Psychological Assessment

Applied Ambulatory Assessment: Integrating Idiographic and Nomothetic Principles of Measurement

Aidan G. C. Wright and Johannes Zimmermann

Online First Publication, March 21, 2019. http://dx.doi.org/10.1037/pas0000685

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

1040-3590/19/$12.00

http://dx.doi.org/10.1037/pas0000685

Applied Ambulatory Assessment: Integrating Idiographic and Nomothetic Principles of Measurement

Aidan G. C. Wright
University of Pittsburgh

Johannes Zimmermann
Psychologische Hochschule Berlin

Ambulatory assessment (AA; also known as ecological momentary assessment) has enjoyed enthusiastic implementation in psychological research. The ability to assess thoughts, feelings, behavior, physiology, and context intensively and repeatedly in the moment in an individual's natural ecology affords access to data that can answer exciting questions about sequences of events and dynamic processes in daily life. AA also holds unique promise for developing personalized models of individuals (i.e., precision or person-specific assessment) that might be transformative for applied settings such as clinical practice. However, successfully translating AA from bench to bedside is challenging because of the inherent tension between idiographic and nomothetic principles of measurement. We argue that the value of applied AA will be most fully realized by balancing the ability to develop personalized models with ensuring comparability among individuals.

Public Significance Statement

Psychological research is increasingly measuring human functioning in life as it is lived, by assessing thoughts, feelings, behavior, physiology, and context using tools like smartphones and other mobile sensors. This article proposes some basic principles for adapting these research tools for applied clinical settings.

Keywords: ambulatory assessment, ecological momentary assessment, personalized models, precision assessment

Ambulatory assessment (AA) refers to a range of techniques (e.g., ecological momentary assessment, experience sampling methodology, ambulatory psychophysiology, daily diaries, passive sensing) that are designed to assess some combination of subjective experience, physiology, and context in an individual's natural environment, intensively and repeatedly over time. AA has enjoyed a dramatic increase in use in psychological research (Hamaker & Wichers, 2017). Enthusiasm for AA is primarily driven by three unique strengths: the ability to assess individuals in their natural environment (enhancing ecological validity), during or temporally near to specific events (minimizing retrospective bias), and intensively and repeatedly as dynamic psychological processes unfold (increasing measurement precision). Using AA to follow an individual in daily life as it is commonly lived provides the ideal vehicle for capturing individuals' unique dynamic processes, guiding clinical decision making, and facilitating real-time interventions. These features have spurred a renewed interest in idiographic methods and a push to develop and implement *precision* assessment and *personalized* (i.e., person-specific or idiographic) models for use in clinical settings (Epskamp, van Borkulo, et al., 2018; Fisher, 2015; Roche, Pincus, Rebar, Conroy, & Ram, 2014; van Os, Delespaul, Wigman, Myin-Germeys, & Wichers, 2013; Wright et al., 2016; Zimmermann et al., in press).

Idiographic models of behavior are not new (e.g., Cattell & Luborsky, 1950; Zevon & Tellegen, 1982), but the technological and methodological advances that AA has fostered have reinvigorated interest in them. Intensive data can now be collected, stored, and analyzed with relative ease. Idiographic models exploit frequent assessments of a single individual by capturing not just the level of behaviors, but also their temporal profile or patterning. This allows for the assessment of dynamic constructs such as

Aidan G. C. Wright, Department of Psychology, University of Pittsburgh; Johannes Zimmermann, Psychologische Hochschule Berlin.

Johannes Zimmermann is now at the Department of Psychology, University of Kassel.

Aidan G. C. Wright's effort on this project was supported by the National Institute of Mental Health (L30 MH101760). The opinions expressed are solely those of the authors and not those of the funding source. We sincerely thank many colleagues who offered critical thoughts and suggestions: Laura Bringmann, Lauren Bylsma, Katie Gates, Ellen Hamaker, Chris Hopwood, Carissa Low, Mike Roche, Jennifer Silk, Brian Suffoletto, and Marieke Wichers. Any deficiencies in logic or objectionable aspects of the paper are entirely our responsibility, and do not reflect the comments of these helpful colleagues.

Correspondence concerning this article should be addressed to Aidan G. C. Wright, Department of Psychology, University of Pittsburgh, 4121 Sennott Square, 210 South Bouquet Street, Pittsburgh, PA 15260. E-mail: aidan@pitt.edu

duration of behavior, instability, inertia, situation–behavior contingencies, and, ultimately, multivariate personalized psychological models. That behavior varies across time and circumstances is the rule, not the exception, and clinically the focus is often on when and under what conditions critical behaviors occur. For example, depressive episodes are marked by persistent low positive affect, borderline personality pathology is reflected in instability of affect, and compulsions manifest when anxiety rises. Conducting a behavior chain in dialectical behavior therapy (Linehan, 1993) involves walking the patient through in sequence, establishing the chain of events leading up to some targeted outcome (e.g., self-injurious behavior). The logic is similar in AA, with the goal to capture behaviors in real time or close to it, so that these sequences can be ascertained and possibly intervened upon in situ. These might offer important hypotheses for collaborative exploration (Bak, Drukker, Hasmi, & van Os, 2016; Beltz, Wright, Sprague, & Molenaar, 2016; Kroeze et al., 2017), assist in treatment planning (Fisher & Boswell, 2016; Schiepek, Stöger-Schmidinger, Aichhorn, Schöller, & Aas, 2016), or even signal a worsening of functioning (Wichers, Groot, Psychosystems, ESM Group, & EWS Group, 2016).

Traditionally, idiographic measurement has been placed in opposition to nomothetic, or group-based, principles of measurement. For instance, a fully idiographic approach to measurement involves "the measurement of variables and functional relations that have been individually selected, or derived from assessment stimuli or contexts that have been individually tailored, to maximize their relevance for the particular individual" (Haynes, Mumma, & Pinson, 2009, p. 179). In this scheme, little and possibly nothing would be consistent between assessments applied to different individuals. This contrasts with "nomothetic assessment, in which judgments about a person are based on comparison with other persons using data from the same assessment instrument administered in a standardized manner" (p. 180). These approaches are not as incompatible as they might seem, and each can contribute valuable information. The value in idiographic assessment is high fidelity to the specific individual, but at the cost of being able to make comparisons to others. However, if an assessment requires reference to normative functioning, as is almost always the case in clinical or applied settings, a purely idiographic approach will be found lacking. This is because, as Sullivan (1954) so eloquently put it, "We all show everything that any mental patient shows, except for the pattern, the accents, and so on" (p. 183). In other words, most of what we are interested in understanding in a clinical assessment is problems of general processes that all engage in to some degree. To illustrate, we all experience daily stress, and almost all of us experience negative emotions as a result. What are the features of those processes that would signal a problem? Frequency? Intensity? Duration? The strength of the stress-negative affect coupling? Triggering subsequent maladaptive behavior (e.g., binge drinking)? These types of questions are best understood in the context of what the normative stress process looks like, and how any given individual's process departs from it. To understand an individual's model, it needs to be understood in reference to what is normative.

Therefore, to make personalized psychological models viable, idiographic methods need to be wedded to nomothetic principles. Here, we advocate for using fully standardized protocols to assess within-person processes or personalized psychological models. We see two major challenges. First, with few exceptions, our understanding of the temporal patterning of constructs of interest is limited. More systematic research is needed on how to design assessments, including when, where, how, and with what frequency. Second, few inventories have been developed for and given rigorous psychometric evaluation with AA data collection. Omnibus measures are lacking, and measures for specific constructs are sparse. Intensive longitudinal data can be used to calculate person-specific statistics and models that are not available with cross-sectional data. At the same time, the interpretation of these novel statistical features in single case scenarios requires information on their reliability, validity, and normative distributions that is currently rarely available. Taken together, these issues leave assessment basics, like designing an AA battery that supports appropriate inferences to answer specific referral questions, currently beyond the reach for practitioners in applied settings. Addressing the challenges of AA requires integrating theory, sampling, and analytic concerns much more tightly than it is usually done in traditional cross-sectional assessment (Collins, 2006).

Several excellent reviews have extensively covered AA in psychological research (e.g., Carpenter, Wycoff, & Trull, 2016; Mohr, Zhang, & Schueller, 2017; Moskowitz, Russell, Sadikaj, & Sutton, 2009; Shiffman, Stone, & Hufford, 2008; Trull & Ebner-Priemer, 2013; Wrzus & Mehl, 2015). Our goal here is not to recapitulate these, but rather give critical consideration to what is needed to make applied AA viable. Because the notion of using AA for personalized assessment has really accelerated in the last 5–10 years, our review emphasizes recent publications. Our proposition is that idiographic principles of measurement, which emphasize the estimation of person-specific statistics, need to be infused with nomothetic principles, which provide a firm basis for developing reliable, valid, standardized and normed assessment measures and procedures. We start with a case presentation that exemplifies the type of assessment AA facilitates, while also providing a vehicle to introduce many of the relevant statistical and modeling features available and how they link to clinically relevant concepts. Next, we review the fundamentals of data collection and psychometrics in AA, and end with recommendations for future research and current implementation.

Case Example

We selected this case from a recent study in which psychotherapeutic inpatients completed the Personality Dynamics Diary (PDD) every evening during their inpatient treatment (Zimmermann et al., in press; Study 2). The PDD is a 32-item questionnaire that assesses major dimensions of adaptive and maladaptive behaviors as well as situational experiences. Here, we focus on the following six scales: Positive Events (including agentic and communal reward; four items), Social Stress (including hostility and disappointments; four items), and Workload (two items) capture daily fluctuations in situation experiences; Sociability (including outgoingness and agreeableness; four items), Internalizing Symptoms (including negative affect, detachment, and emotional instability; six items), and Externalizing Symptoms (including impulsivity and aggression; four items) capture daily fluctuations in behaviors. Subjects endorse whether each item applied to them during the last 24 hr using 4-point scales ranging from 0 (*very false*) to 3 (*very true*). The case example we selected was a 54-year old man who received treatment for 84 days and provided diaries for 80 days (95% response rate). The data used here is deidentified and is therefore exempt from our institutional review.

The time series and density distributions for the six scales are presented in Figure 1. They provide a first impression of the typical experiences as well as ups and downs of this particular patient during his hospital stay. For example, this patient hardly experienced any social stress or externalizing symptoms, whereas positive events and internalizing symptoms fluctuated considerably across days. A range of person-specific univariate and bivariate statistics can efficiently capture these impressions. Table 1 summarizes the psychological meaning of these statistics, and

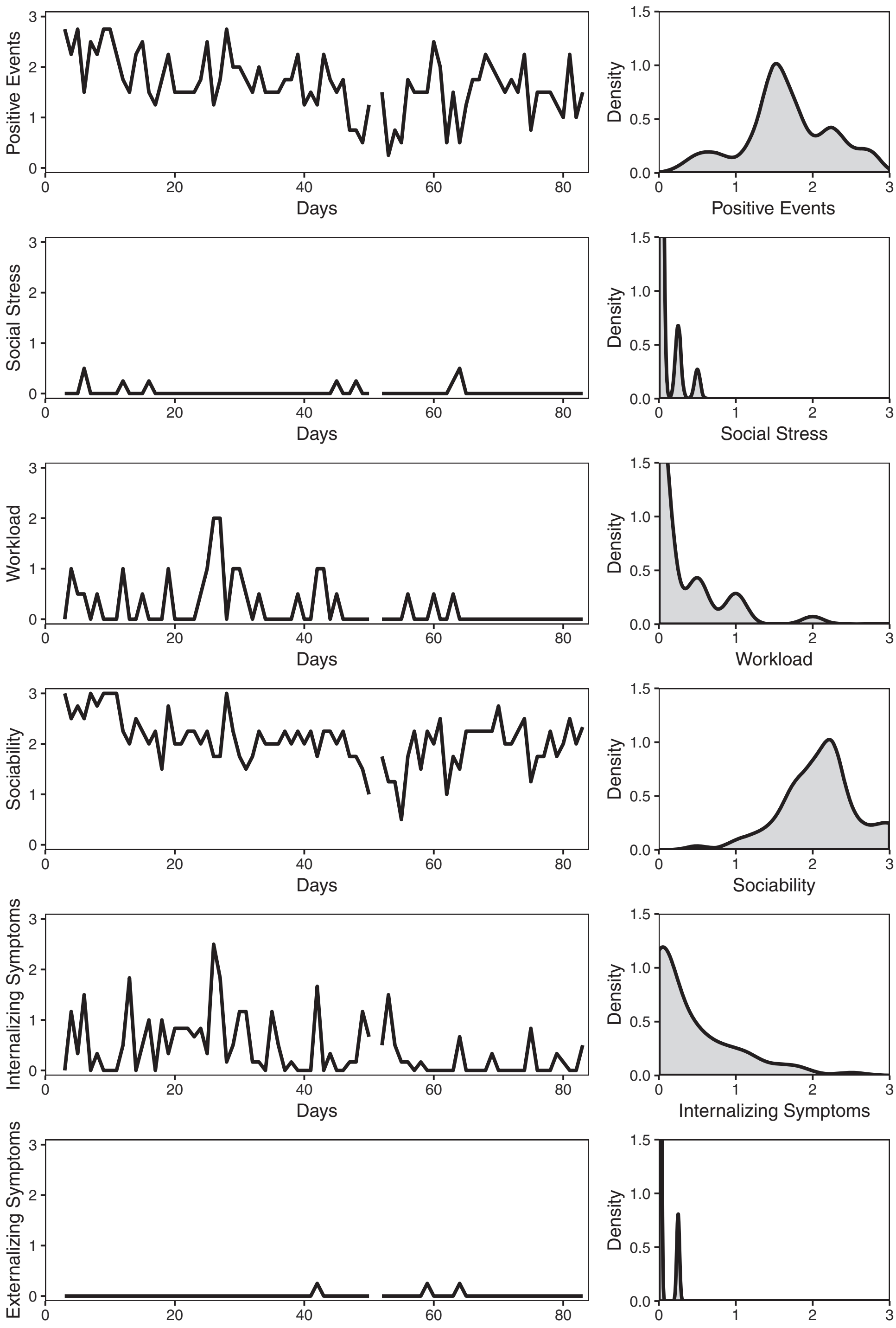

Figure 1. Time series and density distribution of six daily diary scales in the case example.

Table 1
Summary of Commonly Applied Person-Specific Statistics

Statistic	Psychological meaning
Individual mean (i*M*)	The baseline or general tendency of the time series. It may be interpreted as an estimate of the individual's typical or default behavior.
Linear trend (r_{TIME})	The degree to which the time series is correlated with time. It can be interpreted as an estimate of how much the typical behavior of an individual changes linearly across the observation period. The linear trend seems most informative when AA is coupled with some kind of intervention. In other contexts, this statistic is often not of substantive interest or might reflect reactivity to the protocol.
Individual standard deviation (i*SD*)	The gross fluctuation or "flux" of the time series. It may be interpreted as an estimate of the individual's gross variability in behavior.
Individual root mean square successive difference (iRMSSD)	The amount of observation-to-observation change in the time series. It may be interpreted as an estimate of the individual's instability in behavior. In contrast to i*SD*, this statistic is sensitive to the temporal order of the test scores (Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009).
First-order autocorrelation (r_{AUTO1})	The degree to which a current state can be predicted by the previous state. It may be interpreted as an estimate of the individual's carry-over effect or "inertia" of behavior. In applied settings, the autocorrelation is likely to vary between 0 and 1 for most assessment designs (Hamaker, Grasman, & Kamphuis, 2016). Within this range, a low autocorrelation indicates a time series that returns rapidly to its mean following a deviation. A high autocorrelation indicates a time series that once perturbed tends to remain so, taking a long time to return to baseline.
Within-person correlation (*r*)	The degree to which two time series are associated; that is, how much fluctuation in one variable correlates with fluctuation in the other. It can be interpreted as an estimate of how much two processes run in parallel or are dynamically coupled within an individual across time.

Table 2 presents the results for this particular patient. The individual means (i*M*) confirm the impression that experiencing positive events (iM = 1.66) and behaving in a sociable way (iM = 2.08) is much more typical for this patient than experiencing social stress (iM = 0.03) and behaving in an externalizing way (iM = 0.01). With regard to the linear trend, we find that Workload and Internalizing Symptoms decreased over time, which may indicate that the patient responded to the inpatient treatment. Notably, Positive Events and Sociability also decreased, suggesting that the process of recovery for this particular patient was associated with disengaging from social contacts. The individual standard deviations (i*SD*) indicate almost no variability in Externalizing Symptoms (iSD = 0.05) and Social Stress (iSD = 0.10), but relatively high variability in Positive Events (iSD = 0.58). However, although variability was highest in Positive Events, the iRMSSD suggest that instability was highest in Internalizing Symptoms. According to the first-order autocorrelations, it seems that inertia is relatively higher in the daily dynamics of experiencing Positive Events, Workload, and Sociability as compared to the other scales. Within-person correlations show that experiencing Positive Events and Sociability as well as Workload and Internalizing Symptoms are closely intertwined for this patient. We also observe additional associations, including moderate associations between Social Stress and Externalizing Symptoms, and between low Internalizing Symptoms and Sociability. The full within-person correlation matrix is visualized as a network of associations in Panel A of Figure 2, providing the clinician with a rough graphical summary of the patient's dynamic structure.

The univariate and bivariate statistics presented so far summarize key aspects of an individual's AA data. However, more powerful statistical tools are available to jointly analyze the multivariate time series of an individual. One of the most prominent techniques in this regard is vector autoregressive modeling (VAR; Brandt & Williams, 2007), including its recent extensions graphical VAR (Epskamp, van Borkulo, et al., 2018; Wild et al., 2010) and structural VAR (SVAR; Gates, Molenaar, Hillary, Ram, & Rovine, 2010; Wright et al., in press). A key feature of these techniques is that they move beyond the analysis of bivariate contemporaneous relationships (i.e., within-person correlations with a time lag of zero) and capture how multiple time series predict each other while unfolding over time. That is, within-person effects in VAR (with a time lag of one measurement occasion) indicate the extent to which a variable predicts another variable (or itself) at the next measurement occasion, after controlling for the influence of all other variables at the current occasion. Thus, VAR provides insight into unique effects with a

Table 2
Summary of the Six Time Series in the Case Example

Scale	i*M*	i*SD*	iRMSSD	iα	r_{AUTO1}	r_{TIME}	r_{PE}	r_{SS}	r_{W}	r_{S}	r_{IS}
Positive Events (PE)	1.66	0.58	.65	.73	.36	−.43					
Social Stress (SS)	0.03	0.10	.13	0	.08	−.09	−.22				
Workload (W)	0.22	0.44	.48	.67	.38	−.33	.22	.07			
Sociability (S)	2.08	0.49	.51	.76	.42	−.37	.78	−.07	.01		
Internalizing Symptoms (IS)	0.40	0.55	.70	.90	.18	−.35	−.25	.14	.51	−.31	
Externalizing Symptoms	0.01	0.05	.07	0	−.04	.10	−.20	.28	.05	−.10	.14

Note. N = 80 days. iM = individual mean, with a theoretical range from 0 to 3; iSD = individual standard deviation; iRMSSD = individual root mean square successive difference; iα = within-person consistency; r_{AUTO1} = first-order autocorrelation; r_{TIME} = linear trend.

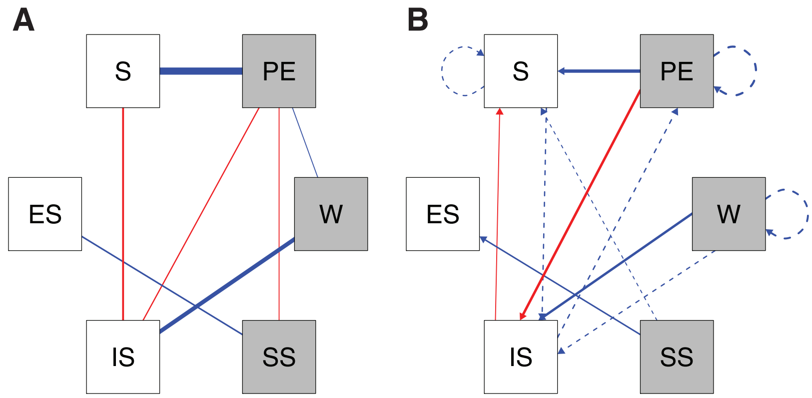

Figure 2. Modeling the within-person structure using raw correlations (Panel A), and structural vector autoregression with a time lag of 1 day (Panel B). Blue (darker) lines represent positive effects, red (lighter) lines represent negative effects. Solid lines represent contemporaneous effects, dotted lines represent lagged effects. The thickness of the lines represents the effect size. Nonsignificant effects ($p > .05$) are omitted. S = Sociability; PE = positive events; W = workload; SS = social stress; IS = internalizing symptoms; ES = externalizing symptoms. See the online article for the color version of this figure.

clear temporal order, thereby highlighting potential causal pathways between variables.

Panel B of Figure 2 presents the results of SVAR for the case example. SVAR can be conceived as an extension of VAR that jointly models lagged and contemporaneous effects between time series as path coefficients. We estimated this model using the "indSEM" function of the R package "gimme" (Lane & Gates, 2017) that automatically performs a stepwise model selection by starting with an empty model and successively adding paths that maximize model fit (i.e., using structural equation modeling modification indices). The final model achieved good fit, with $\chi^2(36) = 43.6$, $p = .18$, root mean square error of approximation = .051, comparative fit index = .975, and standardized root mean square residual = .067. Panel B of Figure 2 reveals a number of interesting findings that go beyond the raw within-person correlation matrix shown in Panel A. For example, Workload and Social Stress seem to act as exogenous predictors that are not influenced by any other variable in the model: Workload predicts current and lagged Internalizing Symptoms, and Social Stress predicts current Externalizing Symptoms and lagged Sociability. The latter path was not visible in Panel A and suggests that the sociable behavior of the patient may in part be driven by trying to repair social conflicts of the past day. On the other hand, Externalizing Symptoms seem to act as an endogenous variable that does not influence any other variable in the model: For this patient, they are simply contingent on contemporaneous Social Stress. The contemporaneous associations between Positive Events, Sociability, and Internalizing Symptoms that are evident from Panel A are now qualified by specific directions: Positive Events predict high Sociability and low Internalizing Symptoms, and Internalizing Symptoms predict low Sociability. However, in addition to these contemporaneous effects we also uncover somewhat surprising lagged effects, with Sociability predicting high Internalizing Symptoms at the next day, and Internalizing symptoms predicting high Positive Events at the next day. This may reflect a pattern of engaging with others one day, only to regret the exchanges and agonize about them the next, and the individual may be motivated to regulate and disrupt internalizing symptoms by seeking out positive experiences after experiencing them. Finally, in line with the pattern of first-order autocorrelations reported in Table 1, we observe significant autoregressive effects for Positive Events, Workload, and Sociability.

Multivariate modeling techniques that are based on VAR and its extensions provide additional summary statistics that may capture relevant features of an individual's psychological structure. For example, graph theory has established a range of centrality indices that putatively capture the importance of a specific node (i.e., variable) in the network (Epskamp, Borsboom, & Fried, 2018). For example, node strength is defined as the sum of the (absolute) path coefficients that are directly related to the node, and thus captures how central fluctuations in this variable are for the whole network. Other indices such as outward degree (i.e., the sum of all outgoing connections of a node) may be especially helpful for selecting variables as leverage points for intervention, because changes in nodes with high outward degree can be expected to lead to maximum changes in other variables. It is also possible to compute indices that summarize aspects of the whole network, for example, the overall residual variance representing the lack of sequential structure (Fisher, Newman, & Molenaar, 2011).

What all the statistics and models presented so far have in common is that they are exclusively based on the data from this particular case. Moreover, all results were interpreted taking this particular case as the frame of reference. This idiographic approach is typical for applied AA and provides a range of insights that appear to be useful for understanding and treating the individual. Nevertheless, this approach leaves several questions unanswered that may have important consequences for interpreting the results: For example, experiencing more Positive Events than Social Stress, and behaving in a relatively Sociable as compared to an Externalizing way might simply be the normative pattern of (self-reported) daily experience that applies to most persons, providing no specific insight about this particular case. The same might be true for most of the person-specific statistics and models, including the seemingly informative associations of Positive

Events and Sociability as well as of Workload and Internalizing Symptoms. One of the reasons for developing the PDD was to address these concerns by providing a standardized measure for assessing major dimensions of within- and between-person differences. That is, instead of tailoring the item pool to the particular patient, the PDD imposes a fixed set of items with a common factor structure across levels, thereby promoting a nomothetic perspective to applied AA. For example, using a sample of 77 inpatients completing the PDD while receiving the same treatment as a frame of reference (see S11 in Zimmermann et al., in press), it turns out that the amount of Positive Events the patient experienced was rather typical (T = 53), whereas the scores on Social Stress (T = 34) and Externalizing Symptoms (T = 37) were indeed quite low, and the scores on Sociability (T = 63) quite high. Moreover, despite representing normative processes that are typical for most persons, the associations of Positive Events and Sociability (T = 64) as well as of Workload and Internalizing Symptoms (T = 60) were indeed clearly pronounced for this particular case. These comparisons should illustrate that the interpretation of person-specific statistics and models can be enriched when based on standardized protocols and linked back to normative (group) data.

This case offers one glimpse of the type of information that might be gleaned with AA in a single case scenario. This example belies a large number of essential considerations that went in to the design of the assessment battery and protocol. In the next section, we turn to summarizing these main considerations with a view toward developing methods and measurement that would support scalable and generalizable AA in clinical settings.

Design Considerations in AA

Unlike traditional cross-sectional assessments, which largely rely on fixed stimuli presented at a single time-point, designing AA protocols requires an explicit consideration of the sampling schedule and frame. Much has been written about these essential design features (see, e.g., Reis, Gable, & Maniaci, 2014), but with few exceptions like the areas of pain (e.g., Stone et al., 2003) or affect, little systematic investigation has studied what happens if the design features are varied. For instance, what differences emerge between event-contingent and signal contingent assessment (Himmelstein, Woods, & Wright, 2018)? Here, we review major design considerations, but emphasize that for the purposes of applied assessment, more basic research is needed. The key considerations can be organized around the five *W*s, taking the first as a given: Who is being assessed? Although presented as distinct, these considerations interdigitate, such that what is assessed may have implications for how, when, and where it is assessed, and so on.

What Is Being Assessed?

In general, the constructs that are most interesting to track using AA are those that vary across time and circumstances on a relatively brief time scale (i.e., moments to days). This is one reason affect has so frequently been the focus of AA research, because it is well understood to fluctuate in reaction to life's vicissitudes, thereby serving its functional purpose (Kuppens, Oravecz, & Tuerlinckx, 2010). In contrast, something that is not presumed to vary much over short periods of time, such as the number of friends in one's social network or one's weight, is less amenable to AA. Constructs commonly sampled in AA include stressors (e.g., Kamarck et al., 2005), situational perceptions (e.g., Sherman, Rauthmann, Brown, Serfass, & Jones, 2015), motivations (e.g., McCabe & Fleeson, 2012), interpersonal behavior (e.g., Moskowitz & Zuroff, 2004), psychiatric symptoms (e.g., Myin-Germeys et al., 2009), specific behaviors (e.g., substance use; Carney, Tennen, Affleck, Del Boca, & Kranzler, 1998; Shiffman, Paty, Gnys, Kassel, & Hickcox, 1996), and personality states (e.g., Fleeson, 2001). Most of these can be assessed using self-report inventories designed to capture states as opposed to traits (i.e., with appropriate wording and temporal referents). Beyond these traditional psychological constructs, AA has a long history of assessing ambulatory psychophysiology (Fahrenberg, 1996), sleep and wake patterns, activity levels, and endocrine functioning (e.g., salivary cortisol samples).

How Is It Assessed?

How data are sampled has implications for which construct and how frequently it can be sampled. Self-report and passive sensors are likely to encompass the large majority of AA data in practical settings. Self-report in AA has the important strength of capturing an individual's subjective experience. It stands to reason that burden, fatigue, and reactivity (i.e., changing one's behavior in response to repeated sampling) could threaten the validity of self-reported AA data. Yet as Shiffman and colleagues (2008) reviewed, scant evidence has been found to support the concern in the aggregate. However, our personal experience running large samples through AA research would suggest that there are marked individual differences in reactivity—some clearly experience it, some do not. As noted, when administered intensively and repeatedly in the moment, self-reports allow for direct assessment of patterns that can otherwise be muddied when using traditional cross-sectional assessment (e.g., due to biases of retrospection, lack of insight, poor memory). The PDD used in the case example here follows the classic self-report measure model.

But other options exist, including having the participant capture qualitative data (e.g., photo, video, audio) that can then be coded in some way either by person or machine, and, although difficult to implement, it is possible to get observer reports or the reports from multiple informants (Roche et al., 2014). One example might be using a yoked smart-phone application, whereby when one person enters data about an event (e.g., target patient), another is prompted to do the same (e.g., patient's spouse). Emerging technologies use sensor arrays, either newly designed to be worn by the individual for the purpose of AA or take advantage of the large number of sensors embedded in the average person's smartphone (Beierle et al., 2018; Ferreira, Kostakos, & Dey, 2015; Harari et al., 2016). Examples of passive sensors that might be relevant include heart-rate monitors, blood-pressure cuffs, accelerometers, global positioning system antennae, microphones, light sensors, Bluetooth, and screen or application use recorders. To illustrate, smartphone screen on and off times can be used to reliably index an individual's sleep–wake cycle (Chen et al., 2013). What should be made clear is that for the most part, passive sensors only indirectly correspond to constructs of interest (Mohr et al., 2017). For instance, a practitioner may be interested in assessing arousal using passive sensors (e.g., heart-rate, skin conductance), but these sensors are unlikely to be specific to arousal states, and algorithms will need to be developed to translate streams of sensor data into interpretable states. In developing these algorithms, which is largely done in a data-driven manner using machine-learning techniques, there is a tension between building models that are highly accurate and those that generalize well. In the published research, small and nonrepresentative samples currently abound, and thus, much work is needed in this area. Nevertheless, the potential to assess individuals using objective measures that continuously record behavior is opening up exciting new avenues of measurement.

In general, though, complex psychological constructs (i.e., motivations, thoughts, feelings) are most likely to be assessed directly with self-report surveys, whereas sensor data may be used to generate objective state categories (e.g., walking, running, sleeping, communicating; Ickin et al., 2012). Passive monitoring has vast potential and can be used to infer relatively complex states (e.g., stress), but likely has an upper limit to what can be inferred. Issues of how data are being collected are tightly intertwined with what is being sampled.

When (and How Frequent) Are Assessments?

Common sense dictates that assessment schedules should match the occurrence and timing of the phenomena of interest. In practice this can be difficult to achieve. In part this is because most theories have given little attention to the precise timing and frequency of relevant events. For instance, if one is self-mutilating to reduce negative affect, does it occur immediately when it is experienced, after some sustained period of time, or threshold (Nock & Prinstein, 2004)? Even if these were well articulated, real world considerations, such as burden, fatigue, and rarity of events impinge upon the assessment enterprise. For one, events and behavior of interest may not occur on the same time-scale. Nightly sleep occurs once per day, whereas stressful events and their evoked response can happen many times within a day. Some high-value events may be quite rare (e.g., interpersonal violence, self-mutilation), such that even with frequent sampling an assessment is unlikely to capture many instances. One might be tempted to reduce the frequency of sampling (e.g., once daily), but this may blur the specific processes leading up to the target behavior if they occur on a briefer time-scale. Selecting a sampling scale often involves trade-offs between frequency and fidelity. Being able to make reliable inferences requires some minimum of observations of an event, whereas making inferences at the right level of granularity requires the correct timing. Both are important but may be at odds with each other in many scenarios. Thus, rules that require sampling more or less frequently than an event can be expected to occur are generally misguided, and instead the exact timing schedule should be selected based on substantive and methodological concerns (both practical and quantitative). In our case example, we chose daily assessments because it is an intuitive time unit that is easy to implement, it is infrequent enough to ask a longer scale, and many of the target behaviors are too rare for hour-to-hour assessments, but not so rare that week-to-week assessments would suffice.

Three general categories of sampling schemes have been developed: event contingent, fixed, and random interval. Event contingent recording is tied to the occurrence of some event, a situation, or internal state. This approach is often used for rare or unpredictable events, as might be the case in meaningful interpersonal interactions (Moskowitz & Zuroff, 2004) or binge-eating episodes (Smyth et al., 2007). The target of the assessment is instructed to complete a survey or begin recording at the start of or soon after an event has occurred. A challenge associated with this approach is the difficulty determining whether all relevant events have been reported. Realistically some degree of missing data is expected, though it is difficult to ascertain the amount and nature of the missing data. Hence, one must hope that the observed cases are representative of the total set of cases (i.e., missing is completely at random). One of the exciting promises of passive sensing is its ability to safeguard against missed events by detecting them and prompting the participant to respond. Alternatively, one sensor could be used to trigger other sensors (e.g., Bluetooth contact between smartphones in proximity could engage the microphone to capture an interpersonal interaction).

An alternative approach is to adopt a fixed or random interval of assessment. Daily diaries commonly use a fixed interval approach, as was done in the case example above. The assessment target is asked to complete a survey each evening. Less common, although also feasible, is an hour or quarter-hour interval. Naturally, the narrower the interval, the more participant burden, therefore brief fixed intervals require strong justification. Fixed intervals need not be symmetrical and building in some asymmetry to more densely sample important parts of a process is conceivable (e.g., few samples during the day but frequent assessments on nights when substance use occurs; Piasecki et al., 2011). Continuous sampling, as is the case with most passive sensors, would be considered a fixed schedule, albeit with very high sampling rate (e.g., every minute or second). A concern with fixed interval recording is that the assessed target may change their natural behavior and rhythms according to the schedule (i.e., reactivity).

Random or pseudorandom prompts have been used to get around these concerns. A typical approach might be to sample six to eight times per day, but at random times or at random within some block (e.g., 2 hr) to avoid large periods without sampling. A limitation of truly random prompts arises when one wants to use statistical models that examine autoregressive effects (i.e., time-series models), which have traditionally relied on an assumption of equal sampling interval. New statistical methods are being developed that circumvent or account for unequal intervals (de Haan-Rietdijk, Voelkle, Keijsers, & Hamaker, 2017). Alternatively, random assessments within defined blocks might be assumed to be equidistant on average, and estimated effects interpreted accordingly. Finally, sampling approaches are not mutually exclusive, such that they can be combined with good effect. For instance, one might combine event-contingent sampling with random prompt sampling to decouple assessments of events and affect (e.g., Greeno, Wing, & Shiffman, 2000; Kockler, Santangelo, & Ebner-Priemer, 2018). This approach, although more burdensome, allows for examining antecedents and consequents of events.

Where Is It Assessed?

Modern AA methods can also be used to tailor assessments to specific settings or contexts, in what has been termed "context-aware" sampling (Intille, 2007). Certain behaviors may not occur frequently in some settings (e.g., interactions with employers) or more frequently in others (e.g., drinking at a bar). Isolating assessments to relevant contexts can reduce burden and the collection of large amounts of irrelevant data. Similar to event contingent sampling, context contingent sampling could be entrusted to the participant (i.e., instructions to complete assessments only while at work, or drinking at a bar) or alternatively might be offloaded to hardware and software (e.g., using geo-fencing to establish "hot-zones" for assessment around relevant establishments or sensors that detect certain contexts).

Psychometric Foundations of Applied AA

Research using AA has yielded valuable new insights, but it has seldom systematically addressed the reliability and validity of either the momentary scales used or the statistics derived from time series of symptoms, behaviors, contexts, and so on. For instance, with some notable exceptions (e.g., Tomko et al., 2014), it is common in basic and applied research to arbitrarily select items that are presumed to tap in to the constructs of interest without complete psychometric evaluation (e.g., David, Marshall, Evanovich, & Mumma, 2018; Schiepek et al., 2016). That is, basic requirements for interpreting and evaluating test results may be not met in many instances of applied AA (International Test Commission, 2013). To place applied AA on firm empirical footing, it is necessary to addresses these issues by both developing new measures that adhere to contemporary assessment standards and refine these standards in line with the specific challenges of AA (Bornstein, 2017). We adapt nomothetic assessment basics for use with AA and person-specific statistics and point to some of the emerging work that may coalesce in the future.

Reliability

Broadly speaking, reliability captures whether a test score is reproducible. If we could apply the same AA twice (i.e., under the same circumstances) on the same person we would hope to get the same, or at least very similar, results. At least three aspects of reliability can be distinguished in an applied AA setting: within-person consistency, precision, and stability.

Within-person consistency refers to the reliability of a single time series. It is defined by the relative amount of true score variability in a time series, or, in other words, the degree to which differences among scores across time reflect true score changes. In cross-sectional measurement, consistency of a scale can be straightforwardly computed using classic measures of internal consistency like Cronbach's alpha or McDonald's omega. Within-person analogues for these statistics have been developed to estimate the degree to which fluctuations across time in constructs is consistent across presumed constituent items (Hu et al., 2016). In research settings it is common to estimate the pooled within-person consistency using multilevel modeling (e.g., Cranford et al., 2006; Geldhof, Preacher, & Zyphur, 2014; Shrout & Lane, 2012), although it is likely that within-person consistency differs between persons. Note that in many AA applications, only single items are used to represent a construct, and thus within-person consistency remains unclear in these cases (but see Schuurman & Hamaker, 2018, for a possible solution). We argue that the number of items per scale has to balance the burden of repeated assessments with the construct of interest, with some constructs only requiring a single item (e.g., cigarettes smoked in the last hour) and more complex constructs requiring several items (e.g., positive or negative affect [Scott et al., 2018] or impulsivity [Tomko et al., 2014]).

Precision refers to the reliability of a single AA statistic. It is based on the assumption that measurement occasions are randomly sampled from the individual's life and can be defined by the standard error of the statistic at hand, with smaller standard errors indicating greater precision. Although this is not a standardized measure of reliability, it can be used to test whether a statistic is significantly different from zero, or to compute a 95% confidence interval for the statistic. For instance, in the case example, the standard error of the individual mean for Positive Events is 0.064, resulting in a 95% confidence interval of [1.53, 1.78]. This suggests that we were able to estimate the i*M* of Positive Events for this particular patient with relatively high precision (given that the full theoretical scale ranges from 0 to 3). Note that this is analogous to the use of the standard error of measurement in cross-sectional assessment, but now is based on the individual's own distribution of scores. The precision of other statistics for which the standard error cannot be formally derived (e.g., centrality indices in SVAR networks) may be estimated using resampling techniques (Epskamp, Borsboom, et al., 2018). It is also possible to incorporate information on within-person consistency (e.g., by using latent variable modeling) when estimating the statistics of interest as well as their precision (Schuurman, Houtveen, & Hamaker, 2015). In any case, we argue that it is crucial to estimate precision of AA statistics before one can interpret them in single case scenarios. This is especially important when dealing with multivariate models such as networks of partial correlations which are likely more susceptible to sampling error (Forbes, Wright, Markon, & Krueger, 2017). Otherwise there is an unknown risk of interpreting something as a signal that is simply noise.

Stability can be interpreted at the within-person and the between-person level. At the within-person level stability refers to the assumption that the statistic of interest is stationary, that is, does not change across the assessment period. This assumption can be explicitly tested using more general models that allow for time-varying parameters (Bringmann et al., 2017). At the between-person level stability refers to the rank-order stability of individuals when assessed twice. Under the assumption that the underlying construct does not change, the rank-order stability of a statistic corresponds to its retest-reliability. In classical psychological testing, this reliability estimate is used to compute the standard error of measurement, which can be thought of as the between-person analogue of the standard error mentioned above. There is some evidence that the rank-order of individual means and standard deviations is relatively stable (Jones, Brown, Serfass, & Sherman, 2017; Wright & Simms, 2016), but this is much less clear for within-person correlations (Mejía, Hooker, Ram, Pham, & Metoyer, 2014) and VAR based networks (Beck & Jackson, 2018). A crucial question for future applied AA research will be to clarify how many data points are needed to obtain stable estimates for a specific statistic.

Validity

We argue that applied AA should be judged by the same established criteria of validity as other approaches in psychological assessment. These criteria include aspects such as structural validity (i.e., a theoretically convincing measurement model), convergent validity and discriminant or incremental validity. However, specific challenges come along with the differentiation of within- and between-person structures that is inherent to applied AA.

Measurement models of popular constructs (e.g., personality traits) have almost exclusively been developed based on cross-sectional designs, and such designs actually confound the within-person and between-person measurement model (Hamaker, Schuurman, & Zijlmans, 2017). Thus, the enterprise of developing multi-item measures for applied AA is faced with the challenge of developing new models that ensure *measurement invariance* across levels of analysis (Adolf, Schuurman, Borkenau, Borsboom, & Dolan, 2014; Hamaker et al., 2017). To take a first step in this direction, the items of the PDD were selected for showing a factor loading pattern that is largely equivalent across the pooled within-person and between-person matrix (Zimmermann et al., in press). However, even when adopting such a test development strategy, it seems easily possible that individual within-person factor structures differ between persons (Adolf et al., 2014; Brose, Voelkle, Lövdén, Lindenberger, & Schmiedek, 2015). Thus, it will be important to investigate (e.g., using simulation studies) whether deviances from measurement invariance across persons have detrimental practical consequences for applied AA (e.g., result in biased estimates of within-person correlations). Such research efforts may also come to the conclusion that in some areas achieving measurement invariance across levels is simply not possible, for example, because within- and between-person processes are too different (Molenaar & Campbell, 2009).

Convergent validity refers to the degree to which two measures of the same (or very similar) constructs are associated. There is considerable research showing that AA statistics exhibit theoretically consistent associations with measures of psychopathology and personality. For example, the levels of emotional and behavioral states are robustly related to their corresponding personality trait (Fleeson & Gallagher, 2009; Wright & Simms, 2016), the variability and instability of negative emotions is related to constructs such as neuroticism (Houben, van den Noortgate, & Kuppens, 2015), negative affectivity (Wright & Simms, 2016), and borderline personality disorder (Trull et al., 2008), inertia in negative emotions is positively associated with low self-esteem and (future) depressive symptoms (Kuppens, Allen, & Sheeber, 2010; van de Leemput et al., 2014), within-person correlations between stress and negative emotions have been interpreted as a measure of emotional (or stress) reactivity and were found to be associated with neuroticism and risk for psychosis (Bolger & Schilling, 1991; Myin-Germeys, van Os, Schwartz, Stone, & Delespaul, 2001), and research using VAR has indicated, for instance, that stronger lagged effects between different negative emotions are associated with Major Depression (Pe et al., 2014) and neuroticism (Bringmann et al., 2016). However, this kind of research is far from being conclusive. For example, even if individual means of states are robustly related to their corresponding traits, this may not generalize to other assessment methods. In fact, a recent study suggests that individual means of personality states do not predict informant reports of these traits beyond global self-reports, questioning the incremental value of this AA statistic (Finnigan & Vazire, 2018). Moreover, research on convergent validity of within-person variance is almost fully lacking. This seems concerning because there may be interesting discrepancies across levels. For instance, Hamaker (2012) provides the evocative and almost axiomatic example of typing speed and errors at the within- (positively correlated) and between-person (negatively correlated) levels. Other examples include exercise and heart rate (positive within, negative between), and search for meaning and wellbeing (positive within, negative between), which all point to the importance of establishing validity at each level.

Discriminant validity refers to the degree to which two measures of different constructs are unrelated, and incremental validity refers to the extent to which a measure increases the predictive ability beyond that provided by other relevant measures. There is increasing evidence that many of the AA statistics are highly overlapping from a between-person perspective, and thus probably lacking discriminant and incremental validity (e.g., i*SD*, iRMSSD, and autocorrelation are not mathematically independent). For example, a recent study suggests that when controlling for individual means, other univariate statistics of the individual distributions add little to the prediction of self-reported personality traits (with the exception of i*SD*s being incrementally related to emotionality and extraversion; Jones et al., 2017). As another example, a reanalysis of data presented as evidence for network connection strength being associated with psychopathology (Wigman et al., 2013) suggests that this was an artifact produced by associations of psychopathology with i*SD*s (Terluin, de Boer, & de Vet, 2016). Thus, it seems highly important to jointly consider multiple AA statistics when predicting relevant outcomes to control for their potential overlap (e.g., Castro-Schilo & Ferrer, 2013). More research is needed using a bottom-up approach on the between-person structure of AA statistics (Wigman et al., 2015).

There are further issues that may challenge the validity of AA statistics. For example, they may have quite different meanings depending on the actual content/behavior of the time series (e.g., inertia of positive emotions may have a different psychological meaning than inertia of negative emotions). Actually, most research on the validity of dynamic statistics has been done only with AA ratings of emotions (Houben et al., 2015); thus, their validity with regard to, for example, other personality trait expression, interpersonal behavior, or situation experiences is still unclear. At least one recent study suggests that, although within-person variation in personality states highly overlaps with within-person variation in positive and negative affect, it also captures some information above and beyond that (Wilson, Thompson, & Vazire, 2017). A further threat to validity is response style. For example, a recent study suggested that some statistics, such as i*SD*, may be highly confounded by response styles, in this case, the unspecific tendency to use extreme responses (Baird, Lucas, & Donnellan, 2017). One solution could be, in the long run, to deemphasize self-report rating scales and either replace or augment them with passive sensing based on behavioral stream data (Mohr et al., 2017).

Recommendations for Future Research

We wish to end with a call for basic research on the application of AA to begin developing an empirical base for its use in practical settings. To answer this call, we see the following as the necessary steps: First, measures need to be developed specifically for AA. This should involve consideration of the various issues specific to AA (e.g., timescale, burden) during item derivation and instrument validation. Similar, if not greater, care should be given to reliability and validity during item auditioning. Validation should include both the within- and between-person level of analysis and should consider the use of person-specific statistics beyond individual means. We perceive the greatest need for broadband or omnibus measures that would provide a suite of scales that could be deployed using a flexible battery approach given a particular patient's issues. That is, because not all scales would be relevant to all patients (e.g., Externalizing and Social Stress showed low reliability and variability in the case example), only those that are relevant could be used in a given case. However, there would be considerable benefit in having scales with shared features (e.g., on a similar response format) and standardization. Additional systems that would be valuable include measures of highly relevant contextual features. Focal measures or systems of assessment (e.g., passive sensing of stress) should be developed following these recommendations.

Second, normative data should be gathered for current and future developed measures. Norms should be gathered not only for specific populations of interest, but also for the various statistics of interest, and any other application variable (e.g., timescales) that might differ when implemented. Then, it is possible to compute all relevant statistics separately for each individual and use the resulting distributions to generate norm tables (e.g., using *T* values). This is the approach we illustrated in the case example above (although we admit that the reference sample was neither large nor representative). A more advanced approach to normalization would be to incorporate assumptions and knowledge about the distributional characteristics of these statistics directly when estimating them for the individual case. This may substantially enhance the precision and robustness of the individual estimates, especially for more complex VAR models. Different methods are available that impose different distributional assumptions. For example, group iterative multiple model estimation (Gates & Molenaar, 2012; Lane & Gates, 2017) makes no assumptions of the distributions of path coefficients in SVAR models across persons, whereas multilevel VAR assumes normal distributions (Bringmann et al., 2013). An even more flexible approach to incorporating prior knowledge about distributions may be Bayesian estimation (Wagenmakers, Morey, & Lee, 2016), which has been successfully applied to single case scenarios (Schuurman et al., 2015; see also Epskamp, van Borkulo, et al., 2018, for a similar suggestion). However, it should be noted that other researchers have expressed concerns about integrating group-level data in the estimation process (Ram, Brinberg, Pincus, & Conroy, 2017), so this issue requires careful attention in future research.

Third, the apparatus of implementation must be developed. We need tools that automatically analyze AA data and provide a user-friendly feedback for patients or clinicians with guides to appropriate interpretations. This includes user friendly software that would sit on the assessment target's device (e.g., smartphone), as well as on the practitioner's computer. The software must be based on user-centered designs that make it intuitive if not pleasing to use as well as be able to provide easy to consume output that distils important information from the AA. Ideally this would be customizable, and we would encourage someone to develop a general framework that could flexibly add and remove AA battery components as opposed to different proprietary applications for each measure. First steps in these directions have been made, including, for instance, automated VAR data analyses with graphical output for ease of interpretation (Kaiser & Laireiter, 2017; van der Krieke et al., 2015).

Fourth, even if we have technical solutions, it is important to make sure that test users find the information provided useful, and that the increased burden is worth the effort (i.e., provide incremental information). In other word, more research is needed to demonstrate the clinical acceptability and utility of applied AA. Similar to treatment research that differentiates between efficacy and effectiveness, we view basic AA research as akin to efficacy, and research on usefulness in applied settings under real world conditions as akin to effectiveness. For example, feasibility in clinical settings must be demonstrated, not taken on faith. Although results for feasibility of AA and usability of personalized feedback is promising for lay users or patients (van der Krieke et al., 2017; Zimmermann et al., in press), mental health professionals may be more reluctant to find AA results worth the effort (Zimmermann et al., in press). Moreover, insofar as this type of assessment provides a more faithful view of clinically relevant processes, and potentially with less bias than retrospection, they would provide ideal treatment targets (e.g., examine changes in within-person autocorrelation before and after treatment). Whether this is the case is currently unclear, for example, one study did not find treatment effects on network structure (Snippe et al., 2017). It may also be possible to show that personalized feedback itself can reduce symptoms. However, research in this area is also just emerging and results are preliminary. For example, there is some tentative evidence that symptoms can be reduced (Kramer et al., 2014) or positive affect increased (van Roekel et al., 2017), although it is likely that the utility of feedback to effect change will vary widely across different populations. Finally, it would be compelling if personalized models lead to more rapid treatment gains by allowing for bespoke interventions. Although this question naturally follows from the notion of personalized models, little work has examined it (cf. Fernandez, Fisher, & Chi, 2017). We note, however, that unfortunately the ability for traditional in-depth psychological assessment to significantly improve treatment outcomes is a woefully understudied topic.

We recognize that these are broad recommendations, and they are purposefully so, intended to guide general efforts. The specific steps and decisions may differ considerably depending on the purpose of the measure and its intended use in practice. We hope the review leading up to these points will stimulate the thinking needed to address the challenging but potentially fruitful work needed to place applied AA on a firm scientific foundation.

Conclusion

We have covered some of the main considerations for moving toward empirically based AA in applied settings. Enthusiasm for using AA to develop personalized models is palpable in the field.

We share this enthusiasm, but also believe that although AA and accompanying developments in technology and methods facilitate idiographic assessment and analyses, there is value in hewing closely to the lessons learned from nomothetic assessment as we cut a new path forward. Thus, we have argued that developing measures and models that simultaneously satisfy considerations of AA design and traditional notions of reliability and validity are paramount.

Many additional issues are in need of attention. Some that come to mind include: What should be done with improper distributions or values in an individual case, such as when variables are highly skewed, or even have a variance of zero? What are reasonable expectations for power and precision (i.e., how many data points per person are needed to obtain reliable person-specific estimates)? How should burden in the form of number of items, number of times per day, and total length of protocol be handled? What are reasonable upper bounds on these demands before data quality begins to degrade? What tools can be appropriated to ensure individuals remain engaged and motivated? And, critically, is the effort and expense of AA worth it? Can people simply self-report their "dynamic" patterns given a reasonable prompt? Does AA provide incrementally useful information beyond cross-sectional assessment? How effective would just in time interventions be, and might that justify the effort? Each of these is intertwined with the points we covered in detail above, but also require additional consideration that goes beyond this summary. Indeed, in reviewing the literature we found it much easier to generate questions than answers. Much impressive work has been done over the past several years, and much is ongoing, but there is much left to be done. We are excited for the next steps.

References

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology, 5,* 883. http://dx.doi.org/10.3389/fpsyg.2014.00883

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality, 69,* 170–179. http://dx.doi.org/10.1016/j.jrp.2016.06.015

Bak, M., Drukker, M., Hasmi, L., & van Os, J. (2016). An n=1 clinical network analysis of symptoms and treatment in psychosis. *PLoS ONE, 11,* e0162811. http://dx.doi.org/10.1371/journal.pone.0162811

Beck, E. D., & Jackson, J. J. (2018). *Consistency and change in idiographic personality: A longitudinal ESM network study*. Manuscript submitted for publication.

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., . . . Zimmermann, J. (2018). Context data categories and privacy model for mobile data collection apps. *Procedia Computer Science, 134,* 18–25. http://dx.doi.org/10.1016/j.procs.2018.07.139

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment, 23,* 447–458. http://dx.doi.org/10.1177/1073191116648209

Bolger, N., & Schilling, E. A. (1991). Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality, 59,* 355–386. http://dx.doi.org/10.1111/j.1467-6494.1991.tb00253.x

Bornstein, R. F. (2017). Evidence-based psychological assessment. *Journal of Personality Assessment, 99,* 435–445. http://dx.doi.org/10.1080/00223891.2016.1236343

Brandt, P. T., & Williams, J. T. (2007). *Multiple time series models: Quantitative applications in the social sciences* (Vol. 148). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412985215

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods, 22,* 409–425. http://dx.doi.org/10.1037/met0000085

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., . . . Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment, 23,* 425–435. http://dx.doi.org/10.1177/1073191116645909

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., . . . Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE, 8,* e60188. http://dx.doi.org/10.1371/journal.pone.0060188

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality, 29,* 55–71. http://dx.doi.org/10.1002/per.1961

Carney, M. A., Tennen, H., Affleck, G., Del Boca, F. K., & Kranzler, H. R. (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time "electronic interviews." *Journal of Studies on Alcohol, 59,* 447–454. http://dx.doi.org/10.15288/jsa.1998.59.447

Carpenter, R. W., Wycoff, A. M., & Trull, T. J. (2016). Ambulatory Assessment. *Assessment, 23,* 414–424. http://dx.doi.org/10.1177/1073191116632341

Castro-Schilo, L., & Ferrer, E. (2013). Comparison of nomothetic versus idiographic-oriented methods for making predictions about distal outcomes from time series data. *Multivariate Behavioral Research, 48,* 175–207. http://dx.doi.org/10.1080/00273171.2012.736042

Cattell, R. B., & Luborsky, L. B. (1950). P-technique demonstrated as a new clinical method for determining personality and symptom structure. *The Journal of General Psychology, 42,* 3–24. http://dx.doi.org/10.1080/00221309.1950.9920145

Chen, Z., Lin, M., Chen, F., Lane, N., Cardone, G., Wang, R., . . . Cambell, A. (2013). Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE. http://dx.doi.org/10.4108/pervasivehealth.2013.252148

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57,* 505–528. http://dx.doi.org/10.1146/annurev.psych.57.102904.190146

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32,* 917–929. http://dx.doi.org/10.1177/0146167206287721

David, S. J., Marshall, A. J., Evanovich, E. K., & Mumma, G. H. (2018). Intraindividual dynamic network analysis: Implications for clinical assessment. *Journal of Psychopathology and Behavioral Assessment, 40,* 235–248. http://dx.doi.org/10.1007/s10862-017-9632-8

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology, 8,* 1849. http://dx.doi.org/10.3389/fpsyg.2017.01849

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology, 118,* 195–202. http://dx.doi.org/10.1037/a0014868

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50,* 195–212. http://dx.doi.org/10.3758/s13428-017-0862-1

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science, 6,* 416–427. http://dx.doi.org/10.1177/2167702617744325

Fahrenberg, J. (1996). Ambulatory assessment: Issues and perspectives. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 3–20). Seattle, WA: Hogrefe & Huber.

Fernandez, K. C., Fisher, A. J., & Chi, C. (2017). Development and initial implementation of the Dynamic Assessment Treatment Algorithm (DATA). *PLoS ONE, 12,* e0178806. http://dx.doi.org/10.1371/journal.pone.0178806

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile context instrumentation framework. *Frontiers in ICT*. Advance online publication. http://dx.doi.org/10.3389/fict.2015.00006

Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology, 115,* 321–337. http://dx.doi.org/10.1037/pspp0000136

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology, 83,* 825–836. http://dx.doi.org/10.1037/ccp0000026

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment, 23,* 496–506. http://dx.doi.org/10.1177/1073191116638735

Fisher, A. J., Newman, M. G., & Molenaar, P. C. M. (2011). A quantitative method for the analysis of nomothetic relationships between idiographic structures: Dynamic patterns create attractor states for sustained posttreatment change. *Journal of Consulting and Clinical Psychology, 79,* 552–563. http://dx.doi.org/10.1037/a0024069

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distribution of states. *Journal of Personality and Social Psychology, 80,* 1011–1027. http://dx.doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology, 97,* 1097–1114. http://dx.doi.org/10.1037/a0016786

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology, 126,* 969–988. http://dx.doi.org/10.1037/abn0000276

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63,* 310–319. http://dx.doi.org/10.1016/j.neuroimage.2012.06.026

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage, 50,* 1118–1125. http://dx.doi.org/10.1016/j.neuroimage.2009.12.117

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19,* 72–91. http://dx.doi.org/10.1037/a0032138

Greeno, C. G., Wing, R. R., & Shiffman, S. (2000). Binge antecedents in obese women with and without binge eating disorder. *Journal of Consulting and Clinical Psychology, 68,* 95–102. http://dx.doi.org/10.1037/0022-006X.68.1.95

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: Guilford Press.

Hamaker, E. L., Grasman, R. P. P. P., & Kamphuis, J. H. (2016). Modeling BAS dysregulation in bipolar disorder. *Assessment, 23,* 436–446. http://dx.doi.org/10.1177/1073191116632339

Hamaker, E. L., Schuurman, N. K., & Zijlmans, E. A. O. (2017). Using a few snapshots to distinguish mountains from waves: Weak factorial invariance in the context of trait-state research. *Multivariate Behavioral Research, 52,* 47–60. http://dx.doi.org/10.1080/00273171.2016.1251299

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26,* 10–15. http://dx.doi.org/10.1177/0963721416666518

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science, 11,* 838–854. http://dx.doi.org/10.1177/1745691616650285

Haynes, S. N., Mumma, G. H., & Pinson, C. (2009). Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clinical Psychology Review, 29,* 179–191. http://dx.doi.org/10.1016/j.cpr.2008.12.003

Himmelstein, P. H., Woods, W. C., & Wright, A. G. C. (2018). *A comparison of signal- and event-contingent recording methodologies in ambulatory assessment*. Manuscript submitted for publication.

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, 141,* 901–930. http://dx.doi.org/10.1037/a0038822

Hu, Y., Nesselroade, J. R., Erbacher, M. K., Boker, S. M., Burt, S. A., Keel, P. K., . . . Klump, K. (2016). Test reliability at the individual level. *Structural Equation Modeling, 23,* 532–543. http://dx.doi.org/10.1080/10705511.2016.1148605

Ickin, S., Wac, K., Fiedler, M., Janowski, L., Hong, J.-H., & Dey, A. (2012). Factors influencing quality of experience of commonly used mobile applications. *IEEE Communications Magazine, 50,* 48–56. http://dx.doi.org/10.1109/MCOM.2012.6178833

International Test Commission. (2013). *ITC guidelines on test use* (Version 1.2; Document Reference Number ITC-G-TU-20131008). Unpublished manuscript, International Test Commission, Geneva, Switzerland.

Intille, S. S. (2007). Technological innovations enabling automatic, context-sensitive ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 308–337). United Kingdom: Oxford University Press.

Jones, A. B., Brown, N. A., Serfass, D. G., & Sherman, R. A. (2017). Personality and density distributions of behavior, emotions, and situations. *Journal of Research in Personality, 69,* 225–236. http://dx.doi.org/10.1016/j.jrp.2016.10.006

Kaiser, T., & Laireiter, A. R. (2017). DynAMo: A modular platform for monitoring process, outcome, and algorithm-based treatment planning in psychotherapy. *JMIR Medical Informatics, 5,* e20. http://dx.doi.org/10.2196/medinform.6808

Kamarck, T. W., Schwartz, J. E., Shiffman, S., Muldoon, M. F., Sutton-Tyrrell, K., & Janicki, D. L. (2005). Psychosocial stress and cardiovascular risk: What is the role of daily experience? *Journal of Personality, 73,* 1749–1774. http://dx.doi.org/10.1111/j.0022-3506.2005.00365.x

Kockler, T. D., Santangelo, P. S., & Ebner-Priemer, U. W. (2018). Investigating binge eating using ecological momentary assessment: The importance of an appropriate sampling frequency. *Nutrients, 10,* 10. http://dx.doi.org/10.3390/nu10010105

Kramer, I., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., . . . Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry, 13,* 68–77. http://dx.doi.org/10.1002/wps.20090

Kroeze, R., van der Veen, D. C., Servaas, M. N., Bastiaansen, J. A., Voshaar, R. C. O. V., Borsboom, D., . . . Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research, 3,* 1–11. http://dx.doi.org/10.17505/jpor.2017.01

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science, 21,* 984–991. http://dx.doi.org/10.1177/0956797610372634

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99,* 1042–1060. http://dx.doi.org/10.1037/a0020962

Lane, S. T., & Gates, K. M. (2017). Automated selection of robust individual-level structural equation models for time series data. *Structural Equation Modeling, 24,* 768–782. http://dx.doi.org/10.1080/10705511.2017.1309978

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: Guilford Press.

McCabe, K. O., & Fleeson, W. (2012). What is extraversion for? Integrating trait and motivational perspectives and identifying the purpose of extraversion. *Psychological Science, 23,* 1498–1505. http://dx.doi.org/10.1177/0956797612444904

Mejía, S., Hooker, K., Ram, N., Pham, T., & Metoyer, R. (2014). Capturing intraindividual variation and covariation constructs: Using multiple time-scales to assess construct reliability and construct stability. *Research in Human Development, 11,* 91–107. http://dx.doi.org/10.1080/15427609.2014.906728

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology, 13,* 23–47. http://dx.doi.org/10.1146/annurev-clinpsy-032816-044949

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18,* 112–117. http://dx.doi.org/10.1111/j.1467-8721.2009.01619.x

Moskowitz, D. S., Russell, J. J., Sadikaj, G., & Sutton, R. (2009). Measuring people intensively. *Canadian Psychology, 50,* 131–140. http://dx.doi.org/10.1037/a0016625

Moskowitz, D. S., & Zuroff, D. C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology, 86,* 880–893. http://dx.doi.org/10.1037/0022-3514.86.6.880

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine, 39,* 1533–1547. http://dx.doi.org/10.1017/S0033291708004947

Myin-Germeys, I., van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional reactivity to daily life stress in psychosis. *Archives of General Psychiatry, 58,* 1137–1144. http://dx.doi.org/10.1001/archpsyc.58.12.1137

Nock, M. K., & Prinstein, M. J. (2004). A functional approach to the assessment of self-mutilative behavior. *Journal of Consulting and Clinical Psychology, 72,* 885–890. http://dx.doi.org/10.1037/0022-006X.72.5.885

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., . . . Gotlib, I. H. (2014). Emotion-network density in major depressive disorder. *Clinical Psychological Science, 3,* 292–300. http://dx.doi.org/10.1177/2167702614540645

Piasecki, T. M., Jahng, S., Wood, P. K., Robertson, B. M., Epler, A. J., Cronk, N. J., . . . Sher, K. J. (2011). The subjective effects of alcohol–tobacco co-use: An ecological momentary assessment investigation. *Journal of Abnormal Psychology, 120,* 557–571. http://dx.doi.org/10.1037/a0023033

Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in Human Development, 14,* 253–270. http://dx.doi.org/10.1080/15427609.2017.1340052

Reis, H. T., Gable, S. L., & Maniaci, M. R. (2014). Methods for studying everyday experience in its natural context. In H. T. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 373–403). New York, NY: Cambridge University Press.

Roche, M. J., Pincus, A. L., Rebar, A. L., Conroy, D. E., & Ram, N. (2014). Enriching psychological assessment using a person-specific analysis of interpersonal processes in daily life. *Assessment, 21,* 515–528. http://dx.doi.org/10.1177/1073191114540320

Schiepek, G. K., Stöger-Schmidinger, B., Aichhorn, W., Schöller, H., & Aas, B. (2016). Systemic case formulation, individualized process monitoring, and state dynamics in a case of dissociative identity disorder. *Frontiers in Psychology, 7,* 1545. http://dx.doi.org/10.3389/fpsyg.2016.01545

Schuurman, N. K., & Hamaker, E. L. (2018). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000188

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n = 1 psychological autoregressive modeling. *Frontiers in Psychology, 6,* 1038. http://dx.doi.org/10.3389/fpsyg.2015.01038

Scott, S. B., Sliwinski, M. J., Zawadzki, M., Stawski, R. S., Kim, J., Marcusson-Clavertz, D., . . . Smyth, J. M. (2018). A coordinated analysis of variance in affect in daily life. *Assessment*. Advance online publication. http://dx.doi.org/10.1177/1073191118799460

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109,* 872–888. http://dx.doi.org/10.1037/pspp0000036

Shiffman, S., Paty, J. A., Gnys, M., Kassel, J. A., & Hickcox, M. (1996). First lapses to smoking: Within-subjects analysis of real-time reports. *Journal of Consulting and Clinical Psychology, 64,* 366–379. http://dx.doi.org/10.1037/0022-006X.64.2.366

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4,* 1–32. http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York, NY: Guilford Press.

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology, 75,* 629–638. http://dx.doi.org/10.1037/0022-006X.75.4.629

Snippe, E., Viechtbauer, W., Geschwind, N., Klippel, A., de Jonge, P., & Wichers, M. (2017). The impact of treatments for depression on the dynamic network structure of mental states: Two randomized controlled trials. *Scientific Reports, 7,* 46523. http://dx.doi.org/10.1038/srep46523

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain, 104,* 343–351. http://dx.doi.org/10.1016/S0304-3959(03)00040-X

Sullivan, H. S. (1954). *The psychiatric interview*. New York, NY: Norton.

Terluin, B., de Boer, M. R., & de Vet, H. C. W. (2016). Differences in connection strength between mental symptoms might be explained by differences in variance: Reanalysis of network data did not confirm staging. *PLoS ONE, 11,* e0155205. http://dx.doi.org/10.1371/journal.pone.0155205

Tomko, R. L., Solhan, M. B., Carpenter, R. W., Brown, W. C., Jahng, S., Wood, P. K., & Trull, T. J. (2014). Measuring impulsivity in daily life: The momentary impulsivity scale. *Psychological Assessment, 26,* 339–349. http://dx.doi.org/10.1037/a0035083

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9,* 151–176. http://dx.doi.org/10.1146/annurev-clinpsy-050212-185510

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology, 117,* 647–661. http://dx.doi.org/10.1037/a0012532

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., . . . Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 87–92. http://dx.doi.org/10.1073/pnas.1312114110

van der Krieke, L., Blaauw, F. J., Emerencia, A. C., Schenk, H. M., Slaets, J. P. J., Bos, E. H., . . . Jeronimus, B. F. (2017). Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic Medicine, 79,* 213–223. http://dx.doi.org/10.1097/PSY.0000000000000378

van der Krieke, L., Emerencia, A. C., Bos, E. H., Rosmalen, J. G., Riese, H., Aiello, M., . . . de Jonge, P. (2015). Ecological momentary assessments and automated time series analysis to promote tailored health care: A proof-of-principle study. *JMIR Research Protocols, 4,* e100. http://dx.doi.org/10.2196/resprot.4000

van Os, J., Delespaul, P., Wigman, J., Myin-Germeys, I., & Wichers, M. (2013). Beyond DSM and ICD: Introducing "precision diagnosis" for psychiatry using momentary assessment technology. *World Psychiatry, 12,* 113–117. http://dx.doi.org/10.1002/wps.20046

Van Roekel, E., Vrijen, C., Heininga, V. E., Masselink, M., Bos, E. H., & Oldehinkel, A. J. (2017). An exploratory randomized controlled trial of personalized lifestyle advice and tandem skydives as a means to reduce anhedonia. *Behavior Therapy, 48,* 76–96. http://dx.doi.org/10.1016/j.beth.2016.09.009

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science, 25,* 169–176. http://dx.doi.org/10.1177/0963721416643289

Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics, 85,* 114–116. http://dx.doi.org/10.1159/000441458

Wigman, J. T. W., van Os, J., Borsboom, D., Wardenaar, K. J., Epskamp, S., Klippel, A., . . . the MERGE. (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. http://dx.doi.org/10.1017/S0033291715000331

Wigman, J. T. W., 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. http://dx.doi.org/10.1371/journal.pone.0059559

Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology, 10,* 28. http://dx.doi.org/10.1186/1471-2288-10-28

Wilson, R. E., Thompson, R. J., & Vazire, S. (2017). Are fluctuations in personality states more than fluctuations in affect? *Journal of Research in Personality, 69,* 110–123. http://dx.doi.org/10.1016/j.jrp.2016.06.006

Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (in press). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*.

Wright, A. G. C., Hallquist, M. N., Stepp, S. D., Scott, L. N., Beeney, J. E., Lazarus, S. A., & Pilkonis, P. A. (2016). Modeling heterogeneity in momentary interpersonal and affective dynamic processes in Borderline Personality Disorder. *Assessment, 23,* 484–495. http://dx.doi.org/10.1177/1073191116653829

Wright, A. G. C., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology, 125,* 641–656. http://dx.doi.org/10.1037/abn0000169

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality, 29,* 250–271. http://dx.doi.org/10.1002/per.1986

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111–122. http://dx.doi.org/10.1037/0022-3514.43.1.111

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., . . . Wright, A. G. C. (in press). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a Personality Dynamics Diary. *Psychological Assessment*.

Received December 4, 2017
Revision received October 25, 2018
Accepted November 1, 2018 ■