Evaluating Individual Differences in Emotion Regulation in Response to Sadness Using Digital Phenotyping

Colin M. Bosma[1], Curtis Wojcik[2], and Emily A. P. Haigh[3]

[1]Department of Psychology, Providence Medical Group

[2]Department of Psychology, University of Maine

[3]Department of Psychology, University of Victoria

Author Note

Colin M. Bosma https://orcid.org/0000-0002-4828-6023
Curtis Wojcik
Emily A. P. Haigh https://orcid.org/0000-0003-1763-554X
We have no known conflict of interest to disclose.
Correspondence concerning this article should be addressed to Colin M. Bosma, Providence Medical Group, Department of Psychology. Email: colin.bosma@providence.org

Abstract

Background: The majority of research on emotion regulation processes has been restricted to controlled laboratory settings that use experimental paradigms to investigate short-term outcomes. A true understanding of emotion regulation requires an unobtrusive, ecologically valid assessment of the construct as it occurs in the environment. Digital phenotyping is a novel method for evaluating human behavior in naturalistic settings.

Objective: This study aimed to evaluate whether smartphone-based digital phenotyping data predicts individual differences in emotion regulation in both in-lab and naturalistic settings.

Methods: During an in-lab session, unselected university student participants (N = 69) completed self-report questionnaires measuring trait emotion regulation as well as state emotion regulation and state affect following a baseline period, a negative mood induction, and a recovery period. Smartphone-based digital phenotyping data were then collected over the course of a 7-day follow-up. Variation in global positioning system (GPS) distance and mobile power state level were examined as predictors of longitudinal variation in negative affect, emotion regulation, and depression.

Results: Results showed that variation in GPS distance was significantly associated with variation in state cognitive reappraisal (b = -0.0004, SE = 0.0002, p = .02) and negative state affect (b = 0. 005, SE = 0.002, p = .01) over time. Variation in power state level was also significantly associated with variation in cognitive reappraisal over time (b = -4.98, SE = 1.72, p = .005) and marginally associated with variation in negative state affect (b = -29.58, SE = 16.73, p = .08) over time. Cluster and classification analyses showed both power state level and GPS distance accurately classified two trait emotion regulation clusters with high sensitivity (.95 and .96 respectively) and specificity (.86 and .97 respectively). Variation in power state level and GPS distance together with trait and state emotion regulation did not predict current depressive symptoms (ps > .05).

Conclusions: Overall, the findings provide initial and foundational data on the use of digital phenotyping data in predicting individual differences in state and trait emotion regulation in both in-lab and naturalistic settings. The results suggest that operationalizations of digital phenotyping data and modeling approaches are particularly important factors to consider when implementing digital phenotyping methodology in the study of mental health processes such as emotion regulation.

Keywords: Emotion regulation; digital phenotyping; ecological momentary assessment; ecological validity; mental health; psychopathology; depression

Evaluating Individual Differences in Emotion Regulation in Response to Sadness Using Digital Phenotyping

Emotion regulation refers to the strategic and automatic processes that impact the magnitude, occurrence, and duration of emotional responses (Gross, 2015a). The ability to effectively regulate our emotions is fundamental to well-being, as emotion dysregulation has been shown to increase the likelihood of developing and maintaining a range of mental health problems (Aldao et al., 2010, Aldao et al., 2016; Joormann & Vanderlind, 2014). Research broadly supports the transdiagnostic role of emotion regulation such that both negative affect and deficits in emotion regulation cut across psychopathology (Aldao et al., 2016; Joormann & Siemer, 2014; Mennin & Fresco, 2015; Sheppes et al., 2015; Quinn et al., 2014). Overall, the evidence suggests that emotion regulation is key to understanding the onset and course of psychopathology.

While our understanding of emotion regulation has benefited from a strong theoretical framework (e.g., the extended process model; Gross, 2015b) that has yielded an impressive body of research (Aldao et al., 2010; Aldao et al., 2016; Sheppes et al., 2015), significant gaps remain. Our current understanding of emotion regulation is limited by a near exclusive reliance on tightly controlled laboratory-based paradigms and measurement of outcomes exclusively over the course of these paradigms (on the order of minutes) (Aldao, 2013; Gross, 2015a). It is unclear how explicitly directing participants to elicit or regulate emotions impacts subjective reporting, or how accurately these findings predict real-world consequences (Berking & Wupperman, 2012). While several, validated self-report measures of emotion regulation have increased our conceptual understanding of emotion regulation processes, these measures are typically administered cross-sectionally (Fernandez et al., 2016; Mauss & Robinson, 2009), which precludes evaluation of how emotion regulation processes unfold over time. In short, our traditional approaches to measurement fail to assess variability in dynamic processes associated with emotion regulation (Kuppens & Verduyn, 2015; Kuppens & Verduyn, 2017). Although

ecological momentary assessment (EMA), or the repeated sampling of participants' behaviors in real-time and in their natural environments, is an important advancement for improving ecological validity in psychological research, findings remain limited by an exclusive reliance on self-report ratings of emotional experiences, goals, and behaviors.

Research on emotion regulation has examined how individuals regulate emotions in laboratory settings; however, a true understanding of emotion regulation depends upon an unobtrusive, ecologically valid assessment of the construct as it occurs in daily life. Digital sensors in smartphones can facilitate this type of assessment by enabling the collection of moment-by-moment data from individuals in naturalistic settings (Onnela & Rauch, 2016; Torous et al., 2016). Quantifying individual-level human behavior using data from smartphone sensors—or digital phenotyping—is a novel method for evaluating human behavior, such as emotion regulation, in naturalistic settings (Onnela & Rauch, 2016).

Digital Phenotyping and Emotion Regulation

Digital sensors in smartphones enable high-resolution data collection from individuals as data can be collected every few seconds rather than a few times a day (Onnela & Rauch, 2016). Continuous data of metrics from smartphones can provide accurate predictions of psychopathology in relation to naturalistic behavioral patterns making it possible to use these data to develop precise markers of well-being and illness at the individual and population level (Insel, 2017; Onnela & Rauch, 2016). For example, degrees of social interaction may be reflected in call and text logs as well as temporal mobility patterns derived from global positioning system (GPS) data. Phone sensors allow for continuous telemetry of psychological health and generate comprehensive longitudinal data with high accuracy (Torous et al., 2016; Torous et al., 2017).

An important advantage of digital phenotyping data is that it is exceptionally granular, and therefore potentially more sensitive to individual differences in behavior trajectories, compared to

traditional longitudinal sampling methods such as EMA. This type of data can also be used to generate a digital profile of behavioral patterns, or a digital phenotype (Insel, 2017; Torous et al., 2016). For example, Saeb and colleagues (Saeb et al., 2015) conducted an exploratory study using GPS smartphone data to distinguish participants with depressive symptoms from healthy controls. The results showed that mobility between favorite locations and general location variance classified those with significant depressive symptoms from those without with 86.5% accuracy. Changes in depressive symptoms among those with Major Depressive Disorder (Zhang et al., 2022) as well as those with comorbid Generalized Anxiety Disorder (Meyerhoff et al.,2021) have also been reliably predicted by similar phone-based GPS variables, such as time spent at home, travel frequency, and time spent at different locations. Adjacent mobility-based data can be further derived from link-ups between smartphones and Bluetooth devices, with changes in the number and regularity of these link-ups aligning with changes in self-reported depressive symptoms (Zhang et al., 2021). In addition to these mobility indicators, call duration and text message frequency have been used to signify social interactions in adolescents with Major Depressive Disorder, with these data correlating strongly with multiple measures of symptomatology (Cao et al., 2020). These findings demonstrate how smartphone data, such as GPS, can function as a behavioral marker of depressive symptom severity (see Seppälä et al., 2019 and Yim et al., 2020 for an additional review of similar findings). Several other studies have found smart phone sensor data can predict psychiatric symptoms in other diagnostic groups as well, such as those with schizophrenia or bipolar disorder (Seppälä et al., 2019).

To date, no research has specifically examined the relationship between emotion regulation and digital phenotyping. Experimental designs using smartphone-based digital phenotyping can offer sophisticated assessments of emotion regulation and well-being. Digital phenotyping, through improved detection, is poised to positively impact clinical decision-making, ultimately in service of personalized treatment for mental health problems.

Aims and Hypotheses

As digital phenotyping is a novel approach for examining psychological phenomena, and to our knowledge, this is the first project to evaluate the associations between digital phenotyping and emotion regulation, the aims and hypotheses are exploratory. Given the transdiagnostic properties of sustained negative affect and emotion dysregulation across mental health problems, the study investigated the relationship between emotion regulation in response to transient negative emotions and emotional well-being independently. To keep consistent with well-established emotion regulation paradigms evaluating transient negative emotions (Aldao, 2013, Gross, 2014), we chose to elicit and measure sadness, as changes in affective experiences have been considered indicative of mood repair, or the up-regulation of negative emotions (Fernandez et al., 2016). As sustained negative affective experiences and difficulties up-regulating negative emotions (e.g., sadness) is characteristic of longer-lasting, chronic conditions such as Major Depressive Disorder (Aldeo et al., 2010), we investigated current depressive symptoms as an indicator of general emotional well-being.

The first aim was to investigate the digital behavior correlates of self-report (state and trait) emotion regulation. It was hypothesized that 1) trait emotion regulation (cognitive reappraisal, expressive suppression, rumination, and difficulties in emotion regulation) would relate to power state level and GPS distance over time, and 2) state emotion regulation over time (variation in negative affect and spontaneous emotion regulation) would be related to power state and GPS distance over time. The second aim investigated whether self-reported state and trait emotion regulation interacted with digital behaviors (variation in power state level and GPS distance over time) to predict emotional well-being at baseline. Accordingly, we hypothesized that 1) trait emotion regulation and digital phenotyping data together would predict baseline depressive symptoms, and 2) state emotion regulation in response to a negative mood induction and digital phenotyping data together would predict baseline depressive symptoms.

Method

Recruitment

We recruited participants from introductory psychology courses at a mid-sized New England university from 02/12/2020 to 03/10/2020, providing course credit as compensation. Eligibility requirements were 18 years of age or older and ownership of a smartphone with either Apple iOS or Android operating systems. All participants provided an electronic agreement to participate following informed consent and each participant received debriefing information about the study. A power analysis conducted using G*Power version 3.1.9.3 (Faul et al., 2007) indicated that approximately 68 participants would be sufficient to find a medium effect size (e.g., $f^2 = .15$) with sufficient power (e.g., $\beta = .80$) for multiple regression with two predictors. We planned a total sample of 100 participants to account for attrition and non-compliance, but participant recruitment ended early due to COVID-19-related university closures.

Study Design

All participants completed the same in-lab experimental paradigm as part of a larger study (see Bosma, 2020b for details), which included an initial battery of self-report measures and a negative mood induction. At the end of the in-lab session, participants installed the Beiwe mobile device application on their smartphones and were asked to keep the application installed for seven consecutive days. Over these seven days, participants were prompted via the application to complete brief questionnaires regarding affect, emotion regulation, and current activities. These prompts were provided twice a day (once at 10 AM and once at 6 PM), for a total of 14 times. GPS distance and power state level data were passively collected from participants' smartphone sensors via the application during the seven-day follow up period. After the follow-up period, participants were sent an email indicating they should uninstall the Beiwe application from their smartphone.

Mood Induction

After completing the questionnaires, participants watched a series of video clips on a desktop computer, presented approximately 24 inches in front of them. To establish a neutral baseline, participants viewed an emotionally neutral 3-minute nature film clip from Alaska's Denali National Park. For the sad mood induction, the neutral film was followed by a 2-minute and 51-second film depicting a boy who is distraught at the death of his father from the movie "The Champ." This paradigm has been validated in previous research to elicit transient sadness (Rottenberg et al., 2007). After the sad mood induction, participants were instructed to sit quietly for a 5-minute recovery period. To ensure procedural standardization, directions, videos, and audio were presented on a computer using E-Prime 2.0 (2015) computer software.

Measures

Cognitive Reappraisal and Expressive Suppression

Participants completed the Emotion Regulation Questionnaire (ERQ; Gross &. John, 2003) to assess individual differences in the tendency to use two emotion regulation strategies: cognitive reappraisal and expressive suppression. This 10-item self-report questionnaire has a subscale for cognitive reappraisal (six items) and expressive suppression (four items). The response format of each item is a 7-point Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The items pertain to regulating both positive and negative emotion. Higher scores indicate greater tendencies to implement the respective emotion regulation strategy. The ERQ has demonstrated adequate internal consistency for both scales, adequate test-retest reliability across three months for both scales, adequate convergent validity with specific measures of inauthenticity, coping, rumination, and negative mood regulation, and adequate discriminant validity with measures of broad personality, impulse control, cognitive ability, and desirability (Gross &. John, 2003). In the present sample, internal reliability was acceptable ($\alpha = .79$) for the full scale, good ($\alpha = .89$) for the cognitive reappraisal subscale, and acceptable ($\alpha = .70$) for the expressive suppression subscale.

Trait Rumination

The Ruminative Responses Scale (RRS; Treynor et al., 2003) was used to assess trait rumination. The RRS is a widely used 10-item questionnaire designed to measure individual differences in rumination, or the tendency to engage in repetitive and passive thinking about problems, negative events, and negative feelings. The response format of each item is a 4-point Likert-type scale ranging from 1 (*almost never*) to 4 (*almost always*) indicating the frequency with which an individual engages in a ruminative activity. The measure yields a total score and separate scores for a pondering and brooding subscale. The RRS was shown to have good internal consistency and moderate test-retest reliability (Treynor et al., 2003). Internal reliability for our sample was acceptable ($\alpha = .70$).

Emotion Regulation Implementation

The Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2008) was used to measure trait difficulties with implementing emotion regulation. The DERS was designed to assess trait-level perceived emotion regulation ability across four domains, including awareness and understanding of emotions, acceptance of emotions, the ability to control impulses and behave in accordance with goals in the presence of negative affect, and access to emotion regulation strategies that are perceived to be effective for feeling better. Items are rated on a scale of 1 (*almost never* [0-15%]), 2 (*sometimes* [11-35%]), 3 (*about half the time* [36-65%]), 4 (*most of the time* [55-90%]), to 5 (*almost always* [91-100%]). The DERS can be interpreted as a total score, or each subscale can be scored separately. Higher scores indicate more difficulty in emotion regulation. The DERS has demonstrated strong internal consistency (Gratz & Roemer, 2008). The full-scale internal reliability for our sample was excellent ($\alpha = .96$).

State Emotion Regulation

The Spontaneous Affect Regulation Scale (SARS; Egloff et al., 2006) is a self-report questionnaire based on the format of the ERQ (Gross & John, 2003) and measures the degree to which individuals spontaneously (i.e., without instruction) implemented emotion regulation strategies during a specified

time. The SARS is comprised of six items and two subscales, including cognitive reappraisal and expressive suppression. Higher scores on the subscales represent a greater degree of implementation of that emotion regulation strategy. The SARS subscales have demonstrated adequate internal consistency in previous studies (Egloff et al., 2006; Gruber et al., 20012; Stange et al., 2017). In our study, the internal consistency for the cognitive reappraisal subscale at the first EMA timepoint was questionable ($\alpha = 0.66$) and acceptable for the expressive suppression subscale ($\alpha = 0.72$). Across EMA timepoints, internal consistency ranged from poor ($\alpha = 0.16$) to acceptable ($\alpha = 0.76$) for the reappraisal subscale and from poor ($\alpha = 0.30$) to acceptable ($\alpha = 0.82$) for the expressive suppression subscale. Multiple imputation was not used for missing data prior to calculating internal consistency on the SARS; as such, these values may have been impacted by participant non-response and attrition during the EMA monitoring period.

State Affect

A visual analog scale (VAS; Feinstein, 1987) was used to assess state negative affect. Using horizontal 100 count lines with anchors at the 0 (*not at all*) and 100 (*extremely*) endpoints, participants rated the degree to which they were currently experiencing sadness. Although the results of studies evaluating the validity and reliability of the VAS have been mixed, the scales are useful for capturing state affective experiences because of their simplicity, especially for paradigms with short intervals between assessments (Ahearn, 1997).

Depressive Symptoms

The Patient Health Questionnaire 9 (PHQ-9; Kroenke et al., 2001) was used to measure current depressive symptoms. The PHQ-9 is a widely used, 9-item, multipurpose instrument for screening, diagnosing, monitoring, and measuring severity of depression. The response format of each item is a 4-point Likert-type scale ranging from 0 (*not at all*) to 3 (*nearly every day*) indicating the frequency with which an individual endorses each symptom for the previous 2 weeks. Higher scores indicate greater

depressive symptom severity. The PHQ-9 has demonstrated strong internal consistency (Kroenke et al., 2001). The internal consistency for the present sample was excellent (α = .91).

Digital Phenotyping

The Beiwe mobile device application used in this study is part of a cloud-based research platform developed and maintained by the Onnela Lab at the Harvard T.H. Chan School of Public Health. The application is designed for continuously collecting digital phenotyping data from smartphone sensors, operating unobtrusively as the individual uses their smartphone (Torous et al., 2016). GPS distance and power state level were the digital phenotyping data used in the present study. These were chosen because their accessibility and measurement via the Beiwe platform was consistent across both Apple and Android devices. Moreover, GPS distance and power state level are more easily re-analyzed for replication and extension studies compared to other types of digital phenotyping data. GPS distance was calculated as the difference between one GPS coordinate and its preceding coordinate (in meters) using a great circle distances approach, which accounts for the curvature of the earth. The accuracy of GPS data from smartphone devices is on average within 4.9 meters and is impacted by satellite geometry, atmospheric conditions, receiver design and quality, and the position of buildings. Power state level (0%-100%) was to the nearest whole percentage point. Based on pilot testing, the sampling rate for GPS position and power state level ranged from approximately 10-seconds to 15-minutes, with GPS data sampled more frequently than power state level, on average.

Data Analysis

We first conducted descriptive analyses to provide information on primary study variables. To capture the range and amplitude of emotion and emotion regulation processes occurring longitudinally over the seven-day monitoring period, we operationalized variables as the *SD* of the values over time (Kuppens & Verduyn, 2015, 2017). Specifically, variation in state emotion regulation across time was operationalized as the *SD* of the cognitive appraisal and expressive suppression subscales of the SARS, as

well as the *SD* of sadness VAS ratings (i.e., Hypothesis 2). Variation in digital behaviors over time was operationalized as the *SD* of GPS distance and power state level values across all sampled time points (i.e., all four hypotheses).

We conducted primary analyses for each aim using simple and multiple linear regressions to evaluate both directional relationships and interactions. We also employed cluster analyses to determine grouping patterns in numerous continuous variables, followed by classification analyses to determine whether digital phenotyping data accurately predicted the groups identified in the cluster analyses. All reported 95% confidence intervals (CIs) were the result of bootstrapping with 2000 bootstrap replicates. When investigating the relationship between digital behaviors over time and trait emotion regulation using cluster and classification analyses, the time stamps of each GPS distance and power state level entry were used to evaluate variation in digital behaviors over time (i.e., Hypothesis 1).

Given the longitudinal design of the research project, attrition was expected. To reduce bias and increase efficiency in statistical modeling, a multiple imputation approach was implemented for missing values using the Multivariate Imputation via Chained Equations R package (MICE; van Buuren & Groothuis-Oudshoorn, 2011).

Results

Participants

N = 69 participants (Table 1) were recruited and included in the final analyses. Descriptive statistics for the digital phenotyping data collected during the EMA phase are presented in Table 5. While none of the 69 students dropped out of the study during the EMA phase, 21 did not reply to any prompts to complete self-report measures during the EMA phase. It is notable that 22% of the power level data were missing. An investigation of patterns contributing to the missing data revealed the 16

participants without power state data to each be using an Android phone from multiple manufacturers and running various versions of Android (versions 7, 8, 8.1, 9, and 10).

Table 1

Sample Demographic Characteristics

Variable	*M*	*SD*
Age (N = 69)	19.7	1.72
Variable	*n*	*%*
Gender Identity		
Male	33	48
Female	36	52
Race		
Asian	1	1.44
Black	1	1.44
Multiple	2	2.9
Native American	2	2.9
White	62	89
Relationship Status		
Never Married/Single	69	100
Education Level		
High School	43	62
1 Year of College	14	20
2 Years of College	9	13
Bachelors	1	1.44
Associates or Other	1	1.44

Trait Emotion Regulation and Digital Phenotyping

Power State

Variation in power state level did not predict trait cognitive reappraisal and explained less than one percent of the variance in trait cognitive reappraisal ($_{adj}R^2$ = -.005, $F(1, 67) = 0.65$, $p = .42$). Variation in power state level did not predict trait expressive suppression and explained less than one percent of the variance in trait cognitive reappraisal ($_{adj}R^2$ = -.003, $F(1, 67) = 0.81$, $p = .37$). With regard to trait rumination, variation in power state level did not predict trait rumination and explained less than one percent of the variance in trait cognitive reappraisal ($_{adj}R^2$ = .01, $F(1, 67) = 0.04$, $p = .85$).

GPS

Variation in GPS distance did not predict trait cognitive reappraisal and explained less than one percent of the variance in trait cognitive reappraisal ($_{adj}R^2$ = -.02, F(1, 67) = 0.02, p = .89). Variation in GPS distance did not predict trait expressive suppression and explained less than one percent of the variance in trait cognitive reappraisal ($_{adj}R^2$ = -.006, F(1, 67) = 0.60, p = .44). Variation in GPS distance did not predict trait rumination and explained less than one percent of the variance in trait cognitive reappraisal ($_{adj}R^2$ = -.004, F(1, 67) = 0.71, p = .40).).

Unsupervised Learning Models

K-means clustering analysis was conducted to examine how self-reported trait cognitive emotion regulation scores grouped together. Scores from the ERQ-R, ERQ-S, RRS, and the DERS were included in the clustering analysis. As *K*-means clustering requires *k* to be specified, the optimal *k* was determined through observing results from the Elbow method, Silhouette method, and the gap statistic (James et al., 2013). Across optimal *k* methods, *k* = 2 was optimal (2 clusters). The two trait emotion regulation clusters were used as the dependent variable in subsequent *K*-nearest neighbors classification analyses.

Two k-Nearest Neighbor Classifier models were conducted to examine the extent that digital phenotyping data predicted the previously determined two trait emotion regulation clusters from the *k*-means cluster analysis. The first model evaluated power state level and time as classifiers, and the second model evaluated GPS distance and time as classifiers. Model performance indices including accuracy, specificity, and sensitivity for both *k*-Nearest Neighbors models are presented in Table 2. Optimal *k*-neighbors were selected for each model based on maximizing specificity and sensitivity.

Table 2.

Results and Confusion Matrices for Trait Emotion Regulation K-Nearest Neighbors Classification Models

				k Neighbors	Accuracy	Specificity	Sensitivity
Power State Level and Time[a]							
	Predicted			5	0.93	0.86	0.95
Actual	Cluster 1	Cluster 2	% Correct				
Cluster 1	19653	893	95.65				
Cluster 2	997	6171	86.03				
GPS Distance and Time[b]							
	Predicted			8	0.98	0.97	0.96
Actual	Cluster 1	Cluster 2	% Correct				
Cluster 1	272550	2611	99.01				
Cluster 2	2993	79677	96.37				

Note: [a]Matthews correlation coefficient = .82. [b]Matthews correlation coefficient = .96.

State Emotion Regulation and Digital Phenotyping

Power State

Variation in power state level marginally predicted variation in negative affect (b = -29.58, t(67) = -1.76, p = .08, 95% CI [-66.47, -6.72]) and explained three percent of the variance in state negative affect over time ($_{adj}R^2$ = .03, F(1, 67) = 3.13, p = .08). Variation in power state level over time significantly predicted variation in state cognitive reappraisal (b = -4.98, t(67) = -2.89, p = .005, 95% CI [-8.09, -0.21]) and explained eleven percent of the variance in state cognitive reappraisal over time ($_{adj}R^2$ = .11, F(1, 67) = 8.37, p = .005). Variation in power state level over time did not predict variation in state expressive suppression and explained less than one percent of the variance in state expressive suppression over time ($_{adj}R^2$ = -.003, F(1, 67) = 0.82, p = .37).

GPS

Variation in GPS distance over time significantly predicted variation in state negative affect (b = 0.005, t(67) = 2.66, p = .01, 95% CI [0.002, 0.009]) and explained nine percent of the variance in state

negative affect ($_{adj}R^2$ = .09, F(1, 67) = 7.05, p = .01). Variation in GPS distance over time significantly predicted variation in state cognitive reappraisal over time (b = 0.0004, t(67) = -2.38, p = .02, 95% CI [-0.0007, -0.0001]) and explained seven percent of the variance in state cognitive reappraisal over time ($_{adj}R^2$ = .07, F(1, 67) = 5.68, p = .02). Variation in GPS distance over time did not predict variation in state expressive suppression over time and explained less than one percent of the variance in state expressive suppression over time ($_{adj}R^2$ = -.01, F(1, 67) = 0.10, p = .75).

See Table 3 for full results of the models evaluating the relationship between digital phenotyping and both trait and state emotion regulation.

Table 3.

Regression Models for Power State Level and GPS Distance Predicting Trait and State Emotion Regulation

Predictor	*b*	*SE*	95% CI	*p*	Predictor	*b*	*SE*	95% CI	*p*
Hypothesis 1									
Trait Cognitive Reappraisal Model: Power Level					**Trait Cognitive Reappraisal: GPS**				
(Intercept)	32.30	5.08	17.11, 39.34	<.0001	(Intercept)	28.41	1.40	25.95, 30.73	<.0001
Power Level SD	-16.58	20.53	-45.30, 39.55	.42	GPS Distance SD	-0.0002	0.002	-0.004, 0.002	.89
Trait Expressive Suppression: Power Level					**Trait Expressive Suppression: GPS**				
(Intercept)	18.12	2.93	10.74, 22.65	<.0001	(Intercept)	16.01	0.81	14.30, 17.50	<.0001
Power Level SD	-10.64	11.84	-30.86, 18.64	.37	GPS Distance SD	-0.001	0.001	-0.003, 0.002	.44
Rumination: Power Level					**Rumination: GPS**				
(Intercept)	26.34	4.07	18.15, 31.24	<.0001	(Intercept)	26.36	1.12	24.18, 28.60	<.0001
Power Level SD	3.05	16.43	-16.35, 34.39	.85	GPS Distance SD	0.001	0.002	-0.001, 0.004	.40
Hypothesis 2									
State Cognitive Reappraisal SD: Power Level					**State Cognitive Reappraisal SD: GPS**				
(Intercept)	2.73	0.42	1.53, 3.47	<.0001	(Intercept)	1.72	0.13	1.44, 2.00	<.001

Power Level SD	-4.98	1.72	-8.09, -0.21	.005	GPS Distance SD	-0.0004	0.0002	-0.0007, -0.0001	.02
State Expressive Suppression SD: Power Level					**State Expressive Suppression SD: GPS**				
(Intercept)	1.35	0.29	0.85, 1.62	<.0001	(Intercept)	1.07	0.09	0.89, 1.26	<.0001
Power Level SD	-1.08	1.19	-2.31, 0.98	.37	GPS Distance SD	0.0000	0.0000	-0.0003, 0.0002	.75
State Affect SD: Power Level					**State Affect SD: GPS**				
(Intercept)	15.25	4.12	10.20, 24.51	<.0001	(Intercept)	6.62	1.47	4.22, 9.53	<.001
Power Level SD	-29.58	16.73	-66.47, -6.72	.08	GPS Distance SD	0.005	0.002	0.002, 0.009	.01

Depressive Symptoms, Emotion Regulation, and Digital Phenotyping

In the models predicting depressive symptoms (Table 4), the main effects for variation in power state level and GPS distance were non-significant (all *p*s > .05). Moreover, in the same models, the interaction between these digital phenotyping indices and trait and state emotion regulation were also non-significant (all *p*s > .05). See Table 4 for full results of the models evaluating the relationship between digital depressive symptoms, trait and state emotion regulation, and digital phenotyping.

Table 4.

Regression Models Predicting Depressive Symptoms

Predictor	*b*	*SE*	95% CI	*p*	Predictor	*b*	*SE*	95% CI	*p*
Hypothesis 3									
Model 1					**Model 2**				
(Intercept)	13.37	0.89	11.65, 15.25	<.0001	(Intercept)	13.25	0.87	11.61, 15.08	<.0001
ERQ-R	-0.24	0.12	-0.45, 0.02	.04	ERQ-R	-0.23	0.11	-0.43, 0.009	.05
Power Level SD	-17.27	13.56	-40.61, 16.96	.21	Distance SD	0.001	0.002	-0.002, 0.004	.52
ERQ-R*Power Level SD	2.23	2.51	-3.11, 6.10	.38	ERQ-R*Distance SD	-0.0005	0.0003	-0.009, 0.0003	.14
Model 3					**Model 4**				

(Intercept)	13.72	0.98	11.75, 15.76	<.0001
ERQ-S	-0.20	0.23	-0.70, 0.27	.38
Power Level SD	-26.97	18.11	-66.63, 6.85	.14
ERQ-S*Power Level SD	4.73	3.97	-2.99, 14.41	.23

(Intercept)	13.28	0.91	11.62, 14.93	<.0001
ERQ-S	-0.04	-.21	-0.44, 0.01	.86
Distance SD	0.001	0.002	-0.002, 0.004	0.67
ERQ-S*Distance SD	0.0001	0.0004	-0.0009, 0.0003	0.49

Model 5

(Intercept)	13.27	0.85	11.63, 15.05	<.0001
RRS	0.42	0.14	0.06, 0.66	.004
Power Level SD	-13.74	12.4	-37.61, 9.46	.27
RRS*Power Level SD	-0.79	2.15	-4.41, 6.40	.71

Model 6

(Intercept)	13.29	0.87	11.77, 15.29	<.0001
RRS	0.41	0.15	0.06, 0.66	.007
Distance SD	0.00	0.002	-0.004, 0.005	.96
RRS*Distance SD	0.00	0.00	-0.0009, 0.0005	.77

Hypothesis 4

Model 7

(Intercept)	15.79	4.81	5.59, 29.20	.002
SARS-R SD	0.11	2.36	-8.11, 8.06	.96
Power Level SD	-10.66	19.82	-63.62, 33.68	.59
SARS-R*Power Level SD	0.98	9.75	-33.33, 31.86	.92

Model 8

(Intercept)	13.57	1.38	11.10, 16.28	<.0001
SARS-R SD	1.40	0.88	-0.27, 3.04	.09
Distance SD	-0.001	0.002	-0.004, 0.004	.76
SARS-R SD*Distance SD	-0.002	0.001	-0.004, 0.001	.14

Model 9

(Intercept)	14.73	4.82	4.59, 27.46	.003
SARS-S SD	0.86	3.26	-9.19, 9.49	.79
Power Level SD	-6.16	20.01	-55.63, 37.00	.76
SARS-S SD*Power Level SD	2.65	14.01	-35.60, 43.77	.85

Model 10

(Intercept)	13.09	1.33	10.61, 15.83	<.0001
SARS-S SD	1.28	1.14	-1.82, 3.50	.26
Distance SD	0.00	0.002	-0.004, 0.004	.87
SARS-S SD*Distance SD	0.00	0.002	-0.005, 0.005	.91

Model 11

(Intercept)	16.74	4.84	6.91, 30.57	.001
VAS Negative SD	0.30	0.27	-0.73, 1.32	.27

Model 12

(Intercept)	13.06	1.34	10.54, 15.49	<.0001
VAS Negative SD	0.14	0.09	-0.02, 0.31	.11

Power Level SD	-14.85	20.02	-71.00, 24.72	.46	Distance SD	0.0003	0.002	-0.003, 0.004	.86
VAS SD*Power Level SD	-1.02	1.12	-5.19, 3.16	.37	VAS SD*Distance SD	0.0002	0.0001	-0.0004, 0.0002	.25

Note: Abbreviations: ERQ-R, Emotion Regulation Questionnaire cognitive reappraisal subscale; ERQ-S, Emotion Regulation Questionnaire expressive suppression subscale; RRS, Ruminative Responses Scale; SARS-R, Spontaneous Affect Regulation cognitive reappraisal subscale; SARS-S, Spontaneous Affect Regulation expressive suppression subscale; VAS, Visual Analog Scale; SD, Standard Deviation.

Table 5

Digital Phenotyping Data Characteristics

Variable	*N Observations*	% Missing	*M*	*SD*	Range
Power state					
Power State level	186593	22	0.48	0.27	0-1
GPS					
Distance	1789156	0	33.2	1108.38	0-485472

Note. GPS distance was transformed to meters. Power level was measured in percentage of battery level ranging from 0% to 100%.

Discussion

This study sought to expand the nomological network of emotion regulation by evaluating the associations between digital phenotyping data and emotion regulation, guided by two principle aims. The first was to examine digital behavior correlates of self-report emotion regulation (state and trait). The second aim was to determine whether self-reported emotion regulation interacts with digital phenotypes to predict emotional well-being.

Results showed that variation in mobile power state level and GPS distance were associated with variation in negative state affect and state cognitive reappraisal over time, with greater variation in these digital behaviors possibly reflecting less successful down regulation of negative emotions. These findings broadly align with previous work by demonstrating that digital phenotyping data (e.g., text and phone use patterns) can be predictive of EMA-based measures of mood (Asselbergs et al., 2016; Hung et al., 2016). While replication is needed, our study advances this prior work by also showing GPS-based indices predict affective variation in daily life. Additionally, this is the first study to demonstrate an

association between digital phenotyping data and emotion regulation implementation; it would behoove researchers to further examine state emotion regulation (particularly cognitive reappraisal) as it relates to other, previously established digital phenotyping predictors of mood (i.e., SMS, accelerometer). As we operationalized the two digital phenotyping indices (i.e., GPS and power state) as variation over time, direct inference regarding specific digital phenotyping behaviors and emotion regulation is limited. Future research can build on these findings by further isolating specific digital phenotyping behaviors as they predict individual differences in emotion regulation broadly. Future studies using digital phenotyping may also seek to explore the discrepancy between state-based EMA and global trait-based measures of emotion regulation. Consistent with our study, recent research has suggested these may frequently be discrepant (Koval et al., 2023), as trait emotion regulation questionnaires may not fully capture individual differences in how strategy selection and implementation occur in the lab or in naturalistic settings. The discrepancy between state-based EMA and trait measures of emotion regulation may additionally be reflective of emotion regulation choice in response to context, as individuals may flexibly implement different emotion regulation strategies based on contextual cues and perceived severity of an environmental stressor (Sheppes et al., 2014).

Power state level and GPS distance across time accurately classified two trait emotion regulation clusters. One cluster was characterized by less difficulty in implementing emotion regulation strategies, greater cognitive reappraisal use, and less expressive suppression use. The other cluster was characterized by greater difficulty with implementing emotion regulation strategies, low cognitive reappraisal use, and high expressive suppression. Both clusters were characterized by relatively low use of rumination on average. In line with the present study, Saeb et al. (2015) used GPS data operationalized as favorite locations and location variance to accurately classify individuals with significant depressive symptoms from those without. Meyerhoff et al. (2021) also used similar GPS (e.g., location variability and number of locations visited) and phone usage indicators (e.g., texting and email)

to predict changes in symptoms among four groups with distinct symptom clusters—they found differential associations between digital phenotyping indicators and symptom change in those with comorbid depression and anxiety and those without comorbidities. Digital phenotyping data has also been shown to predict differences in reported positive and negative affect, anxiety, and energy over time between individuals with Bipolar Disorder and non-psychiatric controls (Ortiz et al., 2017). Although Ortiz and colleagues (2017) did not specify which digital phenotype indices were used, they noted the relationship between digital phenotyping data and psychological variables was not linear, which is consistent with the results of our classification analysis and the numerous non-significant results for our linear models.

Variation in mobile phone power level and GPS distance over time together with state and trait emotion regulation did not predict current depressive symptoms. To date, the modeling approaches and operationalizations of digital phenotyping data have resulted in mixed findings in terms of predicting psychopathology. For instance, while many have found significant relationships between various GPS-based indices and depressive symptoms (Cao et al., 2020; Meyerhoff et al., 2021; Zhang et al., 2022), some have yielded more equivocal results in which it was unclear whether GPS-data incrementally predicted self-reported symptomatology beyond clinician ratings, daily smaller surveys of well-being and symptoms, and/or physiological indices (Currey & Torous, 2022; Pedrelli et al., 2020; Pellegrini et al., 2022). Considering these findings, individual digital phenotyping measures may best be viewed as offering a more complete picture of depressive behavior without independently signifying the entire set of one's symptoms. Alternatively, mixed findings may be a matter of differences in operationalization of digital phenotyping indices. For instance, in a comparable pilot project examining a sample of outpatients diagnosed with Schizophrenia Spectrum Disorder (Barnett et al., 2018), digital phenotyping indices were operationalized as frequency of unique daily measurements. Lower frequency of accelerometer and GPS data was weakly associated with higher (worse) scores on psychopathology

outcomes, while slower survey completion rates were associated with worse psychopathology symptom outcomes. Compared to the present study, the use of frequency of data-point acquisition may have been more reflective of the participant's amount of on-person phone use (i.e., accelerometer, or device movement) or general study adherence.

The present study has several strengths that meaningfully contribute to prior findings as well as methodology in emotion regulation research. The current study addressed methodological issues present in previous emotion regulation research. Although prior research has employed self-report, behavioral, neurological, genetic, and physiological assessments of emotion regulation, the majority of research on emotion regulation has been circumscribed to controlled laboratory settings that use experimental paradigms to investigate short-term outcomes, and EMA studies largely rely on self-report of emotional experiences (Berking & Wupperman, 2012). To date, the present study is the first to apply digital phenotyping methodology to the investigation of emotion regulation phenomena. Using a digital phenotyping approach, the current study concurrently collected passive behavioral data and subjective affective experiences in naturalistic settings. One strength of this approach is that it is consistent with the RDoC framework for capturing individual differences in psychological phenomena (Cuthbert, 2014; Torous et al., 2017). Furthermore, the high sampling rate of digital phenotyping data allowed for moment-by-moment quantification of digital behavior and higher fidelity of changes in behavioral patterns (Insel, 2017). Since digital phenotyping acquisition is passive, it addressed potential participant burden from too frequently responding to self-report inquiries.

Another strength of the current study was the incorporation of measuring spontaneous cognitive emotion regulation strategies. The majority of prior research conventionally operationalizes spontaneous emotion as change in state affect or by instructing participants to engage in a particular strategy following exposure to an emotion-eliciting stimuli (Dixon-Gordon et al., 2015). Spontaneous emotion regulation was measured using both change in affect and spontaneous cognitive emotion

regulation strategy implementation (i.e., cognitive reappraisal and expressive suppression) in response to a validated negative mood induction as well as sampled in naturalistic settings using EMA methodology. Building on recent research on spontaneous emotion regulation (e.g., Gruber et al., 2012; Stange et al., 2017), this is the first study to incorporate spontaneous emotion regulation strategy implementation using EMA.

These findings indicate that operationalizations of digital phenotyping data and modeling methods are especially important to consider when using digital phenotyping methodology broadly and in the study of emotion regulation. Unsupervised learning approaches (i.e., clustering) predicted individual differences in emotion regulation while supervised learning (i.e., linear regression) was less predictive. Given that unsupervised learning statistical analyses provide relatively less inference compared to supervised approaches, future research using digital phenotyping methods can build on these findings by investigating operationalizations of digital phenotyping data that better fit supervised learning statistical models. Alternatively, there may be more utility to implementing unsupervised learning modeling methods as the associations between digital phenotypes and mental health constructs may be better explained by within-individual variation over time.

Limitations

There are several caveats to interpreting our digital phenotyping data that are important to note for future applications. First, GPS technology cannot detect whether an individual is inside or outside a building and it is not accurate enough to detect movement within a small space, so it may not be sensitive enough to capture all individual differences in behavior associated with psychopathology. Future work that plans to use GPS data from digital phenotyping platforms would benefit from developing research questions that take into account the inherent limitations of GPS data acquisition in smartphone technology. Since GPS is not sensitive to within-building movement, researchers can create algorithms to detect when and for how long individuals are in certain geographical areas and determine

whether the individual is likely at home, work, or another frequently visited location. Second, even though power state level and GPS digital phenotyping data were sampled at a high rate, operationalizing variation over time as *SD* limited the analyzed sample to 69 observations; however, the thousands of digital phenotyping observations contributed to better estimations of *SD* for each participant. Third, digital phenotyping data indices are sampled at different rates and at different times than self-report data, generating variables of varying lengths and time stamps that do not match. The differences in number of observations between digital phenotyping indices and non-matching timestamps preclude the ability to conduct direct time series analyses comparing digital phenotype indices. The same issue exists when attempting to conduct time series analyses comparing self-report measures collected concurrently with digital phenotyping data. Future studies can take further advantage of the longitudinal nature of digital phenotyping data by creating time index variables to approximately match observations for time series analyses. Additionally, the study participants were university students, potentially limiting the extent to which the results can be generalized to other populations.

It is notable that data collection occurred during a period when awareness, and institutional response to, the COVID-19 pandemic began in the United States. Data collection ended the week preceding university closures and the CDC declaring COVID-19 as a pandemic. Since its onset, the COVID-19 pandemic has had extensive and lasting global impacts on emotional well-being. Although the university where the study was conducted was one of the earlier adopters of in-person school closure and transition to remote learning compared to others in the United States, participants may have already been experiencing COVID-19 pandemic-related stressors. The impacts of the COVID-19 pandemic on emotion regulation is worthy of further investigation; however, the time frame of the data collection for the present study precludes generalizing the results to the consequences of the COVID-19 pandemic. Nonetheless, the study was designed to investigate whether digital phenotyping was associated with individual differences in emotion regulation in naturalistic settings, with participants

responding to any variety of life stressors, which would include the impact of any COVID-19 pandemic-stressors.

Conclusions

Emotion regulation has been proposed as an RDoC domain for transdiagnostic criteria for psychopathology (Fernandez et al., 2016). As the field of emotion regulation has relatively robust foundational knowledge of individual differences in emotion regulation in laboratory settings, a complete understanding of normative ranges in emotion regulation requires an unobtrusive, ecologically valid assessment of the construct as it occurs in real-world settings. Therefore, comprehensive emotion regulation assessment depends on our ability to harness innovative interdisciplinary methodology to advance our understanding of the implicit and passive experience of emotion. Digital phenotyping is a promising methodological approach that can help accurately identify individual differences in emotion regulation implementation, which can improve our ability to identify and treat emotion regulation deficits and promote or teach adaptive emotion regulation skills. In a more applied sense, a robust understanding of normative ranges of emotion regulation, as well as reliable measurement, could contribute to clinical applications where healthcare providers could more accurately identify, treat, and monitor mental health issues. The applications are potentially analogous to how physicians have patients wear heart monitors to detect specific cardiovascular issues. To our knowledge, the present study was the first to implement digital phenotyping in the investigation of emotion regulation. Results from the current investigation suggest that digital behavior can predict individual differences in trait emotion regulation implementation. Building on the findings of this study, future research can investigate additional digital phenotyping indices (e.g., call logs, app usage, screen on/off tracking, alternative operationalizations of digital phenotyping data, and their relationship to other markers of emotion regulation. For example, additional digital phenotyping indices could be collected using wearable devices that can acquire psychophysiological data or by connecting supplementary sensors to

the outside of smartphones. In line with the RDoC initiative (Torous et al., 2017). these findings provide a foundation for future research using digital phenotyping with the ultimate goal of accurately identifying normative ranges in emotion regulation associated with mental well-being. The study findings offer initial and fundamental insights into how digital phenotyping data can predict individual differences in state and trait emotion regulation in both controlled and real-world settings. They highlight that the way digital phenotyping data is operationalized, and the modeling techniques used, are crucial considerations when applying this methodology to mental health research, particularly in the study of emotion regulation.

Open Science

An additional strength of the current study is that it was conducted following open-science practices. The project was pre-registered on the Open Science Framework (Bosma, 2020a). The open-access version of the Beiwe Research Platform was used for collecting digital phenotyping data. In addition, the code for processing data collected using the Beiwe Research Platform is hosted on a public repository with supporting documentation (Bosma, 2020b).

References

Ahearn, E. P. (1997). The use of visual analog scales in mood disorders: A critical review. *Journal of Psychiatric Research, 31*(5), 569-579. https://doi.org/10.1016/s0022-3956(97)00029-0

Aldao, A. (2013). The future of emotion regulation research. *Perspectives on Psychological Science, 8*(2), 155-172. https://doi.org/10.1177/1745691612459518

Aldao, A., Gee, D. G., De Los Reyes, A., & Seager, I. (2016). Emotion regulation as a transdiagnostic factor in the development of internalizing and externalizing psychopathology: Current and future directions. *Development and Psychopathology, 28*(4pt1), 927-946. https://doi.org/10.1017/s0954579416000638

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review, 30*(2), 217-237. https://doi.org/10.1016/j.cpr.2009.11.004

American Psychiatric Association. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Vol. 167*. https://doi.org/10.1176/appi.ajp.2010.09091379

Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M., & Riper, H. (2016). Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: An explorative study. *Journal of Medical Internet Research, 18*(3), e72. https://doi.org/10.2196/jmir.5505

Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J. P. (2018). Relapse prediction in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology, 43*(8), 1660-1666. https://doi.org/10.1038/s41386-018-0030-z

Berking, M., & Wupperman, P. (2012). Emotion regulation and mental health. *Current Opinion in Psychiatry, 25*(2), 128-134. https://doi.org/10.1097/yco.0b013e3283503669

Bosma, C. M. (2020a). beiwe-scripts. Retrieved from github.com/cmbosma/beiwe-scripts.

Bosma, C. M. (2020b). Ecological validity of emotion regulation. Retrieved from osf.io/wdsvj.

Cao, J., Truong, A. L., Banu, S., Shah, A. A., Sabharwal, A., & Moukaddam, N. (2020). Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: Development and usability study. *JMIR Mental Health, 7*(1), e14045.

Currey, D., & Torous, J. (2022). Digital phenotyping correlations in larger mental health samples: Analysis and replication. *BJPsych Open, 8*(4), e106.

Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, *13*(1), 28–35. https://doi.org/10.1002/wps.20087

Dixon-Gordon, K. L., Aldao, A., & De Los Reyes, A. (2015). Emotion regulation in context: Examining the spontaneous use of strategies across emotional intensity and type of emotion. *Personality and Individual Differences*, *86*(December), 271–276. https://doi.org/10.1016/j.paid.2015.06.011

Egloff, B., Schmukle, S. C., Burns, L. R., & Schwerdtfeger, A. (2006). Spontaneous emotion regulation during evaluated speaking tasks: Associations with negative affect, anxiety expression, memory, and physiological responding. *Emotion, 6*(3), 356-366. https://doi.org/10.1037/1528-3542.6.3.356

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/bf03193146

Feinstein, A. R. (1987). Clinimetric perspectives. *Journal of Chronic Diseases, 40*(6), 635-640. https://doi.org/10.1016/0021-9681(87)90027-0

Fernandez, K. C., Jazaieri, H., & Gross, J. J. (2016). Emotion regulation: A transdiagnostic perspective on a new RDoC domain. *Cognitive Therapy and Research, 40*(3), 426-440. https://doi.org/10.1007/s10608-016-9772-2

Gratz, K. L., & Roemer, L. (2008). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment, 30*(4), 315-321. https://doi.org/10.1007/s10862-008-9102-4

Gross, J. J. (2014). Emotion Regulation: Conceptual and Empirical Foundations. *Handbook of Emotion Regulation*, 3–20. https://doi.org/10.1080/00140130600971135

Gross, J. J. (2015a). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1-26. https://doi.org/10.1080/1047840x.2014.940781

Gross, J. J. (2015b). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry, 26*(1), 130-137. https://doi.org/10.1080/1047840X.2015.989751

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348-362. https://doi.org/10.1037/0022-3514.85.2.348

Gruber, J., Harvey, A. G., & Gross, J. J. (2012). When trying is not enough: Emotion regulation and the effort–success gap in bipolar disorder. *Emotion, 12*(5), 997-1003. https://doi.org/10.1037/a0026822

Hung, G. C. L., Yang, P. C., Chang, C. C., Chiang, J. H., & Chen, Y. Y. (2016). Predicting negative emotions based on mobile phone usage patterns: An exploratory study. *JMIR Research Protocols, 5*(3), e160. https://doi.org/10.2196/resprot.5551

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA - Journal of the American Medical Association*, *318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Statistical learning. *Springer Texts in Statistics*. https://doi.org/10.1007/978-1-4614-7138-7_2

Joormann, J., & Siemer, M. (2014). Emotion regulation in mood disorders. In *Handbook of Emotion Regulation* (2nd ed.). https://doi.org/10.1016/2014

Joormann, J., & Vanderlind, W. M. (2014). Emotion regulation in depression. *Clinical Psychological Science, 2*(4), 402-421. https://doi.org/10.1177/2167702614536163

Koval, P., Kalokerinos, E. K., Greenaway, K. H., Medland, H., Kuppens, P., Nezlek, J. B., Hinton, J. D. X., & Gross, J. J. (2023). Emotion regulation in everyday life: Mapping global self-reports to daily processes. *Emotion, 23*(2), 357–374. https://doi.org/10.1037/emo0001097

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kuppens, P., & Verduyn, P. (2015). Looking at emotion regulation through the window of emotion dynamics. *Psychological Inquiry, 26*(1), 72-79. https://doi.org/10.1080/1047840x.2015.960505

Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology, 17*, 22-26. https://doi.org/10.1016/j.copsyc.2017.06.004

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209-237. https://doi.org/10.1080/02699930802204677

Mennin, D. S., & Fresco, D. M. (2015). Advancing emotion regulation perspectives on psychopathology: The challenge of distress disorders. *Psychological Inquiry, 26*(1), 80-92. https://doi.org/10.1080/1047840X.2015.969624

Meyerhoff, J., Liu, T., Kording, K. P., et al. (2021). Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: Longitudinal cohort study. *Journal of Medical Internet Research, 23*(9), e22844.

Onnela, J. P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41*(7), 1691-1696. https://doi.org/10.1038/npp.2016.7

Ortiz, A., Bradler, K., Garnham, J., MacLean, S., & Alda, M. (2017). Studying mood in those who do not get sick: Nonlinear dynamics of mood regulation in first-degree healthy relatives of bipolar disorder patients. *Biological Psychiatry, 81*(10), S131-S132. https://doi.org/10.1016/j.biopsych.2017.02.335

Pedrelli, P., Fedor, S., Ghandeharioun, A., Howe, E., Ionescu, D. F., Bhathena, D., Fisher, L. B., Cusin, C., Nyer, M., Yeung, A., Sangermano, L., Mischoulon, D., Alpert, J. E., Picard, R. W. (2020). Monitoring changes in depression severity using wearable and mobile sensors. *Frontiers in Psychiatry*, *11*, 584711. https://doi.org/10.3389/fpsyt.2020.584711

Pellegrini, A. M., Huang, E. J., Staples, P. C., Hart, K. L., Lorme, J. M., Brown, H. E., Perlis, R. H., & Onnela, J. P. (2022). Estimating longitudinal depressive symptoms from smartphone data in a transdiagnostic cohort. *Brain and Behavior*, *12*(2), e02077.

Quinn, M. E., Keil, D. C., Utke, S., & Joormann, J. (2014). Trait rumination moderates the effect of executive control training. *Journal of Experimental Psychopathology, 5*(3), 289-301.

Rottenberg, J., Ray, R. D., & Gross, J. (2007). Emotion elicitation using films. *Handbook of Emotion Elicitation and Assessment*, *9*.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research*, *17*(7), 1–11. https://doi.org/10.2196/jmir.4273

Seppälä, J., De Vita, I., Jämsä, T., Miettunen, J., Isohanni, M., Rubinstein, K., Feldman, Y., Grasa, E., Corripio, I., Berdun, J., D'Amico, E., M-RESIST Group, Bulgheroni, M. (2019). Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review. *JMIR Mental Health*, *6*(2), e9819. https://doi.org/10.2196/mental.9819

Sheppes, G., Scheibe, S., Suri, G., Radu, P., Blechert, J., & Gross, J. J. (2014). Emotion regulation choice: A conceptual framework and supporting evidence. *Journal of Experimental Psychology: General*, *143*(1), 163–181. https://doi.org/10.1037/a0030831

Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual Review of Clinical Psychology, 11*(1), 379-405. https://doi.org/10.1146/annurev-clinpsy-032814-112739

Stange, J. P., Hamilton, J. L., Fresco, D. M., & Alloy, L. B. (2017). Flexible parasympathetic responses to sadness facilitate spontaneous affect regulation. *Psychophysiology, 54*(7), 1054-1069. https://doi.org/10.1111/psyp.12856

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J. P. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data-driven smartphone research. *JMIR Mental 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

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research, 27*(3), 247-259. https://doi.org/10.1023/A:1023910315561

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1-67. https://doi.org/10.18637/jss.v045.i03

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

Zhang, Y., Folarin, A. A., Sun, S., et al. (2021). Predicting depressive symptom severity through individuals' nearby Bluetooth device count data collected by mobile phones: Preliminary longitudinal study. *JMIR mHealth and uHealth, 9*(7), e29840.

Zhang, Y., Folarin, A. A., Sun, S., et al. (2022). Longitudinal relationships between depressive symptom severity and phone-measured mobility: Dynamic structural equation modeling study. *JMIR Mental Health, 9*(3), e34898.