Published in final edited form as:

J Affect Disord. 2019 December 01; 259: 325–336. doi:10.1016/j.jad.2019.08.060.

Using Ambulatory Assessment to Measure Dynamic Risk Processes in Affective Disorders

Jonathan P. Stange, Ph.D.[1], **Evan M. Kleiman, Ph.D.**[2], **Robin J. Mermelstein, Ph.D.**[1], **Timothy J. Trull, Ph.D.**[3]

[1]University of Illinois at Chicago

[2]Rutgers, The State University of New Jersey

[3]University of Missouri

Abstract

Background: Rapid advances in the capability and affordability of digital technology have begun to allow for the intensive monitoring of psychological and physiological processes associated with affective disorders in daily life. This technology may enable researchers to overcome some limitations of traditional methods for studying risk in affective disorders, which often (implicitly) assume that risk factors are distal and static – that they do not change over time. In contrast, ambulatory assessment (AA) is particularly suited to measure dynamic "real-world" processes and to detect fluctuations in proximal risk for outcomes of interest.

Method: We highlight key questions about proximal and distal risk for affective disorders that AA methods (with multilevel modeling, or fully-idiographic methods) allow researchers to evaluate.

Results: Key questions include between-subject questions to understand *who* is at risk (e.g., are people with more affective instability at greater risk than others?) and within-subject questions to understand *when* risk is most acute among those who are at risk (e.g., does suicidal ideation increase when people show more sympathetic activation than usual?). We discuss practical study design and analytic strategy considerations for evaluating questions of risk in context, and the benefits and limitations of self-reported vs. passively-collected AA.

Limitations: Measurements may only be as accurate as the observation period is representative of individuals' usual life contexts. Active measurement techniques are limited by the ability and willingness to self-report.

Conclusions: We conclude by discussing how monitoring proximal risk with AA may be leveraged for translation into personalized, real-time interventions to reduce risk.

Introduction

Until recently, the study of risk in affective disorders has been handcuffed by the methods available for measuring these factors and processes (aan het Rot et al., 2012; Armey et al.,

Correspondence concerning this article should be addressed to Jonathan P. Stange, Center on Depression and Resilience, Department of Psychiatry, University of Illinois at Chicago, 1601 W Taylor St., Chicago, IL 60612; Phone: 312-355-1106; jstange@uic.edu.

2015). For example, over the past several decades research on cognitive risk factors typically has measured risk factors and outcomes retrospectively across intervals of months to years, with some success in identifying which individuals have an elevated likelihood of experiencing future outcomes such as depression (Abela and Hankin, 2011; Alloy et al., 2017, 2006). These methods have had some utility for examining theoretical questions related to static, distal risk factors that do not change over time. However, many cognitive and affective processes are dynamic, occurring at a rapid pace as they unfold in the context of daily life. These risk processes are challenging to measure without intensive, repeated sampling throughout the day.

Fortunately, rapid advances in the capability and affordability of digital technology over the past decade has begun to allow for the intensive monitoring of psychological and physiological processes in daily life (Malhi et al., 2017). Furthermore, the increasing ubiquity of smartphones (carried throughout the day by >90% of Americans ages 18–49 (Pew Research Center, 2018) means that questions can be administered instantaneously in "real-world" contexts (Trull and Ebner-Priemer, 2013), rather than waiting for participants to return to the lab or clinic to complete assessments. Known collectively as ambulatory assessment (AA) (Trull and Ebner-Priemer, 2013), these technologies afford novel avenues for the assessment of dynamic processes related to risk, but have been underutilized to date. AA techniques can involve ecological momentary assessment (EMA) of responses to surveys on smartphones, passive recording from smartphones (e.g., keyboard typing behavior, movement with accelerometer, location information with GPS), and other wearable devices for measuring physiological processes such as electrocardiogram, respiration, and sleep actigraphy (for a more detailed review of AA methods, see (Trull and Ebner-Priemer, 2013)). In contrast with traditional assessments, AA allows for the assessment of what is happening *in the moment*, helping to avoid asking participants to report "averages" over a long period of time (e.g., mood over the past month), within which there likely was variability that cannot be captured with such methods.

Importantly, studying people intensively over time may elucidate *proximal* antecedents of outcomes of interest, including risk factors that vary over short periods of time and have utility for indicating *when* an undesirable outcome is likely to occur in the future. This developing field holds great promise for generating novel insights into the mechanistic processes of affective disorders and improving personalized clinical care (Malhi et al., 2017). AA may be particularly important for the study of risk for affective disorders and related phenomena such as suicide (Allen et al., 2019; Kleiman et al., 2018), given existing limitations in the ability to detect *which individuals* are at greatest risk (Chang et al., 2016; Franklin et al., 2017), as well as in identifying *when* people are most likely to be at risk for outcomes such as depression and suicide. Specifically, AA allows for the modeling of dynamic, proximal risk (because assessments happen close to the outcome of interest), and for modeling of individual differences in changes over short periods of time (because assessments happen frequently). In the present report, we highlight some of the key questions about proximal and distal risk for affective disorders that AA methods and multilevel statistical modeling can allow researchers to evaluate. We also discuss how monitoring proximal risk with AA can be leveraged for translation into personalized, real-time interventions to reduce risk.

Types of Risk Factors and Timescales

The identification of risk factors for affective disorders is critical for prevention, early detection, and intervention to improve illness course. Before highlighting the utility of AA for addressing questions of risk, it is worth defining what is meant by a risk factor. Kraemer (Kraemer, 1997) has defined risk factors as measurable individual difference characteristics that precede an outcome of interest (Haggerty and Mrazek, 1994). Whereas some risk factors (e.g., genetic vulnerabilities) cannot change (or be changed by intervention) and may be called fixed risk markers, others (e.g., thoughts or behaviors) can change and can be called variable risk factors (Cicchetti and Toth, 2009; Kraemer, 1997). In this manuscript, we focus primarily on the use of AA for measuring fluctuations in variable risk factors that may be associated with proximal risk. Among variable risk factors, if manipulation of the risk factor changes the associated outcome, it may be a causal risk factor (Kazdin, 2007; Kraemer, 1997) (assuming other measured risk factors are held constant), which has important implications for prevention and intervention. Questions may still remain about the *mechanism or process* by which these causal risk factors operate (Kazdin, 2007; Kraemer, 1997), another area that AA may help to address.

Historically, longitudinal studies of affective disorders have focused primarily on distal risk factors (e.g., female sex (Hankin et al., 1998), rumination (Stange et al., 2016a)), or follow-up assessments that are spaced far (e.g., months or years) apart (Alloy et al., 2017, 2006). Although helpful for some applications (e.g., identifying stable, distal risk factors), many processes involved in risk for affective disorders may be dynamic, occurring on a more proximal time scale (i.e., over the course of seconds or minutes). By studying processes *in situ* or "in the moment," AA allows for the identification of dynamic proximal risk factors. It also may facilitate the development of interventions that can be delivered proximal to an increase in detected risk (e.g., prompting individuals to use regulation strategies when they are needed), rather than only once an outcome has occurred (e.g., waiting for a patient to seek treatment for a mood episode).

What Types of Questions About Risk Does Ambulatory Assessment Allow Us to Ask?

What are the questions that are most relevant to improving detection of risk, and to preventing and treating mood disorders, and how can AA help to answer these questions?

Within-Subjects Questions: When is Risk Most Acute?

One of the most important clinical and empirical questions is about *when* individuals are most at risk for an outcome of interest. By the time a clinical outcome occurs (e.g., a depressive episode, a suicide attempt), it may too late for preventative intervention to take place. This is especially true for clinical outcomes whose onset is rapid (e.g., suicidal thinking, manic episode). AA can allow us to examine what happens *just before* an outcome occurs (e.g., an increase in suicidal ideation, negative affect, or alcohol consumption). Examining proximal antecedents of these outcomes, using frequent sampling with AA, may ultimately allow for interventions to be delivered before moments of crisis, potentially preventing these outcomes from occurring. By naturalistically sampling individuals

repeatedly over time, researchers can examine the extent to which fluctuations in potential risk factors (i.e., variable risk factors) may be temporally related to outcomes of interest.

Many theoretical models of depression propose that individuals with an existing vulnerability factor are most likely to experience depression when they experience negative life events (Abramson et al., 1989; Beck and Bredemeier, 2016). Implicitly, this means that depression (or its antecedents, such as hopelessness) should appear shortly after the occurrence of life events. However, traditionally, empirical tests of these models often have only measured life events at one or two time points retrospectively with a relatively long period between timepoints (e.g., life events over the past four weeks) (Hankin, 2012). This type of limited design necessitates a between-subjects (or nomothetic) approach to measuring variability in negative events – that is, examining whether individuals who experience more negative events are more likely to experience depression than are those who experience fewer negative events (Abela and Hankin, 2008). Although comparing negative event exposure across individuals may be a worthy question of its own, it does not actually test the question that may be of most theoretical relevance – that is, *when* are people most at risk for depression? Perhaps the answer is: *When* they are exposed to more negative events than usual (Abela and Hankin, 2008). Answering this, however, would require repeated assessments of life events to determine what is "usual."

Person-Centered Multilevel Models.—In contrast to the between-subjects approach necessitating nomothetic models that treat all people the same, repeated sampling of negative events over time (as is possible with AA) allows for a within-subjects (or idiographic) approach to examining variability in putative risk factors (Abela and Hankin, 2008). Idiographic approaches are especially important because recent research using AA shows that group-level findings do not often apply to specific individuals from that group (Fisher et al., 2018). With enough samples of a within-subjects variable over time, an estimate of *each person's own mean level* can be computed. By centering each observation of the variable around each person's own mean, researchers can conduct a multilevel analysis that is *person-centered*, examining how each observation of the risk factor compares to that person's own norm (Figure 1). With repeated observations of an outcome variable, a multilevel analysis then can test, for example, whether increases in negative life events (compared to one's usual level) are associated with subsequent increases in depression severity. Thus, a person-centered analysis can examine theoretical questions about *when* outcomes are most likely to occur, which may be a more precise test of what theories actually propose (Abela and Hankin, 2008). More generally, person-centered analyses allow researchers to examine within-subject temporal relationships between two variables (i.e., slopes) that can be compared across people. The person-centered approach to modeling is possible with multi-wave data that is not obtained from AA (Abela and Hankin, 2008; Stange et al., 2017c), but AA can measure affective processes much closer in time to events of interest (Trull and Ebner-Priemer, 2013), which is a necessity for some factors that vary over a short time period. For example, a recent AA study measured daily sleep quality using actigraphy along with awakening levels of suicidal ideation across the course of a week (Littlewood et al., 2019). Person-centered multilevel models indicated that individuals experienced higher levels of ideation after nights *when* they experienced poorer sleep

quality. Recent expansions of person-centered multi-level modeling approaches, such as idiographic network analysis and person-specific factor analysis, offer promise to gain even further insight into both the phenomenology (Fisher et al., 2017) and treatment (Fisher et al., 2019) of affective disorders.

As individuals can differ not only in their mean levels of variables but in the variance in observations around their own mean, one strategy for conducting person-centered analyses is to standardize each individual's (person-centered) observations. This allows individuals with different levels of variance to be compared to one another on person-centered scales that are similar to one another (in which a standardized observation score of "1" represents 1 standard deviation above one's own mean). For example, an investigator could examine whether individuals are at greater risk for depressive symptoms when they experience stress at a level that is greater than one standard deviation above their own usual level. Thus, this method accounts for individual differences in both mean and variance when conducting analyses at the within-subject level. The person-standardizing approach might be useful for understanding levels of confidence around deviations from individuals' own means, which may facilitate the development of "cut points" for intervention (e.g., at > 1 SD from the mean).

Fully Idiographic Models.—Multilevel modeling approaches to person-centered analyses, as discussed above, may only be appropriate if certain assumptions are met. Multilevel modeling assumes homogeneity across participants – that is, that there are certain similarities that generalize amongst individuals. For example, most individuals experience increases in negative affect when they encounter a negative event. Multilevel modeling assumes that the shape of the relationships between variables (e.g., linear) is similar across participants. Random effects can then be used to model person-specific deviations from these common structures in the data (e.g., person-specific slope and intercept). However, multilevel modeling may not be appropriate when these homogeneity requirements are violated, such as when individuals vary in the way that their affect, thoughts, and behaviors fluctuate over time. Repeating features, such as day of the week, time of day, or diurnal cycles, often differ substantially between individuals (Fisher and Bosley, 2019).

In the face of such heterogeneity between participants, *fully idiographic* modeling can be used to compute specific combinations of risk factors that are relevant to each individual. Using each person's own data to generate a person-specific model has utility for maximally understanding individual behavior (Fisher et al., 2018). As the number and nature of predictors often varies across individuals (Fisher and Soyster, 2019), to tailor interventions toward each person's own risk factors requires constructing prediction models on a case-by-case basis. Machine learning approaches can be used to define the most relevant variables (sometimes called the "feature space"), and to reduce the dimensionality of the full feature space to the predictors relevant to each person. For example, elastic net regression (Zou and Hastie, 2005) helps to select variables that are significantly associated with the outcome, and protects against model overfitting. Random forest approaches also can be used for feature selection, in conjunction with naïve Bayes classification (Hand and Yu, 2001) to identify the most relevant factors for each person (for an example of this approach, see Fisher and Soyster, 2019). One downside of fully idiographic modeling is that it does not allow for the

generalization of results across individuals (e.g., in the case that there are some commonalities among most individuals of a certain type, such as those with depression). However, between-subject modeling of risk factors also can be conducted, so the two sources of information may be complementary (although idiographic research may be more generalizable to individuals than between-subject/group-level factors; Fisher et al., 2018). Additionally, although the specific risk factors identified by fully idiographic modeling will vary between individuals, the *process* of using fully idiographic methods (e.g., using machine learning to select person-specific risk factors) may be generalizable across individuals (Fisher and Soyster, 2019), thus representing a promising direction for future research.

Lagged (vs. Contemporaneous) Analyses.—Although the naturalistic study of risk phenomena in daily life does not replace the utility of true experimental designs for determining causality, experimental designs often are infeasible for ethical or practical reasons (Alloy et al., 1999). Despite the limitations of correlational designs, using AA to assess temporal associations between factors on a tight time scale allows us to address within-subject questions (such as *when* proximal risk is greatest), and essentially can bring us one step closer to inferring possible causality between risk factor and outcome. Obtaining repeated observations of potential risk factors and outcomes is particularly useful for addressing questions of possible causality, when manipulation of the risk factor is infeasible. This approach allows for regression modeling of *lagged* effects, whereby observations of the potential risk factor can be modeled at time t, predicting the outcome at time $t+1$, while controlling for levels of the outcome at time t (sometimes referred to as autoregressive modeling or Granger causality). If this resulting beta weight of the lagged risk factor is significantly associated with the outcome variable at the next time point, the risk factor can be interpreted as predicting residual change in the outcome (i.e., variance in the outcome at $t+1$ that remains after accounting for levels of the outcome variable at time t). Lagged effects provide a more rigorous test of the presence of a risk factor (which requires that the risk factor precede the outcome variable) than do tests of contemporaneous associations between variables (i.e., a potential risk factor at time t predicting the outcome at time t), which leave open the possibility that there are bidirectional relationships (i.e., that changes in the outcome variable could actually occur prior to changes in the predictor), or that the predictor and outcome could change simultaneously without one preceding the other (in which case one would not be a risk factor for the other).

Prediction Modeling.—In addition to testing the strength of associations between potential risk factors and outcomes, prediction modeling (containing discrete outcome variables, such as whether or not an event occurred) can allow for testing the sensitivity, specificity, and accuracy of risk factors in predicting events of clinical interest (i.e., assessing for the *quality* of the predictions). Although prediction modeling can be conducted on the same data set used to define the predictors, ideally the predictions identified in one model (e.g., regression weights specific to the first study, called “training” data set) are cross-validated in an independent data set (the “testing” set), providing important information about the replicability and generalizability of the results. At an idiographic level of analysis, this can mean sampling data from one person during two different periods of

time. For example, a recent study gathered EMA time series data that were used to identify episodes of smoking behavior (Fisher and Soyster, 2019). After dividing the sample of observations into training and testing portions, idiographic predictions of smoking yielded mean accuracies near 80%. Prediction modeling is critical to determining the clinical utility of identified risk factors before translating studies of risk into personalized interventions.

Between-Subjects Questions: Which Individuals are Most at Risk?

Ambulatory assessment also can be useful for elucidating questions about risk at the between-subjects level of analysis. These approaches may involve comparing individuals to one another in terms of estimates of each person's average level of a factor across a measurement interval, relationships between within-subject factors (i.e., cross-level interactions), or within-subject variability over time.

Intercepts: Are Individuals with Higher Average Levels of a Risk Factor More Likely to Experience an Outcome?—Studying people in their natural environments over time has the potential to provide more accurate estimates of (as well as potentially revised conceptualizations of) *trait-like behaviors or risk factors*, compared to one-time self-report measures, due to various biases associated with past recall over longer periods of time (e.g., depressed individuals may show less accuracy in recall of negative affect (Ben-Zeev et al., 2012); participants report higher levels of suicidal ideation during AA than on weekly retrospective assessments (Torous et al., 2015). First, measures of behavioral characteristics as they occur in the real world should have enhanced ecological validity compared to measures that rely on recall or are measured in artificial environments such as the lab. Second, measuring constructs repeatedly over time allows for the use of latent variable modeling approaches to making inferences about the underlying constructs. For example, if researchers want to know how much individuals engage in rumination when they feel sad, they could (a) ask individuals how much they typically ruminate when they feel sad (Stange et al., 2016a; Treynor et al., 2003); (b) experimentally induce sadness in the lab and then ask individuals how much they ruminated in the lab; or (c) use AA to ask individuals, several times throughout the day across the course of a week, how much they have ruminated in the past few hours. The latter approach is ecologically valid and allows for the computation of a latent intercept value, conceptually representing the average amount that each individual ruminated across the week of sampling (Figure 3). Of course, one limitation of using AA to assess trait-like constructs is that the measurements may only be as accurate as the period is representative of their usual life contexts (e.g., AA during a particularly stressful period might yield unrepresentative estimates of usual patterns). In some ways, however, this may not be a limitation if a study period is long enough to capture multiple periods in one's life, allowing insight into how trait-like vulnerabilities actually interact with different daily contexts.

Slopes: Which Individuals are More at Risk Than Others When They Experience an Increase in a Proximal Risk Factor?—Above, we discussed how within-subject modeling of AA can be used to identify temporal relationships between variables over time (e.g., determining whether an outcome is more likely to occur after an increase in a potential risk factor than after a decrease in that factor). Multilevel modeling

also allows researchers to examine individual differences in relationships between within-subject variables (i.e., individual differences in slopes, or cross-level interactions). For example, theories of depression risk posit that some individuals are more vulnerable than others, and that this vulnerability is more apparent when individuals experience higher (rather than lower) levels of negative events (Alloy et al., 2017, 1999). Thus, the slope of the relationship between negative events and subsequent depression may differ between individuals (as depicted in Figure 2; those vulnerable have a stronger relationship between negative events and depression). This hypothesis can be tested empirically with multilevel modeling using a "slopes-as-outcomes" model in which a between-subject variable (e.g., cognitive vulnerability) is used to predict the slope between two within-subject variables (e.g., negative events and depressive symptoms), which itself is a between-subject variable (i.e., each individual's within-person data is used to calculate one slope per person).

Between-subjects questions such as these would be difficult to answer without the intensive sampling provided by AA, which enables the estimation of a slope of the relationship between two variables for each individual. For example, one AA study (Pe et al., 2013) repeatedly sampled engagement in cognitive reappraisal, along with negative affective experiences, throughout the day. This approach allowed for the estimation of a slope, at the within-subject level, of the relationship between reappraisal and downregulation of negative affect. However, individuals' slopes varied: those individuals who had poorer cognitive control ability showed a weaker relationship between reappraisal and decreases in negative affect – that is, people with poorer cognitive control were less successful (than those with better control) in regulating their negative affect *when* they used reappraisal.

Within-Subject Variability: Are Individuals with More Instability at Greater Risk?—AA also enables the measurement of within-subject dynamics, which may not be observable with single-time point assessments or with assessments that are spaced farther apart (Solhan et al., 2009). Individuals then can be compared to one another in terms of their (possibly trait-like) patterns of variability over time. Longitudinal designs can then test whether variability is associated with future outcomes of interest, such as which individuals will develop a mood episode. One metric of variability in negative affect that has shown utility in identifying individual differences in risk is the mean square successive differences (MSSD) score (Jahng et al., 2008; Scheiderer et al., 2016; Trull et al., 2008), which provides a measure of intra-individual assessment-to-assessment fluctuations that accounts for the magnitude and temporal order of changes (Figure 4). Existing research suggests that individuals with bipolar disorder who have greater variability in negative affect are at risk for a poorer course of depression (Gershon and Eidelman, 2015; Mason et al., 2017; Stange et al., 2018, 2016b) and individuals who have more recently attempted suicide show more variability in suicidal thinking (Kleiman et al., 2018) (although see section on Sampling Frequency, Temporal Scaling, Stationarity, and Generalizability below for a discussion of how temporal scaling influences the interpretation of the MSSD metric). When the MSSD was applied to a measure of daily typing speed on a smartphone keyboard, individuals with bipolar disorder who had greater day-to-day instability of typing speed also had greater prospective risk for depression (Stange et al., 2018). Although little empirical work has examined the area to date using AA, scholars have proposed that variability (or "flexibility")

of emotion regulation strategies, such as the ability to implement a variety of strategies over time, may facilitate adaptive outcomes (Aldao et al., 2015; Bonanno and Burton, 2013; Hollenstein, 2015; Hollenstein et al., 2013; O'Toole et al., 2017; Stange et al., 2017a). For more detailed discussions of examining variance in intensive longitudinal AA data with multilevel modeling, see (de Haan-Rietdijk et al., 2016; Hedeker et al., 2012; Hedeker D, 2008; Jahng et al., 2008).

Within-Subject Dynamic Systems: Are Individuals Whose Negative Affective States are More Strongly Temporally Associated at Greater Risk?—AA also has utility for measuring theories about complex dynamical systems that involve interrelations between affective states, but that are difficult to assess without intensive sampling. These dynamical systems frameworks suggest that associations between affective states may increase in the background over time without overt changes in psychopathological symptoms (Wichers et al., 2015), which increases the chance that activation of a single node (i.e., negative affective state) triggers the other affective states. In the context of vicious circles of relationships between affective states, individuals may become affectively stuck, which may ultimately elicit a downslide into depression (Wigman et al., 2015). These types of theories can be tested using AA, to compare individuals to one another in terms of relative risk for affective disorders (Rogers and Joiner, 2019). Indeed, there is evidence that there are stronger temporal interconnections between negative affective states among individuals with and vulnerable to psychopathology compared to those who are healthy and resilient (Pe et al., 2015; Wigman et al., 2013). Relatedly, AA also allow for the measurement of affective inertia, or the autocorrelation of affect over time, and the extant literature has suggested that individuals whose affect is insufficiently dynamic (e.g., those who are emotionally "stuck") may also be at risk for depression (Hollenstein, 2015; Koval et al., 2012, 2013b; Kuppens et al., 2010, 2012; Trull et al., 2015).

Identifying Subtypes: Digital Phenotyping and Mixture Modeling to Identify Subgroups at Differential Risk.—The exploration of subgroups represents a "middle ground" between treating all individuals as the same (which often is done in traditional research using only one assessment) and treating all individuals as completely different from one another (which is implied when separate models are computed for each individual using AA data). Exploring subgroups could be particularly useful for providing a starting point when determining more individually tailored solutions. One method for identifying subgroups is called digital phenotyping or digital footprinting (Bidargaddi et al., 2017; Hussein et al., in press; Insel, 2017; Kleiman et al., 2018; Onnela and Rauch, 2016; Smets et al., 2018b; Torous et al., 2015; Zulueta et al., 2018), which refers to the concept of using many streams of participants' data to determine latent or underlying subgroups of similar individuals. Digital phenotyping often refers to using a combination of passively- and actively-reported data streams (e.g., self-rated affect, typing speed, GPS location, accelerometer, sleep quality with actigraphy) that can be used together to identify specific individuals from their (observable) typical behavioral patterns. This work is based on the assumption that with enough sources of data and enough observations in time, there are enough idiosyncratic tendencies for a variety of outcomes to be determined (e.g., current mood state (Huang et al., 2018; Zulueta et al., 2018)).

Mixture modeling (e.g., latent class analysis, latent profile analysis) can be a particularly useful tool to carry out digital phenotyping (Bernanke et al., 2017). Latent class analysis can be used when the indictors used for digital phenotyping are categorical, or latent profile analyses when the indicators are continuous. Mixture modeling could also be used outside of a digital phenotyping framework or in combination with one. For example, person-level variables extracted from AA methods typically used in digital phenotyping could be included with trait-like variables (e.g., impulsivity), demographic variables (e.g., gender), or other non-AA data streams (e.g., electronic health records) to help further refine and separate phenotypes. Supporting the utility of this approach, one recent paper found that machine learning models that used EMA data along with electronic health record data were more accurate at predicting suicidal ideation than electronic health records alone (Peis et al., 2019).

One area within affective disorders in which digital phenotyping has demonstrated some initial promise has been within classifying subtypes of suicidal thinking (Bernanke et al., 2017). One study that applied a digital phenotyping approach to EMA data of suicidal thinking found that different profiles of suicide risk could be determine using metrics of variability (e.g., MSSD) and average levels of suicidal thinking (Kleiman et al., 2018). Specifically, those who had most recently attempted suicide tended to be members of a subgroup typified by a high average level of suicidal thinking over the study period, with low variability around their average. Although mixture models are useful for identifying subgroups of individuals, these analyses only allow exploration of subgroups, but do not provide information on individual variability within these subgroups. Recently introduced novel models such as Group Iterative Multiple Model Estimation (GIMME) framework can address this limitation by simultaneously modeling group- and individual-level processes (Lane et al., 2019; Wright et al., in press).

Other Considerations in the Ambulatory Assessment of Risk

Measuring Context and Validating Laboratory Measures

One of the obvious potential benefits of AA is the ability to examine constructs of interest across a variety of contexts. For example, one can measure questions about context involving where (e.g., at work, at a bar), when (e.g., during positive and negative events, when using substances), and with whom (e.g., with one's partner) relationships of interest may exist (Epstein et al., 2009; Lane et al., 2016). AA thus has the potential for greater ecological validity than lab contexts, which may not closely parallel real-world environments and individuals' responses to those contexts. Within the context of risk, AA might be used to examine the ecological validity of traditional markers of distal risk – that is, whether we really are studying the constructs we think we are studying within the lab (Raugh et al., 2019; Trull and Ebner-Priemer, 2013). For example, with sufficient temporal resolution within the lab and with AA, researchers might study how well stress responses in the lab map onto physiological and psychological responses to self-reported stressors in daily life, across contexts. It is important to note, however, that it is necessary to test the assumption that AA measures will be better than laboratory measures for detecting which individuals are at greatest risk. AA also can be used to provide real-world validation of

measures proposed to be potential physiological mechanisms of risk-related processes. For example, several recent studies have combined fMRI and AA approaches to examining neural correlates of real-world rumination (Ismaylova et al., 2018), affective inertia (Schwartz et al., 2019), positive (Forbes et al., 2010; Heller et al., 2015) and negative affect (Forbes et al., 2011), and sleep disruption with actigraphy (Holm et al., 2009), as well as psychophysiological correlates of affective instability (Koval et al., 2013a).

Despite the benefits of AA for measuring real-world factors in context, one ramification of measuring data outside of the lab is that signals (both active-reported and passively-recorded) are more variable, due in part to context (Raugh et al., 2019). If this variability in context is of theoretical interest to the question at hand, investigators can choose to examine different contexts (e.g., home vs. work) as moderator variables, provided that there is sufficient variability in such contexts to examine them reliably. If the contextual variability is not of interest or there is insufficient power to examine contexts as moderators (because contextual moderators are at a higher level than the observations [observations within contexts within people], power is decreased when using these variables), contexts can serve as covariates of non-interest in statistical models. With respect to ambulatory ECG data, for example, to avoid artifactual influence of physical activity that is not due to affective processes, investigators might decide to only analyze data that were measured when participants were not engaged in physical activity such as walking (based on accelerometer signal or EMA reports); alternatively, they might choose to covary for physical activity and thus examine ECG indices above and beyond the influence of activity level (Brown et al., 2017; Smets et al., 2018a; Valenza et al., 2015, 2014; Verkuil et al., 2016). Investigators should consider examining (and modeling) whether missing data are systematically related to contexts and outcomes of interest, which may improve model fit (Gao et al., 2016; Lin et al., 2018).

Investigators also can examine how much risk factors that are considered to be “trait-like” actually vary across time and context. For example, with repeated observations of a risk factor, an intraclass correlation coefficient (ICC) can be computed within a multilevel modeling framework to determine how much of the sample variance in the factor occurs between participants (e.g., as a result of individual differences), as opposed to within individuals (varying over time and across context). It should be noted, however, that because ICC is computed on a sample-level, interpretations of ICC are sample-dependent (e.g., a non-depressed sample, who rarely experience any symptoms of depression at all, may show very little within-person variability in depressive symptoms relative to a depressed sample). Ambulatory psychophysiological data also can contain more variability than lab data as a result of artifacts (or “noise”), due to issues with signal quality (Raugh et al., 2019). Signal quality issues can occur either as a result of inferior sensor quality compared to those used in the lab, or as a result of artifacts (e.g., due to motion). Although there is no universal “gold standard” for determining when ambulatory physiological data are usable versus when they should be discarded, typically data should be inspected to ensure sufficient quality for use in analysis, and to ensure a reliable data stream with good signal-to-noise ratio is being used.

Active vs. Passive Assessment

Active Measurement.—Given the range of methods that comprise AA, a useful distinction can be made between sources of data that are provided actively by participants (e.g., self-reported responses with EMA), as compared to sources that can be collected passively without input from participants (e.g., psychophysiology, location, and keyboard data). One of the primary benefits of active assessment such as EMA is that it is likely to have face validity – that is, questions can be asked that clearly map onto psychological constructs of interest, compared to passive modes of data that may provide less interpretable information about context or about subjective aspects of people's experiences.

However, EMA has several notable limitations (Ebner-Priemer and Trull, 2009; Holmlund et al., 2019; Trull and Ebner-Priemer, 2013). First, because the measures are self-reported, they are subject to potential biases if participants are unable or unwilling to provide accurate responses. Participants may be unable to provide responses because of context (e.g., while driving, while at work), because they lack awareness of their current subjective state (e.g., individuals with alexithymia have difficulty accurately reporting on their emotional states (Mason et al., 2005)), or because of other cognitive and memory biases (Ebner-Priemer and Trull, 2009; Fredrickson and Kahneman, 1993), some of which may be exacerbated among individuals who are depressed (Ben-Zeev et al., 2009). There is also the possibility of confounds between periods in which responses would be most likely to be indicative of risk (e.g., during periods of stress) and participants' willingness or ability to complete prompts received during these times. For example, people may be less likely to complete surveys when they are distressed or preoccupied with other more immediate goals, such as during an argument with a partner. There are limits to the frequency with which EMA can be sampled due to participant burden issues (Trull and Ebner-Priemer, 2013); some studies also have suggested that patterns of variability in responses may change with increased length of sampling, perhaps due to burden or familiarization (Vachon et al., 2016). Finally, although participants may be able to accurately report on certain aspects of subjective experience (Haeffel and Howard, 2010) such as periods of acute risk (e.g., suicidal ideation), they might have less awareness of some of the potential proximal antecedents of these periods of risk (e.g., autonomic changes, sleep quality), and thus might be less able to provide such information with active responses.

Passive Assessment.—Passive methods of assessment with AA avoid some of the limitations of active assessments. These methods can be measured without active awareness of the participant, provided that they have provided consent for data collection and are using equipment properly. Data can be collected "on-line" continuously, without additional requirements for participation, even during times of stress when risk may be increasing. Thus, passive assessment avoids some of the restrictions regarding the frequency of sampling due to participant burden that may occur with active assessment. Passive assessment therefore has particular promise for use in the identification of periods (and biomarkers) of proximal risk, provided that measurements can be obtained reliably and with sufficient sensitivity. For example, recent studies have suggested that wearable devices for measuring parasympathetic psychophysiology (such as heart rate variability, a marker of regulatory capacity and depression risk (Beauchaine and Thayer, 2015; Hamilton and Alloy,

2016; Stange et al., 2017a, 2017d, 2017b; Yaroslavsky et al., 2014)) may have predictive utility in detecting mood states (Kappeler-Setz et al., 2013; Valenza et al., 2015, 2014), although measures of sympathetic activity (such as electrodermal activity) can detect arousal but not valence, meaning there is a possibility of a high rate of false-positives in detecting risk. It is possible that such devices might be deployed in the future to determine periods of proximal risk for declining mood states, representing opportunities for "just-in-time" interventions to ameliorate such risk.

One of the limitations of passive assessment is that to "validate" whether a physiological signal is indicative of the experience of interest (e.g., stress), as opposed to some other issue (e.g., noise from context), it is helpful to have a linked EMA response (Raugh et al., 2019). For example, a researcher might confirm that low heart rate variability is associated with a period of self-reported increased stress (Dikecligil and Mujica-Parodi, 2010; Ottaviani et al., 2008) (mirroring lab results (Hamilton and Alloy, 2016; Stange et al., 2017a)), or obtain a more accurate estimate of sleep by using sleep diary along with actigraphy (Lauderdale et al., 2008). Thus, combining active and passive methods of assessment might yield better prediction of periods of risk, although active assessment may not be feasible for long-term monitoring of risk. Therefore, combining these methods might be useful for validating passive assessments as indices of risk, and for training and testing predictive models within the context of machine learning (Carpenter et al., 2016; Valenza et al., 2014) – models that could be conducted across participants (to identify indices of risk that are true of risk overall), and perhaps more importantly, models conducted within individual participants' data to identify each individual's own indices of risk (Lewis and Ridenour, 2019). By continually updating and refining the predictive model as more data are obtained from each person over time, models may be able to continually improve their accuracy in predicting risk. Once such models can be reliably replicated with sufficient sensitivity and specificity to periods of risk, passive measures might be used for longer-term follow-up to continue to monitor risk in a minimally invasive way.

Design Considerations

There are several factors related to study design that are important to consider when developing studies that aim to address questions about risk.

Base Rates.—One such consideration is the expected base rate of any events of interest. Base rates must be high enough that there is sufficient variability in outcomes of interest to examine relationships between contextual risk factors and the occurrence (or non-occurrence) of such outcomes. For example, to address within-subject questions about proximal risk factors for suicidal ideation, studies must be designed so that there is sufficient within-subject variability in suicidal ideation across the course of the study, so that proximal risk factors can be compared under conditions when the outcome occurred as well as when it did not occur. One way this might be achieved is by selecting a study sample that has sufficient severity that ideation may be expected (e.g., recruiting people who recently have made a suicide attempt, as opposed to a remitted depressed sample with low levels of symptoms). Alternatively, if base rates are expected to be relatively low, researchers might choose to follow individuals for a longer period of time to ensure a sufficient sample of the

events of interest. Considering base rates also is important for between-subjects questions that involve comparing individuals in terms of their variability of a potential risk factor of interest (e.g., negative affective instability); individuals must be followed for long enough to establish reliable estimates of within-subject variability before comparing participants to one another.

Pseudo-Random vs. Event-Related Sampling.—Another factor to consider is whether to use a pseudo-random sampling approach, in which individuals receive EMA prompts randomly throughout the day within pre-defined windows, or an event-related approach, in which participants decide when to report information to the investigators based on an event of interest (e.g., participant indicates when a stressor, distress, or suicidal ideation occurs by pressing a button or completing a survey). The pseudo-random sampling approach is most common and has the relative advantage of some experimental control, as all participants receive the same number of prompts. One disadvantage of this approach, however, is that the times when prompts are completed may not be immediately proximal to the event of interest (e.g., depending on the sampling schedule, a prompt might be received several hours after an acute stressor has passed), and thus participants' responses may be subject to some of the same types of recall biases discussed earlier (Ben-Zeev et al., 2012; Torous et al., 2015). Moreover, participants may not always be able to respond to a pseudo-random prompt (e.g., during school, while driving). Event-related sampling avoids some of these limitations by enabling the researcher to know exactly when an event of interest occurred, which might lead to more accurate self-reports (e.g., of affective symptoms during a crisis), and may allow for a better ability to identify the most relevant physiological signals that occur immediately proximal to events of interest.

One obvious downside to the event-related approach is that the sampling is biased toward times when participants are aware of, and are willing and able to report, an event of interest. This approach may be subject to its own set of biases and confounds, as some individuals may be more aware (or more conscientious) than others and therefore may be more likely to report events, even if they did not actually experience more events than other individuals. Thus, there may be more between-subjects variability in the number of observations collected when using event-related versus pseudo-random sampling. An alternative approach used by some researchers is to use both types of sampling – a pseudo-random prompting schedule with the option for participants to conduct additional surveys when events occur. This approach may allow researchers to "catch" more features of interest at the time they actually occur, particularly features that are not always self-initiated through random prompting. A downside of using both approaches (and for using the event-related approach alone) is the potential for additional participant burden, as it may be more challenging for participants to complete surveys in the time surrounding events of interest (e.g., during moments of crisis). Alternatively, for some questions it may be possible to use a pseudo-random prompting schedule, with branching logic that converts to event-based questions if an event of interest is in progress (e.g., asking additional event-related questions about substance use if currently engaged in substance use).

Sampling Frequency, Temporal Scaling, Stationarity, and Generalizability.— Several issues related to the frequency of AA sampling are worth discussing. Researchers should consider the naturalistic time course of the variables of interest, and match the sampling rate to their hypotheses about the temporal dynamics of the phenomena under investigation. For example, if affect is likely to fluctuate throughout the day, then sampling several times per day would allow for the modeling of these dynamics (e.g., Fisher and Newman, 2016). In practice, studies often have based sampling rates (e.g., frequency of EMA survey prompts) at least partly on convenience (the rate at which participants can handle, or the rate used in previous studies). To the extent that the ideal sampling rate is not known for the phenomena of interest, future studies might also consider manipulating the sampling rate, either experimentally in the study design (e.g., randomizing participants to different sampling rates), or when selecting observations for analyzing data (e.g., if six EMA prompts were sent daily, comparing results when analyzing one survey from each day compared to all six surveys each day).

The issue of sampling frequency is of particular importance when data are likely to demonstrate cyclic local temporal variation, such as diurnal cycles, days of the week, or time of day. Data should be sampled with enough temporal precision to accurately model the shape of the variable's temporal variation. For example, a recent study (Fisher and Newman, 2016) highlighted that the interpretation of constructs such as instability (e.g., the MSSD metric) that are sampled repeatedly over time may be influenced substantially by the timing of the measurements, particularly if data show a fixed, periodic pattern over time, such as a sinusoidal waveform. Using the example of anxiety levels across the day, which showed a sinusoidal pattern (rising and falling at stable intervals throughout the day), the authors demonstrated that opposite conclusions would be drawn if computing instability based on once-daily measurements (which might always measure anxiety at the same point in the waveform, yielding estimates of high stability), compared to when sampling four times daily (which would measure anxiety at varying parts of the waveform, yielding estimates of high instability).

Another issue related to patterns of fluctuations in data is stationarity, or the consistency of the mean and variance over time, which tells us how representative data are of the individual more generally. Stationarity is particularly important for idiographic analyses, which typically have the goal of generalizing across time, rather than across subjects (as in the case of nomothetic or between-subject analyses). As one of the goals of idiographic analyses is to use person-specific data to predict each person's future outcomes, data must be representative and consistent (and not undergoing substantive changes in mean or variance over time) to be relevant to predicting these future outcomes. Particularly problematic for data analysis are nonrepeating *global trends* that result in shifts in mean or variance, such as linear, quadratic, or cubic growth trends, which may be specific to the measurement period and thus are unlikely to generalize to future periods. In contrast, *cycles* represent local temporal patterns of variation that fluctuate at regular intervals, such as sleep/wake cycles (Waterhouse et al., 2012). For example, *sinusoidal cycles* represent the rising and falling of a factor over a fixed interval. Other *interval-based cycles* occur at a specific time of day or day of the week (Fisher and Bosley, 2019). To the extent that these repetitive cycles represent variable risk factors, they can be useful for predicting future fluctuations in outcomes of

interest. Future work can test whether detrending nonstationary data prior to predicting future outcomes improves prediction, or whether data with excessive trends must be disregarded in favor of prediction models defined by repetitive cycles and intervals or non-cyclic data.

Implications for Designing Real-Time Translational Interventions

Once predictive models of risk have been refined and are able to reliably identify periods of proximal risk, how can this information be used to reduce risk and improve functioning? An exciting direction for AA is the ability to implement interventions to ameliorate risk processes in real time in daily life. Referred to as just-in-time adaptive interventions (JITAIs), these interventions can be delivered on smartphones and can be "pushed" so that they are delivered during times when risk is predicted to be increasing, based on ongoing monitoring of physiological or behavioral events of interest (Ebner-Priemer and Trull, 2009; Heron and Smyth, 2010; Intille, 2007; Kumar et al., 2013; Nahum-Shani et al., 2016). Ideally, these interventions are based on person-centered models of proximal risk, so that each individual receives tailored interventions at times when the interventions are most likely to be successful in reducing risk (Nahum-Shani et al., 2016). From the perspective of risk, delivering JITAIs with AA can allow investigators to try to manipulate variable risk factors and examine effects on outcomes of interest (i.e., to test whether variable risk factors may also be *causal* risk factors (Kraemer, 1997)).

In theory, JITAIs have the potential to have superior effectiveness compared to traditional interventions, as JITAIs involve changing real-world behavior. In contrast, interventions delivered in the clinic or laboratory may have greater experimental control (a psychotherapist can guide a patient through use of a new skill in the intended manner), but at the great expense of limited generalizability to the real world (Ebner-Priemer and Trull, 2009). Indeed, patients often have difficulty utilizing skills practiced in psychotherapy, or remembering to do so, in the situations when they need it the most outside of treatment sessions, and this may interfere with treatment progress (Helbig and Fehm, 2004; Mausbach et al., 2010; Rees et al., 2005). In a recent study, a biofeedback app was developed and used to facilitate recovery of autonomic stress responses during a laboratory-based psychosocial stressor (Plans et al., 2019). By leveraging recent AA methods for monitoring autonomic functioning *in situ* in daily life, participants might be monitored for risk (e.g., following a drop in heart rate variability that is not attributable to metabolic demand, based on accelerometer (Brown et al., 2017; Smets et al., 2018a; Valenza et al., 2015, 2014; Verkuil et al., 2016)), at which point they could be prompted to use the biofeedback app to modulate their stress responses. Another study examined effects of a smartphone-based JITAI for patients leaving treatment for alcohol use disorders, and found that delivering interventions when patients were detected to be at high risk (e.g., when approaching high-risk locations like a bar) was associated with better outcomes compared to treatment-as-usual (Gustafson et al., 2014). Of course, the success of such interventions may depend on participants' willingness and ability to use the information provided in the context at hand; thus, tailoring interventions so that information is provided when it is most useful may be key (Nahum-Shani et al., 2016). For example, it may be unwise to step away to attend to an intervention

while giving presentation at work; it may be difficult to effectively use certain regulatory strategies like reappraisal in the presence of strong negative affect (Sheppes et al., 2014).

These types of technological interventions also hold promise for reducing risk, by helping to bridge the gap between what is learned in face-to-face therapy sessions and one's behavior in everyday life (Fairburn and Patel, 2017; Heron and Smyth, 2010; Kazdin, 2015; Kleiman and Nock, 2017). It often is difficult for patients to remember *when*, and *how*, to use skills learned in therapy. Passive monitoring of risk may enable individuals to be prompted *when* skills are likely to be needed, and also could provide guidance on *how* to use the skills in these moments (Kleiman and Nock, 2017). AA also has the potential to make treatments more accessible to individuals who have difficulty accessing treatment, given the increasing affordability and acceptability of electronic tools (Fairburn and Patel, 2017; Kazdin, 2015) for managing mental health. However, given that the field of JITAIs is still in its infancy, it is worth noting that relatively little evidence exists to date for the efficacy of interventions administered in daily life (Kaplan and Stone, 2013; Kumar et al., 2013; Nahum-Shani et al., 2016). Future work in this promising area should continue to empirically test JITAIs, and to integrate such interventions with the continual improvements being made in the ability to monitor risk processes in daily life with AA.

Future Directions in the AA of Risk in Affective Disorders

AA is increasingly being used as a tool for identifying risk factors across the lifespan. Thus, there is a need for future studies to test assumptions about whether risk factors identified with AA in adults generalize to children (Heron et al., 2017) and elderly individuals (Brose and Ebner-Priemer, 2015). In addition to the potential for risk factors to differ across development, cohorts at different stages of development also may have different levels of familiarity with technology, which could influence compliance with actively-reported data such as EMA, and measurements of passively-collected data (e.g., typing speed might be slower among the elderly or those who are less familiar with smartphones). The use of AA in these groups certainly has potential implications for measuring risk and developing JITAIs. For example, risk factors could be passively and actively monitored among children at high risk for a mood disorder (e.g., as a function of having a parent with a mood disorder), to identify potential windows of intervention when help may be needed. In older adults, AA could be used to monitor fluctuations in potential proximal risk factors for late-life depression, such as changes in physical activity or sleep patterns. When studying samples with broad age ranges, future work can test the relative utility of using multilevel modeling to examine age-related deviations from mean effects in the sample, versus using fully idiographic models that can be person-specific, and hence would be adaptable even in the presence of age-related differences in risk factors.

The long-term use of AA to monitor variability in risk will be especially feasible if data are passively collected, and may be especially fruitful if variables measured are theoretically grounded and are appropriately temporally scaled. However, more research is needed to validate (and test the generalizability of) such passive markers of risk before rolling them out for assessment and prevention efforts at a larger scale. Future work in the AA of risk should continue to increase the focus on testing prediction (accuracy, sensitivity, and specificity) to

determine the potential clinical utility of identified risk factors. This work also should seek to cross-validate results from predictive models in other samples (e.g., testing the stability of regression weights in multilevel models), or within the same individual during a different period of time (in the case of fully idiographic models). Once AA metrics of risk are validated, and are suggestive of clinical utility in prediction of affective outcomes, JITIAs can be developed. These new treatments can be compared to treatment as usual to determine whether targeting these risk factors *in situ* actually is more effective in preventing or reducing symptoms, is more cost effective, and/or whether it helps some individuals (e.g., those living in rural areas) to obtain access to care.

Conclusions

The rapid developments being made in technology allow for monitoring dynamic risk processes in daily life like never before. We have outlined here some of the key methods that can be used for studying individuals intensively over time, and how AA can be used to address key theoretical questions about risk processes in affective disorders. These methods have utility for understanding within subject dynamics (such as *when* individuals are at risk), and for obtaining a window into real-world processes that allow individuals to be compared to one another, to determine *which individuals* are at greatest risk. As the field continues to progress, we look forward to seeing how these tools will be used to improve our ability to understand risk, and to improve well-being by intervening in the moments when help is needed the most.

Acknowledgments

Jonathan P. Stange was supported by grant 1K23MH112769-01A1 from NIMH.

References

aan het Rot M, Hogenelst K, Schoevers RA, 2012 Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. Clinical Psychology Review 32, 510–523. 10.1016/j.cpr.2012.05.007 [PubMed: 22721999]

Abela JRZ, Hankin BL, 2011 Rumination as a vulnerability factor to depression during the transition from early to middle adolescence: A multiwave longitudinal study. Journal of Abnormal Psychology 120, 259–271. 10.1037/a0022796 [PubMed: 21553940]

Abela JRZ, Hankin BL, 2008 Cognitive vulnerability to depression in children and adolescence: A developmental psychopathology perspective, in: Abela JRZ, Hankin BL (Eds.), Handbook of Depression in Chldren and Adolescents. Guilford Press, New York, pp. 35–78.

Abramson LY, Metalsky GI, Alloy LB, 1989 Hopelessness depression: A theory-based subtype of depression. Psychological Review 96, 358–372. 10.1037/0033-295X.96.2.358

Aldao A, Sheppes G, Gross JJ, 2015 Emotion Regulation Flexibility. Cognitive Therapy and Research 39, 263–278. 10.1007/s10608-014-9662-4

Allen NB, Nelson BW, Brent D, Auerbach RP, 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. 10.1016/j.jad.2019.03.044 [PubMed: 30856493]

Alloy LB, Abramson LY, Raniere D, Dyller IM, 1999 Research methods in adult psychopathology, in: Kendall PC, Butcher JN, Holmbeck GN (Eds.), Handbook of Research Methods in Clinical Psychology. John Wiley & Sons Inc., Hoboken, NJ, pp. 466–498.

Alloy LB, Abramson LY, Whitehouse WG, Hogan ME, Panzarella C, Rose DT, 2006 Prospective incidence of first onsets and recurrences of depression in individuals at high and low cognitive risk for depression. Journal of Abnormal Psychology 115, 145–156. 10.1037/0021-843X.115.1.145 [PubMed: 16492105]

Alloy LB, Salk R, Stange JP, Abramson LY, 2017 Cognitive vulnerability and unipolar depression, in: DeRubeis RJ, Strunk DR (Eds.), The Oxford Handbook of Mood Disorders pp. 142–153.

Armey MF, Schatten HT, Haradhvala N, Miller IW, 2015 Ecological momentary assessment (EMA) of depression-related phenomena. Current Opinion in Psychology 4, 21–25. 10.1016/j.copsyc.2015.01.002 [PubMed: 25664334]

Beauchaine TP, Thayer JF, 2015 Heart rate variability as a transdiagnostic biomarker of psychopathology. International Journal of Psychophysiology 98, 338–350. 10.1016/j.ijpsycho.2015.08.004 [PubMed: 26272488]

Beck AT, Bredemeier K, 2016 A Unified Model of Depression: Integrating Clinical, Cognitive, Biological, and Evolutionary Perspectives. Clinical Psychological Science 4, 596–619. 10.1177/2167702616628523

Ben-Zeev D, Young MA, Depp CA, 2012 Real-time predictors of suicidal ideation: Mobile assessment of hospitalized depressed patients. Psychiatry Research 197, 55–59. 10.1016/j.psychres.2011.11.025 [PubMed: 22397912]

Ben-Zeev D, Young MA, Madsen JW, 2009 Retrospective recall of affect in clinically depressed individuals and controls. Cognition and Emotion 23, 1021–1040. 10.1080/02699930802607937

Bernanke JA, Stanley BH, Oquendo MA, 2017 Toward fine-grained phenotyping of suicidal behavior: the role of suicidal subtypes. Molecular Psychiatry 22, 1080–1081. 10.1038/mp.2017.123 [PubMed: 28607457]

Bidargaddi N, Musiat P, Makinen V-P, Ermes M, Schrader G, Licinio J, 2017 Digital footprints: facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. Molecular Psychiatry 22, 164–169. 10.1038/mp.2016.224 [PubMed: 27922603]

Bonanno GA, Burton CL, 2013 Regulatory Flexibility: An Individual Differences Perspective on Coping and Emotion Regulation. Perspectives on Psychological Science 8, 591–612. 10.1177/1745691613504116 [PubMed: 26173226]

Brose A, Ebner-Priemer UW, 2015 Ambulatory Assessment in the Research on Aging: Contemporary and Future Applications. GER 61, 372–380. 10.1159/000371707

Brown SBRE, Brosschot JF, Versluis A, Thayer JF, Verkuil B, 2017 New methods to optimally detect episodes of non-metabolic heart rate variability reduction as an indicator of psychological stress in everyday life. International Journal of Psychophysiology. 10.1016/j.ijpsycho.2017.10.007

Carpenter RW, Wycoff AM, Trull TJ, 2016 Ambulatory Assessment: New Adventures in Characterizing Dynamic Processes. Assessment 23, 414–424. 10.1177/1073191116632341 [PubMed: 26887808]

Chang BP, Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Nock MK, 2016 Biological risk factors for suicidal behaviors: a meta-analysis. Transl Psychiatry 6, e887 10.1038/tp.2016.165 [PubMed: 27622931]

Cicchetti D, Toth SL, 2009 The past achievements and future promises of developmental psychopathology: the coming of age of a discipline. Journal of Child Psychology and Psychiatry 50, 16–25. 10.1111/j.1469-7610.2008.01979.x [PubMed: 19175810]

de Haan-Rietdijk S, Kuppens P, Hamaker EL, 2016 What's in a Day? A Guide to Decomposing the Variance in Intensive Longitudinal Data. Frontiers in Psychology 7 10.3389/fpsyg.2016.00891

Dikecligil GN, Mujica-Parodi LR, 2010 Ambulatory and Challenge-Associated Heart Rate Variability Measures Predict Cardiac Responses to Real-World Acute Emotional Stress. Biological Psychiatry, Amygdala Activity and Anxiety: Stress Effects 67, 1185–1190. 10.1016/j.biopsych.2010.02.001

Ebner-Priemer UW, Trull TJ, 2009 Ambulatory Assessment: An Innovative and Promising Approach for Clinical Psychology. European Psychologist 14, 109–119. 10.1027/1016-9040.14.2.109

Epstein DH, Willner-Reid J, Vahabzadeh M, Mezghanni M, Lin J-L, Preston KL, 2009 Real-Time Electronic Diary Reports of Cue Exposure and Mood in the Hours Before Cocaine and Heroin

Craving and Use. Arch Gen Psychiatry 66, 88–94. 10.1001/archgenpsychiatry.2008.509 [PubMed: 19124692]

Fairburn CG, Patel V, 2017 The impact of digital technology on psychological treatments and their dissemination. Behaviour Research and Therapy, Expanding the impact of cognitive behaviour therapy: A special edition in honor of G. Terence Wilson Expanding the impact of cognitive behaviour therapy: A special edition in honor of G. Terence Wilson 88, 19–25. 10.1016/j.brat.2016.08.012

Fisher AJ, Bosley HG, 2019 Identifying the Presence and Timing of Discrete Mood States Prior to Therapy. 10.31219/osf.io/2jrhf

Fisher AJ, Bosley HG, Fernandez KC, Reeves JW, Soyster PD, Diamond AE, Barkin J, 2019 Open trial of a personalized modular treatment for mood and anxiety. Behaviour Research and Therapy 116, 69–79. 10.1016/j.brat.2019.01.010 [PubMed: 30831478]

Fisher AJ, Medaglia JD, Jeronimus BF, 2018 Lack of group-to-individual generalizability is a threat to human subjects research. Proceedings of the National Academy of Sciences 115, E6106–E6115. 10.1073/pnas.1711978115

Fisher AJ, Newman MG, 2016 Reductions in the diurnal rigidity of anxiety predict treatment outcome in cognitive behavioral therapy for generalized anxiety disorder. Behaviour Research and Therapy 79, 46–55. 10.1016/j.brat.2016.02.006 [PubMed: 26953959]

Fisher AJ, Reeves JW, Lawyer G, Medaglia JD, Rubel JA, 2017 Exploring the idiographic dynamics of mood and anxiety via network analysis. Journal of Abnormal Psychology 126, 1044–1056. 10.1037/abn0000311 [PubMed: 29154565]

Fisher AJ, Soyster P, 2019 Generating Accurate Personalized Predictions of Future Behavior: A Smoking Exemplar. 10.31234/osf.io/e24v6

Forbes EE, Phillips ML, Silk JS, Ryan ND, Dahl RE, 2011 Neural systems of threat processing in adolescents: role of pubertal maturation and relation to measures of negative affect. Dev Neuropsychol 36, 429–52. [PubMed: 21516542]

Forbes EE, Ryan ND, Phillips ML, Manuck SB, Worthman CM, Moyles DL, Tarr JA, Sciarrillo SR, Dahl RE, 2010 Healthy Adolescents' Neural Response to Reward: Associations With Puberty, Positive Affect, and Depressive Symptoms. Journal of the American Academy of Child & Adolescent Psychiatry 49, 162–172.e5. 10.1016/j.jaac.2009.11.006 [PubMed: 20215938]

Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, Musacchio KM, Jaroszewski AC, Chang BP, Nock MK, 2017 Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychological Bulletin 143, 187–232. 10.1037/bul0000084 [PubMed: 27841450]

Fredrickson BL, Kahneman D, 1993 Duration neglect in retrospective evaluations of affective episodes. Journal of Personality and Social Psychology 65, 45–55. 10.1037/0022-3514.65.1.45 [PubMed: 8355141]

Gao W, Hedeker D, Mermelstein R, Xie H, 2016 A scalable approach to measuring the impact of nonignorable nonresponse with an EMA application. Statistics in Medicine 35, 5579–5602. 10.1002/sim.7078 [PubMed: 27538504]

Gershon A, Eidelman P, 2015 Inter-episode affective intensity and instability: Predictors of depression and functional impairment in bipolar disorder. Journal of Behavior Therapy and Experimental Psychiatry 46, 14–18. 10.1016/j.jbtep.2014.07.005 [PubMed: 25164093]

Gustafson DH, McTavish FM, Chih M-Y, Atwood AK, Johnson RA, Boyle MG, Levy MS, Driscoll H, Chisholm SM, Dillenburg L, Isham A, Shah D, 2014 A Smartphone Application to Support Recovery From Alcoholism: A Randomized Clinical Trial. JAMA Psychiatry 71, 566–572. 10.1001/jamapsychiatry.2013.4642 [PubMed: 24671165]

Haeffel G, Howard GS, 2010 Self-Report: Psychology's Four-Letter Word. The American Journal of Psychology 123, 181 10.5406/amerjpsyc.123.2.0181 [PubMed: 20518434]

Haggerty R, Mrazek P, 1994 Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research. National Academies Press.

Hamilton JL, Alloy LB, 2016 Atypical reactivity of heart rate variability to stress and depression across development: Systematic review of the literature and directions for future research. Clinical Psychology Review 50, 67–79. 10.1016/j.cpr.2016.09.003 [PubMed: 27697746]

Hand DJ, Yu K, 2001 Idiot's Bayes--Not So Stupid After All? International Statistical Review 69, 385–398. 10.1111/j.1751-5823.2001.tb00465.x

Hankin BL, 2012 Future Directions in Vulnerability to Depression Among Youth: Integrating Risk Factors and Processes Across Multiple Levels of Analysis. Journal of Clinical Child & Adolescent Psychology 41, 695–718. 10.1080/15374416.2012.711708 [PubMed: 22900513]

Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE, 1998 Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. Journal of Abnormal Psychology 107, 128–140. 10.1037/0021-843X.107.1.128 [PubMed: 9505045]

Hedeker D, Mermelstein RJ, Demirtas H, 2012 Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Statistics in Medicine 31, 3328–3336. 10.1002/sim.5338 [PubMed: 22419604]

Hedeker DMR, 2008 An application of a mixed-effects location scale model for analysis of Ecological Momentary Assessment (EMA) data. Biometrics 64, 627–634. [PubMed: 17970819]

Helbig S, Fehm L, 2004 PROBLEMS WITH HOMEWORK IN CBT: RARE EXCEPTION OR RATHER FREQUENT? Behavioural and Cognitive Psychotherapy 32, 291–301. 10.1017/S1352465804001365

Heller AS, Fox AS, Wing EK, McQuisition KM, Vack NJ, Davidson RJ, 2015 The Neurodynamics of Affect in the Laboratory Predicts Persistence of Real-World Emotional Responses. Journal of Neuroscience 35, 10503–10509. 10.1523/JNEUROSCI.0569-15.2015 [PubMed: 26203145]

Heron KE, Everhart RS, McHale SM, Smyth JM, 2017 Using Mobile-Technology-Based Ecological Momentary Assessment (EMA) Methods With Youth: A Systematic Review and Recommendations. J Pediatr Psychol 42, 1087–1107. 10.1093/jpepsy/jsx078 [PubMed: 28475765]

Heron KE, Smyth JM, 2010 Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. British Journal of Health Psychology 15, 1–39. 10.1348/135910709X466063 [PubMed: 19646331]

Hollenstein T, 2015 This Time, It's Real: Affective Flexibility, Time Scales, Feedback Loops, and the Regulation of Emotion. Emotion Review 7, 308–315. 10.1177/1754073915590621

Hollenstein T, Lichtwarck-Aschoff A, Potworowski G, 2013 A Model of Socioemotional Flexibility at Three Time Scales. Emotion Review 5, 397–405. 10.1177/1754073913484181

Holm SM, Forbes EE, Ryan ND, Phillips ML, Tarr JA, Dahl RE, 2009 Reward-Related Brain Function and Sleep in Pre/Early Pubertal and Mid/Late Pubertal Adolescents. Journal of Adolescent Health 45, 326–334. 10.1016/j.jadohealth.2009.04.001 [PubMed: 19766936]

Holmlund TB, Foltz PW, Cohen AS, Johansen HD, Sigurdsen R, Fugelli P, Bergsager D, Cheng J, Bernstein J, Rosenfeld E, Elvevåg B, 2019 Moving psychological assessment out of the controlled laboratory setting: Practical challenges. Psychological Assessment 31, 292–303. 10.1037/pas0000647 [PubMed: 30802115]

Huang H, Cao B, Yu PS, Wang C, Leow AD, 2018 dpMood: Exploiting Local and Periodic Typing Dynamics for Personalized Mood Prediction, in: 2018 IEEE International Conference on Data Mining (ICDM). Presented at the 2018 IEEE International Conference on Data Mining (ICDM), pp. 157–166. 10.1109/ICDM.2018.00031

Hussein F, Stange JP, Langenecker SA, McInnis MG, Zulueta J, Piscitello A, Cao B, Huang H, Yu PS, Ajilore OA, Leow A, in press Passive sensing of affective and cognitive functioning in mood disorders by analyzing keystroke kinematics and speech dynamics, in: Mobile Sensing and Psychoinformatics. Springer.

Insel TR, 2017 Digital Phenotyping: Technology for a New Science of Behavior. JAMA 318, 1215 10.1001/jama.2017.11295 [PubMed: 28973224]

Intille SS, 2007 Technological innovations enabling automatic, context-sensitive ecological momentary assessment, in: The Science of Real-Time Data Capture: Self-Reports in Health Research. Oxford University Press, pp. 308–337.

Ismaylova E, Di Sante J, Gouin J-P, Pomares FB, Vitaro F, Tremblay RE, Booij L, 2018 Associations Between Daily Mood States and Brain Gray Matter Volume, Resting-State Functional Connectivity and Task-Based Activity in Healthy Adults. Frontiers in Human Neuroscience 12 10.3389/fnhum.2018.00168

Jahng S, Wood PK, Trull TJ, 2008 Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. Psychological Methods 13, 354–375. 10.1037/a0014173 [PubMed: 19071999]

Kaplan RM, Stone AA, 2013 Bringing the Laboratory and Clinic to the Community: Mobile Technologies for Health Promotion and Disease Prevention. Annual Review of Psychology 64, 471–498. 10.1146/annurev-psych-113011-143736

Kappeler-Setz C, Gravenhorst F, Schumm J, Arnrich B, Tröster G, 2013 Towards Long Term Monitoring of Electrodermal Activity in Daily Life. Personal Ubiquitous Comput. 17, 261–271. 10.1007/s00779-011-0463-4

Kazdin AE, 2015 Technology-Based Interventions and Reducing the Burdens of Mental Illness: Perspectives and Comments on the Special Series. Cognitive and Behavioral Practice 22, 359–366. 10.1016/j.cbpra.2015.04.004

Kazdin AE, 2007 Mediators and Mechanisms of Change in Psychotherapy Research. Annual Review of Clinical Psychology 3, 1–27. 10.1146/annurev.clinpsy.3.022806.091432

Kleiman EM, Nock MK, 2017 Advances in Scientific Possibilities Offered by Real-Time Monitoring Technology. Psychiatry 80, 118–124. 10.1080/00332747.2017.1325661 [PubMed: 28767336]

Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, Nock MK, 2018 Digital phenotyping of suicidal thoughts. Depression and Anxiety 35, 601–608. 10.1002/da.22730 [PubMed: 29637663]

Koval P, Kuppens P, Allen NB, Sheeber L, 2012 Getting stuck in depression: The roles of rumination and emotional inertia. Cognition & Emotion 26, 1412–1427. 10.1080/02699931.2012.667392 [PubMed: 22671768]

Koval P, Ogrinz B, Kuppens P, Van den Bergh O, Tuerlinckx F, Sütterlin S, 2013a Affective Instability in Daily Life Is Predicted by Resting Heart Rate Variability. PLoS ONE 8, e81536 10.1371/journal.pone.0081536 [PubMed: 24312315]

Koval P, Pe ML, Meers K, Kuppens P, 2013b Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? Emotion 13, 1132–1141. 10.1037/a0033579 [PubMed: 23914765]

Kraemer HC, 1997 Coming to Terms With the Terms of Risk. Archives of General Psychiatry 54, 337 10.1001/archpsyc.1997.01830160065009 [PubMed: 9107150]

Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, Riley WT, Shar A, Spring B, Spruijt-Metz D, Hedeker D, Honavar V, Kravitz R, Craig Lefebvre R, Mohr DC, Murphy SA, Quinn C, Shusterman V, Swendeman D, 2013 Mobile Health Technology Evaluation: The mHealth Evidence Workshop. American Journal of Preventive Medicine 45, 228–236. 10.1016/j.amepre.2013.03.017 [PubMed: 23867031]

Kuppens P, Allen NB, Sheeber LB, 2010 Emotional Inertia and Psychological Maladjustment. Psychol Sci 21, 984–991. 10.1177/0956797610372634 [PubMed: 20501521]

Kuppens P, Sheeber LB, Yap MBH, Whittle S, Simmons JG, Allen NB, 2012 Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. Emotion 12, 283–289. 10.1037/a0025046 [PubMed: 21988744]

Lane SP, Carpenter RW, Sher KJ, Trull TJ, 2016 Alcohol Craving and Consumption in Borderline Personality Disorder: When, Where, and With Whom. Clinical Psychological Science 4, 775–792. 10.1177/2167702615616132 [PubMed: 28042520]

Lane ST, Gates KM, Pike HK, Beltz AM, Wright AGC, 2019 Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. Psychological Methods 24, 54–69. 10.1037/met0000192 [PubMed: 30124300]

Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ, 2008 Self-reported and measured sleep duration: how similar are they? Epidemiology 19, 838–845. 10.1097/EDE.0b013e318187a7b0 [PubMed: 18854708]

Lewis KC, Ridenour JM, 2019 The Integration of EMA and Single-Occasion Multimethod Assessment Data for a Complex Psychiatric Patient. Assessment 1073191118825313. 10.1177/1073191118825313

Lin X, Mermelstein R, Hedeker D, 2018 A shared parameter location scale mixed effect model for EMA data subject to informative missing. Health Serv Outcomes Res Method 18, 227–243. 10.1007/s10742-018-0184-5

Littlewood DL, Kyle SD, Carter L-A, Peters S, Pratt D, Gooding P, 2019 Short sleep duration and poor sleep quality predict next-day suicidal ideation: an ecological momentary assessment study. Psychological Medicine 49, 403–411. 10.1017/S0033291718001009 [PubMed: 29697037]

Malhi GS, Hamilton A, Morris G, Mannie Z, Das P, Outhred T, 2017 The promise of digital mood tracking technologies: are we heading on the right track? Evidence Based Mental Health 20, 102–107. 10.1136/eb-2017-102757 [PubMed: 28855245]

Mason L, Eldar E, Rutledge RB, 2017 Mood Instability and Reward Dysregulation—A Neurocomputational Model of Bipolar Disorder. JAMA Psychiatry 74, 1275 10.1001/jamapsychiatry.2017.3163 [PubMed: 29049438]

Mason O, Tyson M, Jones C, Potts S, 2005 Alexithymia: Its prevalence and correlates in a British undergraduate sample. Psychology and Psychotherapy: Theory, Research and Practice 78, 113–125. 10.1348/147608304X21374

Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL, 2010 The Relationship Between Homework Compliance and Therapy Outcomes: An Updated Meta-Analysis. Cogn Ther Res 34, 429–438. 10.1007/s10608-010-9297-z

Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, Murphy SA, 2016 Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Annals of Behavioral Medicine. 10.1007/s12160-016-9830-8

Onnela J-P, Rauch SL, 2016 Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 41, 1691–1696. 10.1038/npp.2016.7 [PubMed: 26818126]

O'Toole MS, Zachariae R, Mennin DS, 2017 Social anxiety and emotion regulation flexibility: considering emotion intensity and type as contextual factors. Anxiety, Stress, & Coping 30, 716–724. 10.1080/10615806.2017.1346792

Ottaviani C, Shapiro D, Davydov DM, Goldstein IB, 2008 Autonomic Stress Response Modes and Ambulatory Heart Rate Level and Variability. Journal of Psychophysiology 22, 28–40. 10.1027/0269-8803.22.1.28

Pe ML, Kircanski K, Thompson RJ, Bringmann LF, Tuerlinckx F, Mestdagh M, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Kuppens P, Gotlib IH, 2015 Emotion-Network Density in Major Depressive Disorder. Clinical Psychological Science 3, 292–300. 10.1177/2167702614540645 [PubMed: 31754552]

Pe ML, Raes F, Koval P, Brans K, Verduyn P, Kuppens P, 2013 Interference resolution moderates the impact of rumination and reappraisal on affective experiences in daily life. Cognition & Emotion 27, 492–501. 10.1080/02699931.2012.719489 [PubMed: 22966838]

Peis I, Olmos PM, Vera-Varela C, Barrigon ML, Courtet P, Baca-Garcia E, Artes-Rodriguez A, 2019 Deep Sequential Models for Suicidal Ideation from Multiple Source Data. IEEE Journal of Biomedical and Health Informatics 1–1. 10.1109/JBHI.2019.2919270

Pew Research Center, 2018 Demographics of Mobile Device Ownership and Adoption in the United States. URL http://www.pewinternet.org/fact-sheet/mobile/ (accessed 2.17.19).

Plans D, Morelli D, Sütterlin S, Ollis L, Derbyshire G, Cropley M, 2019 Use of a Biofeedback Breathing App to Augment Poststress Physiological Recovery: Randomized Pilot Study. JMIR Formative Research 3, e12227 10.2196/12227 [PubMed: 30684437]

Raugh IM, Chapman HC, Bartolomeo LA, Gonzalez C, Strauss GP, 2019 A Comprehensive Review of Psychophysiological Applications for Ecological Momentary Assessment in Psychiatric Populations. Psychological Assessment 31, 304–317. [PubMed: 30802116]

Rees CS, McEvoy P, Nathan PR, 2005 Relationship Between Homework Completion and Outcome in Cognitive Behaviour Therapy. Cognitive Behaviour Therapy 34, 242–247. 10.1080/16506070510011548 [PubMed: 16319035]

Rogers ML, Joiner TE, 2019 Exploring the temporal dynamics of the interpersonal theory of suicide constructs: A dynamic systems modeling approach. Journal of Consulting and Clinical Psychology 87, 56–66. 10.1037/ccp0000373 [PubMed: 30570302]

Scheiderer EM, Wang T, Tomko RL, Wood PK, Trull TJ, 2016 Negative Affect Instability among Individuals with Comorbid Borderline Personality Disorder and Posttraumatic Stress Disorder. Clin Psychol Sci 4, 67–81. 10.1177/2167702615573214 [PubMed: 26904388]

Schwartz J, Ordaz SJ, Kircanski K, Ho TC, Davis EG, Camacho MC, Gotlib IH, 2019 Resting-state functional connectivity and inflexibility of daily emotions in major depression. Journal of Affective Disorders 249, 26–34. 10.1016/j.jad.2019.01.040 [PubMed: 30743019]

Sheppes G, Scheibe S, Suri G, Radu P, Blechert J, Gross JJ, 2014 Emotion regulation choice: A conceptual framework and supporting evidence. Journal of Experimental Psychology: General 143, 163–181. 10.1037/a0030831 [PubMed: 23163767]

Smets E, De Raedt W, Van Hoof C, 2018a Into the Wild: The Challenges of Physiological Stress Detection in Laboratory and Ambulatory Settings. IEEE Journal of Biomedical and Health Informatics 1–1. 10.1109/JBHI.2018.2883751

Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, Cornelis J, Janssens O, Van Hoecke S, Claes S, Van Diest I, Van Hoof C, 2018b Large-scale wearable data reveal digital phenotypes for daily-life stress detection. npj Digital Medicine 1 10.1038/s41746-018-0074-9

Solhan MB, Trull TJ, Jahng S, Wood PK, 2009 Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. Psychological Assessment 21, 425–436. 10.1037/a0016869 [PubMed: 19719353]

Stange JP, Alloy LB, Fresco DM, 2017a Inflexibility as a Vulnerability to Depression: A Systematic Qualitative Review. Clinical Psychology: Science and Practice 24, 245–276. 10.1111/cpsp.12201

Stange JP, Connolly SL, Burke TA, Hamilton JL, Hamlat EJ, Abramson LY, Alloy LB, 2016a Inflexible cognition predicts first onset of major depressive episodes in adolescence. Depression and Anxiety 33, 1005–1012. 10.1002/da.22513 [PubMed: 27093215]

Stange JP, Hamilton JL, Fresco DM, Alloy LB, 2017b Flexible parasympathetic responses to sadness facilitate spontaneous affect regulation. Psychophysiology 54, 1054–1069. 10.1111/psyp.12856 [PubMed: 28334441]

Stange JP, Hamilton JL, Olino TM, Fresco DM, Alloy LB, 2017c Autonomic reactivity and vulnerability to depression: A multi-wave study. Emotion 17, 602–615. 10.1037/emo0000254 [PubMed: 27991819]

Stange JP, Hamilton JL, Olino TM, Fresco DM, Alloy LB, 2017d Autonomic reactivity and vulnerability to depression: A multi-wave study. Emotion 17, 602–615. 10.1037/emo0000254 [PubMed: 27991819]

Stange JP, Sylvia LG, da Silva Magalhães PV, Miklowitz DJ, Otto MW, Frank E, Yim C, Berk M, Dougherty DD, Nierenberg AA, Deckersbach T, 2016b Affective instability and the course of bipolar depression: Results from the STEP-BD randomised controlled trial of psychosocial treatment. British Journal of Psychiatry 208, 352–358. 10.1192/bjp.bp.114.162073 [PubMed: 26795426]

Stange JP, Zulueta J, Langenecker SA, Ryan KA, Piscitello A, Duffecy J, McInnis MG, Nelson P, Ajilore O, Leow A, 2018 Let your fingers do the talking: Passive typing instability predicts future mood outcomes. Bipolar Disorders 20, 285–288. 10.1111/bdi.12637 [PubMed: 29516666]

Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, Onnela J-P, 2015 Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. JMIR Mental Health 2, e8 10.2196/mental.3889 [PubMed: 26543914]

Treynor W, Gonzalez R, Nolen-Hoeksema S, 2003 Rumination Reconsidered: A Psychometric Analysis. Cognitive Therapy and Research 13.

Trull TJ, Ebner-Priemer U, 2013 Ambulatory Assessment. Annual Review of Clinical Psychology 9, 151–176. 10.1146/annurev-clinpsy-050212-185510

Trull TJ, Lane SP, Koval P, Ebner-Priemer UW, 2015 Affective Dynamics in Psychopathology. Emotion Review 7, 355–361. 10.1177/1754073915590617 [PubMed: 27617032]

Trull TJ, Solhan MB, Tragesser SL, Jahng S, Wood PK, Piasecki TM, Watson D, 2008 Affective instability: Measuring a core feature of borderline personality disorder with ecological

momentary assessment. Journal of Abnormal Psychology 117, 647–661. 10.1037/a0012532 [PubMed: 18729616]

Vachon H, Bourbousson M, Deschamps T, Doron J, Bulteau S, Sauvaget A, Thomas-Ollivier V, 2016 Repeated self-evaluations may involve familiarization: An exploratory study related to Ecological Momentary Assessment designs in patients with major depressive disorder. Psychiatry Research 245, 99–104. 10.1016/j.psychres.2016.08.034 [PubMed: 27541343]

Valenza G, Citi L, Gentili C, Lanata A, Scilingo EP, Barbieri R, 2015 Characterization of Depressive States in Bipolar Patients Using Wearable Textile Technology and Instantaneous Heart Rate Variability Assessment. IEEE Journal of Biomedical and Health Informatics 19, 263–274. 10.1109/JBHI.2014.2307584 [PubMed: 25561449]

Valenza G, Nardelli M, Lanata A, Gentili C, Bertschy G, Paradiso R, Scilingo EP, 2014 Wearable Monitoring for Mood Recognition in Bipolar Disorder Based on History-Dependent Long-Term Heart Rate Variability Analysis. IEEE Journal of Biomedical and Health Informatics 18, 1625–1635. 10.1109/JBHI.2013.2290382 [PubMed: 24240031]

Verkuil B, Brosschot JF, Tollenaar MS, Lane RD, Thayer JF, 2016 Prolonged Non-metabolic Heart Rate Variability Reduction as a Physiological Marker of Psychological Stress in Daily Life. Annals of Behavioral Medicine 50, 704–714. 10.1007/s12160-016-9795-7 [PubMed: 27150960]

Waterhouse J, Fukuda Y, Morita T, 2012 Daily rhythms of the sleep-wake cycle. Journal of Physiological Anthropology 31, 5 10.1186/1880-6805-31-5 [PubMed: 22738268]

Wichers M, Wigman JTW, Myin-Germeys I, 2015 Micro-Level Affect Dynamics in Psychopathology Viewed From Complex Dynamical System Theory. Emotion Review 7, 362–367. 10.1177/1754073915590623

Wigman JTW, van Os J, Borsboom D, Wardenaar KJ, Epskamp S, Klippel A, MERGE, Viechtbauer W, Myin-Germeys I, Wichers M, 2015 Exploring the underlying structure of mental disorders: cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. Psychological Medicine 45, 2375–2387. 10.1017/S0033291715000331 [PubMed: 25804221]

Wigman JTW, van Os J, Thiery E, Derom C, Collip D, Jacobs N, Wichers M, 2013 Psychiatric Diagnosis Revisited: Towards a System of Staging and Profiling Combining Nomothetic and Idiographic Parameters of Momentary Mental States. PLoS ONE 8, e59559 10.1371/journal.pone.0059559 [PubMed: 23555706]

Wright AGC, Gates KM, Arizmendi C, Lane S, Woods WC, Edershile EA, in press Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. Psychological Assessment.

Yaroslavsky I, Rottenberg J, Kovacs M, 2014 Atypical patterns of respiratory sinus arrhythmia index an endophenotype for depression. Development and Psychopathology 26, 1337–1352. 10.1017/S0954579414001060 [PubMed: 25422965]

Zou H, Hastie T, 2005 Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320. 10.1111/j.1467-9868.2005.00503.x

Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, McInnis M, Ajilore O, Nelson PC, Ryan K, Leow A, 2018 Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study. J Med Internet Res 20 10.2196/jmir.9775

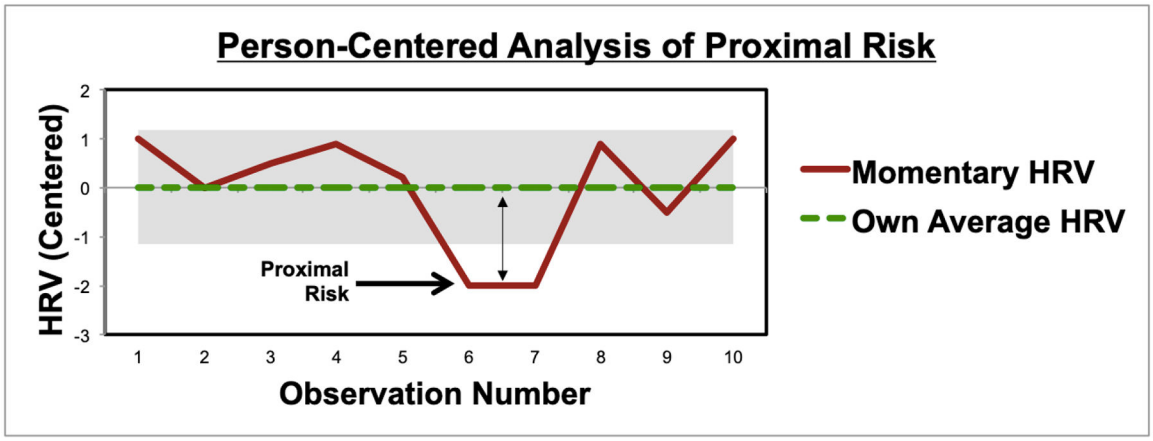

Figure 1.

Example of a single participant's data using a person-centered approach to measuring a risk factor (heart rate variability, HRV, a marker of regulatory capacity and depression risk (Beauchaine and Thayer, 2015; Hamilton and Alloy, 2016; Stange et al., 2017a, 2017d, 2017b; Yaroslavsky et al., 2014) repeatedly over time using ambulatory assessment. Each observation of the risk factor can be compared to the individual participant's own average (one standard deviation from participant's mean is shaded). Deviations from the average at each time point can be used as a within-subject predictor of outcomes of interest (e.g., negative affect at the next time point) within a multilevel statistical model.

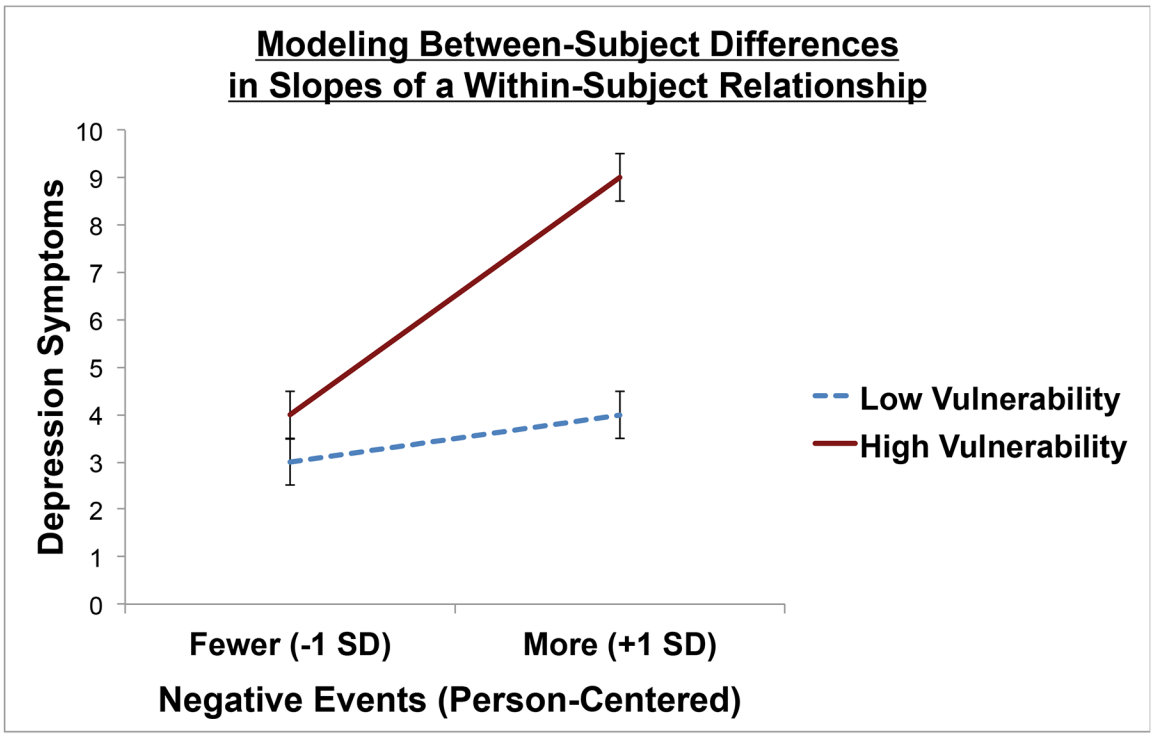

Figure 2.

Example of modeling individual differences (at the between-subject level) in the slope of the within-subject relationship between person-centered negative events and subsequent depression symptoms.

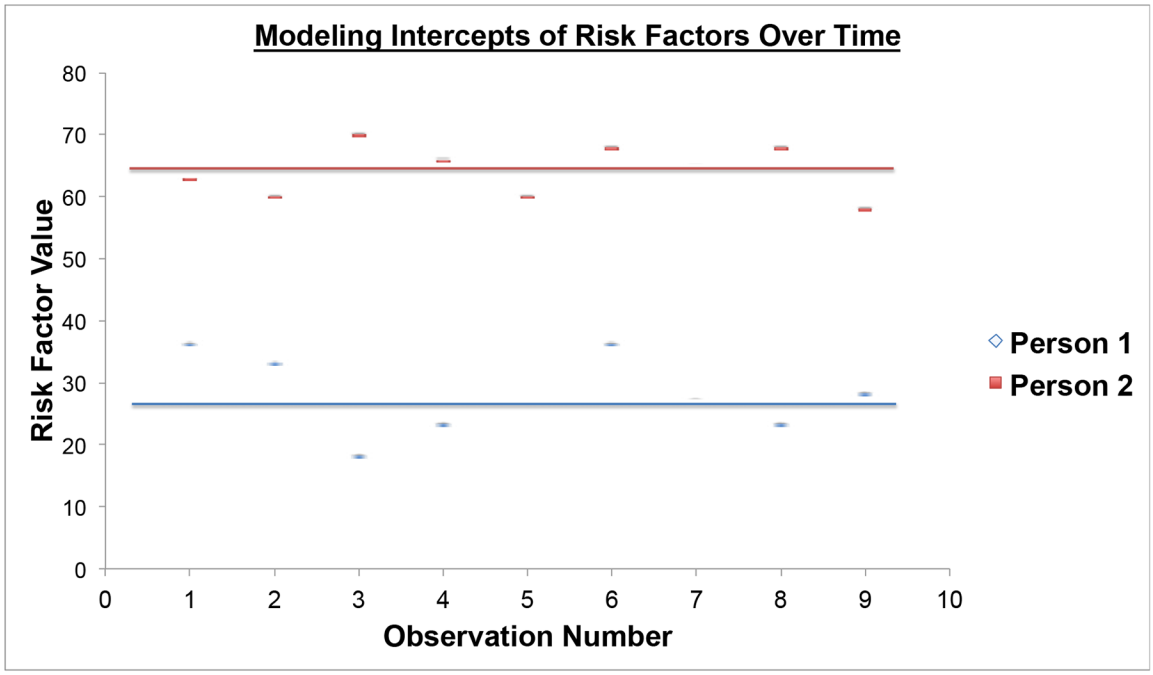

Figure 3.

Example of modeling individual differences (at the between-subject level) in the intercepts of a risk factor, measured repeatedly over time with ambulatory assessment. Two hypothetical participants' data are shown here.

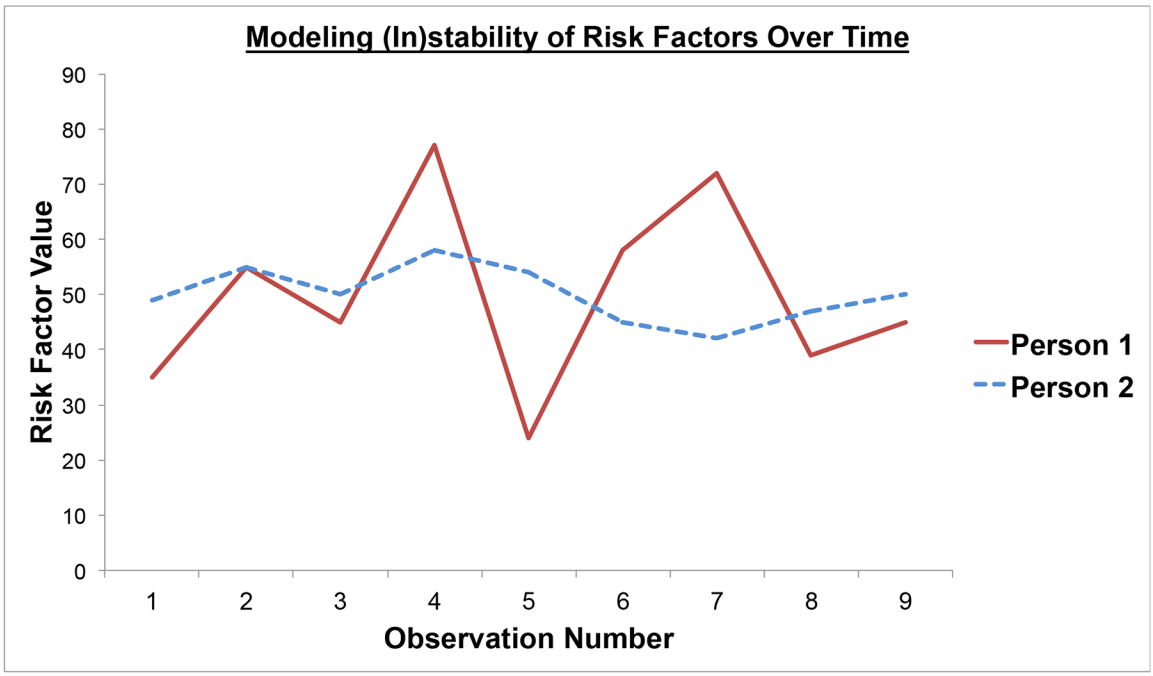

Figure 4.

Example of two hypothetical participants with high instability (Person 1), and low instability (Person 2), of a potential risk factor measured with ambulatory assessment. The two participants have the same mean value of the risk factor (50), but they differ in levels of instability (mean square successive difference).