Title: Concurrent and Prospective Associations Between Fitbit Wearable Derived RDoC Arousal and Regulation Constructs and Adolescent Internalizing Symptoms

Authors: Benjamin W. Nelson, Ph.D.[a], Jessica E. Flannery, Ph.D.[a], Natasha Duell, Ph.D.[a], John Flournoy, Ph.D.[b], Mitchell J. Prinstein, Ph.D.[a], & Eva Telzer, Ph.D.[a]

Affiliations:

a. Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

b. Department of Psychology, Harvard University, Cambridge, MA, USA

Corresponding Author:

Benjamin W. Nelson

Department of Psychology and Neuroscience

235 E. Cameron Ave

Chapel Hill, NC 27599 USA

Email: bwn@unc.edu

Phone: (310) 801-4595

Abstract

Background: Adolescence is characterized by alterations in biobehavioral functioning, during which individuals are at heightened risk for onset of psychopathology, particularly internalizing disorders. Researchers have proposed using digital technologies to index daily biobehavioral functioning, yet there is a dearth of research examining how wearable metrics are associated with mental health.

Methods: We preregistered analyses using the Adolescent Brain Cognitive Development Study dataset using wearable data collection in 5,686 adolescents (123,862 person days or 2,972,688 person hours) to determine whether wearable indices of resting heart rate (RHR), step count, and sleep duration as well as variability in these measures were cross-sectionally associated with internalizing symptomatology. All models were also run controlling for age, sex, body mass index, socioeconomic status, and race. We then performed prospective analyses on a subset of this sample (n = 143) across 25 months that had Fitbit data available at Baseline and Follow Up in order to explore directionality of effects.

Results: Cross-sectional analyses revealed a small, yet significant effect size (R^2=0.053) that higher RHR, lower step count and step count variability, and greater variability in sleep duration were associated with greater internalizing symptoms. Cross-lagged panel model analysis revealed that there were no prospective associations between wearable variables and internalizing symptoms (partial R^2=0.026), but greater internalizing symptoms and higher RHR predicted lower step count 25 months later (partial R^2=0.010), while higher RHR also predicted lower step count variability 25 months later (partial R^2=0.008).

Conclusions: Findings indicate that wearable indices concurrently associate with internalizing symptoms during early adolescence, while a larger sample size is likely required to accurately

assess prospective or directional effects between wearable indices and mental health. Future research should capitalize on the temporal resolution provided by wearable devices to determine the intensive longitudinal relations between biobehavioral risk factors and acute changes in mental health.

Keywords: Adolescence; Fitbit; Heart Rate; Internalizing Symptoms; Sleep; Steps; Wearables

Abbreviations:

ABCD- Adolescent Brain Cognitive Development

BMI- Body Mass Index

Bpm- beats per minute

CBCL- Child Behavior Checklist

CLPM- Cross-Lagged Panel Model

EMA- Ecological Momentary Assessment

HR- Heart Rate

HRV- Heart Rate Variability

mHealth- Mobile Health

PPG- photoplethysmography

RDoC- Research Domain Criteria

RHR- Resting Heart Rate

SD- Standard Deviation

SES- Socioeconomic Status

Introduction

Adolescence is a developmental period characterized by alterations in affective functioning, during which youth are at heightened risk for the onset of internalizing disorders, such as depression and anxiety (Dahl et al., 2018; Zisook et al., 2007). Increased risk during this time coincides with developmental changes in biobehavioral functioning, including increased stress sensitivity, onset of sleep difficulties, and changes in behavioral patterning (Dahl et al., 2018). Recently, researchers have proposed the use of mobile health (mHealth) technologies, such as wearable and smartphone devices, to passively index mental health in everyday life (Torous et al., 2016, 2017). Despite many review articles delineating the potential promise of these devices (Miller, 2012; Onnela & Rauch, 2016; Torous et al., 2016; Torous & Roberts, 2017) and some studies emerging on smartphone sensor data and mental health (Ben-Zeev et al., 2015; Jacobson & Chung, 2020; Pratap et al., 2019; Saeb et al., 2016), there is currently a dearth of research examining 1) how passively collected wearable biobehavioral features (e.g., heart rate, step count, sleep duration and their variability) are associated with mental health functioning, 2) with even fewer research studies conducted during adolescence (Sequeira et al., 2020), and 3) virtually no research conducted with traditional hypothesis testing as opposed to computational psychiatry approaches (i.e., machine learning) that are prevalent in the field. Note that while machine learning approaches are valuable and may increase predictive validity, traditional hypothesis testing is also needed to provide an explanation of associations among mHealth variables of interest and mental health. This may be particularly important as recent research indicates that machine learning approaches often inflate predictive performance in mental health research (Jacobucci et al., 2021), although this is not found in all studies (Jacobson et al., 2021). The present study used a novel multimethod approach with wearable indices of

biobehavioral functioning as cross-sectional and prospective predictors of adolescent internalizing symptoms across early adolescence.

Wearable devices provide the opportunity for continuous, scalable, unobtrusive, and ecologically valid measurement of multiple transdiagnostic biobehavioral units of analysis relevant to the National Institutes of Mental Health Research Domain Criteria (RDoC) Arousal and Regulation Systems, in real-world environments (Insel et al., 2010; Torous et al., 2017). Markers of the Arousal RDoC domain (i.e., increased sensitivity to external and internal stimuli) can be assessed at the level of resting heart rate (HR), while regulation (i.e., homeostatic functions of allostasis) can be captured with both physical activity via step count and sleep-wakefulness via sleep duration (Insel et al., 2010; Torous et al., 2017). As discussed in depth below, these wearable derived RDoC units of analysis conveniently map onto the previously mentioned biobehavioral changes that occur during adolescence, including increased stress sensitivity, onset of sleep difficulties, changes in behavioral patterning (Casey et al., 2014; Dahl et al., 2018). Importantly, these biobehavioral arousal and regulatory markers have long been used as clinical indicators of overall health and have been independently found to be associated with psychiatric functioning in laboratory and medical settings (Farmer et al., 1988; Harvey et al., 2011; Jacquart et al., 2019; Latvala et al., 2016). Together, each wearable metric may allow for the real-world assessment of transdiagnostic RDoC Arousal and Regulatory Systems that can provide clinically relevant data during this vulnerable period of development.

Heart Rate and Mental Health. Adolescence is associated with increased physiological reactivity to stressors when compared to childhood and adulthood (Stroud et al., 2009). Heart rate (HR), measured in beats per minute (bpm), is one unit of analysis of the RDoC Arousal System and is predominantly influenced by the coordination of the sympathetic and

parasympathetic branches of the autonomic nervous system, which activate when an organism faces perceived, imagined, or actual environmental threat and challenge as well as psychosocial stress (Cacioppo et al., 2017). Cardiovascular dysfunction, as reflected in higher cardiac arousal at rest (e.g., high resting HR; see Deutz et al., 2019), has been proposed to be a putative mechanism associated not only with morbidity and mortality (Khan et al., 2015; Qiu et al., 2017; Zhang et al., 2015), but also with a range of psychiatric disorders (Alvares et al., 2015; Clamor et al., 2014; Kandola et al., 2019; Kemp, Brunoni, et al., 2014; Kemp, Quintana, et al., 2014; Latvala et al., 2016; Paulus et al., 2013). This link may be due to the fact that cardiac responses are regulated by prefrontal and subcortical brain regions that are associated with affect and emotion regulation (Lemogne et al., 2011), regions specifically undergoing structural and/or functional connectivity maturational changes during adolescence (Dahl et al., 2018). More specifically, in a large prospective study of over 1 million Swedish males, higher resting HR during late adolescence was associated with increased risk for internalizing disorders in adulthood (Latvala et al., 2016)— a finding that has been replicated in smaller samples (Kemp et al., 2014) and in adolescents at risk for internalizing disorders (Nelson et al., 2020). Furthermore, lower heart rate variability (HRV, i.e., less variability in beat to beat intervals) has been proposed to be a putative marker of psychopathology in children and adolescents (Beauchaine & Thayer, 2015; Koenig et al., 2016), which may be due to its effect on emotion regulation brain networks (Mather & Thayer, 2018). Therefore, greater arousal via high resting HR (RHR) and lower variability in daily RHR may potentiate risk for internalizing problems.

Physical Activity and Mental Health. Greater physical activity is consistently associated with both lower internalizing diagnoses (Korczak et al., 2017; McDowell et al., 2019; Schuch et al., 2017) and symptoms (Hoare et al., 2014) across the lifespan, so much so that behavioral

planning to increase physical movement and exercise is incorporated within evidence-based treatments for internalizing disorders (Spruit et al., 2016). In contrast, sedentary behavior is associated with greater internalizing diagnoses (e.g., depression) during adolescence (Hoare et al., 2014). In turn, low physical activity, as indexed by low step count, may be one behavioral manifestation of psychomotor retardation or social withdrawal, which are key diagnostic criteria and symptoms of depression (American Psychiatric Association, 2013).

Sleep and Mental Health. Lastly, adolescence is associated with alterations in sleep regulation and patterning (Crowley et al., 2007; Dahl & Lewin, 2002), which play an important regulatory role in arousal (see HR above) and affect (Dahl, 1996). Sleep quantity is associated with future internalizing symptoms (Roberts et al., 2009) with both a lack of sleep (Short et al., 2020) and excessive sleep (Mullin et al., 2017) being associated with internalizing disorders as well as internalizing symptoms during adolescence (Liu et al., 2020). Such findings coincide with key internalizing diagnostic criteria (hypersomnia and insomnia in major depressive disorder, see American Psychiatric Association, 2013). Furthermore, variability in sleep duration may be particularly important as greater variability in sleep duration is associated with higher internalizing symptoms in adolescents (Bustamante et al., 2020) even after controlling for sleep duration (Fuligni et al., 2018). Therefore, similar to HR and physical activity, sleep duration and variability in sleep duration can be conceptualized as transdiagnostic mechanisms for internalizing difficulties (Blake & Allen, 2020; Harvey et al., 2011).

The current study was designed to address the role of wearable derived resting HR, step count, and sleep duration and variability in these measures play in cross-sectional and prospective internalizing symptoms during early adolescence.

The Current Study

The current study used preregistered with open code (https://osf.io/ysdfk/) for secondary data analysis of the Adolescent Brain Cognitive Development (ABCD) Study (Bagot et al., 2018). This study was designed to examine cross-sectional and prospective associations between multiple wearable derived RDoC indices of Arousal (i.e., resting HR) and Regulation (i.e., step count and sleep duration) and variability in these biobehavioral variables with internalizing symptoms during adolescence. For cross-sectional analyses, we hypothesized that higher arousal reflected in greater RHR and lower variability in daily RHR as well as lower regulation as reflected in lower sleep duration, greater variability in sleep duration, lower step count, and higher variability in step count would be associated with greater concurrent internalizing symptomatology. For prospective analyses, we hypothesized that higher RHR, lower variability in daily RHR, greater sleep duration and variability in sleep duration, and lower step count and variability in step count would be associated with increases in internalizing symptoms from Baseline to 2 Year Follow-Up.

Materials and Methods

Participants and Recruitment

We received IRB approval from the local IRB committee for secondary data analysis of ABCD Study data. Participants in this study were part of the nationally-representative, longitudinal ABCD study, which is thought to be poised to be the largest long-term longitudinal single-cohort study in the United States on neurodevelopment and child health. Information regarding recruitment sites, investigators, and project organization are available at http://abcdstudy.org. A baseline cohort of 11,878 children between the ages of 9-10 years and their parents/guardians were recruited across 21 data collection sites (Garavan et al., 2018). At

each site, participants were recruited through local elementary and charter schools. Exclusion criteria for youth participants included lack of English proficiency and severe sensory, intellectual, or neurological issues or any contraindications to magnetic resonance imaging scanning (Thompson et al., 2019). Eligible parents were fluent in either English or Spanish. Data for cross-sectional analyses came from the 2 Year Follow-Up with 5,686 adolescents ($mean_{age}$ = 11.96, sd = 0.65). Data for prospective analyses came from two timepoints: Baseline ($mean_{age}$ = 9.96, sd = 0.60) and 2 Year Follow-Up. The ABCD study recruited a subset of 150 participants at Baseline to examine the feasibility and acceptability of Fitbit Charge HR 2 devices at three sites, including University of California San Diego, SRI International, and Virginia Commonwealth University (Bagot et al., 2018). During Baseline, two Fitbit devices were lost, four were damaged, and after selecting Fitbit devices worn on protocol days, 143 adolescents remained across Baseline and 2 Year Follow-Up, which was the total sample size for prospective analyses. For participant demographics see Table 1.

Assessment Procedures

As described by Bagot et al. (2018), after informed consent, participants were provided a Fitbit Charge HR 2 and continuously wore this device over 21 days at Baseline and then again at 2 Year Follow-Up , except while bathing or engaging in water activities. The Fitbit app was downloaded to either the parent's or participant's phone to allow for data collected by the Fitbit to sync and upload when the Fitbit was within proximity of the phone. If a participant's Fitbit data were not uploaded for four days, an ABCD research assistant contacted the participant family because there is loss of data granularity after 6 days when devices have not been synced. Fitabase (www.fitabase.com) was used to monitor and retrieve Fitbit data for participants. Parent-report questionnaires, including reports of children's internalizing symptoms, were

completed using an iPad provided by the research team. At each time point, parents completed several questionnaires using an iPad provided by the research team. For the purposes of this study, parents reported on demographic information and their children's internalizing symptoms.

Measures

Internalizing Symptoms. Parents completed the Child Behavior Checklist (CBCL), which is a school-aged assessment form of 113 questions that are scored on a 3-point Likert Scale (0 – absent, 1 – occurs sometimes, 2 – occurs often) and is used to screen for emotional and behavioral difficulties (Achenbach, 1991). The CBCL is comprised of 8 subscales including anxious/depressed, depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking, and aggressive behavior. These subscales can then be grouped into an internalizing scale as was done in the current study and externalizing scale. Note that only scored variables were available in this open dataset, so we were unable to compute alpha reliability.

Fitbit Charge HR 2. Fitbit Charge HR 2 devices use proprietary algorithms to continuously measure biobehavioral features at up to a 1 second sampling rate (Bagot et al., 2018) with photoplethysmography (PPG) to calculate HR, accelerometer to calculate step count, and a combination of PPG and accelerometer to calculate sleep duration. HR, especially when averaged over longer time period as was done in the current study (Bent et al., 2020; Nelson et al., 2020; Nelson & Allen, 2019; Wang et al., 2017); step count (Alinia et al., 2017; Diaz et al., 2015; Tedesco et al., 2019); and sleep duration (de Zambotti et al., 2015, 2016; Mantua et al., 2016; Toon et al., 2016) have each been validated against research-grade, gold-standard devices (i.e., electrocardiogram for PPG data, behavioral observation for step count, and polysomnography and research actigraphy for sleep). These validations have also occurred in younger developmental samples of children and adolescents by the University of California San

Diego (UCSD) in anticipation of incorporation into ABCD (Bagot et al., 2018; Wing et al., 2017), but accuracy of sleep metrics have not been replicated in all studies (Meltzer et al., 2015).

Covariates. We included covariates that have been previously associated with internalizing symptoms in adolescence as well as wearable sensor accuracy, including age, biological sex, race, socioeconomic status (SES; i.e., family income), and body mass index (BMI) calculated with participant height and weight (weight (lb) / [height (in)]^2 x 703) (Nelson et al., 2020). Skin tone, wrist circumference, wear on dominant or nondominant hand, and firmware version were not controlled for because they were not collected. In addition, we did not control for cardioactive medication use, as data were not available on whether medication was consumed on Fitbit Charge HR 2 wear days. Models were run with and without covariate inclusion to reduce potential bias of selected covariates (York, 2018).

Statistical Analyses

All statistical analyses were conducted with R Studio, version 1.3.959. Daily Fitbit metrics were each collapsed into average and standard deviation metrics within Baseline and 2 Year Follow-Up as has been suggested in past research to increase wearable sensor accuracy (Nelson & Allen, 2019). The ABCD study did not collect data that would allow for examining more fine-grained associations between Fitbit Charge HR 2 metrics and internalizing symptoms, such as the level of hours or days (e.g., ecological momentary assessment).

Descriptive Analyses. To assess differences in Fitbit measures and internalizing symptoms by sex and race, we used the ggstats plot package (Patil, 2018) with all statistics Bonferroni corrected for multiple comparisons with significance level for sex differences defined as $p < .01$ (.05/4) and race differences defined as $p < .003$ (.05/18). In addition, we provided a

correlation plot among numeric variables of interest using the corrplot package (Wei & Simko, 2016). See supplemental materials for correlations between measures.

Cross-Sectional Analyses. To assess whether data were missing completely at random (MCAR) we performed parametric ($p < 0.001$) and nonparametric tests ($p < 0.001$). Despite that data were not MCAR, based on recent recommendations (Matta et al., 2018), in order to account for missing data, we used multiple imputation (10 imputations) using the mice package (van Buuren, 2020) (see Supplemental Material for missing data by variable). For analyses examining cross-sectional associations between Fitbit indices and internalizing symptoms at 2 Year Follow-Up, a series eight separate multiple regression models were run with each individual predictor alone and then in combination with all covariates, which included (1a) unadjusted model with resting HR, (1b) unadjusted model with variability resting HR, (2a) unadjusted model with step count, (2b) unadjusted model with variability in step count (3a) unadjusted model with sleep duration, (3b) unadjusted model with variability in sleep duration, (4) adjusted model with all Fitbit predictors predicting internalizing symptoms, and (5) final adjusted model with all Fitbit predictor interactions and covariates including age, biological sex, SES, race, and BMI predicting internalizing symptoms. In these models all continuous predictor variables were centered using the QuantPsych package (Fletcher, 2012), and final models were standardized to improve interpretability. The final model was selected based on model fit statistics (i.e., AIC, BIC) using the performance package (Lüdecke et al., 2020) (see Supplemental Materials for model fit statistics for each model). Sensitivity analyses were performed with analyses run as listwise deletion and with all variables winsorized, which did not change results (see supplemental materials).

Prospective Analyses. Analyses examining prospective associations between Fitbit units of analysis and internalizing symptoms, were conducted on a preliminary subset of participants limited to subjects that participated in an initial feasibility component of the ABCD study at baseline and that had follow-up data in the larger sample at follow-up. For these analyses, we used a Cross-Lagged Panel Model (CLPM; Kenny, 2014) using the lavaan package (Rosseel, 2020) with only significant Fitbit indices from cross-sectional models (i.e., RHR, step count, variability in step count, and variability in sleep duration) to predict internalizing symptoms across Baseline and 2 Year Follow-Up.

Results

Descriptive Statistics.

Males had significantly lower RHR and sleep duration as compared to females, while males had higher step count and internalizing symptoms when compared to females (see Figure 1 and note effect sizes). There was a significant and small effect of race on RHR, step count and internalizing symptoms and a significant and medium effect of race on sleep duration (see Figure 2 for Bonferroni corrected simple contrasts). Correlations between all study variables can be found in Figure 3. When comparing sample differences between participants in the larger cross-sectional sample at W2 and the smaller sample of pilot participants that had data across W1 and 2 we found some significant groups differences. While the smaller pilot prospective analysis subsample did not differ from the cross-sectional analyses in terms of participant sex $X^2(1) = 1.674$, $p = 0.196$, there was a difference in participant race, $X^2(17) = 84.558$, $p < 0.001$, such that there is greater racial diversity in the larger cross-sectional sample, and participants in the prospective analyses had significantly higher SES (M = 8.21, SD = 2.05), when compared to the participants in the cross-sectional analyses (M = 7.47, SD = 2.20) $t(111.05) = -3.689$, $p < 0.001$.

Cross-Sectional Analyses

As shown in Table 2, in the final model adjusted for covariates (i.e., age, sex, body mass index, socioeconomic status, and race) was the best model fit and there was a significant effect for higher RHR, lower step count, lower variability in step count, and higher variability in sleep duration with higher internalizing symptoms (see Figure 4). Please see supplementary materials for each unadjusted main effect model.

Prospective Analyses.

Covariance Parameters. Higher RHR and step count at 2 Year Follow-Up, but not Baseline, were associated with higher internalizing symptoms. Greater variability in sleep duration was associated with higher internalizing at Baseline, but not 2 Year Follow-Up, and variability in step count was not associated with internalizing symptoms at either time point (see Table 3).

Autoregressive Parameters. Internalizing symptoms, RHR, variability in step count and variability in sleep duration predicted themselves across time points, but this was not found for step count (see Table 3).

Cross-Lagged Parameters. None of the Fitbit indices at Baseline predicted internalizing symptoms at 2 Year Follow-Up. In contrast, internalizing symptoms and resting HR at Baseline predicted lower step count at 2 Year Follow-Up, while resting HR at Baseline also predicted lower step count variability at 2 Year Follow-Up (see Table 3).

Constraining the paths from W1 Fitbit measures to W2 internalizing symptoms to zero did not reduce model fit ($X^2[4] = 3.65$, $p = .455$), while further constraining correlations between within-wave Fitbit measures and internalizing symptoms to zero did reduce model fit ($X^2[12] = 26.05$, $p = .004$).

Discussion

The current study utilized data from the ABCD Study to investigate whether wearable markers including RHR, step count, sleep duration, and variability in each of these measures, were cross-sectionally and prospectively associated with internalizing symptoms during adolescence. This study was the first to examine the association between wearable Arousal and Regulatory RDoC units of analysis during this sensitive period of development. Study results provided partial support for hypothesized associations.

There were a number of results that were consistent with our cross-sectional hypotheses. First, higher RHR was associated with greater internalizing symptoms, which is consistent with a large body of research indicating that higher RHR signifies greater levels of arousal and is associated with internalizing symptoms (Alvares et al., 2015; Kandola et al., 2019; Kemp, Quintana, et al., 2014; Latvala et al., 2016; Nelson et al., 2020; Paulus et al., 2013). In contrast, variability in RHR was not associated with internalizing symptoms, which isn't particularly surprising as HRV or the variability in beat to beat intervals has been found to be associated with internalizing symptoms (Koenig et al., 2016), while no studies to the authors' knowledge have examined the association between variability in daily RHR and internalizing symptoms. While HRV and RHR are related measures, such that when HRV is high, RHR tends to be lower, HRV is more strongly influenced by the parasympathetic branch of the autonomic nervous system and has long been used as an index of emotion regulation (Beauchaine, 2015), while RHR is influenced by both the parasympathetic and sympathetic branches of the autonomic nervous system. Second, lower step count and variability in step count were both associated with greater internalizing symptoms, which coincides with findings indicating that individuals experiencing depressive symptoms experience behavioral withdrawal and psychomotor retardation (American

Psychiatric Association, 2013). Lastly, greater variability in sleep duration was associated with higher internalizing symptoms, which is consistent with past research (Fuligni et al., 2018) as well as coinciding with the finding that those with internalizing disorders have greater sleep difficulties (American Psychiatric Association, 2013; Bai et al., 2020; Gregory et al., 2011). In contrast, sleep duration was not associated with internalizing disorders, which is in contrast to some research (Liu et al., 2020; Short et al., 2020), but does fall in line with previously mentioned findings that variability in sleep duration is associated with internalizing symptoms over and above sleep duration alone (Fuligni et al., 2018). As a follow-up, we ran an exploratory analysis to examine whether there was a possible u-shaped relationship between both short and long sleep duration and internalizing symptoms as has been indicated in past research (Mullin et al., 2017; Short et al., 2020), but there was no significant association ($p > .05$).

Despite the small effect size for the cross-sectional model, research indicates that small effect sizes that are maintained over long periods of time can have a large effect on outcomes of interest (i.e., cummulative effect; Funder & Ozer, 2019). This indicates that while higher RHR, lower step count and variability in step count, and greater variability in sleep duration may have a small impact on internalizing symptoms at any one point in time, a cumulative impact on internalizing symptoms may be significant if high RHR, low step count and step count variability, and sleep duration variability are maintained across time. Overall, these sets of findings indicate a multifactorial constellation of wearable units of analysis that are associated with greater levels of internalizing symptoms and may comprise a “digital phenotype” of internalizing symptoms (Insel, 2017; Torous et al., 2016, 2017). Interestingly, males had significantly higher internalizing symptoms as compared to female adolescence; however, we caution the over-interpretation of the findings, as the effect size difference was small and both

scores were within 3 points of the 50th percentile for norm-referenced T-scores. Together, this would suggest the difference in mean T-score would not result in clinically significant behavioral profiles.

In terms of CLPM prospective associations, results were less consistent with our hypotheses, such that there were no prospective associations between wearable indices and internalizing symptoms across early adolescence, except for an autoregressive effect of internalizing symptoms, resting HR, and sleep duration variability predicting themselves across time. These results should be interpreted with caution as they were conducted with an extremely small subsample of participants that had data available between a baseline designed for feasibility on a small number of participants and this sample had important differences in SES and race as compared to the larger W2 sample. Interestingly, there was a prospective association between higher internalizing symptoms and higher resting HR at Baseline and lower step count at 2 Year Follow-Up and higher resting HR at Baseline and lower step count variability at 2 Year Follow-Up, which is again in line with findings that individuals experiencing depressive symptoms experience behavioral withdrawal and psychomotor retardation (American Psychiatric Association, 2013). The lack of prospective findings indicates limited feasibility and utility of a single snapshot of wearable biobehavioral features predicting future internalizing mental health difficulty 25 months later. These null findings are not surprising nor discouraging. First, it would be assumed that the association between dynamic risk factors at an earlier timepoint would have diminishing predictive validity with an outcome at a subsequent timepoint. Second, internalizing symptoms are dynamic and adolescence is a period of particularly rapid change in internalizing symptoms (Zisook et al., 2007). Therefore, it is unsurprising that wearable metrics fail to associate with internalizing symptoms 25 months later. Lastly, wearable devices by their very

nature are designed to be continuously worn over long time periods, which may provide sufficient novel capture of data in order to index time-sensitive alterations in affective states and therefore in real-world settings it is unlikely that wearables would be used at discrete timepoints to predict future functioning, but rather worn continuously to offer more relevant health information.

Limitation and Future Directions

Although the present study had a number of notable strengths, including being the first study to examine commercially available wearable metrics' association with adolescent internalizing symptoms in such a large and nationally representative sample, there were a number of limitations that should be noted. First, wearable devices' inherent strength is their ability to allow for the multimodal collection of intensive longitudinal data (i.e., multiple data points per minute, hour, or day). Unfortunately, the ABCD Study did not have intensive self-report data (i.e., ecological momentary assessment or EMA) that would allow for capitalizing on the dynamics between self-reported internalizing symptoms and wearable biobehavioral features at the hourly or daily level. Instead, we were limited to averaging Fitbit Charge HR 2 sensor features within each time point to examine these variables in relation to self-reported internalizing symptoms. Although this was a limitation, concurrent findings presented in this study indicate that future intensive longitudinal studies that collect mental states at more fine-grained details across hours or days may have even stronger effect sizes that could inform prevention and intervention efforts. Second, the CLPM prospective analyses had a smaller sample size (n = 143) and was less racially and economically diverse as compared to the cross-sectional analyses (n = 5,686), which likely limited the power and generalizability of this model. Therefore, it is possible that with a larger and more representative sample could have significant

effects emerge. In addition, due to only two waves of data, we were not able to disambiguate between- versus within-person effects (Curran et al., 2014; Hamaker et al., 2015). Future studies, potentially using upcoming ABCD Study data releases, should run random intercept CLPM with a larger sample in order to determine convergence or divergence of findings based on power considerations and with more than two waves in order to disambiguate between- versus within-person effects. Third, in an attempt to par down outcomes of interest, we did not examine the association between wearable biobehavioral risk factors and other mental or physical health outcomes as well as other wearable features such as sleep onset variability or other important aspects of development, such as pubertal status. Future studies should consider using data-driven approaches by performing specification curve analyses in order to determine the most relevant wearable biobehavioral factors (e.g., heart rate, VO2 max, heart rate variability, step count, exercise minutes, distance traveled, sleep duration, nighttime awakenings, restlessness, etc.), temporal dynamics (e.g., minutes, hours, days buns), temporal lag between variables, and developmental characteristics that are most predictive of mental health outcomes. The identification of the most relevant wearable metrics may then assist in variable selection for computational data-driven classification and prediction of mental health dynamics at shorter time scales that may eventually lead to early identification, prevention, and just-in-time interventions (Allen et al., 2019). Similarly, genetic markers could play an important role in links between Fitbit indices and mental health outcomes and should be examined in future research.

Conclusion

This study investigated whether wearable derived biobehavioral risk factors, including RHR, step count, sleep duration, and variability in each of these variables were cross-sectionally and prospectively associated with internalizing symptoms during adolescence. Findings indicated

that higher RHR, lower step count and variability in step count, and greater variability in sleep duration were cross-sectionally associated with higher internalizing disorders. These findings indicate a multifactorial constellation of wearable Arousal and Regulation RDoC units of analysis that are associated with greater levels of internalizing symptoms and may comprise a "digital phenotype" of internalizing symptoms during early adolescence. In contrast, preliminary prospective analyses that were limited to a subsample of participants due to ABCD study design, wearable biobehavioral risk factors were not associated with internalizing symptoms approximately 25 months later, although greater internalizing symptoms and higher resting HR at Baseline did predict lower step count at 2 Year Follow-Up and higher resting HR at Baseline also predicted lower step count variability at 2 Year Follow-Up. A larger and more racially and economically diverse sample size is likely required to probe this prospective association further to accurately assess longitudinal associations. Overall, findings indicate that commercially available wearable devices, such as the Fitbit Charge HR 2, can provide pertinent biobehavioral information representing RDoC Arousal and Regulation Systems signifying risk when examined in close temporal proximity to current internalizing symptomatology. Future research should capitalize on the inherent temporal resolution provided by wearable devices using both computational and hypothesis driven approaches to improve the prediction and explanation of mental health dynamics, respectively. These approaches may eventually lead to early identification, prevention, and just-in-time interventions for those experiencing mental health difficulties, particularly during sensitive periods of development.

Key Points

- Adolescence is characterized by heightened risk for first onset of psychopathology.
- Wearable devices collect multiple transdiagnostic biobehavioral units of analysis relevant to the NIMH RDoC Arousal/Regulation Systems, which map onto adolescent biobehavioral changes (e.g., increased stress sensitivity, onset of sleep difficulties, altered behavioral patterning).
- Wearable data collection in 5,686 adolescents was used to determine whether wearable sensor data were cross-sectionally and prospectively associated with internalizing symptoms.
- Higher RHR, lower step count and step count variability, and greater variability in sleep duration were associated with greater internalizing symptoms.
- These sets of findings indicate a multifactorial constellation of wearable units of analysis comprising a “digital phenotype” of internalizing symptoms that may lead to prevention and just-in-time interventions.

Table 1. Participant Demographics

Variable	N	Percentage	Mean (SD)
Age	5685		T1: 9.96 (0.60) T2: 11.96 (0.65)
Sex			
Female	2725	47.92%	
Male	2961	52.08%	
Race			
White	3942	69.33%	
Black	641	11.27%	
Other	237	4.17%	
Native American	195	3.43%	
Native Alaskan	96	1.69%	
Missing	86	1.51%	
Refuse to Answer	78	1.37%	
Filipino	68	1.20%	
Don't Know	60	1.06%	
Chinese	57	1.00%	
Vietnamese	47	0.83%	
Korean	38	0.67%	
Guamanian	31	0.55%	
Asian Indian	29	0.51%	
Asian Other	26	0.46%	
Japanese	25	0.44%	
Native Hawaiian	19	0.33%	
Pacific Islander Other	9	0.16%	
Samoan	2	0.04%	
Ethnicity			
Not Hispanic or Latino	4548	79.99%	
Hispanic or Latino	1063	18.70%	
Missing	75	1.32%	
SES (family income)			
< $5,000	121	2.13%	
$5,000 - $11,999	158	2.78%	
$12,000 -$15,999	103	1.81%	
$16,000 - $24,999	201	3.53%	
$25,000 - $34,999	309	5.43%	
$35,000 - $49,999	452	7.95%	
$50,000 - $74,999	761	13.38%	
$75,000 - $99,999	853	15.00%	
$100,000 - $199,999	1725	30.34%	
$200,000 and greater	634	11.15%	
NA	369	6.49%	
Body Mass Index	5676		20.52 (4.92)

Table 2. Cross-Sectional Model

Predictors	***β***	***SE***	***95% CI***	***P-Value***
(Intercept)	48.401	0.167	48.075 – 48.728	**<0.001**
Resting Heart Rate	1.534	0.299	0.948 – 2.121	**<0.001**
Resting Heart Rate SD	0.223	0.281	-0.327 – 0.773	0.427
Step Count	-0.772	0.357	-1.472 – -0.071	**0.031**
Step Count SD	-0.877	0.351	-1.564 – -0.189	**0.013**
Sleep Duration	0.514	0.302	-0.077 – 1.105	0.088
Sleep Duration SD	1.569	0.302	0.978 – 2.160	**<0.001**
Sex	1.665	0.287	1.102 – 2.227	**<0.001**
Age	-0.558	0.282	-1.111 – -0.005	**0.048**
SES	-0.630	0.315	-1.248 – -0.012	**0.046**
Race- Black	-5.010	0.485	-5.961 – -4.059	**<0.001**
Race- Native American	0.483	0.758	-1.004 – 1.970	0.524
Race- Native Alaskan	3.548	1.050	1.489 – 5.607	**0.001**
Race- Native Hawaiian	-3.199	2.346	-7.798 – 1.401	0.173
Race- Guamanian	4.702	1.838	1.098 – 8.305	**0.011**
Race- Samoan	4.941	7.182	-9.138 – 19.021	0.491
Race- Pacific Islander Other	-5.570	3.388	-12.212 – 1.072	0.100
Race- Asian Indian	-4.940	1.892	-8.650 – -1.231	**0.009**
Race- Chinese	-1.487	1.369	-4.172 – 1.197	0.277
Race- Filipino	-2.927	1.243	-5.364 – -0.490	**0.019**
Race- Japanese	0.511	2.036	-3.481 – 4.503	0.802
Race- Korean	-0.542	1.654	-3.785 – 2.701	0.743
Race- Vietnamese	-0.479	1.489	-3.398 – 2.439	0.747
Race- Asian Other	-1.354	2.001	-5.277 – 2.570	0.499

Race- Other	-0.710	0.695	-2.073 – 0.653	0.307
Race- Refuse to Answer	1.559	1.162	-0.720 – 3.838	0.180
Race- Don't Know	-1.417	1.343	-4.049 – 1.216	0.292
BMI	1.254	0.294	0.676 – 1.831	**<0.001**
Observations	5597			
R^2 / R^2 adjusted	0.053 / 0.049			

Table 3. CLPM for Prospective Analyses

Autoregressive and Cross-Lagged Parameters			***b***	***SE***	***z***	***p-value***
CBCL Internalizing W2	~	CBCL Internalizing W1 +	0.567	0.080	7.083	**0.000**
		HR W1 +	0.111	0.106	1.039	0.299
		Step Count W1 +	-1.597	1.119	-1.428	0.153
		Step Count Variability W1 +	0.603	0.964	0.625	0.532
		Sleep Duration Variability W1 +	-0.027	0.033	-0.802	0.422
HR W2	~	CBCL Internalizing W1 +	0.070	0.056	1.242	0.214
		HR W1 +	0.813	0.073	11.105	**0.000**
		Step Count W1 +	0.470	0.814	0.577	0.564
		Step Count Variability W1 +	0.043	0.668	0.064	0.949
		Sleep Duration Variability W1 +	-0.025	0.024	-1.059	0.290
Step Count W2	~	CBCL Internalizing W1 +	-0.020	0.009	-2.201	**0.028**
		HR W1 +	-0.036	0.012	-3.041	**0.002**
		Step Count W1 +	-0.031	0.124	-0.247	0.805
		Step Count Variability W1 +	0.102	0.108	0.943	0.346
		Sleep Duration Variability W1 +	-0.004	0.004	-0.972	0.331
Step Count Variability W2	~	CBCL Internalizing W1 +	-0.010	0.009	-1.042	0.297
		HR W1 +	-0.035	0.012	-2.842	**0.004**
		Step Count W1 +	-0.066	0.129	-0.509	0.611
		Step Count Variability W1 +	0.213	0.115	1.863	0.063
		Sleep Duration Variability W1 +	-0.004	0.004	-1.087	0.277
Sleep Duration Variability W2	~	CBCL Internalizing W1 +	0.366	0.423	0.866	0.386
	~	HR W1 +	0.118	0.557	0.211	0.833
	~	Step Count W1 +	3.034	6.335	0.479	0.632
	~	Step Count Variability W1 +	-5.240	5.924	-0.885	0.376
	~	Sleep Duration Variability W1 +	0.607	0.179	3.390	**0.001**

Covariance Parameters			***r***	***Covariance***	***SE***	***z***	***p-value***
CBCL Internalizing W1	~~	HR W1	.08	6.191	6.560	0.944	0.345
CBCL Internalizing W1	~~	Step Count W1	-.08	-0.800	0.833	-0.961	0.337
CBCL Internalizing W1	~~	Step Count Variability W1	-.03	-0.294	0.862	-0.341	0.733
CBCL Internalizing W1	~~	Sleep Duration Variability W1	-.18	-54.149	26.556	-2.039	**0.041**
CBCL Internalizing W2	~~	HR W2	.22	10.332	4.636	2.228	**0.026**
CBCL Internalizing W2	~~	Step Count W2	-.22	-1.638	0.717	-2.285	**0.022**
CBCL Internalizing W2	~~	Step Count Variability W2	-.17	-1.353	0.753	-1.797	0.072
CBCL Internalizing W2	~~	Sleep Duration Variability W2	-.20	-60.279	35.677	-1.690	0.091

Overall Model: Test Statistic 0.000, df = 0
Free Parameters = 65, Number of Observations = 143, Missing Patterns = 15.

R^2

CBCL Internalizing W2	HR W2	Step Count W2	Step Count Variability W2	Sleep Duration Variability W2
0.334	0.539	0.147	0.149	0.189

Partial R^2

CBCL Internalizing W2	HR W2	Step Count W2	Step Count Variability W2	Sleep Duration Variability W2
0.026	0.013	0.010	0.008	0.008

Note: CBCL = Child Behavior Checklist; HR = Heart Rate; SE = Standard Error; z = z-score, W1 = Baseline; W2 = 2 Year Follow-Up; ~~ = is correlated with; ~ = is regressed on.

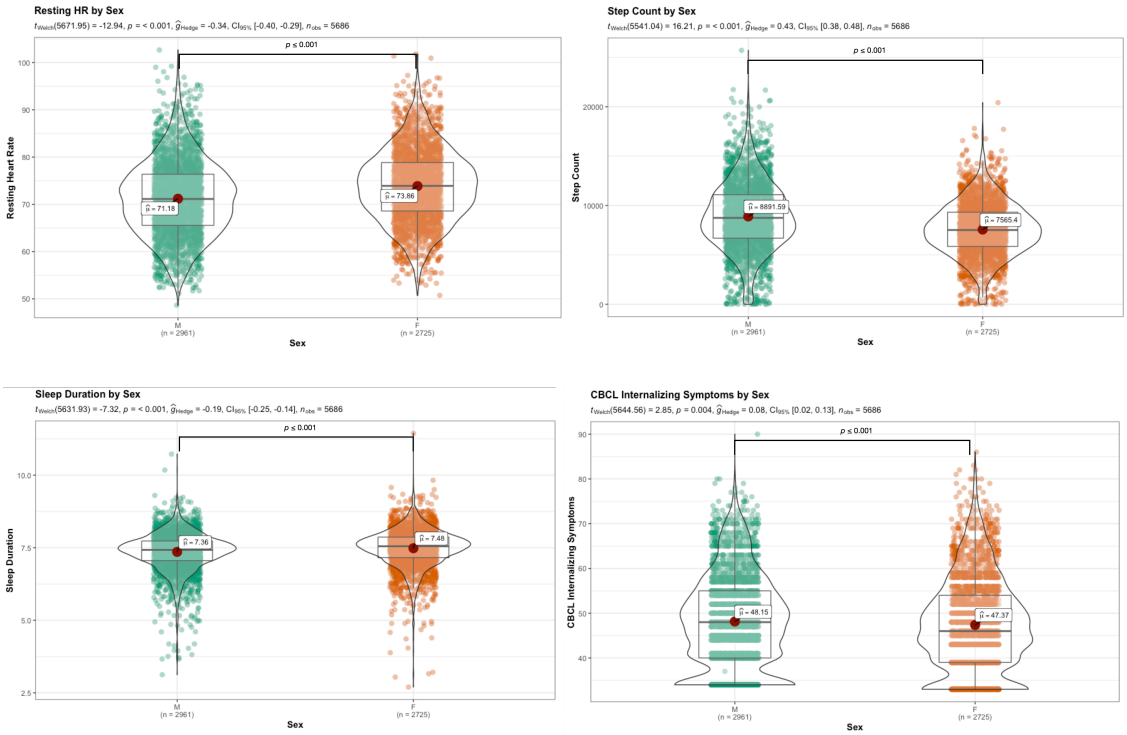

Figure 1. Wearable Metrics and CBCL Internalizing Symptoms by Sex.

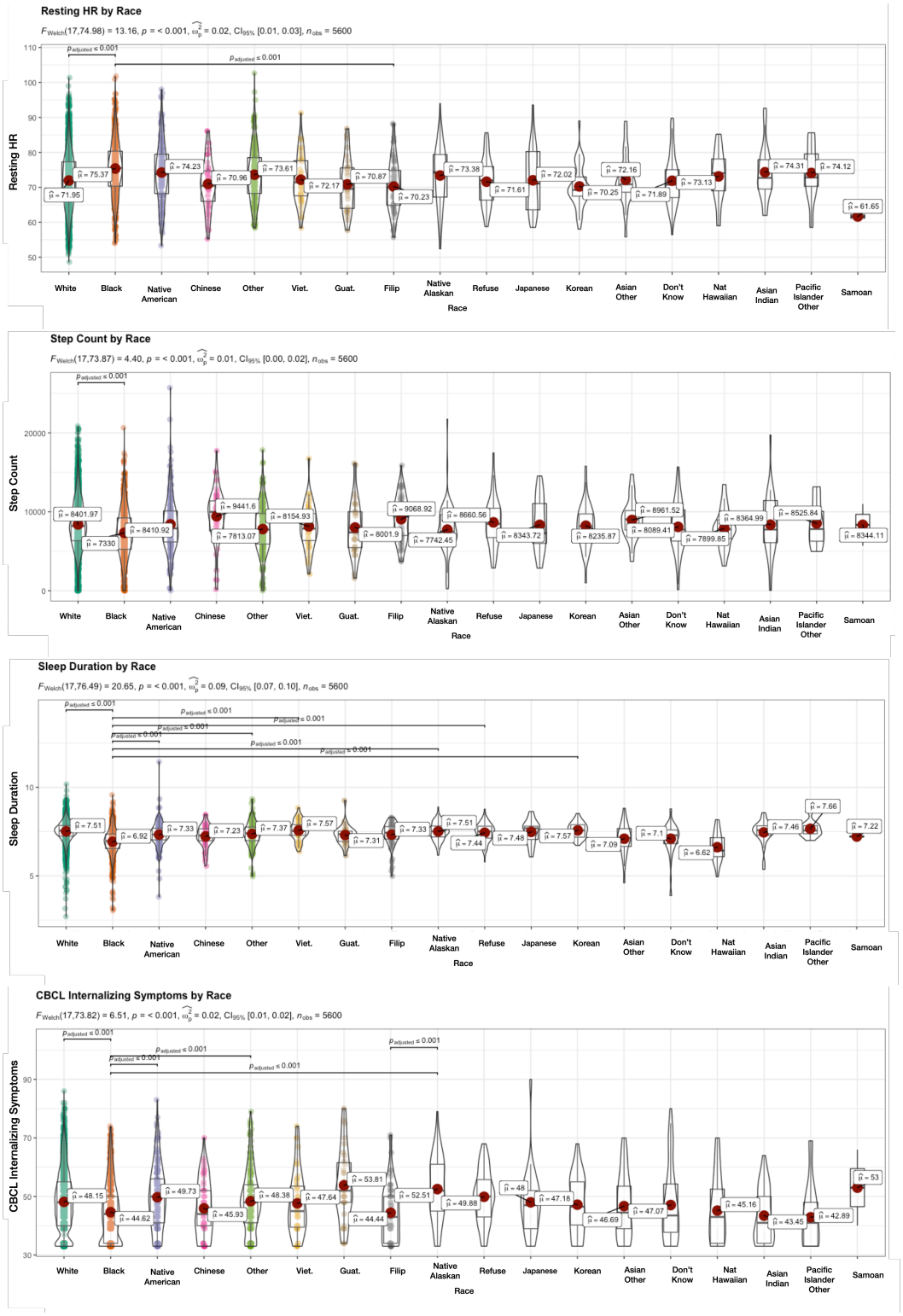

Figure 2. Wearable Metrics and CBCL Internalizing Symptoms by Race. Note: Viet = Vietnamese, Guat = Guatemalan, Filip = Filipino, Nat = Native.

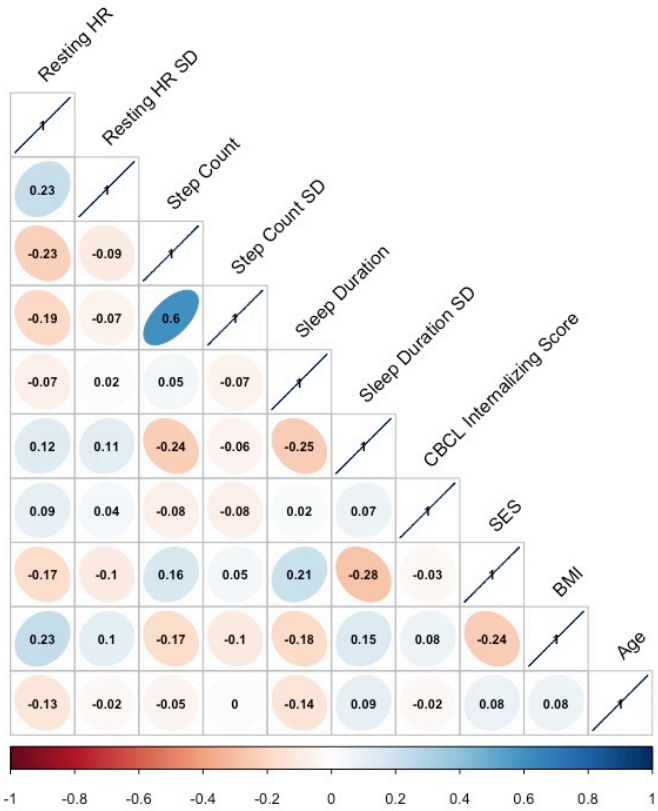

Figure 3. Correlations Between Study Variables

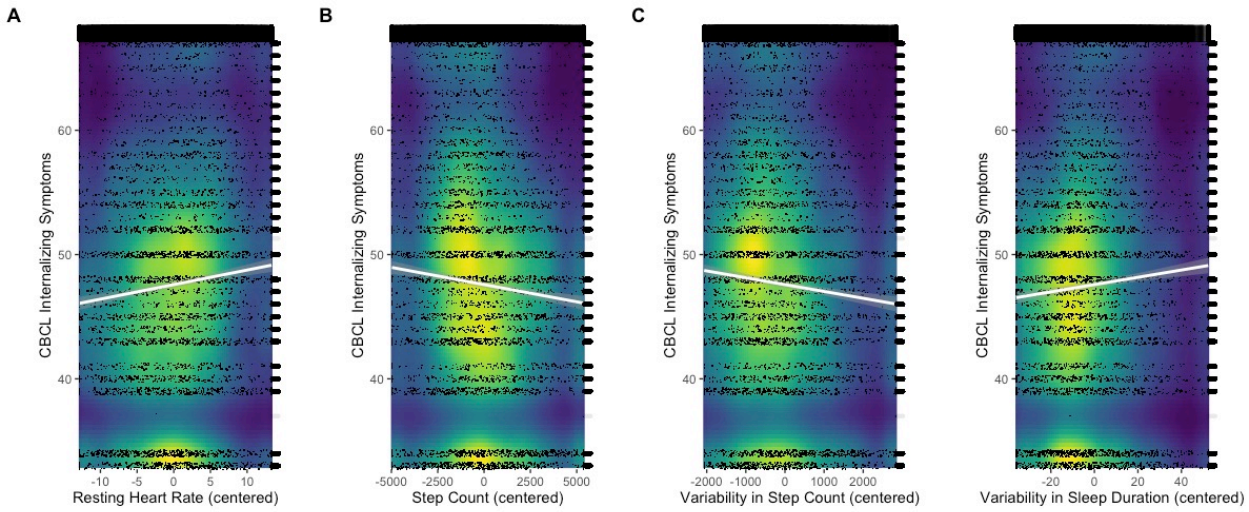

Figure 4. Association Between A. Resting HR, B. Step Count, C. Variability in Step Count, and D. Variability in Sleep Duration with CBCL Internalizing Symptoms. Note: CBCL = Child Behavior Checklist.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 4-18 and 1991 Profile*. University of Vermont Department of Psychiatry.

Alinia, P., Cain, C., Fallahzadeh, R., Shahrokni, A., Cook, D., & Ghasemzadeh, H. (2017). How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities. *JMIR MHealth and UHealth*, *5*(8), e106. https://doi.org/10.2196/mhealth.6321

Allen, N. B., Nelson, B. W., Brent, D., & Auerbach, R. P. (2019). Short-term prediction of suicidal thoughts and behaviors in adolescents: Can recent developments in technology and computational science provide a breakthrough? *Journal of Affective Disorders*, *250*, 163–169. https://doi.org/10.1016/j.jad.2019.03.044

Alvares, G. A., Quintana, D. S., Hickie, I. B., & Guastella, AmJ. (2015). Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: A systematic review and meta-analysis. *Journal of Psychiatry & Neuroscience*, *40*(6), 1–16. https://doi.org/10.1503/jpn.140217

American Psychiatric Association. (2013). *DSM 5*. https://doi.org/10.1176/appi.books.9780890425596.744053

Bagot, K. S., Matthews, S. A., Mason, M., Squeglia, L. M., Fowler, J., Gray, K., Herting, M., May, A., Colrain, I., Godino, J., Tapert, S., Brown, S., & Patrick, K. (2018). Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. *Developmental Cognitive Neuroscience*, *32*, 121–129. https://doi.org/10.1016/j.dcn.2018.03.008

Bai, S., Ricketts, E. J., Thamrin, H., Piacentini, J., Albano, A. M., Compton, S. N., Ginsburg, G. S., Sakolsky, D., Keeton, C. P., Kendall, P. C., & Peris, T. S. (2020). Longitudinal Study of Sleep and Internalizing Problems in Youth Treated for Pediatric Anxiety Disorders. *Research on Child and Adolescent Psychopathology*, *48*(1), 67–77. https://doi.org/10.1007/s10802-019-00582-x

Beauchaine, T. P. (2015). Respiratory Sinus Arrhythmia: A Transdiagnostic Biomarker of Emotion Dysregulation and Psychopathology. *Current Opinion in Psychology*, *3*, 43–47. https://doi.org/10.1016/j.copsyc.2015.01.017

Beauchaine, T. P., & Thayer, J. F. (2015). Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology*, *98*(2), 338–350. https://doi.org/10.1016/j.ijpsycho.2015.08.004

Bent, B., Goldstein, B. A., Kibbe, W. A., & Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-0226-6

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal*, *38*(3), 218–226. https://doi.org/10.1037/prj0000130

Blake, M. J., & Allen, N. B. (2020). Prevention of internalizing disorders and suicide via adolescent sleep interventions. *Current Opinion in Psychology*, *34*, 37–42. https://doi.org/10.1016/j.copsyc.2019.08.027

Bustamante, C. M. V., Rodman, A. M., Dennison, M. J., Flournoy, J. C., Mair, P., & McLaughlin, K. A. (2020). Within-person fluctuations in stressful life events, sleep, and anxiety and depression symptoms during adolescence: A multiwave prospective study.

Journal of Child Psychology & Psychiatry, *61*(10), 1116–1125. https://doi.org/10.1111/jcpp.13234

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.). (2017). *Handbook of Psychophysiology* (4th ed.). Cambridge University Press. https://doi.org/10.1017/9781107415782

Casey, B. J., Oliveri, M. E., & Insel, T. (2014). A Neurodevelopmental Perspective on the Research Domain Criteria (RDoC) Framework. *Biological Psychiatry*, *76*(5), 350–353. https://doi.org/10.1016/j.biopsych.2014.01.006

Clamor, A., Hartmann, M. M., Köther, U., Otte, C., Moritz, S., & Lincoln, T. M. (2014). Altered autonomic arousal in psychosis: An analysis of vulnerability and specificity. *Schizophrenia Research*, *154*(1–3), 73–78. https://doi.org/10.1016/j.schres.2014.02.006

Crowley, S. J., Acebo, C., & Carskadon, M. A. (2007). Sleep, circadian rhythms, and delayed phase in adolescence. *Sleep Medicine*, *8*(6), 602–612. https://doi.org/10.1016/j.sleep.2006.12.002

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*(5), 879–894. https://doi.org/10.1037/a0035297

Dahl, R. E. (1996). The regulation of sleep and arousal: Development and psychopathology. *Development and Psychopathology*, *8*(1), 3–27. https://doi.org/10.1017/S0954579400006945

Dahl, R. E., Allen, N. B., Wilbrecht, L., & Suleiman, A. B. (2018). Importance of investing in adolescence from a developmental science perspective. *Nature*, *554*(7693), 441–450. https://doi.org/10.1038/nature25770

Dahl, R. E., & Lewin, D. S. (2002). Pathways to adolescent health sleep regulation and behavior. *Journal of Adolescent Health*, *31*(6), 175–184. https://doi.org/10.1016/S1054-139X(02)00506-2

de Zambotti, M., Baker, F. C., & Colrain, I. M. (2015). Validation of Sleep-Tracking Technology Compared with Polysomnography in Adolescents. *Sleep*, *38*(9), 1461–1468. https://doi.org/10.5665/sleep.4990

de Zambotti, M., Baker, F. C., Willoughby, A. R., Godino, J. G., Wing, D., Patrick, K., & Colrain, I. M. (2016). Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiology & Behavior*, *158*, 143–149. https://doi.org/10.1016/j.physbeh.2016.03.006

Deutz, M. H. F., Woltering, S., Vossen, H. G. M., Deković, M., van Baar, A. L., & Prinzie, P. (2019). Underlying Psychophysiology of Dysregulation: Resting Heart Rate and Heart Rate Reactivity in Relation to Childhood Dysregulation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(6), 589–599. https://doi.org/10.1016/j.jaac.2018.09.434

Diaz, K. M., Krupka, D. J., Chang, M. J., Peacock, J., Ma, Y., Goldsmith, J., Schwartz, J. E., & Davidson, K. W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, *185*, 138–140. https://doi.org/10.1016/j.ijcard.2015.03.038

Farmer, M. E., Locke, B. Z., Mościcki, E. K., Dannenberg, A. L., Larson, D. B., & Radloff, L. S. (1988). PHYSICAL ACTIVITY AND DEPRESSIVE SYMPTOMS: THE NHANES I EPIDEMIOLOGIC FOLLOW-UP STUDY. *American Journal of Epidemiology*, *128*(6), 1340–1351. https://doi.org/10.1093/oxfordjournals.aje.a115087

Fletcher, T. D. (2012). *QuantPsyc: Quantitative Psychology Tools*. https://cran.r-project.org/web/packages/QuantPsyc/index.html

Fuligni, A. J., Arruda, E. H., Krull, J. L., & Gonzales, N. A. (2018). Adolescent Sleep Duration, Variability, and Peak Levels of Achievement and Mental Health. *Child Development*, *89*(2), e18–e28. https://doi.org/10.1111/cdev.12729

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*, 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Gregory, A. M., Buysse, D. J., Willis, T. A., Rijsdijk, F. V., Maughan, B., Rowe, R., Cartwright, S., Barclay, N. L., & Eley, T. C. (2011). Associations between sleep quality and anxiety and depression symptoms in a sample of young adult twins and siblings. *Journal of Psychosomatic Research*, *71*(4), 250–255. https://doi.org/10.1016/j.jpsychores.2011.03.011

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Harvey, A. G., Murray, G., Chandler, R. A., & Soehner, A. (2011). Sleep disturbance as transdiagnostic: Consideration of neurobiological mechanisms. *Clinical Psychology Review*, *31*(2), 225–235. https://doi.org/10.1016/j.cpr.2010.04.003

Hoare, E., Skouteris, H., Fuller-Tyszkiewicz, M., Millar, L., & Allender, S. (2014). Associations between obesogenic risk factors and depression among adolescents: A systematic review: Obesogenic risk and depression: a review. *Obesity Reviews*, *15*(1), 40–51. https://doi.org/10.1111/obr.12069

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2017). Digital Phenotyping: Technology for a New Science of Behavior. *Jama*, *94301*, E1–E2. https://doi.org/10.1001/JAMA.2017.11295

Jacobson, N. C., & Chung, Y. J. (2020). Passive Sensing of Prediction of Moment-To-Moment Depressed Mood among Undergraduates with Clinical Levels of Depression Sample Using Smartphones. *Sensors*, *20*(12), 3572. https://doi.org/10.3390/s20123572

Jacobson, N. C., Lekkas, D., Huang, R., & Thomas, N. (2021). Deep learning paired with wearable passive sensing data predicts deterioration in anxiety disorder symptoms across 17–18 years. *Journal of Affective Disorders*, *282*, 104–111. https://doi.org/10.1016/j.jad.2020.12.086

Jacobucci, R., Littlefield, A. K., Millner, A. J., Kleiman, E. M., & Steinley, D. (2021). Evidence of Inflated Prediction Performance: A Commentary on Machine Learning and Suicide Research. *Clinical Psychological Science*. https://doi.org/10.1177/2167702620954216

Jacquart, J., Dutcher, C. D., Freeman, S. Z., Stein, A. T., Dinh, M., Carl, E., & Smits, J. A. J. (2019). The effects of exercise on transdiagnostic treatment targets: A meta-analytic review. *Behaviour Research and Therapy*, *115*, 19–37. https://doi.org/10.1016/j.brat.2018.11.007

Kandola, A., Ashdown-Franks, G., Stubbs, B., Osborn, D. P. J., & Hayes, J. F. (2019). The association between cardiorespiratory fitness and the incidence of common mental health disorders: A systematic review and meta-analysis. *Journal of Affective Disorders*, *257*, 748–757. https://doi.org/10.1016/j.jad.2019.07.088

Kemp, A. H., Brunoni, A. R., Santos, I. S., Nunes, M. A., Dantas, E. M., De Figueiredo, R. C., Pereira, A. C., Ribeiro, A. L. P., Mill, J. G., Andreão, R. V., Thayer, J. F., Benseñor, I. M., & Lotufo, P. A. (2014). Effects of depression, anxiety, comorbidity, and antidepressants on resting-state heart rate and its variability: An ELSA-Brasil cohort baseline study. *American Journal of Psychiatry*, *171*(12), 1328–1334. https://doi.org/10.1176/appi.ajp.2014.13121605

Kemp, A. H., Quintana, D. S., Quinn, C. R., Hopkinson, P., & Harris, A. W. F. (2014). Major depressive disorder with melancholia displays robust alterations in resting state heart rate and its variability: Implications for future morbidity and mortality. *Frontiers in Psychology*, *5*, 1–9. https://doi.org/10.3389/fpsyg.2014.01387

Kenny, D. A. (2014). Cross-Lagged Panel Design. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online* (p. stat06464). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat06464

Khan, H., Kunutsor, S., Kalogeropoulos, A. P., Georgiopoulou, V. V., Newman, A. B., Harris, T. B., Bibbins-Domingo, K., Kauhanen, J., Gheorghiade, M., Fonarow, G. C., Kritchevsky, S. B., Laukkanen, J. a., & Butler, J. (2015). Resting heart rate and risk of incident heart failure: Three prospective cohort studies and a systematic meta-analysis. *Journal of the American Heart Association*, *4*(1), e001364. https://doi.org/10.1161/JAHA.114.001364

Koenig, J., Kemp, A. H., Beauchaine, T. P., Thayer, J. F., & Kaess, M. (2016). Depression and resting state heart rate variability in children and adolescents—A systematic review and meta-analysis. *Clinical Psychology Review*, 15.

Korczak, D. J., Madigan, S., & Colasanto, M. (2017). Children's Physical Activity and Depression: A Meta-analysis. *Pediatrics*, *139*(4), e20162266. https://doi.org/10.1542/peds.2016-2266

Latvala, A., Kuja-Halkola, R., Rück, C., D'Onofrio, B. M., Jernberg, T., Almqvist, C., Mataix-Cols, D., Larsson, H., & Lichtenstein, P. (2016). Association of Resting Heart Rate and Blood Pressure in Late Adolescence With Subsequent Mental Disorders: A Longitudinal Population Study of More Than 1 Million Men in Sweden. *JAMA Psychiatry*, *73*(12), 1268. https://doi.org/10.1001/jamapsychiatry.2016.2717

Lemogne, C., Thomas, F., Consoli, S. M., Pannier, B., Jégo, B., & Danchin, N. (2011). Heart Rate and Completed Suicide: Evidence From the IPC Cohort Study. *Psychosomatic Medicine*, *73*(9), 731–736. https://doi.org/10.1097/PSY.0b013e3182365dc7

Liu, B.-P., Wang, X.-T., Liu, Z.-Z., Wang, Z.-Y., An, D., Wei, Y.-X., Jia, C.-X., & Liu, X. (2020). Depressive symptoms are associated with short and long sleep duration: A longitudinal study of Chinese adolescents. *Journal of Affective Disorders*, *263*, 267–273. https://doi.org/10.1016/j.jad.2019.11.113

Lüdecke, D., Makowski, D., Waggoner, P., Patil, I., & Ben-Shachar, M. S. (2020). *performance: Assessment of Regression Models Performance*. https://cran.r-project.org/web/packages/performance/index.html

Mantua, J., Gravel, N., & Spencer, R. (2016). Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography. *Sensors*, *16*(5), 646. https://doi.org/10.3390/s16050646

Mather, M., & Thayer, J. F. (2018). How heart rate variability affects emotion regulation brain networks. *Current Opinion in Behavioral Sciences*, *19*, 98–104. https://doi.org/10.1016/j.cobeha.2017.12.017

Matta, T. H., Flournoy, J. C., & Byrne, M. L. (2018). Making an unknown unknown a known unknown: Missing data in longitudinal neuroimaging studies. *Developmental Cognitive Neuroscience*, *33*, 83–98. https://doi.org/10.1016/j.dcn.2017.10.001

McDowell, C. P., Dishman, R. K., Gordon, B. R., & Herring, M. P. (2019). Physical Activity and Anxiety: A Systematic Review and Meta-analysis of Prospective Cohort Studies. *American Journal of Preventive Medicine*, *57*(4), 545–556. https://doi.org/10.1016/j.amepre.2019.05.012

Meltzer, L. J., Hiruma, L. S., Avis, K., Montgomery-Downs, H., & Valentin, J. (2015). Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents. *Sleep*, *38*(8), 1323–1330. https://doi.org/10.5665/sleep.4918

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Mullin, B. C., Pyle, L., Haraden, D., Riederer, J., Brim, N., Kaplan, D., & Novins, D. (2017). A Preliminary Multimethod Comparison of Sleep Among Adolescents With and Without

Generalized Anxiety Disorder. *Journal of Clinical Child & Adolescent Psychology*, *46*(2), 198–210. https://doi.org/10.1080/15374416.2016.1220312

Nelson, B. W., & Allen, N. B. (2019). Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study. *JMIR MHealth and UHealth*, *7*(3), e10828. https://doi.org/10.2196/10828

Nelson, B. W., Byrne, M. L., Sheeber, L., & Allen, N. B. (2017). Does Context Matter? A Multi-Method Assessment of Affect in Adolescent Depression Across Multiple Affective Interaction Contexts. *Clinical Psychological Science*, 216770261668006. https://doi.org/10.1177/2167702616680061

Nelson, B. W., Low, C. A., Jacobson, N., Areán, P., Torous, J., & Allen, N. B. (2020). Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *Npj Digital Medicine*, *3*(1), 90. https://doi.org/10.1038/s41746-020-0297-4

Nelson, B. W., Sheeber, L., Pfeifer, J., & Allen, N. B. (2020). Psychobiological markers of allostatic load in depressed and nondepressed mothers and their adolescent offspring. *Journal of Child Psychology and Psychiatry*, jcpp.13264. https://doi.org/10.1111/jcpp.13264

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *January*, 1–12. https://doi.org/10.1038/npp.2016.7

Patil, I. (2018). *ggstatsplot: "ggplot2" Based Plots with Statistical Details*. https://CRAN.R-project.org/package=ggstatsplot

Paulus, E. J., Argo, T. R., & Egge, J. A. (2013). The Impact of Posttraumatic Stress Disorder on Blood Pressure and Heart Rate in a Veteran Population. *Journal of Traumatic Stress*, *26*(1), 169–172. https://doi.org/10.1002/jts.21785

Pratap, A., Atkins, D. C., Renn, B. N., Tanana, M. J., Mooney, S. D., Anguera, J. A., & Areán, P. A. (2019). The accuracy of passive phone sensors in predicting daily mood. *Depression and Anxiety*, *36*(1), 72–81. https://doi.org/10.1002/da.22822

Qiu, S., Cai, X., Sun, Z., Li, L., Zuegel, M., Steinacker, J. M., & Schumann, U. (2017). Heart Rate Recovery and Risk of Cardiovascular Events and All-Cause Mortality: A Meta-Analysis of Prospective Cohort Studies. *Journal of the American Heart Association*, *6*(5). https://doi.org/10.1161/JAHA.117.005505

Roberts, R. E., Roberts, C. R., & Duong, H. T. (2009). Sleepless in adolescence: Prospective data on sleep deprivation, health and functioning. *Journal of Adolescence*, *32*(5), 1045–1057. https://doi.org/10.1016/j.adolescence.2009.03.007

Rosseel, Y. (2020). *lavaan: Latent Variable Analysis* (0.6-7) [Computer software]. https://cran.r-project.org/web/packages/lavaan/index.html

Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P., & Mohr, D. C. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*, *4*, e2537. https://doi.org/10.7717/peerj.2537

Schuch, F., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P., Reichert, T., Bagatini, N. C., Bgeginski, R., & Stubbs, B. (2017). Physical activity and sedentary behavior in people with major depressive disorder: A systematic review and meta-analysis. *Journal of Affective Disorders*, *210*, 139–150. https://doi.org/10.1016/j.jad.2016.10.050

Sequeira, L., Perrotta, S., LaGrassa, J., Merikangas, K., Kreindler, D., Kundur, D., Courtney, D., Szatmari, P., Battaglia, M., & Strauss, J. (2020). Mobile and wearable technology for monitoring depressive symptoms in children and adolescents: A scoping review. *Journal of Affective Disorders*, *265*, 314–324. https://doi.org/10.1016/j.jad.2019.11.156

Short, M. A., Booth, S. A., Omar, O., Ostlundh, L., & Arora, T. (2020). The relationship between sleep duration and mood in adolescents: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *52*, 101311. https://doi.org/10.1016/j.smrv.2020.101311

Spruit, A., Assink, M., van Vugt, E., van der Put, C., & Stams, G. J. (2016). The effects of physical activity interventions on psychosocial outcomes in adolescents: A meta-analytic review. *Clinical Psychology Review*, *45*, 56–71. https://doi.org/10.1016/j.cpr.2016.03.006

Stroud, L. R., Foster, E., Papandonatos, G. D., Handwerger, K., Granger, D. A., Kivlighan, K. T., & Niaura, R. (2009). Stress response and the adolescent transition: Performance versus peer rejection stressors. *Development and Psychopathology*, *21*(1), 47–68. https://doi.org/10.1017/S0954579409000042

Tedesco, S., Sica, M., Ancillao, A., Timmons, S., Barton, J., & O'Flynn, B. (2019). Validity Evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in Free-Living Environments in an Older Adult Cohort. *JMIR MHealth and UHealth*, *7*(6), e13084. https://doi.org/10.2196/13084

Toon, E., Davey, M. J., Hollis, S. L., Nixon, G. M., Horne, R. S. C., & Biggs, S. N. (2016). Comparison of Commercial Wrist-Based and Smartphone Accelerometers, Actigraphy, and PSG in a Clinical Cohort of Children and Adolescents. *Journal of Clinical Sleep Medicine*, *12*(03), 343–350. https://doi.org/10.5664/jcsm.5580

Torous, J, Kiang, M., Lorme, J., & Onnela, J. (2016). New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment Health*, *3*(2), e16. https://doi.org/10.2196/mental.5165

Torous, J, Onnela, J.-P., & Keshavan, M. (2017). New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. *Translational Psychiatry*, *7*(3), e1053. https://doi.org/10.1038/tp.2017.25

Torous, John, & Roberts, L. W. (2017). Needed Innovation in Digital Health and Smartphone Applications for Mental Health. *JAMA Psychiatry*, 4–5. https://doi.org/10.1001/jamapsychiatry.2017.0262

van Buuren, S. (2020). *mice: Multivariate Imputation by Chained Equations* (3.12.0) [Computer software]. https://cran.r-project.org/web/packages/mice/index.html

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology*, *2*(1), 104–106. https://doi.org/10.1001/jamacardio.2016.3340

Wei, T., & Simko, V. (2016). *corrplot: Visualization of a Correlation Matrix*. https://cran.r-project.org/package=corrplot

Wing, D., Godino, J. G., de Zambotti, M., Pautz, C., Inkelis, S., Higgins, M., Nichols, J., Brumback, T., Surana, J., Nahar, P. A., Baker, F. C., Colrain, I. M., Patrick, K., & Tapert, S. (2017, May 5). *Validation measures of physical activity and sleep in children from a consumer-level wearable*.

York, R. (2018). Control variables and causal inference: A question of balance. *International Journal of Social Research Methodology*, *21*(6), 675–684. https://doi.org/10.1080/13645579.2018.1468730

Zhang, D., Shen, X., & Qi, X. (2015). *Resting heart rate and all-cause and cardiovascular mortality in the general population: A meta-analysis*. *188*(3), 1–11. https://doi.org/10.1503/cmaj.150535/-/DC1

Zisook, M. D. S., Lesser, M. D. I., Stewart, M. D. J., Wisniewski, Ph. D. S., Balasubramani, Ph. D. G. K., Fava, M. D. M., Gilmer, M. D. W., Dresselhaus, M. D. M. P. H. T., Thase, M. D. M., Nierenberg, M. D. A., Trivedi, M. D. M., & Rush, M. D. a. (2007). Effect of Age at Onset on the Course of Major Depressive Disorder. *American Journal of Psychiatry*, *164*(10), 1539–1546. https://doi.org/10.1176/appi.ajp.2007.06101757