Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Review

Ambulatory assessment in psychopathology research: Current achievements and future ambitions

Merijn Mestdagh and Egon Dejonckheere

Abstract

Ambulatory assessment (AA) – a collection of methods that aim to track individuals in the realm of everyday life via repeated self-reports or passive mobile sensing – is well established in contemporary psychopathology research. Unravelling the dynamic signature of patients' symptoms and emotions over time and in their own personal ecology, AA methodology has improved our understanding of the real-time pathogenic processes that underlie mental ill-being. In this article, we evaluate the current strengths and shortcomings of AA in psychopathology research and spell out important ambitions for next-generation AA studies to consider. Regarding AA's current achievements, a selective review of recent AA studies underscores the ecological qualities of this method, its ability to bypass retrospective biases in self-report and the introduction of a within-person perspective. Regarding AA's future ambitions, we advocate for a stronger idiosyncratic focus, the incorporation of contextual information and more psychometric scrutiny.

Addresses

Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Leuven, 3000, Belgium

Corresponding author: Mestdagh, Merijn (merijn.mestdagh@kuleuven.be)

Current Opinion in Psychology 2021, **41**:1–8

This review comes from a themed issue on **Psychopathology**

Edited by **Peter de Jong** and **Yannick Boddez**

For a complete overview see the Issue and the Editorial

Available online 8 January 2021

https://doi.org/10.1016/j.copsyc.2021.01.004

2352-250X/© 2021 Elsevier Ltd. All rights reserved.

Keywords

Ambulatory assessment, Experience sampling methods, Ecological momentary assessment, Affect dynamics, Symptom networks, Mobile sensing.

Introduction

The traditional conceptualisation of emotions and psychopathological symptoms as monotone or stable entities that are either absent or present, and their simple combination reflecting a consistent clinical syndrome, has recently been replaced by a dynamical systems perspective, in which mental health researchers acknowledge the continuous, time-varying and context-embedded nature of emotion and symptom experiences [1,2,3,4*]. This theoretical shift has partly been enabled by the accrued adoption of *ambulatory assessment* (AA [5]) methods in psychopathology research – an umbrella term for different data collection techniques that entail methods such as *experience sampling* (ESM [6]), *ecological momentary assessment* (EMA [7]), *daily diary methods* [8] and *mobile sensing* [9].

While many AA protocols are unique in their design and the research questions they try to answer, their general premise in psychopathology research is to understand how patients' emotions, appraisals or thoughts, behaviour and psychological complaints naturally unfold in the complexity of everyday life [10]. In AA studies, the focal aim is to capture how these phenomena change within an individual, over time and across contexts. To this end, participants are repeatedly prompted in their daily lives to actively reflect on their momentary experiences via structured self-report surveys [11**]. Passively, participants' subjective evaluations may be complemented by more objective measures of psychophysiology (e.g. wearables that monitor heartrate, sleep cycles, physical activity, etc. [12]), and the detection of other relevant contextual parameters (e.g. smartphone sensors that track location, proximity, online interaction, etc. [9]), often in a continuous and unobtrusive way.

As the use of AA methods is becoming well-established in psychopathology research, this review aims to evaluate to what extent this data collection technique is currently able to live up to its promises. Specifically, we discuss six key features of AA methodology in light of the recent scientific literature and zoom in on both the strengths this method is able to deliver today and the current shortcomings that leave room for future improvement.

Current achievements of ambulatory assessment

Ecological validity

A first principle of AA is to provide researchers with a unique insight into the daily lives of their participants, drawing on the notion that people's behaviour and experiences only have meaning in the context in which

they take place [11**]. As such, AA methodology is able to provide a welcome and essential extension to more experimental research approaches.

Indeed, the extent with which clinical findings from standardised lab studies easily replicate into the realm of ordinary life is a commonly asserted limitation of experimental psychopathology research. For example, the notion that patients who suffer from clinical depression are characterised by blunted emotional responding is frequently supported by lab-based studies [13], yet in naturalistic settings, they typically show stronger emotional reactivity to daily events [14,15]. Relatedly, in the domain of clinical emotion regulation, lab findings typically point to a downregulation of negative emotion following strategies like cognitive reappraisal and distraction [16]. In daily life, however, a natural adoption of these strategies paradoxically predicts an increase in negative affect (even when controlling for initial levels of negative effect [17*]), hinting at a differential impact of experimentally instructed versus spontaneously initiated emotion regulation efforts.

In sum, the ecological qualities of clinical AA studies aptly complement the artificial settings of classic experiments. At the same time, however, the asset of collecting patient data in a real-world environment also implies increasing uncertainty about causal inferences (even when relying on cross-lagged analyses [18**]), which is the crux of standardised lab research. This inherent trade-off between external and internal validity may be partly resolved by quasi-experiments [19**], where the emotional impact of a real-life event is tracked across time (see also section 3.2 [20,21*]).

The reduction of retrospective biases

Second, the momentary and repeated assessment of symptoms and emotions in an AA protocol may effectively counter the impact of retrospective biases inherent to traditional survey methods [5,22]. Rather than patients' self-constructed snapshot of their complaints or the individual beliefs about their persona, AA research allows for a dynamic movie of (close to) real-time evaluations about the experiential phenomena under investigation.

Indeed, in conventional cross-sectional symptom surveys or clinical interviews, patients are typically instructed to retrospectively report on the intensity or frequency of various psychological complaints (e.g. *Over the past two weeks, how often did you feel nervous or on edge?*). Although these instruments are time-efficient and relatively effortless, they assume an accurate and impeccable retrieval of earlier information, free from memory distortions [22]. However, cognitive heuristics such as the peak-end rule [23] or mood-congruent recall [24] are known to undermine the reliability and validity of these retrospective evaluations, obfuscating the true nature of patients' symptom severity or emotion levels [25,26,27*].

In various research areas, recent work highlights the divergence between retrospective self-reflections and momentary AA findings. In the domain of personality, AA research shows that neuroticism, one of the major dimensions underlying individual differences in human personality and with strong ties to the development of affective disorder [28], is associated with the experience of more intense, but not more variable, negative emotions [29**,30], questioning the mainstream idea that this trait reflects emotional variability. Relatedly, in emotion research, dispositional reports of affect intensity and variability show little agreement with momentary measures of emotion in daily life, both in clinical groups [31] and community samples [32]. In the clinical domain, retrospective evaluations of the intensity of trauma-related symptoms show little overlap with the average severity ratings of these complaints during a one-week AA protocol, but rather match the worst day of that sampling week [33]. Similarly, suicidal ideation remains often undetected in retrospective clinical interviews, but is revealed when adopting AA methods [34].

In sum, the instant character of AA ratings fosters a more immaculate evaluation of psychological phenomena by compressing the time window between occurrence and recall. However, their repeated nature may also feature other sources of potential bias [35], such as annoyance with the protocol (leading to low compliance or drop-out [36**]), or reactivity to the measurement. Regarding this latter concern, systematic reviews in the context of psychopathological AA research are still lacking. While some studies find iatrogenic effects of repeatedly being asked to report on symptoms (e.g. alcohol abuse [37]), other research finds evidence for clinical improvement (e.g. depressive symptoms [38]). The vast majority of research, however, finds no evidence for symptom reactivity in clinical AA studies (e.g. pain complaints [39]; or suicidal ideation [40]).

Within-person focus

Finally, the repeated sampling of emotions and symptoms in an AA protocol paves the way for an intra-individual perspective on affect and psychopathology [41]. Contrary to popular cross-sectional paradigms, multiple short-term measurements per person enable researchers to track how these phenomena change over time [2], and how they interrelate within an individual [4*].

Regarding within-person change, AA research regularly reveals considerable temporal variability in previously

assumed static or trait-like psychological concepts. In basic affect science, for example, traditional models of human affect generally hypothesised a fixed and time-invariant relation between positive and negative affect [42], but recent AA work established meaningful within-person variation in this relation, driven by the experience of personally relevant events [21*]. Similarly, emotional granularity, the extent with which people differentiate between various same-valenced emotions, and are able to label their feelings with a high degree of specificity and precision, was typically thought of an important between-person dimension underlying mental well-being [17*]. Remarkably, however, AA research determined more variation at the within-person level, with people showing particularly poor emotion differentiation on days they experience high levels of stress [43*]. Finally, also in clinical research, AA studies highlight the low short-term stability of various symptom types (e.g. trauma complaints; [33], pain-related symptoms; [39], hallucinations and delusions; [44], etc.).

Regarding within-person interrelations, AA methods permit the examination of emotion and symptom configurations (contemporaneous or lagged) within an individual (often the average person in the sample, see also section 3.1 [1,45]). This is in marked contrast with traditional affect and mental health research, which predominantly relies on the analysis of between-person variation [21*,46]. Although between-person studies often make claims about within-person processes [47], Simpson's paradox illustrates that a generalisation between these levels of analysis is only warranted under very specific conditions [46,48**]. In emotion research, for example, there are notable differences in the between- versus within-person structure of affect, with personal and situational factors determining their degree of convergence [49]. Similarly, in the clinical domain, symptom constellations based on between- and within-person variation may markedly differ [50], revealing unique patterns and symptom associations for each level of analysis [51].

Future ambitions of ambulatory assessment

Idiosyncratic approach

Although AA introduces a within-person perspective to psychopathology research, this methodology is still primarily adopted to determine nomothetic patterns or regularities in symptom and emotion fluctuations (although exceptions exist [20,52,53,54**]). This common practice is remarkable, given that AA studies' *in vivo* approach is exceptionally well suited to uncover idiosyncratic associations between different symptoms and emotions and the situations in which they occur. While a discerning benefit of AA is to track patients in their own personal ecology [5,11**], researchers regularly collapse across these unique environments in their analyses to pursue general patterns of within-person association instead [18**,48**].

Statistically, this practice often translates into studying the fixed effect in a multilevel model, disregarding the person-specific heterogeneity in the modelled relation (i.e. the observed variability in random slopes [18**]). When AA researchers do link differences in within-person association to person-level variables, they generally communicate their results as a matter of degree (e.g. *Decreases in negative affect following a positive event are stronger for people who experience more depressive symptoms* [14]). However, when person-specific heterogeneity is substantial, within-person effects may not only vary in magnitude, but also in direction. In these cases, merely interpreting the average within-person effect (which may be net non-significant when different directionalities are distributed equally across the population) conceals the considerable between-person variability in the studied relation [55**]. For example, it has been argued that the ongoing debate about the effect of social media on adolescents' mental health (as a result of inconsistent findings) may be partly explained by an over-reliance on aggregated within-person effects [56*]: While a majority shows an increase in affective well-being following social media use, this relation was absent in an equally large group. A minority, however, felt worse.

In sum, the between-person aggregation of a within-person effect does not adequately represent its individual constituents [48**], and future AA studies could benefit from describing the individual variation around the fixed effect. A person-specific approach, in which AA researchers unravel more fine-grained differences at the level of a single subject [55**], could complement the dominant focus on nomothetic regularities in emotion or symptom time series. Nevertheless, researchers should bear in mind that an idiosyncratic model estimation generally requires longer time series, because person-specific estimations cannot borrow strength from similar cases in the sample distribution.

Incorporating context

Second, although AA researchers aim to understand how people live their lives in their own natural habitat, they actually know surprisingly little about the contexts in which their participants function [57**]. Unlike standardised experiments, where participants are deliberately exposed to specific situations or stimuli, a crucial aspect of AA is to interfere with people's daily routine as little as possible, in order to maximise the ecological nature of the behaviour, feelings or symptoms under study [35]. The downside of this non-interference, however, is that researchers are relatively

blind to the contextual factors that drive participants' feelings or symptoms.

As a result of this contextual black box, AA studies in psychopathology research primarily focus on the dimension of *time*. In the clinical emotion literature, for example, research on *emotion dynamics* [2] or *affective chronometry* [58] has generated a multitude of measures to capture temporal patterns of emotional change that are thought to play a role in psychopathology. However, the majority of these dynamical measures explain little variation in clinical disorder compared to simple differences in average levels of positive and negative affect [59**]. Among other reasons, it is argued that the context insensitivity of traditional AA protocols (and hence also the insensitivity of dynamical measures that are derived from these AA time series) may lie at the basis of the weak association between emotion dynamics and well-being [57**,60**]. Similarly, in the study of *symptomics* [1,3], researchers conceptualise mental disorders as a network of individual symptoms that unfold over time and interact with each other (e.g. *How does fatigue temporally predict increases in concentration difficulties, and how does this, in turn, predict low self-esteem, etc.?*). However, a common limitation of this network approach is that it generally understands clinical disorders as closed systems [1,61**], and hence disregards the role of exogenous input variables.

Even when AA studies do attempt to collect contextual information, a problem is that these assessments usually remain confined to self-report. Not only do these judgements often carry an affect-laden connotation (e.g. *Did something positive/negative happen since the last beep?* [14]), which inevitably produces a confound between stimulus and response, they also determine situational parameters in a very unrefined way (e.g. *Who are you with [Friends/Partner/Strangers/Alone]* [62]).

To pinpoint the role of context over a longer frame, a quasi-experimental AA approach [19**] in which symptoms [63] or emotions [17*,21*] are tracked around the occurrence of real-life meaningful events is known to increase the contextual signal in ecological time series [57**]. To examine the role of short-term contextual shifts on people's symptoms or emotions, a promising solution may be reserved for sensor-based wearables and mobile devices that passively track parameters such as location, movement and interaction [9,64]. The translation and integration of these rich, yet noisy data sources into meaningful situational signals could improve the contextual resolution of conventional AA and determine how situational factors shape people's mood or complaints.

Measurement

Finally, the repeated nature of AA introduces new psychometric challenges. Sampling people's emotions and symptoms over and over again requires unique methodological considerations for survey construction, protocol design and analysis plan.

Regarding survey construction, researchers generally limit the number of items at each measurement occasion to minimise participant burden [65]. This practice often results in single-item measures that may be valid at face value, but prevents their reliability estimation [66]. Moreover, even when AA surveys do rely on multi-item construct operationalisations, items are frequently selected, worded and combined in an *ad hoc* manner, with little theoretical guidance [67**]. For example, in emotion research, the item compositions of positive and negative affect differ remarkably across studies [59**], which endangers the replicability of AA results. A promising development regarding the internal robustness of AA findings is the leave-one-out multiverse analysis [17*,21*]. Here, researchers evaluate how their conclusions change on the basis of different construct operationalisations. Nevertheless, based on recent initiatives in the AA literature (e.g. ESM item repository [68]), we believe that the empirical validation of newly developed AA instruments is underway [69,70], introducing the use of more standardised questionnaires, comparable to cross-sectional research.

A second element regarding survey construction that deserves future attention in AA studies are scaling issues. In bounded measurements (e.g. Likert or continuous scales), floor or ceiling effects due to the low or high prevalence of particular symptoms or emotions (or an extreme phrasing of their respective items; e.g. *being tired* versus *hypersomnia*) limit the variability one can observe near the scale endpoints [71**]. As a consequence, bounded scales create a structural dependency between the mean and variability of a participant's symptom or emotion time series, introducing a confound between these two statistical moments for extremely low and high intensity ratings. In emotion research, for example, low intensity levels of negative affect in healthy populations may explain the relation between negative emotional variability and important dimensions of personality [29**,30] and psychological well-being [59**]. Similarly, when studying symptom networks, low intensity ratings for a specific complaint may misrepresent its relations with other symptoms, as a result of the limited variability that complaint can show [72]. While various measures have been proposed to

bypass this inherent association between mean and variability in bounded measurements [71**], it may be worthwhile for future AA research to explore the use of unbounded symptom or emotion assessments (e.g. forced-choice responding [73]).

Regarding protocol design, determining the appropriate time scale on which specific symptoms and emotions should be assessed will be a major challenge for future AA research. Tracking these phenomena at the wrong time scale runs the risk of missing their deterministic dynamics entirely [74]. As a rule of thumb, the time interval between consecutive prompts should be shorter than the rate of change of the emotions or symptoms under study, in order to adequately pick up on the serial dependency between discrete assessments [35]. Only if emotion or symptom evaluations are sufficiently autocorrelated, we can truly determine their dynamic properties (e.g. in the form of reactivity or recovery functions [57**]). By contrast, for the mean and the variance of an affective or symptom time series, appropriate time sampling is not essential [74], which could explain why these measures outperform more complex dynamics in the prediction of psychological well-being [59**]. Nevertheless, practical constraints prohibit the unlimited increase of the sampling frequency in traditional AA protocols. Endlessly expanding the temporal resolution of emotion or symptom assessments introduces additional participant burden, which may impact data quality [35]. Although unobtrusive and continuous mobile sensing could resolve this inevitable trade-off, relying on passive data sources to adequately assess experiential phenomena is not straightforward. A possible alternative could be the implementation of micro-level measurement burst designs, in which short-term follow-up evaluations are triggered when momentary symptom or emotion levels deviate from a previously specified setpoint.

Finally, regarding analysis plan, current AA studies often disregard the measurement error in momentary symptom or emotion ratings. Noisy assessments conceal the true signal in participants' affective or symptom time series, and neglecting measurement error is known to lead to biased dynamical estimates (e.g. an underestimation of the auto-regressive effects [75]). Decomposing within-person temporal variability into a deterministic (i.e. stimulus-related) and stochastic (i.e. noise) component will be imperative to draw valid conclusions in future AA studies [57**], but current protocols make it hard to directly disentangle their individual contribution. Including identical symptom or emotion items within the same momentary survey allows a direct evaluation of participants' response reliability, but is often dismissed over multiple single-item assessments for different psychological constructs.

Conclusion

Taken together, these outlined strengths and weaknesses describe AA as a field in motion. While AA has opened the door for answering novel and unprecedented research questions in psychopathology research, we believe the potential of this methodology is far from exhausted. As such, we hope that next-generation AA research embraces the outlined ambitions in this article.

Author contributions

MM and ED conceptualised, wrote and revised this article. Both authors approved the final version.

Funding sources

The preparation of this article was supported by the Research Fund of KU Leuven (C14/19/054). Merijn Mestdagh and Egon Dejonckheere are post-doctoral research fellows supported by the Fund for Scientific Research, Flanders (FWO; 1256221N, 1210621N). The funding sources had no involvement in the actual content of this contribution.

Conflict of interest statement

None declared.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Fried EI, Cramer AOJ: **Moving forward: challenges and directions for psychopathological network theory and methodology**. *Perspect Psychol Sci* 2017, **12**:999–1020.

2. Kuppens P, Verduyn P: **Emotion dynamics**. *Curr Opin Psychol* 2017, **17**:22–26.

3. Fried EI, Nesse RM: **Depression sum-scores don't add up: why analyzing specific depression symptoms is essential**. *BMC Med* 2015, **13**:72.

4. * Wigman JTW, de Vos S: **Time for a change: a more dynamic perspective on psychopathology**. In *Clinical Staging in Psychiatry: making diagnosis work for research and treatment*. Edited by Hickie IB, McGorry PD, Cambridge University Press; 2019: 34–66.
Conceptual overview chapter describing the dynamical systems perspective on psychopathology.

5. Trull TJ, Ebner-Priemer U: **Ambulatory assessment**. *Annu Rev Clin Psychol* 2013, **9**:151–176.

6. Csikszentmihalyi M, Larson R: **Validity and reliability of the experience-sampling method**. In *Flow and the foundations of positive psychology: the collected works of mihaly csikszentmihalyi*. Edited by Csikszentmihalyi M, Springer Netherlands; 2014: 35–54.

7. Stone AA, Shiffman S: **Ecological momentary assessment (EMA) in behavorial medicine**. *Ann Behav Med* 1994, **16**: 199–202.

8. Bolger N, Davis A, Rafaeli E: **Diary methods: capturing life as it is lived**. *Annu Rev Psychol* 2003, **54**:579–616.

9. Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT: **A survey of mobile phone sensing**. *IEEE Commun Mag* 2010, **48**:140–150.

10. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J: **Experience sampling research in psychopathology: opening the black box of daily life**. *Psychol Med* 2009, **39**: 1533–1547.

11. ** Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, Reininghaus U: **Experience sampling methodology in mental health research: new insights and technical developments**. *World Psychiatr* 2018, **17**:123–132.
Review paper discussing the advantages of AA in psychopathology research.

12. Blaauw FJ, Schenk HM, Jeronimus BF, van der Krieke L, de Jonge P, Aiello M, Emerencia AC: **Let's get Physiqual – an intuitive and generic method to combine sensor technology with ecological momentary assessments**. *J Biomed Inf* 2016, **63**:141–149.

13. Rottenberg J: **Emotions in depression: what do we really know?** *Annu Rev Clin Psychol* 2017, **13**:241–263.

14. Panaite V, Koval P, Dejonckheere E, Kuppens P: **Emotion regulation and mood brightening in daily life vary with depressive symptom levels**. *Cognit Emot* 2019, **33**:1291–1301.

15. Heininga VE, van Roekel E, Ahles JJ, Oldehinkel AJ, Mezulis AH: **Positive affective functioning in anhedonic individuals' daily life: anything but flat and blunted**. *J Affect Disord* 2017, **218**: 437–445.

16. Liu DY, Thompson RJ: **Selection and implementation of emotion regulation strategies in major depressive disorder: an integrative review**. *Clin Psychol Rev* 2017, **57**:183–194.

17. * Kalokerinos EK, Erbas Y, Ceulemans E, Kuppens P: **Differentiate to regulate: low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies**. *Psychol Sci* 2019, **30**: 863–879.
Example of a quasi-experimental study in which researchers investigate how people regulate their emotions in function of an impactful event (study 2).

18. ** Neubauer AB, Schmiedek F: **Studying within-person variation and within-person couplings in intensive longitudinal data: lessons learned and to be learned**. *Gerontology* 2020, **66**(4): 332–339.
Review paper on within-person associations in AA research.

19. ** Schmiedek F, Neubauer AB: **Experiments in the wild: introducing the within-person encouragement design**. *Multivariate Behav Res* 2020, **55**:256–276.
Methodological paper advocating for experimental manipulation in ecologically valid contexts to detect causal within-person effects.

20. Wichers M, Groot PC, Psychosystems EG: **Critical slowing down as a personalized early warning signal for depression**. *Psychother Psychosom* 2016, **85**:114–116.

21. * Dejonckheere E, Mestdagh M, Verdonck S, Lafit G, Ceulemans E, Bastian B, Kuppens P: **The relation between positive and negative affect becomes more negative in response to personally relevant events**. *Emotion* 2019 [Advanced online publication].
Illustration of a quasi-experimental approach in which researchers investigate how people's emotions change around a real-life contextual event.

22. Schneider S, Stone AA: **Ambulatory and diary methods can facilitate the measurement of Patient Reported Outcomes**. *Qual Life Res* 2016, **25**:497–506.

23. Stone AA, Broderick JE, Kaell AT, Delespaul PA, Porter LE: **Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics?** *J Pain* 2000, **1**:212–217.

24. Matt GE, Vázquez C, Campbell WK: **Mood-congruent recall of affectively toned stimuli: a meta-analytic review**. *Clin Psychol Rev* 1992, **12**:227–255.

25. Blum LH, Vakhrusheva J, Saperstein A, Khan S, Chang RW, Hansen MC, Zemon V, Kimhy D: **Depressed mood in individuals with schizophrenia: a comparison of retrospective and real-time measures**. *Psychiatr Res* 2015, **227**:318–323.

26. Campbell SB, Krenek M, Simpson TL: **The role of patient characteristics in the concordance of daily and retrospective reports of PTSD**. *Behav Ther* 2017, **48**:448–461.

27. * Nahleen S, Nixon RDV, Takarangi MKT: **Current PTSD symptomatology distorts memory for past symptoms**. *Psychiatr Res* 2019, **274**:330–334.
Empirical paper illustrating divergence between retrospective self-report and momentary AA findings in patients' symptom ratings.

28. Watson D, Naragon-Gainey K: **Personality, emotions, and the emotional disorders**. *Clin Psychol Sci* 2014, **2**:422–442.

29. ** Kalokerinos E, Murphy SC, Koval P, Bailen NH, Crombez G, Hollenstein T, Gleeson J, Thompson RJ, Van Ryckeghem DML, Kuppens P, Bastian B: **Neuroticism may not reflect emotional variability**. *Proc Nat Acad Sci* 2020, **117**:9270–9276.
Meta-analysis demonstrating the effect of bounded AA rating scales and how these produce a confound between mean levels and variability in negative affect.

30. Hisler GC, Krizan Z, DeHart T, Wright AGC: **Neuroticism as the intensity, reactivity, and variability in day-to-day affect**. *J Res Pers* 2020, **87**:103964.

31. Solhan MB, Trull TJ, Jahng S, Wood PK: **Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall**. *Psychol Assess* 2009, **21**: 425–436.

32. Sperry SH, Kwapil T: **Comparing static and dynamic measures of affect intensity and affective lability: do they measure the same thing?** *Motiv Emot* 2020, **44**:870–879.

33. Schuler K, Ruggero CJ, Mahaffey B, Gonzalez A, Callahan J L, Boals A, Waszczuk MA, Luft BJ, Kotov R: *When hindsight is not 20/20: ecological momentary assessment of PTSD symptoms versus retrospective report. Assessment*. 2019 [Advanced online publication].

34. Gratch I, Choo T-H, Galfalvy H, Keilp JG, Itzhaky L, Mann JJ, Oquendo MA, Stanley B: **Detecting suicidal thoughts: the power of ecological momentary assessment**. *Depress Anxiety* 2021, **38**:8–16.

35. Ram N, Brinberg M, Pincus AL, Conroy DE: **The questionable ecological validity of ecological momentary assessment: considerations for design and analysis**. *Res Hum Dev* 2017, **14**:253–270.

36. ** Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I: **Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations**. *J Med Internet Res* 2019, **21**: e14475.
Meta-analysis determining the study factors that underlie compliance and retention rates in AA protocols.

37. Buu A, Yang S, Li R, Zimmerman MA, Cunningham RM, Walton MA: **Examining measurement reactivity in daily diary data on substance use: results from a randomized experiment**. *Addict Behav* 2020, **102**:106198.

38. Kramer I, Simons CJ, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, Bemmel AL, Myin-Germeys I, Delespaul P, Os J, Wichers M: **A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial**. *World Psychiatr* 2014, **13**:68–77.

39. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P: **Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction**. *Pain* 2003, **104**:343–351.

40. Coppersmith DDL, Fortgang RG, Kleiman EM, Millner AJ, Yeager AL, Mair P, Nock MK: **Frequent assessment of suicidal thinking does not increase suicidal thinking: evidence from a high-resolution real-time monitoring study**. *PsyArXiv Preprint* 2020, https://doi.org/10.31234/osf.io/6bh82.

41. Hamaker EL: **Why researchers should think 'within-person': a paradigmatic rationale**. In *Handbook of research methods for studying daily life*. Edited by Mehl MR, Conner TS, The Guilford Press; 2012:43–61.

42. Russell JA, Carroll JM: **On the bipolarity of positive and negative affect**. *Psychol Bull* 1999, **125**:3–30.

43. * Erbas Y, Ceulemans E, Kalokerinos EK, Houben M, Koval P, Pe ML, Kuppens P: **Why I don't always know what I'm feeling: the role of stress in within-person fluctuations in emotion differentiation**. *J Pers Soc Psychol* 2018, **115**:179–191.
Empirical paper illustrating short-term temporal variability in previously hypothesised static psychological constructs.

44. Hermans K, van der Steen Y, Kasanova Z, van Winkel R, Reininghaus U, Lataster T, Bechdolf A, Gimpel-Drees J, Wagner M, Myin-Germeys I: **Temporal dynamics of suspiciousness and hallucinations in clinical high risk and first episode psychosis**. *Psychiatr Res* 2020, **290**:113039.

45. Nelson B, McGorry PD, Wichers M, Wigman JTW, Hartmann JA: **Moving from static to dynamic models of the onset of mental disorder: a review**. *JAMA Psychiat* 2017, **74**:528–534.

46. Voelkle MC, Brose A, Schmiedek F, Lindenberger U: **Toward a unified framework for the study of between-person and within-person structures: building a bridge between two research paradigms**. *Multivar Behav Res* 2014, **49**:193–213.

47. Bos EH, De Jonge P: **"Critical slowing down in depression" is a great idea that still needs empirical proof**. *Proc Natl Acad Sci* 2014, **111**:e878.

48. ** Fisher AJ, Medaglia JD, Jeronimus BF: **Lack of group-to-individual generalizability is a threat to human subjects research**. *Proc Natl Acad Sci* 2018, **115**:e6106–e6115.
Review paper that illustrates the lack of generalisability from group to individual statistical estimates and that calls for a stronger idiosyncratic focus in AA research.

49. Brose A, Voelkle MC, Lövdén M, Lindenberger U, Schmiedek F: **Differences in the between-person and within-person structures of affect are a matter of degree**. *Eur J Pers* 2015, **29**:55–71.

50. Hoorelbeke K, Van den Bergh N, Wichers M, Koster EHW: **Between vulnerability and resilience: a network analysis of fluctuations in cognitive risk and protective factors following remission from depression**. *Behav Res Ther* 2019, **116**:1–9.

51. McNally RJ: **Can network analysis transform psychopathology?** *Behav Res Ther* 2016, **86**:95–104.

52. Fisher AJ: **Toward a dynamic model of psychological assessment: Implications for personalized care**. *J Consult Clin Psychol* 2015, **83**:825–836.

53. Fisher AJ, Boswell JF: **Enhancing the personalization of psychotherapy with dynamic assessment and modeling**. *Assessment* 2016, **23**:496–506.

54. ** Wright AGC, Zimmermann J: **Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement**. *Psychol Assess* 2019, **31**:1467–1480.
Perspective paper that illustrates the value of applying personalised modelling with AA data in clinical practice.

55. ** Howard MC, Hoffman ME: **Variable-centered, person-centered, and person-specific approaches: where theory meets the method**. *Organ Res Methods* 2018, **21**:846–876.
Overview paper of nomothetic and idiosyncratic approaches in AA research, discussing their strengths and weaknesses.

56. * Beyens I, Pouwels JL, van Driel II , Keijsers L, Valkenburg PM: **The effect of social media on well-being differs from adolescent to adolescent**. *Sci Rep* 2020, **10**:10763.
Empirical paper that illustrates the added value of investigating person-specific relations rather than merely interpreting the average within-person effect.

57. ** Dejonckheere E, Mestdagh M, Kuppens P, Tuerlinckx F: **Reply to: context matters for affective chronometry**. *Nat Hum Behav* 2020:1–4.
Commentary that discusses the importance of maximising the signal-to-noise ratio in affective time series via strong contextual stimuli, adequate time scales and good construct operationalisations.

58. Hemenover SH: **Individual differences in rate of affect change: studies in affective chronometry**. *J Pers Soc Psychol* 2003, **85**:121–131.

59. ** Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, Tuerlinckx F: **Complex affect dynamic measures add limited information to the prediction of psychological well-being**. *Nat Hum Bevah* 2019, **3**:478–491.
Meta-analysis that shows many complex affect dynamic measures are confounded with average levels of positive and negative affect, questioning their added value in the prediction of mental health. In the discussion, methodological guidelines are provided for future AA studies to unravel unique relations between emotion dynamics and psychological well-being.

60. ** Lapate RC, Heller AS: **Context matters for affective chronometry**. *Nat Hum Behav* 2020:1–2.
Commentary that argues unique associations between emotion dynamics and psychological well-being may reveal when investigating affective responding to strong contextual stimuli (in the form of recovery functions).

61. ** Bringmann LF, Elmer T, Epskamp S, Krause R, Schoch D, Wichers M, Wigman JTW, Snippe E: **What do centrality measures measure in psychological networks?** *J Abnorm Psychol* 2019, **128**:892–903.
Overview paper that explicitly discusses the assumptions underlying the use of centrality measures in symptom networks and shows that many of these assumptions are probably violated in the psychological network literature (e.g. the assumption that clinical syndromes are closed systems).

62. Heininga VE, Dejonckheere E, Houben M, Obbels J, Sienaert P, Leroy B, van Roy J, Kuppens P: **The dynamical signature of anhedonia in major depressive disorder: positive emotion dynamics, reactivity, and recovery**. *BMC Psychiatr* 2019, **19**:59.

63. Fried EI, Bockting C, Arjadi R, Borsboom D, Amshoff M, Cramer AOJ, Epskamp S, Tuerlinckx F, Carr D, Stroebe M: **From loss to loneliness: the relationship between bereavement and depressive symptoms**. *J Abnorm Psychol* 2015, **124**:256–265.

64. Bachmann A, Zetzsche R, Riedel T, Beigl M, Reichert M, Santangelo P, Ebner-Priemer U: **Identification of relevant sensor sources for context-aware ESM apps in ambulatory assessment**. In *Adjunct proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 ACM international symposium on wearable computers*. Association for Computing Machinery; 2015:265–268.

65. van Berkel N, Ferreira D, Kostakos V: **The experience sampling method on mobile devices**. *ACM Comput Surv* 2017, **50**:93.

66. Wilhelm P, Schoebi D: **Assessing mood in daily life: structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood**. *Eur J Psychol Assess* 2007, **23**:258–267.

67. ** Flake JK, Fried EI: **Measurement schmeasurement: questionable measurement practices and how to avoid them**. *Adv Methods Prac Psychol Sci* 2020, **3**(4):456–465.
Conceptual paper that discusses questionable measurement practices and validity issues in the operationalisation of psychological constructs.

68. Kirtley OJ, Hiekkaranta AP, Kunkels YK, Verhoeven D, van Nierop MM, Myin-Germeys I: *The experience sampling method (ESM) item repository*. 2019. retrieved from, www.osf.io/kg376.

69. Medland H, De France K, Hollenstein T, Mussoff D, Koval P: **Regulating emotion systems in everyday life**. *Eur J Psychol Assess* 2020, **36**:437–446.

70. McMahon TP, Naragon-Gainey K: **Ecological validity of trait emotion regulation strategy measures**. *Psychol Assess* 2020, **32**(8):796–802.

71. ** Mestdagh M, Pe M, Pestman W, Verdonck S, Kuppens P, Tuerlinckx F: **Sidelining the mean: the relative variability index as a generic mean-corrected variability measure for bounded variables**. *Psychol Methods* 2018, **23**:690–707.
Methodological paper that introduces an alternative metric, the relative variability index, to bypass the problem of mean and variance confound in affect ratings as a result of bounded AA scales.

72. Bulteel K, Tuerlinckx F, Brose A, Ceulemans E: **Using raw VAR regression coefficients to build networks can be misleading**. *Multivar Behav Res* 2016, **51**:330–344.

73. Brown A, Maydeu-Olivares A: **Item response modeling of forced-choice questionnaires**. *Educ Psychol Meas* 2011, **71**: 460–502.

74. Boker SM, Molenaar PCM, Nesselroade JR: **Issues in intra-individual variability: individual differences in equilibria and dynamics over multiple time scales**. *Psychol Aging* 2009, **24**: 858–862.

75. Schuurman NK, Houtveen JH, Hamaker EL: **Incorporating measurement error in N = 1 psychological autoregressive modeling**. *Front Psychol* 2015, **6**:1038.