



Published in final edited form as:

Assessment. 2016 August ; 23(4): 414–424. doi:10.1177/1073191116632341.

Ambulatory Assessment: New adventures in characterizing dynamic processes

Ryan W. Carpenter, Andrea M. Wycoff, and Timothy J. Trull

University of Missouri

Abstract

In recent years, significant technological advances have changed our understanding of dynamic processes in clinical psychology. A particularly important agent of change has been ambulatory assessment (AA). AA is the assessment of individuals in their daily lives, combining the twin benefits of increased ecological validity and minimized retrospective biases. These benefits make AA particularly well-suited to the assessment of dynamic processes, and recent advancements in technology are providing exciting new opportunities to understand these processes in new ways. In the current article, we briefly detail the capabilities currently offered by smartphones and mobile physiological devices, as well as some of the practical and ethical challenges of incorporating these new technologies into AA research. We then provide several examples of recent innovative applications of AA methodology in clinical research, assessment, and intervention and provide a case example of AA data generated from a study utilizing multiple mobile devices. In this way, we aim to provide a sense of direction for researchers planning AA studies of their own.

Clinical psychologists have long been interested in the assessment of psychological constructs that fluctuate and change over time, known as dynamic processes. Such processes span many areas of psychology, including affect, personality, and behavior. For example, mood is a process that ebbs and flows, often as a result of contextual or environmental factors. Recent technological advances have enabled researchers to examine dynamic processes in ways that previously would have been unthinkable. In many cases, we have found that what appeared to be a stable and unchanging trait, turns out to be otherwise when examined under greater resolution. In many cases, it seems that the more (we think) things stay the same, the more they change.

Perhaps most responsible for recent advancements in how we understand dynamic processes is *ambulatory assessment* (AA). AA is an umbrella term for an array of assessment methods, including self-report, observational, and physiological, biological, and behavior monitoring, used to study people in their natural environment. We prefer the term AA to commonly used alternatives (e.g., ecological momentary assessment [EMA], Stone & Shiffman 1994, experience sampling method [ESM], ecological momentary intervention [EMI]) because it captures the variety of methods, sampling, and data structures involved in the assessment of daily life (Fahrenberg & Myrtek, 2001; Wilhelm et al., 2012). AA differs from traditional

forms of assessment (e.g., self-report measures, diagnostic interviews) in several important ways. Most significantly, AA involves multiple assessments over a study period, in real- or near-real time, in participants' daily lives (Trull & Ebner-Priemer, 2013). As a result, AA is uniquely suited to assess within-individual processes, producing a rich clinical picture, allowing for hypothesis-testing within and between individuals (Curran & Bauer, 2011), and minimizing many heuristic biases of self-report (e.g., peak-end rule; Fredrickson & Kahneman, 1993; Shiffman et al., 1997).

Much has been written about AA methodology (e.g., Mehl & Conner, 2012; Shiffman, Stone, & Hufford, 2008; Stone, Shiffman, Atienza, & Nebeling, 2007) and a number of reviews summarize AA research in areas of clinical psychology, including anxiety disorders (Walz, Nauta, & aan Het Rot, 2014), borderline personality disorder (Nica & Links, 2009; Santangelo, Bohus, & Ebner-Priemer, 2014), mood disorders (aan Het Rot, Hogenelst, & Schoevers, 2012; Telford, McCarthy-Jones, Corcoran, & Rowse, 2012; Wenze & Miller, 2010), and substance use disorders (Morgenstern, Kuerbis, & Muench, 2014; Serre, Fatseas, Swendsen, & Auriacombe, 2015; Shiffman, 2009). Therefore, we do not attempt a comprehensive review of AA here.

Instead, we seek to build on previous work by describing “new adventures” in the assessment of dynamic processes using AA. We first describe some of the recent advancements in AA technology, as well as some of the practical and ethical challenges these advancements bring. We then focus on recent innovative applications of AA methodology in clinical research, assessment, and intervention. In this way, we provide a glimpse of the future of AA, especially in terms of dynamic processes. However, to a greater degree, we also take stock of where we are right now, at the threshold of that rapidly approaching future. Given the rapidly changing nature of the field, there is a certain risk to a review of this kind, as one who attempts to stand on shifting sands will likely quickly sink. Nevertheless, we aim to provide a sense of direction for adventures still to come.

New data collection methods

AA methods have advanced rapidly over the span of a few decades. For the interested reader, Wilhelm et al. (2012) provide an excellent, historical overview of daily life research. The relatively recent development and proliferation of smartphones have been a game-changer for AA. Almost two-thirds of adults in the US own smartphones, and there are approximately 2 billion smartphone users worldwide (Pew Internet Research Project, 2014). The smartphone not only can collect participant self-report, but also “passive” data, using its own built-in sensors and functions. For example, smartphones can collect audio and video, still images, physical activity, geolocation through Global Position System (GPS), the presence of light, phone calls, and text messages. Additionally, every operation initiated by a smartphone can be recorded. Many of these (e.g., when the phone is on/off, when the screen is activated, when an application is opened) may be useful for ensuring participant compliance or for triggering data collection. For example, a phone might be programmed to record audio each time it is exposed to light or to send a prompt when a participant is near a bar. The latter utilizes a technique termed geofencing and may be particularly useful for interventions using AA for conditions such as dementia (Vuong, Chan, & Lau, 2015). There

are, however, some limitations to context-based triggers. For example, GPS cannot determine the difference between a participant entering a bar for a drink or walking past it on the way to work. Indeed, despite the wealth of data available, it necessary to assess the likely validity and reliability of what is collected and to balance this against potential costs (e.g., battery usage). For example, while a smartphone is capable of recording physical activity, it will likely not do so accurately for participants who carry the phone in a purse or bag.

Using Bluetooth, the smartphone can also serve as a wireless hub that collects and transmits data from internal and external sensors (e.g., Alemdar & Ersoy, 2010). A wide range of AA devices and sensors have become available over the past few years, capable of continuously recording an entire suite of physiological data (e.g., physical activity, skin temperature, cardiac activity, respiration, electro-dermal activity). What could previously only be measured in the laboratory can now be captured in the real world. Research that incorporates such devices is still rare, and largely conducted in engineering and computer science fields (for examples, see Glenn & Monteith, 2014; Miller 2012). However, this is likely to change in the near future.

Challenges, obstacles, and pitfalls

Space precludes us from a full review of the steps and challenges involved in conducting AA research and many such resources exist already (e.g., Conner & Lehman, 2012). Therefore, we focus here on challenges that are directly related to mobile devices. Little of what we present is prescriptive, as the answers for how to address these concerns will depend on the individual study. Our goal, instead, is to draw attention to issues that we believe are easy to overlook. While troublesome, these challenges can be mitigated through careful attention and planning.

Device (dys)functionality.

The field of mobile devices remains new and untested and, therefore, considerable time should be spent on device and software selection. It is recommended to contact company representatives to discuss device specifications. An important issue is that many mobile devices have been developed for consumers, who typically require less precision from and put less wear on devices than researchers. There are many variables to weigh, including recording capability (i.e., reliability, validity), user-friendliness (e.g., comfort, size, ease-of-use), durability, connectivity (e.g., Bluetooth, mobile broadband, wireless), battery life, and cost. Trade-offs may need to be made across these variables. For example, a palmar device may provide a more precise recording of electro-dermal activity than a sensor worn on the wrist, but will also be much more cumbersome for participants.

When considering the software used to collect self-report data, another concern is that different companies offer differing degrees of flexibility and programming options, as well as different pricing structures. For example, Company A's software may only allow random prompts, while Company B's software is capable of both random and timed prompts, but is more expensive. Similarly, it is important to consider the ability to make changes to device software. Mobile devices often will not offer the level of customization researchers need.

With sufficient expertise, it may be possible to make modifications, but only if the developer is willing to provide the necessary code or API. Modifications, for example, will often be necessary to make devices communicate with one another or to sync with a smartphone hub. Fortunately, both the number of devices and companies offering AA programming are rapidly growing, providing an increasing number of choices (resource lists are available at www.ambulatory-assessment.org).

Once selected, devices and software must be tested. Pilot testing is necessary to ensure that devices work correctly. Often the more challenging part of testing is quality assurance (QA). Whereas pilot testing generally involves checking that study components perform as expected, QA is an active effort to find flaws and problems. Given the complexity of AA projects, it is easy for mistakes to occur and, for example, if the result is that a question is not displayed unless the preceding question is answered “yes,” the mistake may easily go undetected without repeated testing. Thus, QA requires testing the same program over and over, with special attention paid to conditional triggers or other decision points. This should be done by multiple individuals, who all understand how the software is intended to work and who will be able to recognize errors.

Once in the field, problems with hardware and software will arise. Devices will break, glitch, and perform unexpectedly. Complicating matters are actions by participants, such as turning off a device or failing to charge it, or putting on a sensor incorrectly. Frequently, the cause of a problem will be unclear: A broken device, one turned off, and one that has lost connectivity will all fail to transmit data. Participants may carry study devices unsupervised for long periods of time. Therefore, participants should be thoroughly trained and researchers should explain the consequences (e.g., data loss) of not following instructions. Detection of problems can be facilitated by automatic alerts that are triggered when a relevant condition is met or no data is received for a given time period, but it is also important to have research staff manually check incoming data periodically. Staff should also be available to troubleshoot problems and facilitate equipment replacement. Ultimately, equipment loss should be expected and, if possible, extra devices should be purchased. There is no guarantee that a device will continue to be available after a study has begun. While upgrading to a new or improved device may be tempting, using different physiological devices may problematize between-person comparisons.

Participant burden.

When developing an AA protocol, decisions must be made about the type of reports to be included (e.g., time-based, user-initiated, event triggered), the length of the monitoring, the number of assessments each day, and the number and content of items. With physiological devices, additional questions arise about sampling frequency and continuous versus discrete recording. These decisions depend upon the questions that the project is intended to answer (Ebner-Priemer & Sawitzki 2007; Palmier-Claus et al. 2011), and guidelines exist for aiding researchers in the journey from hypotheses to AA protocol (Conner and Lehman, 2012). However, these decisions must also be made in a way that they minimize *participant burden*, while still allowing researchers to collect sufficient data.

AA studies are inherently disruptive to participants' daily lives. Participants typically complete multiple assessments per day over extended time periods. Depending on the study, they are required to remain vigilant for random assessment alarms, to silence the study smartphone when unable to respond, and to initiate reports when events of interest occur. Participants often must also complete trainings on using study devices, answer pre- and post-monitoring trait-level questionnaires, and attend check-in sessions. If a study includes physiological devices, participants must learn to use these, remember to take them off and put them on as appropriate, and tolerate any discomfort (in our experience, many wearable devices are uncomfortable). They must keep the study devices charged and on their person, and they must put up with quizzical looks from friends and family. AA studies demand a great deal from participants and, the more researchers demand, the more likely participants are to drop out, burn out, or never enroll at all.

AA studies, however, are expensive and time intensive for researchers and often take a long time to complete. It is in researchers' best interests to gather as much data as possible. Thus, there is a need to find a balance between participant burden and the amount and quality of data collected. For example, to reduce burden, one should use as few self-report items as possible. However, to ensure good reliability, constructs generally should be assessed using at least three individual items (Shrout & Lane, 2012; although exceptions, for constructs such as craving, may exist [Sayette, in press]). Another example is the use of an alarm to signal random assessments. To reduce burden, an alarm should be kept relatively quiet, but this may lead to participants not hearing the alarm at all. A solution may be to have the AA device vibrate several seconds before making noise, giving participants a chance to answer it before it becomes a nuisance.

The good news is that, by being mindful during study design, it is possible to reduce burden without substantially reducing the amount of data collected. Ease of participation, which will decrease burden, can be increased by ensuring (e.g., through pilot testing and QA) that software is intuitive and works consistently. The aesthetics of user interfaces (e.g., readable font size, questions that appear on-screen without scrolling, intuitive buttons) are similarly important for ease-of-use. Items taken from global-level questionnaires do not always translate directly to the momentary-level, and items may require revision in order to read clearly. The overall length of assessments should also be kept in proportion with how often participants will be required to answer them. Payment, of course, is a valuable method of offsetting burden, but it should be noted that payment likely will only raise the threshold of what participants are willing to tolerate, rather than actually reduce burden. Researchers should critically examine their protocols for areas where burden can be reduced. There should be a clear purpose for every bit of data collected and aspects of the project without a strong justification should be removed.

Data management.

Another challenge arises from the collected data. Even in a relatively small study, the intensive sampling strategy involved in AA can result in thousands of data points. Given the computing power available today, this, in and of itself, would not pose much of a problem, if data at the time of collection (i.e., raw data) were immediately ready for analysis. However,

this is often not the case. Data management is an overlooked and underreported aspect of AA studies (McCabe, Mack, & Fleeson, 2012; Stone & Shiffman, 2002) and researchers new to AA methodology might be surprised at the amount of work awaiting them when they reach this stage of their study. Data management issues are generally not difficult to resolve, though they may be time-intensive, but they can easily slip under the radar of unsuspecting researchers.

Challenges with data management begin with the manner in which the raw data are saved by the AA device. We offer several examples of issues that have arisen in our own studies. First, if a project is contracted through a third-party company, the data structure that makes the most sense for the company may not be the optimal one for data analysis. It is therefore worthwhile, if possible, to request a sample data file, so that adjustments can be made in advance. Second, the raw data may include technical information (e.g., alarm triggers, battery level, schedule of random prompts, or activation of a suspension mode) important for verifying that the AA device is functioning properly. However, the amount of information included can make it difficult to tell where in the data problems exist and complicate data cleaning, described below. Third, gaps in time must be accounted for when data are lost or incomplete due to devices failures. These gaps can lead to data points appearing to take place one after another, when they did not. This must be accounted for when one wants to examine effects over time. However, these gaps likely will not be flagged or otherwise marked in the data and. Additionally, fixing these gaps and other anomalies requires determining how they were caused, which may not always be evident.

Another key part of data management is the creation of variables needed for data analysis. These can range from straightforward tasks like creating a sum score of scale items, to creating a time of day or day of week variable, to creating a variable that tracks the number of days a participant has been in the study. In some cases decisions must be made, such as deciding when a day is considered to “end.” Midnight may not be appropriate as participants may stay up later than that. Starting the day at 6 AM will create a similar problem with participants who wake up early. Considering a new day to start at “wake-up” is one solution, but requires there to be a consistent record in the data of when wake-up occurs. For example, participants could complete a morning report when they wake-up each day, which can be used to determine day in study. However, what happens when a participant does not complete a morning report? By settling on a definition of day before data collection starts, researchers can avoid being forced to later make compromises due to limitations of the collected data.

A third and final part of data management is data cleaning. Cleaning data requires attention to detail, but AA data are too vast to “eyeball” on the computer screen. This is particularly true of physiological data. There is also no preexisting program that will be able to clean data automatically. Thus, careful and creative use of data analytic software is required to check and double-check that variables add up to what they should, that reports do not happen when they should not, that your efforts to manage the data do not have unintended consequences, and so on. For sensor data, artifacts must be identified and removed and decisions must be made about what to do when data are missing. Although data cleaning is a process that largely occurs after data collection, anticipating potential issues in advance can

save considerable time. It can also be helpful to clean the data as it is collected, rather than waiting until the end of the study. The most important feature of data cleaning, however, is to remember that it is never finished. Therefore, syntax used for cleaning data should be saved, revisited, revised, and added to over time: Assume your data are never clean, only cleaner than they were before.

Multimodal AA data.

Recent developments in technology offer the potential to augment more traditional forms of AA data (e.g., self-report) with passive data captured by phone sensors and physiological devices. However, once collected, the question becomes how to integrate these different modalities of data in a productive manner.

To elucidate some of the challenges of multimodal AA data, we present a case example from a study underway in our own laboratory. In this study, patients diagnosed with a mood or anxiety disorder carry a smartphone for three weeks, reporting on their mood, alcohol craving, and alcohol use. The study design involves a mix of different types of assessments, such that participants are prompted randomly during the day and self-initiate a report if they experience a significant change in mood (mood dysregulation), significant alcohol craving, or finish an initial drink of alcohol. After participants report an initial drink, they receive follow-up prompts on a fixed schedule, in order to intensively sample the drinking episode and to facilitate the report of additional drinks. Participants in the study also wear two devices that record heart rate, breathing rate, electro-dermal activity, skin temperature, and physical activity, and the smartphone records GPS. Thus, participants provide an array of self-report and physiological data.

Figure 1 presents data from one participant from the study over the course of one drinking episode. Several challenges arise immediately upon considering how to analyze this data. First, there is much more physiological data than self-report. Decisions, therefore, must be made about whether to aggregate physiological data and over what time interval, without losing the ability to detect important changes. Decisions must also be made about what time periods are of interest. For example, in this study, participants were asked to report their mood over the past 15 minutes. Should the focus, then, be on the physiological data 15 minutes before the self-report? If lagged effects are expected, then it might make more sense to examine periods of time both before and after this 15 minute period. If so, how far back or forward in time should one go? As in any laboratory psychophysiological experiment, the answers to these questions depend on the phenomena under study. However, it is important to remember that there is less control and more “noise” in the real world than in the lab. Laboratory findings, if available, may be useful for guiding decisions, but thought should also be given to how long laboratory effects might be expected to last in the context of daily life. Ultimately, clear decisions should be made in advance of any analyses and “fishing expeditions” should be avoided.

The relative paucity of the self-report data presents its own challenge. This is particularly the case when the focus of interest is a particular event (e.g., alcohol consumption, mood shift) that may occur only a few times per day, if at all. If the goal of the study is to predict these events of interest using physiological indicators, this may prove difficult with rare events. It

may be important, therefore, to recruit participants who engage in the behaviors of interest a sufficient amount. For example, to be eligible in our study, participants had to report drinking on an average of two occasions per week and report recent emotion dysregulation. Inevitably, however, participants still spend the majority of time not drinking and not experiencing mood shifts. Given that many phenomena of interest to clinical psychologists have low base rates, it is important to select events of interest that are likely to generate a strong physiological signal.

A final challenge presented by Figure 1 is the multiple different physiological variables. It will not always be clear which variables are likely to be associated with constructs of interest. However, incorporating information across multiple variables may greatly improve prediction. As a simplistic example, an elevation in heart rate is more likely to indicate physical exercise if activity and skin temperature are also elevated than when they are depressed, a scenario which may be more likely to indicate alcohol consumption. Machine learning techniques can be used to develop algorithms that “learn from” and then make predictions using the wealth of collected self-report, activity, and physiological data (Flach, 2012; McClermon & Choudhury, 2013). For example, in supervised machine learning a “training” data set (e.g., one day of participant self-report and physiological data) can be used to develop algorithms that will be used to predict important events (e.g., alcohol consumption, craving, mood changes) in a “test” data set (e.g., another day) for that participant. Through an iterative process, algorithms can be optimized such that events can be predicted within and across participants. The importance here is that, once developed and refined, these algorithms may be able to accurately predict events in real- or near-real time, allowing, for example, for the possibility of intervention. This is, of course, only a brief overview of machine learning approaches; interested readers should consult textbooks (e.g., Flach, 2012) and recent applications (Hovsepian et al., 2015) of these methods.

Ethical Issues and Challenges

It is a simple fact that the more that we *can* do, the more carefully we must think about what we *should* do. As the capabilities of technology increase, so too will the ethical questions regarding appropriate use of that technology. We provide an overview here of some of the most commonly experienced ethical issues and challenges.

Informed Consent.

AA participants must be informed about the study protocol, the exact nature of the data to be collected, and potential study-related risks and burdens. Passive AA data can be gathered without any effort on the part of the participant, whether through smartphone sensors or mobile physiological devices. Therefore, it is vital to make the participant aware of all of the data that are being collected, as well as how these data might be used. This also applies to data captured on non-consenting individuals (e.g., via audio, video, photo). Investigators must decide in advance how this should be handled and whether participants should be encouraged to inform others with whom they have contact about the AA study. Additionally, some states may forbid the recording of third parties without their consent. Participants should always have the option to stop AA recording at any point and review their data.

Confidentiality and privacy.

Given that participants are reporting on their behaviors in daily life, it is important to discuss with them the limits of confidentiality (e.g., indications of danger to oneself or to others that will require action on the investigator's part) as well as how the investigators plan to ensure privacy in the AA study. In some cases, it may be advantageous to seek a *Certificate of Confidentiality* from the federal government to protect information regarding illegal activity (e.g., illegal substance use; criminality) from being subpoenaed by third parties, or to protect health information from third parties (e.g., insurance companies). Investigators must also consider the possibility that carrying study devices may identify an individual as a study participant. Being a participant in an AA study per se may not be a risk to privacy, but if the study is limited to those who are undergoing mental health treatment or to those with a certain psychiatric diagnosis, then this essentially reveals health information.

Security of data.

AA presents unique challenges in guaranteeing the security of data. For example, smartphones may be lost, misplaced, or stolen, and wireless transmission of data to servers can be intercepted. This is especially relevant as much of the data collected may be identifying (e.g., geolocation, audio and video recordings, text messages). Several precautions can be taken to maintain data security: (1) all devices used should be password-protected; (2) data on smartphones (whether from answers to surveys or collected from mobile devices) must be encrypted at the time of collection such that data from a lost smartphone cannot be read and identified; (3) in studies using wireless transmission of the data from a smartphone to a server, the data transmission itself must be encrypted; and (4) data should be stored locally on a secure server. In summary, password-protection of the protocol and the devices, as well as several layers of data encryption, are needed to ensure the security of collected AA data.

Assessment Issues.

Collecting data in the natural environment presents challenges for assessment. For example, there are events that are highly relevant for assessment purposes, but their occurrence may require action on the part of the researcher (e.g., nonsuicidal self-injury, suicidality). This can prove challenging if data are not collected in real-time or are only uploaded infrequently. One possible solution is to program immediate alerts that will be sent to designated research staff in real-time so that interventions or responses can be made in a timely manner. A similar concern exists for the collection of biological or physiological data. For example, what are the investigator's responsibilities in the case that adverse physiological events or health conditions are detected (e.g., heart arrhythmias)? These are issues that must be considered before commencing with the study, and protocols to handle these situations must be in place.

Treatment and Intervention Issues.

There are several relatively novel issues concerning the use of AA to trigger or administer interventions in daily life. With the widespread adoption of smartphones and tablets, there has been an explosion of applications purporting to address a litany of mental health issues.

Although there is great promise in this line of intervention (sometimes called ecological momentary interventions [EMIs], just in time interventions, or mHealth interventions), there is currently little empirical evidence for the efficacy of these interventions administered in daily life (e.g., see Kaplan & Stone, 2013). To date, there has been little oversight in the development of these daily life interventions from scientific or professional organizations. Furthermore, many are not clearly based on principles of behavior change, but rather are developed to look “face efficacious.” There is a need for the field of psychology to attend to this proliferation of apps that may serve to confuse and perhaps mislead consumers seeking mental health services. Recently, Schueller, Begale, Penedo, & Mohr (2014) described a framework, Purple, that supports the creation of empirically supported mobile mental health interventions. Such efforts will be crucial if we, as clinical psychologists, are to properly take advantage of the possibilities offered by technology for mental health treatment.

Examples of capturing dynamic processes with AA

Many of the constructs clinical psychologists are interested in vary across time and context and AA is uniquely suited to study such change. However, dynamic processes are complex, and in most cases, there exists no “gold standard” for how to best assess them in the moment. Often, there instead exist many different approaches to studying phenomena in daily life, which can yield different results. Our understanding of dynamic processes will change depending on the type of data collected (e.g., self-report, observational, physiological), the time scale or frequency at which data is collected or reports are made, the questions asked, and the analytical methods used. Time is a particularly important variable to consider, as results will change based upon the amount of time participants are asked to retrospect over (Ebner-Priemer & Sawitzki 2007; Stone, Broderick, Shiffman, & Schwartz, 2004). Dynamic constructs such as physical pain and mood, when measured over time, can provide a wealth of possible outcome measures beyond the traditional mean and standard deviation (e.g., Trull et al., 2008; Stone, Broderick, Schneider, and Schwartz, 2012). The level of analysis is also important, as effects at one level may be different from those at another for the same construct (Curran & Bauer, 2011). We focus here on several examples of recent work in the areas of mood and substance use, constructs frequently studied with AA. These examples represent efforts to utilize innovative methods and technologies to push forward our understanding of these dynamic processes.

Mood.

Mood is a dynamic process of particular interest in clinical psychology, as emotion dysregulation is associated with a large number of mental disorders (e.g., anxiety disorders, mood disorders, personality disorders; APA, 2013). AA, by having participants rate their mood many times over a study, offers the opportunity to define and explore emotion dysregulation in new ways. To cite a recent example from our own lab, Tomko et al. (2015) examined *undifferentiated negative affect* (the simultaneous experience of multiple different negative affects) and its relationship with impulsivity in patients with borderline personality disorder and major depressive disorder. Undifferentiated affect is a construct of increasing attention in AA studies (Lane & Trull, 2015). Previous work, however, has focused on undifferentiated affect as a between-person trait variable (e.g., Barrett, Gross, Christensen, &

Benvenuto, 2001). Tomko et al. (2015), by leveraging the fact that the data they analyzed comprised NA subscales each made up by multiple items, calculated intra-class correlation coefficients (ICC) at the occasion, day, and person level. They found that undifferentiated affect at the level of occasion and day, but not person, was positively associated with momentary impulsivity. Thus, there were specific relationships at the momentary and day level that would have been missed using more traditional trait measures. Their findings demonstrate the value of analyzing effects at multiple levels, including the level at which the data were collected.

AA methodology can also be used to test hypotheses derived from the laboratory. Ruscio et al. (2015) examined the role of rumination in the relationship between stressful life events and their consequents, including dysregulated mood, in individuals with major depressive disorder, generalized anxiety disorder, both, or neither. AA is well suited to test this, as information about stressful life events can be captured as they occur (rather than recalled in the laboratory), as can their effects on moods, behaviors, and cognitions. This makes it possible to test the temporal ordering of these variables and their associations. Ruscio et al. (2015) found that rumination was more common in the groups with current psychopathology and that ruminating predicted more symptoms, more negative affect (NA), less positive affect (PA), and more maladaptive behavior. Furthermore, rumination mediated the effect of stress on NA, PA, and symptoms. Thus, Ruscio et al. (2015) found support for the hypothesis that rumination plays a role in the effects of stress on mood. In this way, AA offers exciting potential for bridging the gap between the laboratory and the real world. This potential is only increasing as AA technology continues to advance.

Substance use and craving.

Substance use has long been a variable of interest in AA studies (Shiffman, 2009). With the advent of new technology, the field has begun to combine more traditional self-report methodology with physiological recording in daily life in an effort to better understand, and even predict, substance use. For example, Kennedy et al. (2015) collected heart rate data and self-reported substance use in 40 individuals currently in treatment for opioid dependence over 3 – 4 weeks. Analyses centered around the two hours before and after heroin ($N = 50$) and cocaine ($N = 85$) use reports and revealed that heart rate was elevated in the period both before and after cocaine use compared to heroin use. Heart rate for these events was also higher and lower, respectively, than heart rate 30 mins before and after random prompts. Additionally, Kennedy et al. (2015) found that heart rate was elevated 30 mins before and after random prompts when participants reported craving for either heroin or cocaine. These findings suggest that, by combining multiple types of AA data, it may be possible predict and, thereby, intervene in substance use immediately before it occurs.

Another technology that is presenting new opportunities in AA research is GPS, which can be used to compare the location of a participant to the locations of places of interest (e.g., liquor stores, bars, friends' homes). Epstein et al. (2014) utilized GPS data to examine the relationship of momentary mood, stress and substance craving with three observer-ratings of neighborhood quality (physical disorder, social disorder, and drug activity) in outpatients receiving methadone maintenance. Contrary to hypotheses, they found that craving, mood,

and stress were positively associated with neighborhood quality, such that lower quality neighborhoods were associated with less craving, negative mood, and stress. As Epstein et al. (2014) report, these results are counterintuitive, given that disadvantaged neighborhoods are typically associated with negative outcomes. It may be the case, then, that there are variables that moderate this association, such as familiarity or feelings of belongingness. Such moderators could be tested using AA. The observed findings also demonstrate how the results of within-person analyses can differ from hypotheses derived from between-person research.

Applications in clinical practice

As AA technology continues to develop, becoming easier to use, more flexible, and less expensive, it will become more and more feasible to include AA methodology routinely in clinical practice. This is important, as AA has the potential not only to improve assessment and intervention, but, perhaps, to bring about fundamental changes. That said, any such changes will occur slowly over time and obstacles exist that will need to be overcome before AA data becomes a major source of information for clinicians. Below, we briefly consider some of the implications that advancements in AA may have in the areas of assessment and intervention.

Assessment.

As mentioned above, one of the advantages of AA self-report over traditional forms of assessment (e.g., self-report, interview) is that it is able to minimize retrospective biases. For example, it can be difficult for a person suffering from depression to look back on the previous two weeks and decide during how many days they experienced depressed mood. However, if the same person reports their mood on each of those fourteen days, the clinician has an objective indicator of the frequency of depressed mood. This, of course, is not to say that a person's trait-level view of their mood does not have value, as global and momentary data provide different information (Conner & Barret, 2012). However, a key advantage offered by AA is that it allows for a more precise measurement of phenotypes.

One way in which AA provides this more precise measurement is by facilitating the direct examination of relevant contextual and environmental influences. A common challenge in clinical assessment is determining the influence of situational factors on symptoms and whether symptoms are present across contexts. This, in essence, is a version of the person-situation debate (Mischel & Shoda, 1995), which remains relevant to this day. AA, given its ability to measure environmental and contextual factors in daily-life, can be used to determine the relative effects of these factors. For example, by collecting information on where a person is, what they are doing, and who they are with, it is possible to detect systematic changes based on context (e.g., increase in negative mood when a romantic partner is nearby) or time (e.g., greater alcohol consumption on the weekend; Lane, Carpenter, Sher, & Trull, in press). When the goal is to better understand the person's contribution to a dynamic process, it can be desirable to remove or control for systematic effects of time or context (e.g., Piasecki et al., 2011; Trull et al., 2008)

Better measurement of phenotypes is critical for assessment if we are to succeed in recent initiatives to improve our classification systems. For example, the Research Domain Criteria (RDoC) project is intended to stimulate the development of new classification systems that better leverage existing research at different units of analysis (e.g., genes, cells, physiology, behavior, self-report; Insel et al., 2010). If the endeavor to link research across these units of analysis is to be successful, we need reliable and valid measurements for each unit. Thus, when we are interested in dynamic processes, we need to be able to capture these as precisely as possible. AA is a means of obtaining such precision and also offers the ability to obtain information from multiple units of analyses at the same time in a real-world context. Additionally, by assessing dynamic processes in the moment, AA has the potential to identify new phenotypes of interest by identifying patterns missed by global assessments.

Intervention.

One of the challenges of therapy has always been the divide between the patient's daily life and the therapist's office. AA offers therapists the ability to bridge this divide. An increasing number of reviews have addressed the topic of using AA for intervention (e.g., Adibi, 2015; Heron & Smyth, 2010; Kumar et al., 2013). AA offers enormous potential for intervention. For example, AA can be used for just-in-time intervention delivered as an event of interest or a relevant contextual event occurs. AA can also be used to deliver intervention materials in the moment (e.g., dialectical behavior therapy skills; Rizvi et al., 2011), increase self-monitoring and homework compliance, and can also inform in-person sessions.

This potential is further increased with the growing availability of mobile devices. For example, intervention can potentially occur not only when patients report treatment-relevant events, but also when physiological indicators and/or GPS location indicate that patients have experienced or are about to experience such events. Returning to Figure 1, a number of possible intervention points present themselves. First, there are points suggested by the self-report data (e.g., the Mood Dysregulation report, the increase in Alcohol Craving beginning at 21:00, or either drink report). Interestingly, this participant reported increases in sadness prior to the Mood Dysregulation report, suggesting a need for intervention prior to the participant deciding to initiate this report. With sufficient events of interest and comparison data, it would be possible to determine a profile of physiological activity (e.g., via machine learning; Flach, 2012; McClernon & Choudhury, 2013) that could be used to predict treatment relevant events.

Intervention could involve a prompt from the smartphone asking the patient to confirm that a target event is impending, an automated message with suggested coping skills, a call from the patient's therapist, or other possibilities. Intervention could also be tailored for the individual. For example, one patient may have the goal of abstinence, while another may only wish to reduce their drinking in response to negative mood. In the first case, intervention would occur prior to the first drink report while, in the second, intervention may not take place until the patient reports elevated sadness after the first drink. To date, such interventions are in their infancy, and despite some exciting initial studies combining physiological and self-report data (e.g. White, Umpfenbach, & Alpers, 2014), more research

is needed to demonstrate the feasibility and efficacy of implementing such interventions in patients' lives (Kaplan & Stone, 2013).

Conclusion

We have briefly reviewed recent advancements in AA technology and some of their potential applications in the in the field of clinical psychology. We also highlighted practical and ethical challenges associated with these advancements. Finally, we provided examples of the use of AA to better understand, assess, and treat dynamic processes. With the continuing development and availability of ever-more complex mobile devices, the field of AA is on the precipice of a period of change and growth that may well dwarf what has come before. These future adventures, guided by the research conducted today, promise to further expand our understanding of dynamic processes.

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. [PubMed: 22721999]
- Adibi S (Ed.). (2015). *Mobile Health: A Technology Road Map* (Vol. 5). Springer.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. (5th ed.). Arlington, VA: Author.
- Alemdar H, & Ersoy C (2010). Wireless sensor networks for healthcare: A survey. *Computer Networks*, 54, 2688–2710.
- Barrett LF, Gross J, Christensen TC, & Benvenuto M (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, 15, 713–724.
- Conner TS, & Barrett LF (2012). Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic medicine*, 74, 327. [PubMed: 22582330]
- Conner TS, Lehman BJ (2012). Getting started: launching a study in daily life In Mehl MR, Conner TS, editors. *Handbook of Research Methods for Studying Daily Life* (pp. 89–107). New York: Guilford.
- Curran PJ, & Bauer DJ (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, 62, 583.
- Ebner-Priemer UW, & Sawitzki G (2007). Ambulatory Assessment of Affective Instability in Borderline Personality Disorder. *European Journal of Psychological Assessment*, 23, 238–247.
- Epstein DH, Tyburski M, Craig IM, Phillips KA, Jobes ML, Vahabzadeh M, ... & Preston KL (2014). Real-time tracking of neighborhood surroundings and mood in urban drug misusers: application of a new method to study behavior in its geographical context. *Drug and alcohol dependence*, 134, 22–29. [PubMed: 24332365]
- Fahrenberg J, & Myrtek M, (2001) editors. *Progress in ambulatory assessment: computer-assisted psychological and psychophysiological methods in monitoring and field studies*. Seattle, WA: Hogrefe & Huber.
- Flach P (2012). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press.
- Fredrickson BL, & Kahneman D (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, 65, 45. [PubMed: 8355141]
- Glenn T, & Monteith S (2014). New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. *Current psychiatry reports* 16 1–10

- 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. [PubMed: 19646331]
- Hovsepian K, al' Absi M, Ertin E, Kamarck T, Nakajima M, & Kumar S (2015, September). cStress: towards a gold standard for continuous stress assessment in the mobile environment In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 493–504). ACM.
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, ... & Wang P (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167, 748–751. [PubMed: 20595427]
- 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.
- Kennedy AP, Epstein DH, Jobes ML, Agage D, Tyburski M, Phillips KA, ... & Preston KL (2015). Continuous in-the-field measurement of heart rate: Correlates of drug use, craving, stress, and mood in polydrug users. *Drug and alcohol dependence*, 151, 159–166. [PubMed: 25920802]
- Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, ... & Hedeker D (2013). Mobile health technology evaluation: the mHealth evidence workshop. *American journal of preventive medicine*, 45, 228–236. [PubMed: 23867031]
- Lane SP, & Trull TJ (2015). Operationalizing Undifferentiated Affect: A Generalizability Theory Approach. Manuscript submitted for publication.
- McCabe KO, Mack L, & Fleeson W (2011). A guide to data cleaning in experience-sampling studies In Mehl MR, Conner TS, editors. *Handbook of Research Methods for Studying Daily Life* (pp. 321–338). New York: Guilford.
- McCleron FJ, & Choudhury RR (2013). I am your smartphone, and I know you are about to smoke: the application of mobile sensing and computing approaches to smoking research and treatment. *Nicotine & tobacco research*, 15, 1651–1654. [PubMed: 23703731]
- Mehl MR.; Conner TS, editors. *Handbook of Research Methods for Studying Daily Life*. New York: Guilford; 2012.
- Miller G (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7, 221–237. [PubMed: 26168460]
- Mischel W, & Shoda Y (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*, 102, 246. [PubMed: 7740090]
- Nica EI, & Links PS (2009). Affective instability in borderline personality disorder: Experience sampling findings. *Current psychiatry reports*, 11, 74–81. [PubMed: 19187713]
- Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PAEG, Lewis SW, & Dunn G (2011). Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatrica Scandinavica*, 123, 12–20 [PubMed: 20712828]
- Pew Research Internet Project (2014). Mobile Technology Fact Sheet. Pew Research Center. <http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/>
- Piasecki TM, Jahng S, Wood PK, Robertson BM, Epler AJ, Cronk NJ, ... & Sher KJ (2011). The subjective effects of alcohol–tobacco co-use: An ecological momentary assessment investigation. *Journal of abnormal psychology*, 120, 557. [PubMed: 21443289]
- Ruscio AM, Gentes EL, Jones JD, Hallion LS, Coleman ES, & Swendsen J (2015). Rumination predicts heightened responding to stressful life events in major depressive disorder and generalized anxiety disorder. *Journal of abnormal psychology*, 124, 17. [PubMed: 25688429]
- Santangelo P, Bohus M, & Ebner-Priemer UW (2014). Ecological momentary assessment in borderline personality disorder: a review of recent findings and methodological challenges. *Journal of personality disorders*, 28, 555–576. [PubMed: 22984853]
- Sayette MA (in press). The role of craving in substance use disorders: Theoretical and methodological issues. *Annual Review of Clinical Psychology*.

- Schueller SM, Begale M, Penedo FJ, & Mohr DC (2014). Purple: a modular system for developing and deploying behavioral intervention technologies. *Journal of medical Internet research*, 16, e181. [PubMed: 25079298]
- Serre F, Fatseas M, Swendsen J, & Auriacombe M (2015). Ecological momentary assessment in the investigation of craving and substance use in daily life: A systematic review. *Drug and alcohol dependence*, 148, 1–20. [PubMed: 25637078]
- Shiffman S (2009). Ecological momentary assessment (EMA) in studies of substance use. *Psychological assessment*, 21, 486. [PubMed: 19947783]
- Shiffman S, Hufford M, Hickcox M, Paty JA, Gnys M, & Kassel JD (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of consulting and clinical psychology*, 65, 292. [PubMed: 9086693]
- Shiffman S, Stone AA, & Hufford MR (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol*, 4, 1–32. [PubMed: 18509902]
- Shrout PE, & Lane SP, (2012) Psychometrics In Mehl MR, Conner TS, editors. *Handbook of Research Methods for Studying Daily Life* (pp. 302–320). New York: Guilford.
- Stone AA, Broderick JE, Shiffman SS, & Schwartz JE (2004). Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between-and within-person consistency, and judged change in weekly pain. *Pain*, 107, 61–69. [PubMed: 14715390]
- Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 1994; 16:199–202.
- Stone AA, & Shiffman S (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24, 236–243. [PubMed: 12173681]
- Stone A, Shiffman S, Atienza A, & Nebeling L (2007). *The science of real-time data capture: Self-reports in health research*. Oxford University Press.
- Telford C, McCarthy-Jones S, Corcoran R, & Rowse G (2012). Experience sampling methodology studies of depression: the state of the art. *Psychol Med*, 42, 1119–1129. [PubMed: 22008511]
- Tomko RL, Lane SP, Pronove LM, Treloar HR, Brown WC, Solhan MB, ... & Trull TJ (2015). Undifferentiated negative affect and impulsivity in borderline personality and depressive disorders: A momentary perspective. *Journal of abnormal psychology*, 124, 740. [PubMed: 26147324]
- Trull TJ, & Ebner-Priemer U (2013). Ambulatory assessment. *Annual review of clinical psychology*, 9, 151.
- 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. *J. Abnorm. Psychol*, 117, 647. [PubMed: 18729616]
- Vuong NK, Chan S, & Lau CT (2015). *mHealth Sensors, Techniques, and Applications for Managing Wandering Behavior of People with Dementia: A Review In Mobile Health* (pp. 11–42). Springer International Publishing.
- Walz LC, Nauta MH, & Aan Het Rot M (2014). Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: A systematic review. *Journal of anxiety disorders*, 28, 925–937. [PubMed: 25445083]
- Wenze SJ, & Miller IW (2010). Use of ecological momentary assessment in mood disorders research. *Clinical psychology review*, 30, 794–804. [PubMed: 20619520]
- White AJ, Umpfenbach K, & Alpers GW (2014). Where have they gone? Tracking movement patterns to document the process of situational exposure in agoraphobia. *Professional Psychology: Research and Practice*, 45, 171.
- Wilhelm P, Perrez M, Pawlik K (2012) Conducting research in daily life: a historical overview In Mehl MR, Conner TS, editors. *Handbook of Research Methods for Studying Daily Life* (pp. 62–86). New York: Guilford.

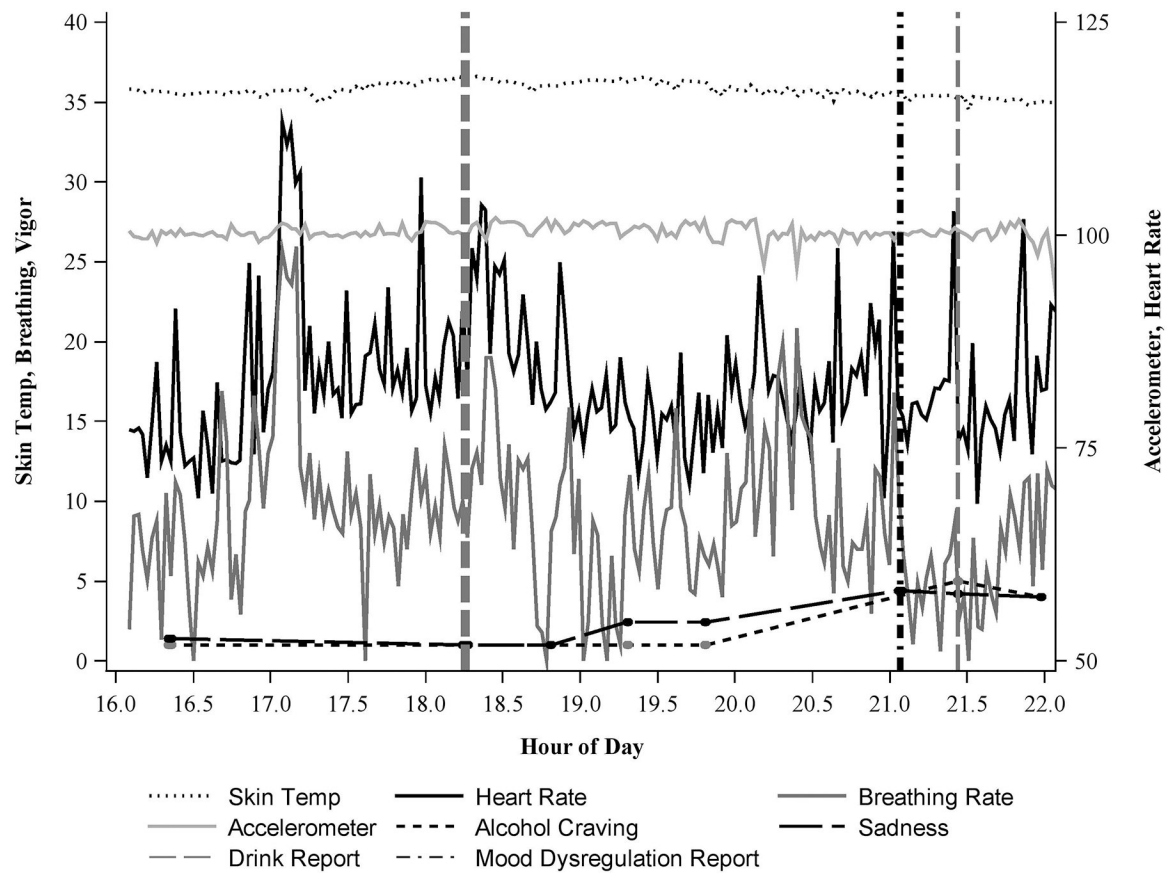


Figure 1. Example physiological and self-report data over the course of one participant's drinking occasion.