: 1254–1265.
- Choi J, Lee JH, Vittinghoff E, et al. Mhealth physical activity intervention: A randomized pilot study in physically inactive pregnant women. *Matern Child Health J* 2016; 20: 1091–1101.
- Darvall JN, Wang A, Nazeem MN, et al. A pedometer-guided physical activity intervention for obese pregnant women (the Fit MUM study): Randomized feasibility study. *JMIR Mhealth Uhealth* 2020; 8: e15112.
- Finkelstein EA, Haaland BA, Bilger M, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): A randomised controlled trial. *Lancet Diabetes Endocrinol* 2016; 4: 983–995.
- Huberty JL, Buman MP, Leiferman JA, et al. Trajectories of objectively-measured physical activity and sedentary time over the course of pregnancy in women self-identified as inactive. *Prev Med Rep* 2016; 3: 353–360.
- Shilah M, Goodale BM, Falco L, et al. Modern fertility awareness methods: Wrist wearables capture the changes in temperature associated with the menstrual cycle. *Biosci Rep* 2018; 38.
- Shilah M, Clerck V de, Falco L, et al. Pulse rate measurement during sleep using wearable sensors, and its correlation with the menstrual cycle phases, a prospective observational study. *Sci Rep* 2017; 7: 1294.
- Shilah M, Kübler F, Goodale B and Goodale B, Capturing the physiological characteristics of early pregnancy using wrist worn wearables. Poster presented at: Annual European Society of Human Reproduction and Embryology Meeting; July 1–4, 2018; Barcelona, Spain. Published online June 25, 2018. doi:10.26226/.morressier.5af300b1738ab10027aa9772.
- Kwak SG and Kim JH. Central limit theorem: the cornerstone of modern statistics. *Korean J Anesthesiol* 2017; 70: 144–156.
- Shaffer F and Ginsberg JP. An overview of heart rate variability metrics and norms. *Front Public Health* 2017; 5: 258.
- Ava Science Inc. What your Ava data means, <https://www.avawomen.com/avaworld/ava-data-means/> (2017, accessed 19 November 2020)
- Lanfranchi PA and Somers VK. Cardiovascular physiology. In: *Principles and practice of sleep medicine*. Amsterdam: Elsevier; 2011. 226–236.
- Dimitriev DA, Saperova EV and Dimitriev AD. State anxiety and nonlinear dynamics of heart rate variability in students. *PLoS One* 2016; 11: e0146131.
- Chen X, Velez JC, Barbosa C, et al. Evaluation of actigraphy-measured sleep patterns among children with disabilities and associations with caregivers' educational attainment: results from a cross-sectional study. *BMJ Open* 2015; 5: e008589.
- Buyssse DJ, Reynolds CFIII, Monk TH, et al. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res* 1989; 28: 193–213.
- Qiu C, Gelaye B, Zhong Q-Y, et al. Construct validity and factor structure of the Pittsburgh sleep quality index among pregnant women in a pacific-northwest cohort. *Sleep Breath* 2016; 20: 293–301.
- Lee E-H. Review of the psychometric evidence of the perceived stress scale. *Asian Nurs Res (Korean Soc Nurs Sci)* 2012; 6: 121–127.
- Spada MM, Nikčević AV, Moneta GB, et al. Metacognition, perceived stress, and negative emotion. *Pers Individ Dif* 2008; 44: 1172–1181.
- Lee PH, Macfarlane DJ, Lam TH, et al. Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *Int J Behav Nutr Phys Act* 2011; 8: 115.
- Sanda B, Vistad I, Haakstad LAH, et al. Reliability and concurrent validity of the international physical activity questionnaire short form among pregnant women. *BMC Sports Sci Med Rehabil* 2017; 9: 7.
- Sjostrom M, Ainsworth B.E., Bauman A., Bull F.C., Hamilton-Craig C.R., and Sallis J.F (2005) Guidelines for data processing analysis of the International Physical Activity Questionnaire (IPAQ) - Short and long forms.
- Forde C. (2016) Scoring the International Physical Activity Questionnaire (IPAQ). Retrieved June 16, 2021 from [https://ugc.futurelearn.com/uploads/files/bc/c5/bcc53b14-ec1e-4d90-88e3-1568682f32ae/IPAQ\\_PDF.pdf](https://ugc.futurelearn.com/uploads/files/bc/c5/bcc53b14-ec1e-4d90-88e3-1568682f32ae/IPAQ_PDF.pdf).

30. Chen E. Development and validation of a new scale to measure the acceptability of mobile health applications among adolescents. Dissertation. University of North Carolina at Chapel Hill; 2019. Accessed December 19, 2020. <https://cdr.lib.unc.edu/concern/dissertations/2b88qh60v>.
31. Sekhon M, Cartwright M and Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Serv Res* 2017; 17: 88.
32. Heale R and Twycross A. Validity and reliability in quantitative studies. *Evid Based Nurs* 2015; 18: 66–67.
33. Martin JL and Hakim AD. Wrist actigraphy. *Chest* 2011; 139: 1514–1527.
34. Pasadyn SR, Soudan M, Gillinov M, et al. Accuracy of commercially available heart rate monitors in athletes: A prospective study. *Cardiovasc Diagn Ther* 2019; 9: 379–385.
35. Samkange-Zeeb F, Ernst SA, Klein-Ellinghaus F, et al. Assessing the acceptability and usability of an internet-based intelligent health assistant developed for use among Turkish migrants: Results of a study conducted in Bremen, Germany. *Int J Environ Res Public Health* 2015; 12: 15339–15351.
36. Evenson KR and Spade CL. Review of validity and reliability of Garmin activity trackers. *J Measurement Phys Behav* 2020; 3: 170–185.
37. Lee J-M, Byun W, Keill A, et al. Comparison of wearable trackers' ability to estimate sleep. *Int J Environ Res Public Health* 2018; 15.
38. Thiebaud RS, Funk MD, Patton JC, et al. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *Digit Health* 2018; 4: 2055207618770322.
39. Renevey P, Delgado-Gonzalo R, Lemkadem A, et al. Respiratory and cardiac monitoring at night using a wrist wearable optical system. *Conf Proc IEEE Eng Med Biol Soc* 2018; 2018: 2861–2864.
40. Nelson BW and Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. *JMIR Mhealth Uhealth* 2019; 7: e10828.
41. Nelson BW, Low CA, Jacobson N, et al. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *npj Digital Med* 2020; 3: 90.
42. Zhang Y, Weaver RG, Armstrong B, et al. Validity of wrist-worn photoplethysmography devices to measure heart rate: a systematic review and meta-analysis. *J Sports Sci* 2020; 38: 2021–2034.
43. Georgiou K, Larentzakis AV, Khamis NN, et al. Can wearable devices accurately measure heart rate variability? A systematic review. *Folia Med (Plovdiv)* 2018; 60: 7–20.
44. Mouritzen NJ, Larsen LH, Lauritzen MH, et al. Assessing the performance of a commercial multisensory sleep tracker. *PLoS One* 2020; 15: e0243214.
45. Evenson KR, Goto MM and Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015; 12: 159.
46. Watson NF, Badr MS, Belenky G, et al. Recommended amount of sleep for a healthy adult: A joint consensus statement of the American Academy of sleep medicine and sleep research society. *Sleep* 2015; 38: 843–844.
47. Williams MA, Miller RS, Qiu C, et al. Associations of early pregnancy sleep duration with trimester-specific blood pressures and hypertensive disorders in pregnancy. *Sleep* 2010; 33: 1363–1371.
48. Shi F, Liu C, Liu K, et al. Female and male sleep duration in association with the probability of conception in two representative populations of reproductive age in US and China. *Sleep Med* 2020; 74: 9–17.
49. Wang L and Jin F. Association between maternal sleep duration and quality, and the risk of preterm birth: A systematic review and meta-analysis of observational studies. *BMC Pregnancy Childbirth* 2020; 20: 125.
50. Myoga M, Tsuji M, Tanaka R, et al. Impact of sleep duration during pregnancy on the risk of gestational diabetes in the Japan environmental and children's study (JECS). *BMC Pregnancy Childbirth* 2019; 19: 483.
51. Martynas Z, Vanda L, Ramune G, et al. Arousals and macrostructure of sleep: Importance of NREM stage 2 reconsidered. *Sleep Sci* 2013; 6: 91.
52. Miyaji T, Kawaguchi T, Azuma K, et al. Patient-generated health data collection using a wearable activity tracker in cancer patients—a feasibility study. *Support Care Cancer* 2020; 28: 5953–5961.
53. Kiani Z, Simbar M, Hajian S, et al. The prevalence of depression symptoms among infertile women: A systematic review and meta-analysis. *Fertil Res Pract* 2021; 7: 6.
54. Bauman BL, Ko JY, Cox S, et al. Vital signs: postpartum depressive symptoms and provider discussions about perinatal depression—United States, 2018. *MMWR Morb Mortal Wkly Rep* 2020; 69: 575–581.

**Appendix. Social media advertisement used for study recruitment**

# **Looking for women to wear the Ava Bracelet!**



This research study is investigating the feasibility and usability of the Ava Bracelet, an FDA-registered fertility tracking wearable.

## **Participants will:**

- Wear the Ava Bracelet for 7 nights
- Complete a 30-minute interview at the end of the study
- Complete a questionnaire at the end of the study

## **Participants must:**

- Be reproductive-aged, pregnant or non-pregnant, women (ages 18-45)
- Have access to a smart phone
- Not wear a pacemaker

**Participants will receive a \$30 Amazon giftcard after the study!**

**Social distancing protocol will be followed.**