The power of individual landscapes

The power of individual landscapes

A clinical exploration of personal experience sampling

and new horizons

Simone J.W. Verhagen

The power of individual landscapes

A clinical exploration of personal experience sampling

and new horizons

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,

op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert,

volgens het besluit van het College van Decanen,

in het openbaar te verdedigen

op vrijdag 10 januari om 10.00 uur

door

Simone Josephine Wilhelmina Verhagen

geboren op 13 september 1991 te Heerlen

Promotoren
Prof. dr. P.A.E.G. Delespaul
Prof. dr. J.J. van Os

Copromotor
Dr. C.J.P. Simons

Beoordelingscommissie
Prof. dr. T.A.M.J. van Amelsvoort (voorzitter)
Prof. dr. N.E. Jacobs (Open Universiteit Heerlen)
Prof. dr. D.E.J. Linden
Prof. dr. P. Kuppens (Katholieke Universiteit Leuven)
Prof. dr. F. Scheepers (Universitair Medisch Centrum Utrecht)

Paranimfen

Suzanne van Bronswijk

Naomi Daniëls

Contents

1

General introduction

Clinical background

Our mental health care system is going through a transition[1]. For years, researchers have invested in the development of extensive classification systems to gain grip on mental suffering. However, frustrations in clinical practice remain[2]. Clinicians and patients alike often find that a categorical diagnostic explanation has marginal relevance for the alleviation of mental suffering[3]. In itself, this is no problem because the ultimate goal is to feel better no matter the cause. Selecting optimal interventions remains relevant, but our knowledge base for treatment choices are organized on those classification systems. It is important to try to bring assessments closer to clinical relevance[2,3].

Knowing why suffering occurs can help to make a treatment choice and provide clues for an adequate solution. In the past years, researchers have spent a tremendous amount of time and resources to find out what causes depression, anxiety or schizophrenia. Although several causes have been identified, no satisfactory explanation to date exists[4,5]. Without a clear etiology, a single treatment target is unrealistic and not enough. The development and revisions of classification systems such as the Diagnostic and Statistical Manual of Mental Disorders is motivated by the assumption that a better delineation will improve etiological research[6]. However, multiple categorical diagnoses with precise definitions and complex decision rules have led to a frustrating reference tool with only limited relevance for research[4,7].

The system was originally meant as a guideline for clinicians to inform treatment choices[8]. One of the problems is that the focus of a categorical system lies on susceptibility for mental health complaints or in other words, vulnerabilities[9]. According to the etiological explanation of the stress-vulnerability model, early life factors such as genetics, stress, and childhood experiences lead to the development of vulnerabilities[10,11]. These vulnerabilities are part of humanity, and human variety is as broad and diverse as all the terms in a mental health spectrum. The majority of people will fit at least one criterion and most do not fit perfectly within a single classification[5]. Knowing that multiple factors cause mental suffering has not led to better health care by definition. Vulnerabilities represent stable components in a specific period of someone's life, but their expression can wax and wane in reaction to the environment[11,12]. Someone can have a (stable) vulnerability for psychosis, but psychotic symptoms can increase after experiencing a stressful event. The deviation from the norm during residual phases is the vulnerability[13]. It is this deviation or variability in vulnerability expressions that can create possibilities for interventions, not the vulnerability itself.

In order to provide appropriate evidence-based treatment and distribute finances, classification is a prerequisite for care in the Netherlands[3]. However, classification is not related to needs and therefore only marginally relevant for prioritizing resources, a concern in a fast growing sector. This results in long waiting lists and specialized care programs that select outpatients but disregard comorbidity, an important factor in severity. The recognition of an inefficient health care system has led to the realization that a transition is needed[3,5]. The time is right to try a different approach, one that shifts the focus from a vulnerability perspective to a health perspective[14]. With the goal to explore what is going on, instead of answering the why question.

Positive health

What is health and what is illness? Historically, health was defined as the lack of illness[15]. In 1948, the World Health Organization (WHO) defined health as "*a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity*"[16]. However, many people feel healthy, while still being treated for health complaints. Curing the disease proved too stringent a criterion and this definition still receives criticism[17]. The WHO defined mental health as "*a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community*"[18]. According to this definition, mental health or well-being is a dynamic concept with a societal impact[19].

Recently, Machteld Huber introduced positive health as "*the ability to adapt and to self-manage, in the face of social, physical and emotional challenges*"[17]. This definition better matches the way we actually think about mental health care. A more holistic view is taken, with a central role for the individual. People suffer from symptoms, but are not defined by those symptoms. The focus shifts to aspects of well-being and resilience[17]. What is the strength of a person? When does someone feel better and when is someone able to participate in meaningful activities and interactions? Positive health warrants a different approach in clinical practice. Often, treatment sessions start with vulnerabilities and assessment instruments focus on defining and monitoring psychological complaints. The new definition invites us to focus on the positive experiences first. The same holds for assessment instruments; they need to have a holistic approach.

The new definition of health changed the role of patients during treatment. Patients are required to play an active role in their own recovery[5]. Mental health treatment often requires a continuous effort from patients; an enduring behavioral change takes time. In these long trajectories, empowerment is especially crucial. Empowerment is the process in which

individuals gain influence on decisions and actions that are related to their personal health[20]. As a psychological concept, it describes the sense of control over one's own health and the feeling of being responsible[21]. In order to be empowered, the role of a health care professional needs to change as well[3,22]. Instead of providing complicated and often mystifying answers from an expert perspective, the professional needs to listen and ask questions. Answers are created together, by thinking along together in a collaborative effort. This is important because it helps to motivate patients to take control of their own situation. Being motivated and feeling empowered will result in enduring health changes.

Recovery

The mental health recovery movement has a long tradition and current views parallel the new health perspective[1,14]. Clinical recovery that is reached by decreasing symptomatic complaints, has only limited impact on well-being and life satisfaction[14]. Other aspects such as societal recovery and personal recovery are equally important. Societal recovery is reached when someone is able to function in the community and personal recovery when someone experiences his or her life as meaningful[14]. The Institute for Positive Health has tried to quantify the different health domains involved[23]. The domains related to recovery include bodily functions (e.g., energy level and physical complaints), mental well-being (e.g., emotional state and cognitive functioning), meaningfulness (e.g., acceptance and future prospects), quality of life (e.g., happiness and balance), social-societal participation (e.g., work and communication skills), and daily functioning (e.g., ability to work and participate). In clinical practice, all domains and aspects are relevant[3].

Recovery is a very personal process; everyone has his or her own tempo and phases[24,25]. Care needs to be tailored to these personal recovery phases and both the client and caregivers need to be aware of this[14]. Recovery as a comprehensive process is reached through different strategies and is not limited to professional care[5,14]. Receiving advice from neighbors or friends, following a restful yoga class, or finding a meaningful job can all improve resilience and life satisfaction.

Daily life functioning

Functioning in daily life is an important outcome measure for recovery. Irrational thoughts about social situations can easily be addressed in the therapy office, but are more challenging in everyday life. Vulnerabilities as well as resilient moments occur in context and this context is missing in the office. Optimal care should therefore have a 24/7 impact. Well-dosed medication has this effect, but interventions such as psychotherapy rely on the generalization of skills. This is often difficult. Detecting automatic depressive thoughts or

relaxing muscle tension for example, are not self-evident when navigating in a busy environment or doing household shores. To increase skill generalization, we need to create a daily life impact and find innovative ways to strengthen the current care system.

eHealth or mHealth

One such innovation is electronic Health or eHealth: a broad term that encompasses the use of internet-related technologies to deliver or enhance health services. A common factor in varying definitions of eHealth is that it increases care efficiency due to improved logistics provided by technology[26]. A computerized blood pressure monitor that can be taken home to collect data and provide automated feedback to the user is an example of eHealth technology. No special skills are required for its use. More specific, but related is mobile Health or mHealth: the use of wireless information and communication technology such as smartphones, tablets, or smartwatches to improve health, and to support and improve the health care system[27]. The crucial element here is that the technology is available to people in different situations. Mobile health can be used in the flow of daily living. It helps to monitor, to remind, or to change health care behavior in different contexts or environments. Advantages are the independency from professionals, fostering active engagement and autonomy and creating empowerment. Furthermore, technology use can be a cost-effective way of distributing resources in mental health care[28].

The above-mentioned benefits have led different stakeholders to invest in health technology, and the government and funding agencies to encourage the development of mHealth tools[28,29]. As a result, the number of mHealth applications has risen exponentially in the last years. However, many tools are developed without a sound scientific background and, while improving care logistics, no fundamental changes are made to care provision[28]. MHealth tools for mental health care need to provide accurate insight into daily life functioning, mood, and cognition in a holistic manner that is relevant to the individual's recovery process[30].

Experience sampling method

The Experience Sampling method (ESM) – or Ecological Momentary Assessment is an empirically validated diary method for repeated assessments in daily life[31-33]. Numerous research studies in the mental health field have used the technology to look into mental states and symptom dynamics, and examine psychopathology at the micro-level. This offered valuable insight on how the expression of vulnerability evolves in the daily context[34-36]. Previously, the logistics of the method were cumbersome, but technology improved and the assessment strategy is now available on a smartphone.

ESM can be used as an mHealth tool in clinical practice[37,38]. Short, repeated assessments on an app allow the collection of information on current mood, behavior, and cognition in different situations. Questions that relate to the current activity, location, and company complement the assessments and contextualize mental state reports. By doing so, information is gathered across relevant recovery domains[37]. When sufficient data is available, the mental state reports and context information can be combined to provide insight in someone's own functioning. This can be done using an online feedback module, as is possible with the PsyMate™ application. In clinical practice, a patient can start to monitor their own functioning and decide to share the results with a clinician. Seeing how mood fluctuates in the personal situation can form the starting point of a treatment session and facilitating shared-decision making[9].

After a while, ESM on a smartphone can function as a therapist in the pocket, increasing autonomy. In order to do so, technology needs to be intuitive, embedded in the flow of daily life with limited burden, and easy to use. Patients and clinicians alike need to be able to understand their data and learn from it[39]. Now, graphs and figures are used to show dynamic patterns between mood states. These data visualizations are not always intuitive and often require some explanation. Although very promising, steps need to be taken in order to make ESM as an mHealth tool widely available and useful for a broad public. An assessment tool is needed that takes into account both vulnerabilities and strengths, and provides an answer to what is going on. In a transdiagnostic world of mental health that is embedded in good clinical practice.

Aims and outline of this thesis

The work in this thesis explores the use of ESM in clinical practice and broadening its scope within mental health care. Attention is payed to the power of individual landscapes: the picture that unfolds after a period of data collection and the lessons we can learn from this picture. Along with the journey of patients exploring their landscapes and engaging in the adventure that can lead to empowerment and recovery. The thesis consists of three parts.

Part I illustrates the usefulness of ESM in clinical practice, approached from different perspectives. **Chapter 2** provides an overview of the use of the experience sampling method in the context of clinical trials. Advantages are highlighted and disadvantages discussed. Guidelines and recommendations are presented for setting up an ESM protocol, analyzing its data, and implementing it in clinical trials. Furthermore, its promise to improve clinical decision-making is considered. In **chapter 3**, data examples are used to illustrate the advantages that ESM provides for routine clinical practice. This chapter gives a

comprehensive overview on applications of the methodology to strengthen individual health care. **Chapter 4** argues that ESM, as a generic transdiagnostic tool, is a reliable instrument for meaningful routine outcome monitoring. ESM is implemented in a general hospital setting and it as psychometric properties are examined. **Chapter 5** illustrates the clinical relevance of self-monitoring during recovery from a first-person perspective. Monitoring for more than two years is possible and yields valuable lessons, showing its strength in providing meaning during darker times in life.

Part II builds on clinical relevance by leveraging the methodological power of ESM data. Providing feedback on an individual basis is very relevant but risky. The feedback needs to reflect the personal situation in a correct way. Now, insufficient knowledge exists to understand when enough data is gathered for a stable interpretation of the results. In **chapter 6**, a first attempt is made to answer this question. In this proof of concept study, a momentary reward-related quality of life statistic is operationalized and tested with Monte Carlo experiments. ESM data from patients suffering from serious mental illness served as seeds for the data simulations.

Part III extents ESM with objective cognitive measures in daily life. **Chapter 7** is a first pilot study in patients suffering from bipolar disorder. A neurocognitive task and a social cognitive task are tested within an ESM paradigm. Initial feasibility is examined and the relation between momentary cognition outcomes and other relevant ESM measures is explored. In **Chapter 8**, the momentary digit-symbol substitution task is translated to a mobile ESM application and tested in the general population. Next to initial feasibility, attention is payed to the assessment of cognitive patterns in daily life. **Chapter 9** is an extension of the research in chapter 8, complementing the digit-symbol substitution task with a visuospatial working memory task. Both tasks are tested in a single ESM questionnaire using a broad general population sample. Special attention is payed to task evaluation, contextualization, and lessons-learned.

In **chapter 10**, the main findings of this thesis are summarized and discussed. This is done in the light of its clinical implications and from personal experience. The thesis will conclude with directions for future ESM use: towards new horizons in clinical care.

References

1. Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. *The Lancet.* 2018;392(10157):1553-1598.

2. Van Os J. De DSM-5 voorbij! Persoonlijke diagnostiek in een nieuwe GGZ. *Tijdschrift voor Psychiatrie.* 2014;56(5):356-357.

3. Delespaul P, Milo M, Schalken F, Boevink W, Van Os J. *Goede GGZ!: Nieuwe concepten, aangepaste taal en betere organisatie.* Springer; 2017.

4. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular psychiatry.* 2012;17(12):1174.

5. van Os J, Guloksuz S, Vijn TW, Hafkenscheid A, Delespaul P. The evidence-based group-level symptom-reduction model as the organizing principle for mental health care: time for change? *World Psychiatry.* 2019;18(1):88-96.

6. Van Os J. *De DSM-5 voorbij!: persoonlijke diagnostiek in een nieuwe ggz.* Springer; 2017.

7. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. In: Am Psychiatric Assoc; 2010.

8. McGorry P, van Os J. Redeeming diagnosis in psychiatry: timing versus specificity. *The Lancet.* 2013;381(9863):343-345.

9. van Os J, Delespaul P, Wigman J, Myin-Germeys I, Wichers M. Beyond DSM and ICD: introducing "precision diagnosis" for psychiatry using momentary assessment technology. *World Psychiatry.* 2013;12(2):113.

10. Ingram RE, Luxton DD. Vulnerability-stress models. *Development of psychopathology: A vulnerability-stress perspective.* 2005;46.

11. Zubin J, Spring B. Vulnerability: a new view of schizophrenia. *Journal of abnormal psychology.* 1977;86(2):103.

12. Green MF. *Schizophrenia from a neurocognitive perspective: Probing the impenetrable darkness.* Allyn & Bacon; 1998.

13. Liberman RP, Green MF. Whither cognitive-behavioral therapy for schizophrenia? *Schizophrenia Bulletin.* 1992;18(1):27-35.

14. Slade M. Mental illness and well-being: the central importance of positive psychology and recovery approaches. *BMC health services research.* 2010;10(1):26.

15. Keyes CL. Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of consulting and clinical psychology.* 2005;73(3):539.

16. WHO. Constitution of the World Health Organisation. 2016; www.who.int/governance/eb/who_constitution_en.pdf. Accessed 09, 2019.

17. Huber M, Knottnerus JA, Green L, et al. How should we define health? *Bmj.* 2011;343:d4163.

18. WHO. Mental Health: New Understanding, New Hope. The World Health Report. In: World Health Organization Geneva; 2001.

19. Barry MM. Addressing the determinants of positive mental health: concepts, evidence and practice. *International Journal of Mental Health Promotion.* 2009;11(3):4-17.

20. Zimmerman MA. Psychological empowerment: Issues and illustrations. *American journal of community psychology.* 1995;23(5):581-599.

21. Aujoulat I, d'Hoore W, Deccache A. Patient empowerment in theory and practice: polysemy or cacophony? *Patient education and counseling.* 2007;66(1):13-20.

22. Van Audenhove C, Vertommen H. A negotiation approach to intake and treatment choice. *Journal of Psychotherapy Integration.* 2000;10(3):287-299.

23. Steekelenburg Ev, Kersten I, Huber M. Positieve gezondheid'in Nederland-Wie, wat, waarom en hoe? Een inventarisatie. In: Amersfoort: ZonMw and Institute for Positive Health; 2016.

24. Anthony WA. Recovery from mental illness: the guiding vision of the mental health service system in the 1990s. *Psychosocial rehabilitation journal.* 1993;16(4):11.

25. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry.* 2011;199(6):445-452.

26. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. *Journal of medical Internet research.* 2005;7(1):e1.

27. Kumar S, Nilsen WJ, Abernethy A, et al. Mobile health technology evaluation: the mHealth evidence workshop. *American journal of preventive medicine.* 2013;45(2):228-236.

28. Marzano L, Bardill A, Fields B, et al. The application of mHealth to mental health: opportunities and challenges. *The Lancet Psychiatry.* 2015;2(10):942-948.

29. Wykes T, Haro JM, Belli SR, et al. Mental health research priorities for Europe. *The Lancet Psychiatry.* 2015;2(11):1036-1042.

30. Torous J, Andersson G, Bertagnoli A, et al. Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry.* 2019;18(1):97.

31. Hektner JM, Schmidt JA, Csikszentmihalyi M. *Experience sampling method: Measuring the quality of everyday life.* Sage; 2007.

32. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1-32.

33. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. In: *Flow and the foundations of positive psychology.* Springer; 2014:35-54.

34. Myin-Germeys I, Delespaul P, van Os J. The experience sampling method in psychosis research. *Current Opinion in Psychiatry.* 2003;16:S33-S38.

35. Walz LC, Nauta MH, aan het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: A systematic review. *Journal of anxiety disorders.* 2014;28(8):925-937.

36. 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. *Psychological medicine.* 2009;39(9):1533-1547.

37. van Os J, Verhagen S, Marsman A, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and anxiety.* 2017;34(6):481-493.

38. van Os J, Delespaul P, Barge D, Bakker RP. Testing an mHealth momentary assessment Routine Outcome Monitoring application: a focus on restoration of daily life positive mood states. *PLoS One.* 2014;9(12):e115254.

39. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of Patients and Clinicians on the Promise of the Experience Sampling Method for Psychiatric Care. *Psychiatric Services.* 2019:appi. ps. 201900050.

Part I

Clinical Relevance

2

Use of the experience sampling method in the context of clinical trials

S. Verhagen, L. Hasmi, M. Drukker, J. van Os, P. Delespaul

Abstract

Objective: The experience sampling method (ESM) is a structured diary technique to appraise subjective experiences in daily life. It is applied in psychiatric patients, as well as in patients with somatic illness. Despite the potential of ESM assessment, the improved logistics and its increased administration in research, its use in clinical trials remains limited. This paper introduces ESM for clinical trials in psychiatry and beyond.

Method: ESM is an ecologically valid method that yields a comprehensive view of an individual's daily life. It allows the assessment of various constructs (e.g., quality of life, psychopathology) and psychological mechanisms (e.g., stress-sensitivity, coping). These constructs are difficult to assess using cross-sectional questionnaires. ESM can be applied in treatment monitoring, as an ecological momentary intervention, in clinical trials, or in single case clinical trials. Technological advances (e.g., smartphone applications) make its implementation easier.

Results: Advantages of ESM are highlighted and disadvantages are discussed. Furthermore, the ecological nature of ESM data and its consequences are explored, including the potential pitfalls of ambiguously formulated research questions and the specificities of ESM in statistical analyses. The last section focuses on ESM in relation to clinical trials and discusses its future use in optimizing clinical decision-making.

Conclusion: ESM can be a valuable asset in clinical trial research and should be used more often to study the benefits of treatment in psychiatry and somatic health.

Introduction

In clinical trials, the effectiveness of a psychiatric intervention is often assessed retrospectively, by asking patients to complete self-report questionnaires or by administering clinical interviews over the past days or weeks[1]. These methods lack detailed insight into momentary processes and context sensitivity relevant to assess stress-reactivity, a notion central to psychopathology. Therefore, the use of momentary assessment techniques is proposed in mental health research to tap into daily life symptom experiences[2,3]. Momentary assessment techniques, such as the Experience Sampling Method (ESM), have a long tradition[4-6]. This paper aims to demonstrate the usefulness of these methods in clinical trial research, in the field of psychiatry and beyond.

What is ESM?

ESM is an umbrella term for a family of momentary assessment techniques that use signals to trigger data collection in daily life[5]. Alternative terms are ecological momentary assessment[6], ambulatory assessment[7], event sampling[8], beeper studies[9], the structured diary method[10], intensive longitudinal assessment[11], and real-time data capture studies[12]. ESM assesses sampled experiences and behaviours as well as moment-to-moment changes in mental states, embedded in (normal) daily life. It is an empirically validated structured diary technique. Typically, participants are asked to repeatedly complete short questionnaires (lasting no more than 2 min) in response to beep prompts. The questionnaires cover current mood, cognitions, perceptions, behaviours and descriptors of the momentary context (e.g., location, company, activity)[4-6]. ESM is useful in psychiatry, as well as in patients with somatic illness. The method focusses on symptoms (ill health), as well as on adaptive functioning (well-being), aiming to map (normal) daily psychological functioning[13].

Originally, paper diaries were used in combination with pagers or electronic wristwatches[4]. As technology became more advanced, data collection logistics and reliability was improved by the use of personal digital assistants[13] and smartphone applications[14] (e.g., the PsyMate, ExperienceSampler or apps by Invivo, lifeData). Typically, assessments are collected using an unpredictable (random) time sampling protocol. Assessments can also be triggered by an event (event sampling)[8]. Questionnaires are designed for quick and easy data collection and use open-ended questions, visual analogue scales, check lists or self-report Likert scales[4,13].

A typical ESM data form

ESM assessments usually comprise a morning questionnaire, a beep questionnaire assessed repeatedly within a day and an evening questionnaire. The content of the items is subject to

the assessment theme. Questions are short and can be rated quickly. An example of such an assessment form is included in table 1.

Table 1
Beep-level assessment sheet from the Psymate standard assessment protocol
(http://www.psymate.eu)

I feel cheerful	1=not at all … 7=very
I feel insecure	1=not at all … 7=very
I feel relaxed	1=not at all … 7=very
I feel annoyed	1=not at all … 7=very
I feel satisfied	1=not at all … 7=very
I feel lonely	1=not at all … 7=very
I feel anxious	1=not at all … 7=very
I feel down	1=not at all … 7=very
I feel guilty	1=not at all … 7=very
(optional additional personal question)	1=not at all … 7=very
What am I doing?	Select: resting, work or study, household, hygiene, eating/drinking, leisure, other, nothing
I would rather do something else	1=not at all … 7=very
Where am I?	Select: at home, at family or friend's place, at work or school, public place, transport, somewhere else
Who am I with?	Select: partner, family resident, family non-resident, friends, colleagues, acquaintances, strangers or others, nobody
I am hungry	1=not at all … 7=very
I am tired	1=not at all … 7=very
I am in pain	1=not at all … 7=very
(automatic recording)	Day and time tag

Why is ESM used in research?

ESM typically has a number of advantages, but its implementation varies.

- ESM has high ecological validity because assessments are made in the natural flow of real life[13,15].
- ESM reduces memory strains and avoids aggregation because only the actual moment is assessed repeatedly over time. This increases accuracy and is comparatively easier[15].
- The repeated assessments are collected in different situations (contextualized), which allows researchers to disentangle and understand the variability in mental states and psychological constructs[16].
- ESM yields a rich data set covering information on mental state, quality of life, mobility, social network and more, which reduces the need to use separate questionnaires measuring different constructs[13].
- Assessment error is reduced by repeated measures over time. This improves the validity, reliability, and transparency of individual pattern assessments, which is helpful in clinical practice[12].
- Sensitivity to detect change increases due to the collection of data at multiple time points[11].
- ESM makes it easier for patients during feedback sessions to acknowledge and translate the findings to daily life practice, because they have rated the items (e.g., anxiety) repeatedly. This avoids mystification. Involvement and collaboration in care are improved and feelings of empowerment in the treatment process are heightened[17].

A possible drawback of ESM is that the method is perceived as time-consuming and demanding[16]. Ideally, assessments are kept as brief as possible, preferably <1 min, with a maximum of 2 min[4]. A study protocol of 10 assessments for 6 consecutive days results in a cumulative time investment of < 120 min. The set of questionnaires needed in cross-sectional studies easily exceeds this time investment. Selection bias is a second concern. Not all patients are willing to participate or comply with the protocol, and this subgroup possibly functions better than the non-responders[3]. However, previous research has shown that the method is feasible in a wide variety of patients, including those suffering from severe mental illness[2,4]. Most studies have reported good compliance[18]. On the other hand, patients are inclined to miss assessments as a response to their current mood, which could also lead to selection bias. However, extensive analysis of missed responses demonstrate that beeps are predominantly missed at logical moments, such as the morning and evening (when participants are most likely sleeping), and lag analysis found no mood-related dropped

responses[19]. Third, ESM can induce reactivity[16]. For instance, participants report increased awareness during ESM data collection[18]. Monitoring is also used as a behavioural intervention, for instance, to reduce cigarette smoking. In this case, behavioural change is maximised by single target assessment of unwanted (automatic) behaviours. ESM, in contrast, is an open exploration of daily life using multiple targets. This open exploration reduces reactivity. Furthermore, repeated (exclusive) assessments of negative mood states can be confrontational and induce negative feelings. To avoid this, a careful and balanced construction of the questionnaire is needed; for example, changing the order of questions so that positive and negative momentary mood items are mixed. Finally, the complexity of ESM data analyses could form a drawback for both researchers and clinicians[11]. Normal statistical methods, such as linear regression analysis and analysis of variance, cannot be used[20]. Aggregation of data per participant removes the multilevel structure, allowing simple analyses. In addition, most statistical packages now have easily accessible multilevel regression tools. The present paper provides guidelines how to analyse the data properly, allowing a reliable assessment of the richness of the ESM data (see below)[20].

Method and statistics

ESM has advantages over classical assessment methods in clinical trials. However, ESM is a complex assessment method and considerations should be taken into account when using this innovative approach.

The nature of ESM data

The nature of ESM data is appealing but its complexity often challenges researchers and clinicians. The data collection parameters (the actual design) that define the nature of the data can be more important than statistical techniques. Aspects to consider are item selection, item order, the timeframe (e.g., number of days), the intensity (e.g., number of beeps per day; number of questions within a beep), the need for additional information (e.g., sleep quality assessed in a morning questionnaire), the signalling algorithm (e.g., random, fixed, beep-free periods, anticipation), the addition of event recording (e.g., of stressors or panic attacks), application type and data storage.

Items from cross-sectional questionnaires are often unsuitable for repeated assessment in daily life. In one-time questionnaires, a reliable assessment of a construct is achieved by redundancy (multiple items in a sum score). Repeated (up to 10 times a day) answering of similar items is frustrating. Often, metaphors are used, but these lack variation within the day[21]. ESM information, such as current mood, activity and company, is assessed with single

items, which can be combined to improve reliability: different aspects at one moment or the same items over time. ESM data is ecologically valid but correlational in nature[22]. The current activity can be a cause as well as an effect of momentary mood states. Furthermore, different mental states have different natural flows. Anxiety, for example, fluctuates and is more contextually reactive than depression. With an ESM sampling frequency of 10 times a day, highly variable states will not be adequately represented. In that case, the process is under-sampled. Other slow changing states are often over-sampled. The actual ESM protocol is usually a compromise.

Event monitoring is often added to ESM protocols. For example, participants are asked to complete a questionnaire when a certain stressor is present, or when the participant has a panic attack. This requires continual prospective monitoring and results in a high workload for participants. The recorded event initiates a questionnaire. Events should be discrete (have a clear beginning and end) and often require coding instructions (what is a panic attack or a social interaction?). There is no correction for rating misinterpretation. Feasibility limits the number of (different) events that can be reported. Having to respond to additional questions after reporting an event can act as a 'punishment' and results in an extinction of the report (not the actual event) under some (stressful) circumstances. The same 'learning curve' can occur in branching when different answers lead to different workloads. Often, time sampling is preferred because it is less burdensome, more reliable, allows the (non-exhaustive) assessment of a larger set of events and reports events as well as non-events (e.g., when a participant smokes, as well as when a participant does not smoke).

When beeps are programmed at fixed times, predictability increases reactivity and this may induce behavioural changes (e.g., postpone shopping or showering). More generally, when beep-free periods can be anticipated (e.g., no beep expected within an hour after a beep), reactivity increases. Random sampling avoids this reactivity. Unexpectedly, true random schedules (e.g., 3 beeps on 1 day and 15 on another) are not ideal, because long periods with no beeps can result in some participants staying at home (not to miss a beep). This behaviour disrupts normal daily life. Therefore, a stratified random schedule with restricted intervals is advised[4]. This list of design modalities is not exhaustive and the actual choices depend on the research question and study population.

Statistical analyses of ESM data
ESM samples daily life and can be used to compute time budgets, for example, the proportion of time spent alone or doing sports. Similarly, the average anxiety level (or anxiety while alone) can be computed, resulting in one statistic per participant as the input for

standard statistical analyses. However, in most situations, statisticians prefer to use all available data. In that case, analyses have to account for the fact that in ESM, multiple participants answer a set of questions multiple times, resulting in a multilevel data set. Hence, ESM observations are not independent and standard linear and logistic regression analysis techniques cannot be used[20]. Multilevel linear and logistic regression is indicated and is now available in most statistical analysis packages (e.g., Stata, R, SPSS). Since observations are nested in participants, the most simple multilevel regression models for ESM data include a random intercept that allows appropriate control for the between participant variability[20]. These models therefore include a second error term at the participant level, in addition to the regular error term (at the assessment level) used in unilevel regression. Even when trends over time are not the primary interest, it is advisable to remove the potentially confounding effects of time in ESM data by adding a time variable (detrending[23]). This is effective when values increase or decrease linearly over the time of the study. Furthermore, since adjacent assessments tend to be more similar to each other than to other assessments further apart in time (autocorrelation), ESM analyses also need more sophisticated adjustment methods[11]. For example, autocorrelation can be taken into account by using appropriate correlational structures for the residuals (e.g., AR, MA, ARMA)[11]. Care must be exercised when using such models, since observations are not evenly spaced due to randomization and data collection usually proceeds over multiple days, resulting in a huge time lag for observations on consecutive days. Therefore, continuous-time structures are usually preferable, although their availability across different software packages differs. It is also highly recommended to allow random slopes in the model where applicable[20]. This yields more realistic estimates of SEs, which otherwise might be underestimated. A random slope means that the regression coefficient is not fixed, but is allowed to vary between participants. This variation is realised by adding an extra error term for each predictor. In these models a simple variance-covariance matrix of the random effects (e.g., which assumes that all covariances are zero: the default in some statistical software packages) is not realistic. Therefore, an unstructured variance-covariance matrix, where all variances and covariances are allowed to be different, is preferable. However, this can lead to rather complex and time-consuming models. Sometimes, the statistical program may even fail to provide output. Alternatively, one can consider only using random slopes for the main independent variable(s). If the analysis includes multiple main independent variables, or if the analysis still does not yield valid results after reducing the number of random slopes, it may be necessary to remove essential random slopes. In that case, resampling methods (e.g., permutation tests, bootstrapping) can be used to obtain valid tests and CIs[24].

Clarifying ESM questions

ESM allows the study of a broad range of research questions. The epidemiological finding that the mental health of people in urban environments is poor[25] (a between subject question), can be explored by assessing whether people in crowded environments feel worse (a within subject question). ESM allows the exploration or 'unpacking' of relevant factors[26]. The multilevel data structure of ESM can confuse researchers. Consider, for example, a study that explores the relation between depression and loneliness. What should be studied? Are depressive persons often alone, do lonely people become depressed, does depression make people lonely, are people in general more depressed when alone, and does this relation hold for depressed people? ESM data analyses require more specific research questions. The pitfall lies in the convergence of person-related questions and situation-related questions, both of which are represented in ESM data and can lead to divergent conclusions[22]. In addition, results from ESM studies are easily misinterpreted. The conclusion that agoraphobia is a bogus diagnosis because ESM data shows that people with this diagnosis feel worse when at home (and better when out), disregards the fact that people may select going out on days they feel best. In correlational relations, causal attributions should be avoided.

Discussion

ESM was developed to sample moments from daily life. Since then, the method has yielded valuable results to better understand mental illness[16]. ESM yields a comprehensive view of the participant's daily life. This comprehensive view is valuable in clinical trials because the same instrument can measure various constructs, such as quality of life, psychopathology, social networks or productive activities. This broad measurement is done with high sensitivity to change[11]. Even more, ESM allows assessment of mechanisms that are difficult to assess with cross-sectional questionnaires, such as stress-sensitivity, coping, reward-sensitivity and resilience[16]. ESM has the potential to replace a larger body of instruments and sensitively assesses new and relevant parameters. It is, however, not yet the gold standard in clinical trials. This fact is mainly due to the heterogeneity in ESM applications[27]. Nevertheless, it has the potential to become the de facto gold standard.

ESM monitoring in clinical trials

ESM data is ideal to inform and optimize personalized treatment in mental health. Feedback modules have been developed that translate ESM data into understandable graphs and figures that help promote patient engagement in treatment. Kramer et al[28] assessed the

benefits of an ESM feedback-intervention in a randomized control clinical trial. An ESM feedback-intervention was studied in addition to treatment as usual in patients diagnosed with depression. They demonstrated that ESM with weekly feedback moments reduced depression more than treatment as usual. ESM monitoring without feedback had a similar association, but this reduction did not last. The majority of ESM studies have short assessment periods, ranging from several days to a few weeks[12]. In clinical practice, it is important to follow a patient for months or years. To reduce the burden, clinicians experiment with interrupted periodic ESM assessments. Psychiatrists and patients can mutually agree to customize ESM to monitor the effectiveness of treatment and update interventions accordingly. There is growing insight in what might be relevant to individual patients. For example, a patient who used ESM for a year had increased associations between symptoms in periods of exacerbation (personal communication, Dr. M. Bak). The patient's subjective feeling of relapse coincided with changes in ESM towards higher severity. Monitoring patients this way provides both the clinician and the patient with additional information that can improve a personalized treatment process.

Ecological momentary intervention (EMI) has been cued as an umbrella term for mobile interventions in the daily living environment of participants. Unlike ESM, EMI aims not to observe, but to intervene. Only recently has mobile technology allows to bridge the gap between the therapist's office and situations in daily life, where acquired skills should generalize and be autonomously applied. Several studies have demonstrated that EMI is feasible, acceptable, and can be successful in mental health treatment[29].

Single case clinical trials
The translation of evidence from clinical trials assumes that patients are a homogeneous group. This is often not true. A large variability exists in treatment responses between subject. Single case designs can be used to customize treatment to individual needs and accommodate to strengths. Since treatment effect (or outcome optimization) can be difficult to detect, single case trials can add (experimental) control to clinical decision making[30-31]. This control requires that the timing of conditions is allocated randomly (independent from the clinical effect). Guidelines are available for the randomization of single case trials[32], and statistical software packages[33] facilitate the analyses. An interesting innovation is a user-initiated program, which implements ESM to monitor the effects of gradual dose-reductions in the user's antidepressant medication[34]. Over the course of a year, the user collected over a thousand ESM observations. Meanwhile, his clinician teamed up with a pharmacist to blindly implement a dose-reduction (tapering) schedule. From the data, early warning signs

were detected for mood changes that predicted relapse. They could relate the mood change patterns to the medication reduction and were able to find the optimal, safe medication dosage[34]. The use of ESM in classical clinical trials is less challenging because only aggregated group findings are pursued. Global constructs can be used to comprehensively describe the mental state of groups of individuals. In single-case studies, the data is highly individual and can benefit from the individual selection of items that maximally reflects the user's variability.

Future of ESM in clinical trials

The advantages of single case trials to optimize treatment are increasingly recognized[31]. In addition, there is increasing recognition of how mobile phones can be used as medical instruments in the treatment of mental illness[35]. The benefits of combining mobile technologies with single case designs are also recognized[36]. Dissemination in regular practice requires freely available user-friendly technology and software that reduces the logistical burden and need for statistical skills. By making this technology available, treatment efficiency can be improved. Most participants are already used to carry mobile phones, so the threshold for using them to collect ESM data via an app is greatly reduced. ESM progresses in parallel with evolutions in mobile technology. Data collection logistics are reduced by innovations in tools and programs, as well as cultural changes related to the use of mobile devices. Furthermore, innovation in mobile tools facilitates the general acceptance of momentary assessment. For broader clinical dissemination, great investments are needed in software that allows on-the-fly comprehensive feedback. More research should focus on the way data become information. This knowledge is important for academically trained professionals, as well as for clinicians who are more reluctant to use and less able to understand statistical analyses. In addition, the patients themselves should be able to comprehend the data. Ideally, treatment planning, implementation, and evaluation should be a collaborative process. ESM can prove its utility both in large-scale studies and at the level of clinical practice. The latter is especially important, as ESM can be used to optimize individual treatment and to help in decision-making. In the future, assessment strategies that are more sensitive in detecting change could be used in all prescriptions of pain or psychotropic medication. Medication should be continued only when efficacy is determined in a specific patient. Sensitive assessment strategies can also be used in the domain of optimal dose finding. Furthermore, recovery oriented care in psychiatry is not restricted to ill health and symptoms, but extends to well-being. Daily life assessment strategies can inform patients and clinicians how to improve resilience by maximizing strengths in patients.

References

1. FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims2009;No. 235:65132-65133. Located at: Fed Regist.

2. DeVries MW. *The experience of psychopathology : investigating mental disorders in their natural settings.* Cambridge ;: Cambridge University Press; 1992.

3. Palmier-Claus JE, Myin-Germeys I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatrica Scandinavica.* 2011;123(1):12-20.

4. Delespaul PAEG. *Assessing schizophrenia in daily life the experience sampling method*. Maastricht :, UPM, Universitaire Pers Maastricht; University Library, Maastricht University [Host]; 1995.

5. Hektner JM, Schmidt JA, Csikszentmihalyi M. *Experience sampling method: Measuring the quality of everyday life.* Sage; 2007.

6. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavorial medicine. *Annals of Behavioral Medicine.* 1994.

7. Fahrenberg J, Myrtek M, Pawlik K, Perrez M. Ambulatory assessment-monitoring behavior in daily life settings. *European Journal of Psychological Assessment.* 2007;23(4):206-213.

8. Reis HT, Gable SL. Event-sampling and other methods for studying everyday experience. *Handbook of research methods in social and personality psychology.* 2000:190-222.

9. Hurlburt R, Heavey CL. To beep or not to beep: Obtaining accurate reports about awareness. *Journal of Consciousness Studies.* 2004;11(7-8):113-128.

10. Bolger N, Davis A, Rafaeli E. Diary methods: Capturing life as it is lived. *Annual review of psychology.* 2003;54(1):579-616.

11. Bolger N, Laurenceau J-P. *Intensive longitudinal methods : an introduction to diary and experience sampling research.* New York, NY :: Guilford Press; 2013.

12. Stone A, Shiffman S, Atienza A, Nebeling L. *The science of real-time data capture: Self-reports in health research.* Oxford University Press; 2007.

13. Csikszentmihalyi M, Mehl MR, Conner TS. *Handbook of research methods for studying daily life.* Guilford Publications; 2013.

14. Doherty ST, Lemieux CJ, Canally C. Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. *Social Science & Medicine.* 2014;106:83-92.

15. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1-32.

16. 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. *Psychological medicine.* 2009;39(09):1533-1547.

17. Simons C, Hartmann J, Kramer I, et al. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. *European Psychiatry.* 2015;30(8):900-906.

18. Johnson EI, Grondin O, Barrault M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. *International journal of methods in psychiatric research.* 2009;18(1):48-57.

19. Silvia PJ, Kwapil TR, Eddington KM, Brown LH. Missed beeps and missing data: Dispositional and situational predictors of nonresponse in experience sampling research. *Social Science Computer Review.* 2013:0894439313479902.

20. Snijders TA. *Multilevel analysis.* Springer; 2011.

21. Kimhy D, Myin-Germeys I, Palmier-Claus J, Swendsen J. Mobile assessment guide for research in schizophrenia and severe mental disorders. *Schizophrenia bulletin.* 2012:sbr186.

22. Larson R, Delespaul P, deVries M. Analyzing experience sampling data: A guidebook for the perplexed. *The experience of psychopathology: Investigating mental disorders in their natural settings.* 1992:58-78.

23. Wang LP, Maxwell SE. On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological methods.* 2015;20(1):63.

24. Davison AC, Hinkley DV. *Bootstrap methods and their application.* Vol 1: Cambridge university press; 1997.

25. van Os J, Rutten BP, Poulton R. Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. *Schizophr Bull.* 2008;34(6):1066-1082.

26. DeVries MW. Recontextualizing psychiatry: Toward ecologically valid mental health research. *Transcultural Psychiatry.* 1997;34(2):185-218.

27. Mitchell V. Mathis MD. *FDA Briefing Document; Psychopharmacologic Drugs Advisory Committee Meeting; topic: Cognitive Dysfunction in Major Depressive Disorder.* FDA; February 3, 2016 2016.

28. Kramer I, Simons CJ, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry.* 2014;13(1):68-77.

29. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology.* 2010;15(Pt 1):1-39.

30. Davidson KW, Peacock J, Kronish IM, Edmondson D. Personalizing Behavioral Interventions Through Single‐Patient (N‐of‐1) Trials. *Social and personality psychology compass.* 2014;8(8):408-421.

31. Dallery J, Raiff BR. Optimizing behavioral health interventions with single-case designs: from development to dissemination. *Translational behavioral medicine.* 2014;4(3):290-303.

32. Heyvaert M, Onghena P. Randomization tests for single-case experiments: State of the art, state of the science, and state of the application. *Journal of Contextual Behavioral Science.* 2014;3(1):51-64.

33. Bulté I, Onghena P. An R package for single-case randomization tests. *Behavior Research Methods.* 2008;40(2):467-478.

34. Wichers M, Groot PC, Psychosystems E, Group E. Critical Slowing Down as a Personalized Early Warning Signal for Depression. *Psychotherapy and psychosomatics.* 2016;85(2):114-116.

35. Gravenhorst F, Muaremi A, Bardram J, et al. Mobile phones as medical devices in mental disorder treatment: an overview. *Personal and Ubiquitous Computing.* 2015;19(2):335-353.

36. Vilardaga R, Bricker J, McDonell M. The promise of mobile technologies and single case designs for the study of individuals in their natural environment. *J Contextual Behav Sci.* 2014;3(2):148-153.

3

The experience sampling method as an mHealth tool to support self-monitoring, self-insight and personalized health care in clinical practice

J. van Os, S. Verhagen, A. Marsman, F. Peeters, M. Bak, M. Marcelis, M. Drukker, U. Reininghaus, N. Jacobs, T. Lataster, C. Simons, ESM-Merge investigators, R. Lousberg, S. Gülüksüz, C. Leue, P. Groot, W. Viechtbauer, P. Delespaul

Abstract

Background: The experience sampling method (ESM) builds an intensive time series of experiences and contexts in the flow of daily life, typically consisting of around 70 reports, collected at 8–10 random time points per day over a period of up to 10 days.

Method: With the advent of widespread smartphone use, ESM can be used in routine clinical practice. Multiple examples of ESM data collections across different patient groups and settings are shown and discussed, varying from an ESM evaluation of a 6-week randomized trial of mindfulness, to a twin study on emotion dynamics in daily life.

Results: Research shows that ESM-based self-monitoring and feedback can enhance resilience by strengthening the capacity to use natural rewards. Personalized trajectories of starting or stopping medication can be more easily initiated and predicted if sensitive feedback data are available in real time. In addition, personalized trajectories of symptoms, cognitive abilities, symptoms impacting on other symptoms, the capacity of the dynamic system of mental health to "bounce back" from disturbance, and patterns of environmental reactivity yield uniquely personal data to support shared decision-making and prediction in clinical practice. Finally, ESM makes it possible to develop insight into previous implicit patterns of thought, experience, and behavior, particularly if rapid personalized feedback is available.

Conclusion: ESM enhances clinical practice and research. It is empowering, providing coownership of the process of diagnosis, treatment evaluation, and routine outcome measurement. Blended care, based on a mix of face-to-face and ESM-based outside-the-office treatment, may reduce costs and improve outcomes.

Introduction

The practice of mental health is changing rapidly, requiring new tools to involve patients in the process of diagnosis and treatment, and to provide perspectives of acceptance and sense of purpose. Two factors are particularly important. First, treatment models propose a more active and empowered role of patients, with active self-management and shared decision making as core elements[1]. Self-monitoring is a main component of self-management, providing an opportunity for collaborative diagnosis and treatment evaluation, and forming the starting point for self-insight and initiation of change in patterns of experience and behavior. Second, the process of assessment and treatment in mental health care is becoming more personalized. Traditional diagnostic and guideline-based practice for the "average" patient is being complemented with more dynamic, personalized formulations of psychopathology as a within-person series of mental states that evolve over time, under the influence of daily life environments and the impact of mental states on each other, requiring personalized solutions. The fit of selected intervention strategies can be assessed, and personalized formulations reevaluated in an iterative monitoring process toward shared decision making.

In this article, we will discuss how the collection of intensive contextual mental state data, known as the experience sampling method (ESM) or ecological momentary assessment (EMA[2]), in the form of an mHealth smartphone application, has the potential to become a mainstream instrument to support self-monitoring, personalized health care and the development of self-insight and avenues for personal change and acceptance of disabilities in clinical practice. Although ESM was initially used as a research tool in social psychology and the study of psychopathology, it was also found to have value in the management of (severe) mental illness[3-8]. Smartphones and customized apps have dramatically simplified the user experience and reduced the logistic burden of data collection and analysis. ESM is now available as a free mHealth smartphone app in Apple App Store and Android Store (e.g., PsyMate, xSample, movisensXS, BeepMe, RealLife). Free availability makes it possible to use ESM in regular clinical practice. Although the multilevel analysis of ESM research data can be complex, even simple graphical feedback of ESM data can represent a source of substantial enrichment of clinical practice. By making use of the data at the personal and the descriptive level, testing $n = 1$ clinical hypotheses about diagnosis, treatment, change, and acceptance becomes possible[4,9].

The practice of ESM

ESM is a structured data collection technique in which participants respond to randomly timed repeated assessments (up to around 10 times per day) over the course of time (up to around 10 days—with extensions of up to 1 year) while functioning within their natural setting[2,4] (Figure 1). The sampling is signal-contingent, meaning that individuals respond to semirandom signals ("beeps"—typically one random beep per 90-min slot). Data collection focusses on (i) emotions—items that reflect the broad dimensions of positive affect (PA; e.g., cheerful, relaxed, satisfied, happy) and negative affect (e.g., down, lonely, anxious, insecure); (ii) contexts—where, with whom, events; (iii) activities—doing what, (iv) food, alcohol, drugs, medication; (v) aspects of psychopathology –psychotic phenomena, suicidality, craving, self-esteem; (vi) somatic symptoms (e.g., pain, fatigue, dizziness, tinnitus, etc.); and (vii) stress—stressful momentary events, stress associated with momentary activity (e.g., "This activity is difficult for me," "I would prefer doing something else"), stress associated with momentary company (e.g., "I find being with these people pleasant[reversely coded to reflect stress]").For use in routine clinical practice, a sampling scheme with eight random signals per day and no more than 30 items to rate is generally feasible[4]. With adequate software, subjects can rate responses within 1 min, only minimally disturbing the flow of normal daily routines. A sampling scheme of less than eight times per day yields insufficient measurements, given a normally 20–30% beep attrition in natural circumstances[4]. Furthermore, care must be taken to include only items that can vary momentarily (e.g., "I am angry") rather than items reflecting trait, that do not vary momentarily ("I am clever").

ESM has a number of advantages[4]. First, it gives the clinician insight in the contingencies of experience and behavior, based on prospective data. Second, ESM observations have, contrary to clinical interactions, ecological validity, reflecting real life variation in response to real life challenges. Third, ESM makes use of in-the-moment reports not subject to memory biases and "global evaluation" distortions. Fourth, although ESM requires considerable motivation on the part of the patient, this motivation likely correlates with motivation for change in the context of treatment. Given that this is an important clinical variable, ESM can be revealing, to both patient and attending clinician with respect to motivation for change. Fifth, ESM allows for a prospective, within-person approach to diagnosis and treatment, contrasting with the between-person, "average patient" approach of traditional evidence-based practice. Sixth, evidence indicates that ESM empowers patients to become actively involved in collecting data, pertinent to treatment and assessment, in their own environments, thus becoming minimally co-participant[10] and possibly part-owner of the process of care[11].

Although ESM can be combined with a range of sensor-based data (movement, position, activity) and ambulatory physiological measures (e.g., salivary cortisol, heart rate, blood pressure, voiding behavior, etc.), more research is required on their added value to ESM ratings of experience and behavior. Furthermore, while a range of physiological measures can be measured in daily life, their relevance for mental health diagnosis and treatment remains to be established.
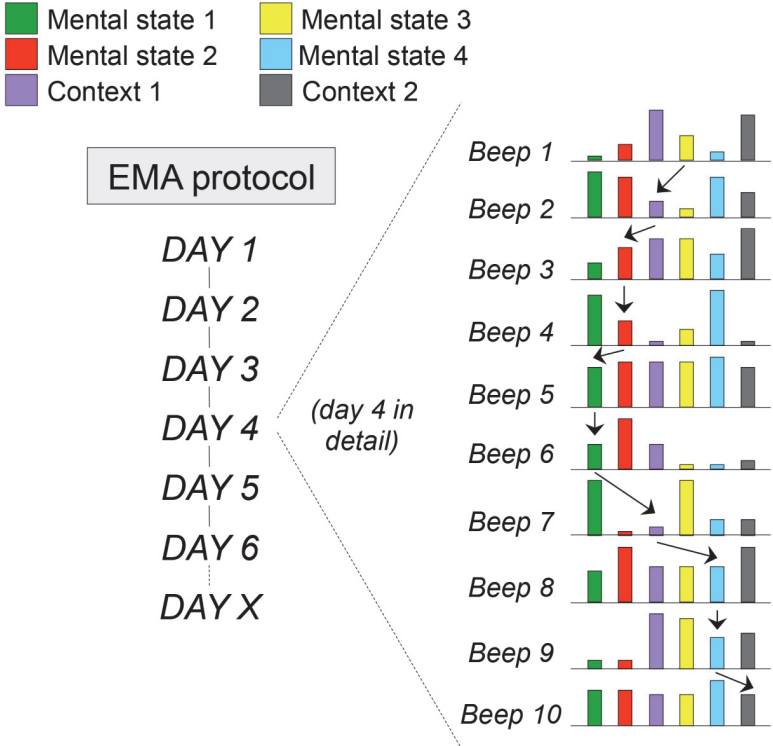
Figure 1

Principles of experience sampling methodology (ESM) or ecological momentary assessment (EMA).

Notes: At 10 random moments during the day, mental states (e.g., anxiety, low mood, paranoia, being happy) and contexts (stress, company, activity, drug use) are assessed. The arrows represent examples of prospectively analyzing the impact of mental states and contexts on each other over time, allowing for the construction of composite measures such as stress-sensitivity (impact of stress on negative affect), reward experience (impact of positive events on positive affect), or the autocorrelation (the impact of the mental state on itself from one moment to the other, resulting in more or less persistence over time).

Clinical applications of ESM

From group-based data to n=1 predictions

A local database with a large number of observations pertaining to different patient groups is helpful as a reference for the interpretation of clinical ESM data in individual patients. Next, a number of clinical applications of ESM will be illustrated using data from the ESM database that is being managed at Maastricht University (version 4.6). The database contains ESM studies that were conducted on different populations with comparable ESM lists, although not all ESM items were available in each group. Version 4.6 of the ESM database contains data on 276 healthy controls, 601 healthy twins from the general population, 178 relatives of patients with a psychotic disorder, 293 patients with psychotic disorder, 115 patients with depression, 129 patients with residual depression, and 106 individuals with variable levels of psychometric risk for psychotic disorder. We will show that ESM patterns at the group level are highly informative and form an attractive basis for testing hypotheses in single patient ESM data.

Assessing psychopathology variation and reactivity across time and place

Figures 2 and 3 display the simple variation over time and place of momentary "feeling low" in controls, individuals with residual depression, and individuals with depression. The graphs show clear differences in terms of severity, pattern of variation over time (Fig. 2), and pattern of reactivity as a function of place (Fig. 3). In controls, the mood item remains constant at low levels, showing little variation as a function of location. From individuals with residual depression to individuals with depression, "feeling low" is progressively more severe and showing progressively more diurnal variation. Furthermore, the data in the depressed groups show more reactivity as a function of location, with the greatest severity being associated with health care locations, as well as during transport.
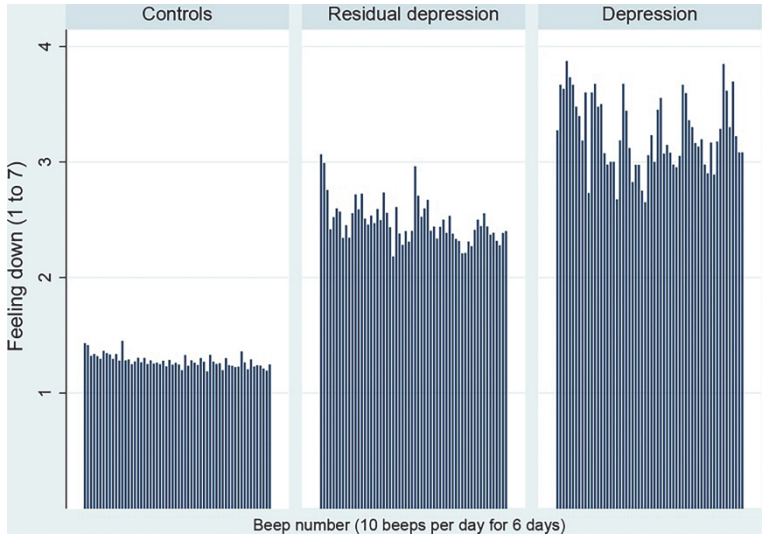
Figure 2

Momentary, beep-level self-reports (maximum of 60 beeps) of the mean of the ESM item "feeling down," collected over 6 days, in three samples of, respectively, controls (N = 251, n = 12,394 beep moments), patients with residual depression (N = 129, n = 6,420 beep moments), and depression (N = 45, n = 2,287 beep moments).
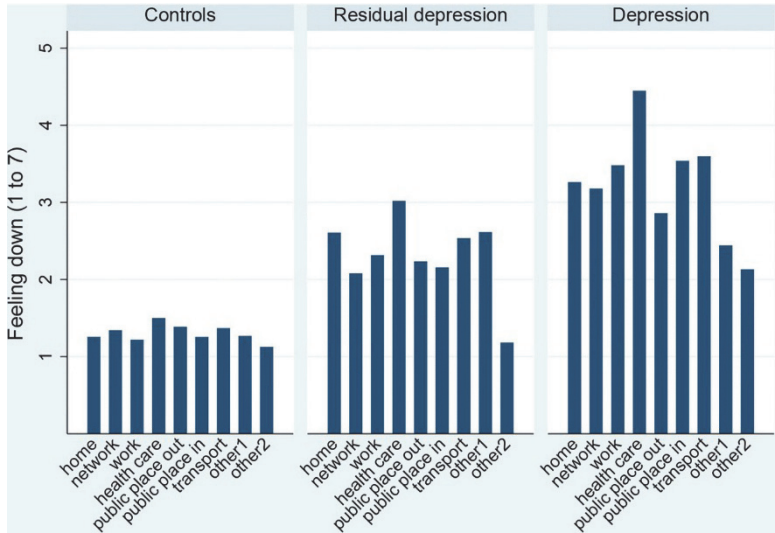
Figure 3

Self-reports of the ESM item "feeling down," collected over 6 days in daily life, summarized as mean level per location in three samples of controls (N = 251, n = 12,394 beep moments), patients with residual depression (N = 129, n = 6,420 beep moments), and depression (N = 45, n = 2,287 beep moments).

Although the Figures 2 and 3 thus suggest clear and meaningful differences between groups, the interpretation of individual ESM data remains a challenge. Figure 4 shows individual trajectories of momentary variation of the mood item "feeling low," separately for each patient in the depression group. These individual trajectories show very substantial variation from beep to beep and day to day, between and within persons. Such variation nevertheless has a meaningful and clinically interpretable pattern within individuals, differing from other persons, and is not entirely random. Therefore, ESM allows for clinical inspection and interpretation of data at both the group and individual levels.
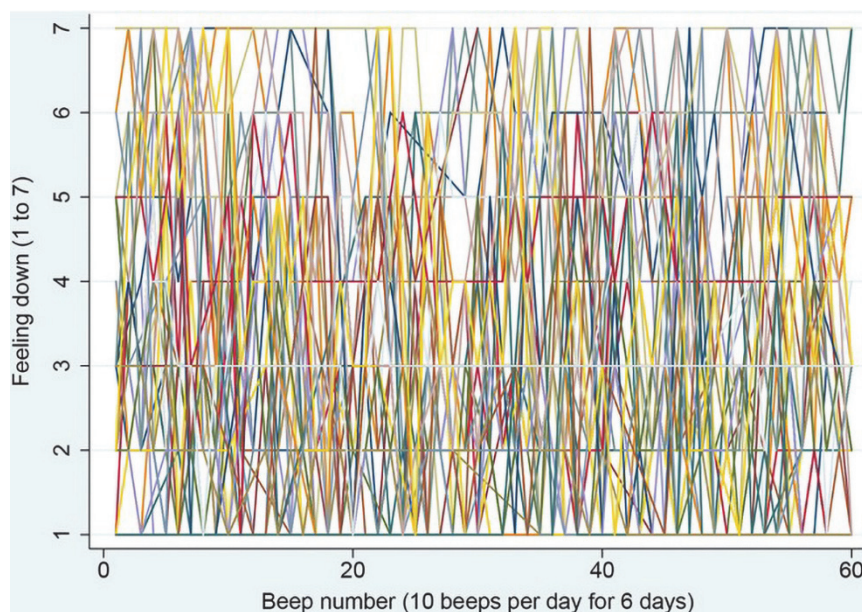
Figure 4

Trajectories of momentary (n = 1), beep-level ESM self-reports of "feeling down," collected over 6 days, in a sample of depressed patients (N = 45, n = 2,287 beep moments).

Assessing dynamic system ability to bounce back from disturbance

An emerging concept in psychopathology is the ability to bounce back given disturbance and impending mental decompensation in a dynamic system model of psychopathology[12,13]. For example, alterations in mood may cause a person to perceive the environment as threatening or alien. Initially, the person may be able to "shake off" this impression of aberrant salience, indicating resilience and stability, with a rapid return of the altered mental state to baseline in the face of disturbing influences. However, in more advanced stages, the

person may not rapidly return to the baseline mental state and progressively spend more and more time in the altered mental state of aberrant salience[14]. ESM data are suitable to capture this process of progressive loss of flexibility and ability to bounce back given system disturbance by plotting the autocorrelation of successive mental states. The autocorrelation of a mental state in ESM is the similarity between observations as a function of the time lag between them, or the tendency of the mental state to persist from one moment to the other over time[14]. In Figure 5a, controls have low variability in the ESM item indexing paranoia, therefore the autocorrelation function is low and not necessarily informative on the ability to bounce back from paranoid mental state deviations. In patient groups, ESM data are less ambiguous in this regard, allowing for interpretation of the autocorrelation. Thus, those with residual depression have altered bounce back ability (less rapid return to baseline), whereas individuals with psychotic disorder show paranoia inflexibility (Figure 5a). Figure 5b shows that from controls to residual depression to depression, there is a decrease in bounce back ability from mental states deviations characterized by anxiety. However, no such differences are seen for the item of "difficulty concentrating" (Figure 5c), suggesting that alterations of mental state flexibility may occur independently across different mental states, and that there may be a "hierarchy" (within and between persons) in the level of volatility in emotions and experiences, ranging from extreme volatility to "trait"-like persistence. In addition, charting the autocorrelation as a function of resilience to "bounce back" from disturbance may inform on environmental or genetic risk factors, as shown in Figure 5d, which depicts altered capacity to bounce back from the mental state of anxiety as a function of parental divorce. Alterations in the ability to bounce back from disturbance can also be studied within individuals, particularly in longer ESM trajectories, as a function of treatment response and prediction of outcome[13]. Conversely, a greater autocorrelation in PA, that is, a greater level of momentary transfer of positive emotions, may be protective, as it was associated with a greater recovery rate and prevention of depressive episodes[15]. To date, however, lengthy ESM-based time series remain rare in the research literature.
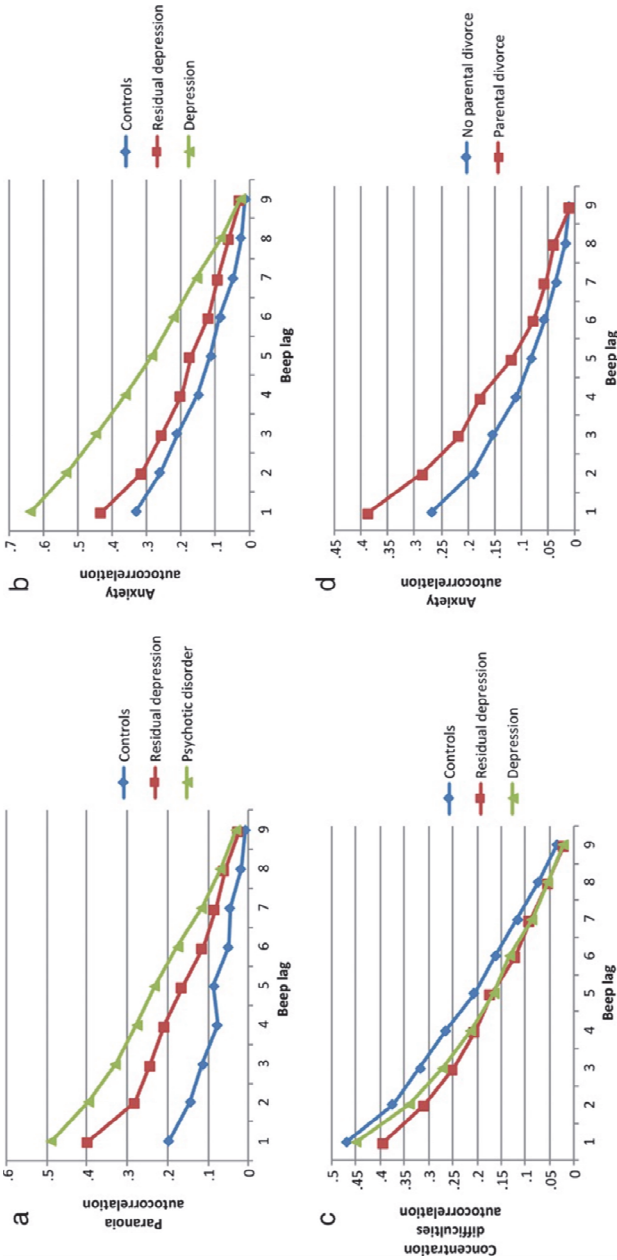
Figure 5

(a) Autocorrelations over successive lags of the ESM mental state of feeling "suspicious" (paranoia), collected over 6 days (10 beeps per day, nine lags), in three samples of controls (N = 212, n = 10,260 beep moments), patients with residual depression (N = 129, n = 6,412 beep moments) and psychotic disorder (N = 293, n = 12,400 beep moments). (b) Autocorrelations over successive lags of the ESM mental state of feeling "anxious," collected over 6 days (10 beeps per day, nine lags), in three samples of controls (N = 276, n = 13,513 beep moments), patients with residual depression (N = 129, n = 6,420 beep moments), and depression (N = 114, n = 5,324 beep moments). (c) Autocorrelations over successive lags of the ESM mental state of "difficulties concentrating," collected over 6 days (10 beeps per day, nine lags), in three samples of controls (N = 276, n = 13,199 beep moments), patients with residual depression (N = 129, n = 6,483 beep moments), and depression (N = 115, n = 5,329 beep moments). (d) Autocorrelations over successive lags of the ESM mental state of "feeling anxious," collected over 5 days (10 beeps per day, nine lags) in a general population sample of twins of whom N = 525 (n = 18,991 beeps) reported no parental divorce and N = 51 (n = 1,852 beeps) reported parental divorce.

Assessing the impact of symptoms on each other

The above example provides indirect evidence of how one mental state (low mood) affects another mental state (paranoia), increasing levels of low mood apparently giving rise to progressively higher levels of paranoia. However, ESM data also allow for a more direct examination of this process, by modeling the prospective relationships between mental states and symptoms. Given the fact that ESM data represent a time series, it is possible to examine the impact of, for example, the level of anxiety at the preceding beep (time point $t − 1$) on low mood at the next beep (time point t). Similar analyses can be carried out at other levels of aggregation, for example, at day-level. These types of analyses have to take into account natural "breaks" in the data, avoiding, for example, that an ESM item pertaining to the first beep of one day is predicted by an ESM item pertaining to the last beep of the previous day. In Figure 6, prospective relationships are assessed between momentary anxiety and depression, and between momentary feeling happy, on the one hand, and anxiety and depression, on the other. In those with low momentary levels of depression at $t - 1$, anxiety at $t - 1$ drives the incidence of depression at t in a dose-response fashion. Vice versa, momentary depression at $t - 1$ drives incident anxiety at t. The two mental states can thus be said to mutually impact each other's momentary incidence in a reinforcing loop. In contrast, feeling happy at $t - 1$ protects, in a dose-response fashion, against the onset of both anxiety and depression (Figure 6). Dynamic prospective relationships between mental states and symptoms can be extended to a full network of mutually impacting momentary mental states, involving multiple symptoms[16]. Although between person analyses are helpful to examine dynamics at the group-level, ESM allows for examination of within-person relationships and their impact on treatment. The concept of ESM *clinical network analysis* was recently introduced to describe the possibility of collecting ESM data for an extended period to help patients gain insight in the dynamics of the underlying psychopathology to decide which are the best connected points in the network where change likely will affect the most other symptoms and to assess the impact of treatment in the network[9]. For example, Bak, Drukker, Hasmi, van Os[9] showed the example of a person with disabling auditory hallucinations that were fueled by a mutually reinforcing loop of paranoia and low mood, allowing for a more focused treatment approach.
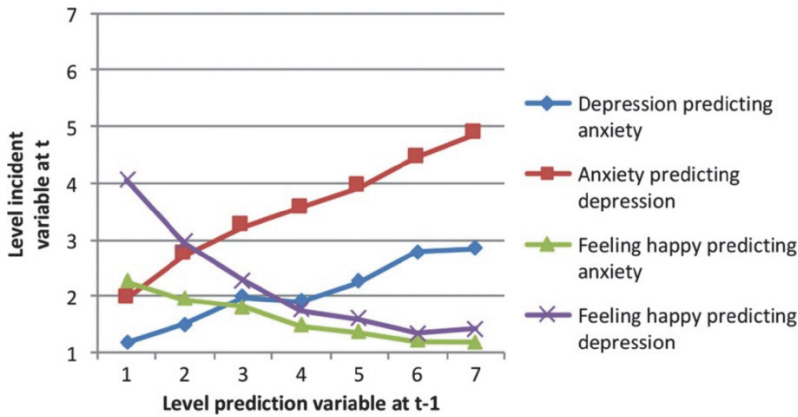
Figure 6

Predicting the incidence (i.e., mean level of mood state at t in those with the lowest score of that mood state at beep t - 1), as a function of the level of another mood state at beep t - 1, in a sample of individuals with residual depression (N = 223, n = 10,582 beeps). For example, in those with the lowest score of "feeling down" at t - 1, higher levels of anxiety at t - 1 predict progressively greater levels of feeling down at t, and higher level of feeling happy at t - 1 predict progressively lower scores of feeling down at t.

Assessing the impact of the environment

One of the important advantages of ESM is the ability to link momentary measures of experience and behavior to self-reports of the environment. In Figure 7a, momentary reports of low mood as a function of activity stress (the rating of the momentary activity in terms of "I would rather do something else") are compared between a control group and a group with residual depression. The results show that low mood in controls has constant low values, not responding much to increases in activity stress, whereas low mood in individuals with residual depression appears to be much more reactive to the stressor. Similarly, momentary events rated as more pleasant neutralize momentary low mood in controls to a small degree, whereas a much greater protective effect is seen in patients with residual depression (Fig. 7b). This latter finding of greater reductions in negative affect following positive events is consistent with an independent report in the literature in patients with depression[17]. The "buffering" effect of positive events is considerable: in Figure 7b, the level of low mood in the patient group at the highest level of pleasant event exposure approaches that of the control group. In other words, individuals with depression have higher levels of low mood, and are also more responsive to perceptions of environmental stress and environmental reward, which is in agreement with findings from other ESM studies[18-20]. The origins of increased momentary mood reactivity in depression have been linked to genetic liability for

depression[20,21] as well as to childhood trauma[22], suggesting that ESM measures of reactivity can provide information on the mechanism whereby genes and environment confer vulnerability and resilience to depression. For example, one study found that in individuals with genetic liability for depression, the ability to generate positive emotions in response to the environment prevented the expression of genetic risk in low mood states[23].
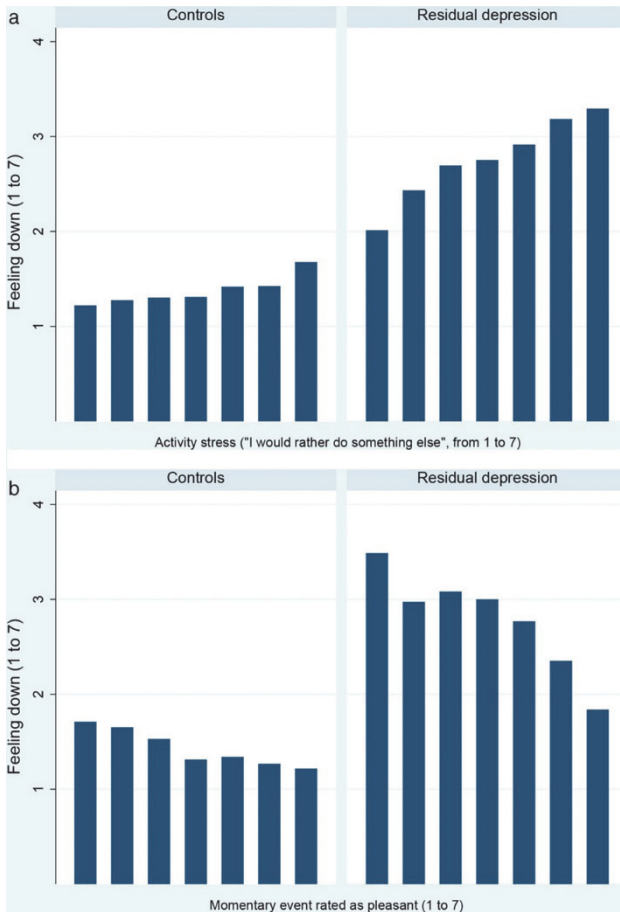
Figure 7

(a) Mean of self-reported ESM item "feeling down," collected over 6 days in daily life, summarized over increasing levels of activity-related stress in a sample of controls (N=212, n=10,142 beep moments) and a sample of patients with residual depression (N=129, n=6,425 beep moments).

(b) Mean of self-reported ESM item "feeling down," collected over 6 days in daily life, summarized over increasing levels of how pleasant momentary events were rated of in a sample of controls (N = 212, n = 9,777 beep moments) and a sample of patients with residual depression (N = 129, n = 6,250 beep moments).

Assessing ecological cognition

Contextual sensitivity is related to the cognitive ability to detect environmental changes. Environmental adaptation strategies or optimized coping assume that subjects are able to detect clues and use these in a learning process (in which PA and avoidance of stress function as rewards). There is a need to understand whether the cognitive skills are influenced by the changing environment, or by the way individuals assess these skills in daily life. It is therefore important to understand changes in cognition within the same time frame as changes in emotions and context. The assessment of variation in cognitive function in daily life is a challenge. The first reason is that the tests used in traditional cross-sectional assessments often rely on repeated trials lasting some minutes. This is difficult to implement in daily life given the time restraints of ESM—disruption should be restricted to 1–2 min maximally. Second, cognitive tests need contextual standardization to be reliable. The circumstances of the daily life assessments are highly variable and make it difficult to fully focus on the test. Third, ESM measurements are typically replicated 10 times each day for days or weeks. Such a sampling scheme requires a battery of parallel tests with limited training effects. More knowledge is required to design tests that can be used to assess different dimensions of cognitive functioning in daily life. Currently, pilots are being conducted to test the implementation of short cognitive tasks in ESM. These pilot studies are relate to memory (i.e., recall and recognition, both with and without intermediate tasks), concentration (i.e., coding tasks or elementary computations), reasoning (i.e., mathematical problem solving), and planning, in addition to tests that are related to cognitive biases and social cognition such as "jumping to conclusions" and "theory of mind".

ESM as a tool for positive emotion enhancement

Given ecological validity, low cost, and smartphone availability, there is increasing interest in using ESM in the context of trials[24]. For example, there is evidence that monotherapy with psychotropic medication may not constitute optimal care for common mental disorders. Instead, treatment combining medication with some form of behavioral intervention appears more effective than treatment with medication alone[25]. Given the fact that (i) behavioral interventions, complex or simple, appear to have similar effect sizes in the treatment of common mental disorder[26] and (ii) simple mHealth self-monitoring interventions can add to the treatment of mental disorders[27], there is interest in the concept of adding self-monitoring interventions to medication in the treatment of common mental disorder as a strategy to enhance the effect that would have been obtained with medication alone. A recent randomized trial added an ESM-based self-monitoring intervention, combined with feedback on how to improve patterns of generating PA in daily life, based on ESM self-monitoring data,

to medication in patients treated for depression. This type of feedback can consist simply of systematically examining instances of positive emotions in daily life, and discussing strategies to increase the frequency of such moments and/or increase the persistence of positive emotions in daily life. Patients receiving the combination of medication and the ESM-based self-monitoring and positive emotion enhancement tool did better than patients receiving medication alone[28]. These data suggest that ESM self-monitoring and positive emotion feedback paradigms may be added routinely to treatment with psychotropic medication in clinical practice. It has been suggested that antidepressant medication may work by facilitating the process of practice-induced plasticity[29]. ESM, as a context-sensitive tool to help individuals optimizing environmental interactions and coping[30], may be well placed to enhance the practice required for the plasticity-enhancing effect of antidepressants. Momentary assessment interventions may be used to enhance nonpharmacological interventions as well. For example, Depp, Ceglowski, Wang, Yaghouti, Mausbach, Thompson, Granholm[31] showed that adding an ecological momentary intervention targeting self-management of mood symptoms increased the effect of a brief psychoeducation on depressive symptoms. However, results also showed that the benefits only lasted as long as the mobile intervention was ongoing[31]. Batink, Bakker, Vaessen, Kasanova, Collip, van Os, Wichers, Germeys, Peeters[32] introduced an ESM-based mHealth paradigm as an add-on intervention to help patients practice with acceptance and commitment therapy (ACT) in their daily lives. The intervention proved practicable and feasible for patients, with high levels of adherence. Although no short-term effects could be demonstrated, the authors hypothesized that effects of ACT may take longer to transpire[32].

ESM as a tool for tapering and dose finding of psychotropic medication

Symptoms of anxiety, depression, and psychosis are sensitive to a range of pharmacological interventions. There are considerable challenges in finding the right dose for a given individual and there is widespread variation in the level of difficulties patients experience when coming off psychotropic medications. There is growing awareness that personalized dosing of psychotropic medication has been— and still is—a neglected area in clinical psychiatry, as is the personalized approach toward reducing or stopping psychotropic medication. ESM studies have shown that intensive sampling of emotions and context represents a sensitive tool to pick up the early effects of changing dose and reducing medications[9,22,33-35], as well as picking up early changes after dose reduction that may predict imminent relapse[13]. Given the potential of ESM to empower the patient to cocreate diagnosis and treatment evaluation, as well as the observed synergistic effects between ESM as a self-management tool and psychotropic medication, described earlier, a case could be made for

the routine use of ESM when initiating or reducing psychotropic medication. For example, when starting antidepressant medication, ESM in the first week can pick up signs of early response in the form of increases in the ability to use natural rewards (variation in positive emotions after positive events in daily life[22]); or reduction of antipsychotic medication can be monitored with ESM for alterations in subtle experiences of aberrant salience as a sign of dopamine supersensitivity syndrome[36].

In Figure 8, we show descriptive ESM data of early response to an antidepressant in a randomized, placebo-controlled trial of imipramine in a group of patients with depressive disorder described previously[22,37,38]. Only participants randomized to the imipramine group with valid ESM data and outcomes at 6 weeks are shown (n = 26). Response at 6 weeks was defined as a reduction of 50% or more in baseline Hamilton Depression Rating Scale (HAMD) score. Figure 8 shows the ESM momentary levels of PA, stratified by future responder status, during the first 3 days of imipramine treatment. Future responders have higher levels of PA (Figure 8). Thus, ESM trajectories may index information on future response in routine clinical practice.
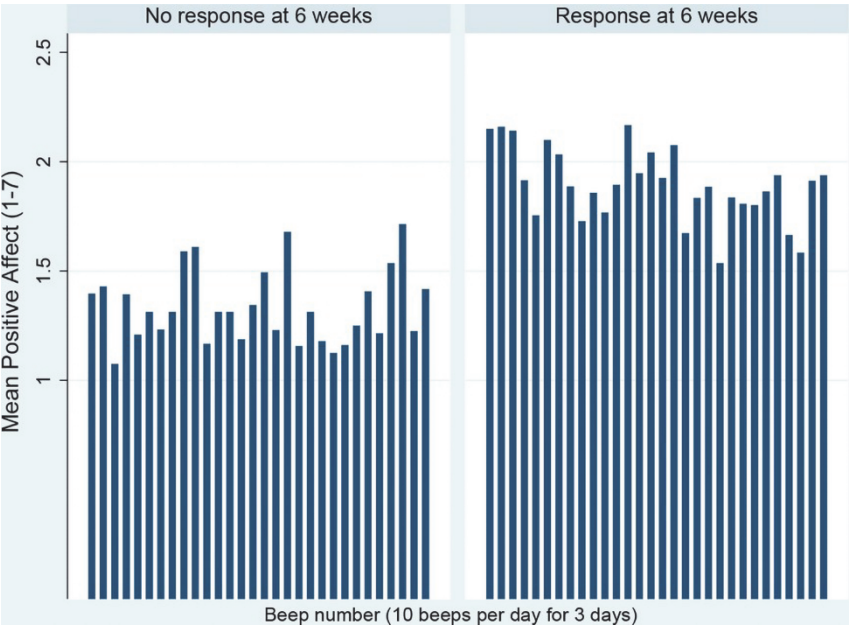
Figure 8

ESM momentary levels of positive affect (PA—the mean of eight momentary positive emotions) in 26 patients with depression during the first 3 days of imipramine treatment, stratified by responder status (50% reduction in HAMD) at 6 weeks (n = 26, 578 beep moments).

ESM as a tool for collaborative care in somatic settings

Experiential and behavioral outcomes are central in the diagnosis and treatment of a range of stress-related somatic syndromes, ranging from Alzheimer and Parkinson's disease to irritable bowel syndrome, pain, tinnitus, Chronic Obstructive Pulmonary Disease (COPD), diabetes, and urinary incontinence. ESM studies show that mood states and contextual factors interact with specific experiential outcomes in these disorders, and that ESM can facilitate management and collaborative care approaches in the hospital[39-46]. For example, deep brain stimulation is used in Parkinson's disease, however, it is often very difficult to assess subtle experiential, cognitive, and behavioral side effects associated with this treatment. A pilot study has shown that ESM can be used to assess both the mental and the motor symptoms in Parkinson's disease[39]; follow-up work is required to assess the effects of deep brain stimulation in this condition. Similarly, the effects of treatment on irritable bowel syndrome are very difficult to capture using conventional methods; it has been suggested that ESM may be used routinely in randomized trials in this disorder[44].

ESM as a form of routine outcome monitoring with a focus on positive mood states

There is consistent evidence that the impact of both pharmacological and nonpharmacological treatments is mediated by enhancing positive emotions, more than reduction of negative emotions, although interventions can be shown to also impact negative emotions and reactivity to negative events[47]. In a randomized controlled trial of antidepressants, improvement in the ability to generate positive emotions differentiated responders and nonresponders to antidepressants medication[21]. Similarly, in another investigation, it was shown that early improvement in ESM-based positive emotions was a better predictor of outcome than improvement in momentary negative emotions[48]. Trial evidence indicates that mindfulness training is effective for the treatment of depression and anxiety[49] and for preventing depression relapse[50]. In a randomized trial of mindfulness training in depression, the underlying emotional mechanism was investigated with ESM. The results revealed that, similar to antidepressant medication, clinical improvement was mediated by greater ability to generate momentary positive emotions, as well as longer momentary persistence of positive emotions[51]. Another study reported evidence that the effect of physical exercise on mood is also mediated by changes in positive emotions[52]. The importance of ESM measures of positive emotions was further shown by a study reporting that experienced PA, even over a single day, displayed a graded relationship with survival in elderly persons that was not caused by baseline health status or other covariates[53]. Thus, positive emotions appear to play a crucial role in facilitating the mobilization of resilience in the face of adversity, with important consequences for coping and health[54].

Given the importance of positive emotions, it has been proposed that ESM is routinely used in clinical practice, not only to improve diagnosis, empower patients, or enhance the effects of medication, but also as a means to collect routine outcome data that, contrary to symptom based measures, yield insight in daily life adaptation[37] and changes in resilience in the form of positive emotions. One study found that of the different measures in ESM that can be used as routine outcome measurement (ROM), changes in positive emotions were the most sensitive to change, suggesting this dimension represents a suitable measure for ROM in mental health services[38].

If ESM is used routinely in a specific mental health treatment setting, it is possible to generate, using anonymized patient data, group-based reference data of ESM patterns of emotions and behavior, as described in this article, against which ESM trajectories pertaining to individual patients can be interpreted, similar to the current analyses[4]. In the foreseeable future, machine learning techniques may be used to help test hypotheses in single patients against a large population base of ESM reference data[55].

Starting to use ESM in clinical practice
To start using ESM in clinical practice, one may commence with 5-day periods of ESM (eight beeps per day) at intake and after 6 weeks of treatment. At both baseline (diagnosis) and after 6 weeks (treatment response), clinician and patient can discuss the online feedback pattern of emotions and reactivity, as shown in the figure for the *PsyMate* app (Figure 9). In addition, the clinician can encourage the patient to engage in ESM over the entire 6-week period at a rate of, for example, 3–4days of ESM per week. ESM can then be combined with a focus on PA and a simple program aimed at enhancing frequency and persistence of PA in daily life. In addition, in PsyMate, aspects of acceptance and commitment and mindfulness exercises can be programmed in the time series[32], thus providing an opportunity to combine pharmacological or psychological treatment with mHealth ACT or mindfulness-based exercises.
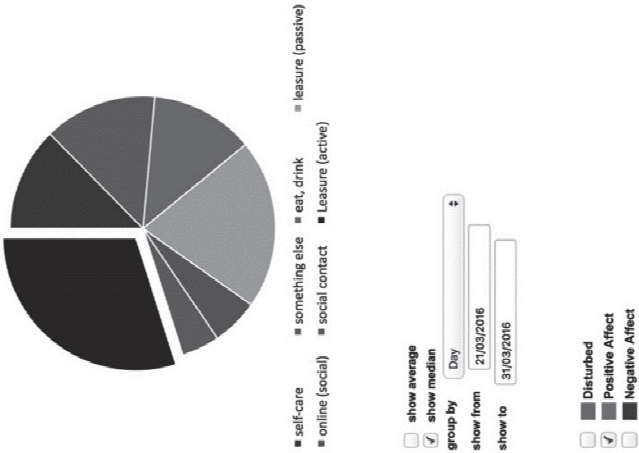
Figure 9

Online contextual feedback on a composite positive affect variable derived from momentary assessment. In the feedback graph, the solid line indicates the mean level of the composite variable "positive affect" (PA) per day. The dotted line is the level of PA if the level of PA during moments of "leisure" (pie chart on the right-hand side), as indicated during the momentary assessment period, is not taken into account—in this case resulting in mostly lower levels of PA.

Conclusion

ESM is of low cost and high impact. Given widespread availability of personal digital devices, intensive data collection based on ESM is now sufficiently practicable to allow widespread use in psychiatry and indeed medicine. ESM is empowering, providing co-ownership of the process of diagnosis, treatment evaluation, and patient-reported ROM. Blended care, based on a mix of face-to-face and ESM-based outside-the-office treatment, has the potential to reduce costs and improve outcomes. ESM makes it possible to develop insight into previous implicit patterns of thought, experience, and behavior, particularly if rapid personalized feedback is available. ESM-based self-monitoring and feedback can enhance resilience by strengthening the capacity to use natural rewards. Personalized trajectories of starting or stopping medication can be more easily initiated and predicted if sensitive feedback data are available in real time. In addition, personalized trajectories of symptoms, cognitive abilities, symptoms impacting on other symptoms, the mental capacity to "bounce back" from dynamic system disturbance and patterns of environmental reactivity yield uniquely personal data to support shared decision making and prognostication in clinical practice.

References

1. Huber M, Knottnerus JA, Green L, et al. How should we define health? *Bmj.* 2011;343:d4163.
2. Csikszentmihalyi M, Larson R, Prescott S. The ecology of adolescent activity and experience. *Journal of youth and adolescence.* 1977;6(3):281-294.
3. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical psychology review.* 2012;32(6):510-523.
4. Delespaul PA. *Assessing schizophrenia in daily life: The experience sampling method.* Maastricht university; 1995.
5. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological assessment.* 2009;21(4):463.
6. 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. *Psychological medicine.* 2009;39(9):1533-1547.
7. Palmier-Claus JE, Ainsworth J, Machin M, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. *BMC psychiatry.* 2012;12(1):172.
8. Walz LC, Nauta MH, aan het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: A systematic review. *Journal of anxiety disorders.* 2014;28(8):925-937.
9. Bak M, Drukker M, Hasmi L, van Os J. An n= 1 clinical network analysis of symptoms and treatment in psychosis. *PLoS One.* 2016;11(9):e0162811.
10. Wichers M, Simons C, Kramer I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatrica Scandinavica.* 2011;124(4):262-272.
11. Simons C, Hartmann J, Kramer I, et al. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. *European Psychiatry.* 2015;30(8):900-906.
12. Odgers CL, Mulvey EP, Skeem JL, Gardner W, Lidz CW, Schubert C. Capturing the ebb and flow of psychiatric symptoms with dynamical systems models. *American Journal of Psychiatry.* 2009;166(5):575-582.
13. Wichers M, Groot PC, Psychosystems E, Group E. Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and psychosomatics.* 2016;85(2):114-116.
14. Wigman JT, Collip D, Wichers M, et al. Altered transfer of momentary mental states (ATOMS) as the basic unit of psychosis liability in interaction with environment and emotions. *PLoS One.* 2013;8(2):e54653.
15. Höhn P, Menne-Lothmann C, Peeters F, et al. Moment-to-moment transfer of positive emotions in daily life predicts future course of depression in both general population and patient samples. *PLoS One.* 2013;8(9):e75655.
16. Borsboom D, Cramer AO. Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology.* 2013;9:91-121.
17. Bylsma LM, Taylor-Clift A, Rottenberg J. Emotional reactivity to daily events in major and minor depression. *Journal of abnormal psychology.* 2011;120(1):155.
18. Audrain-McGovern J, Wileyto EP, Ashare R, Cuevas J, Strasser AA. Reward and affective regulation in depression-prone smokers. *Biological psychiatry.* 2014;76(9):689-697.
19. Myin-Germeys I, Peeters F, Havermans R, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatrica Scandinavica.* 2003;107(2):124-131.
20. Peeters F, Nicolson NA, Berkhof J, Delespaul P, deVries M. Effects of daily events on mood states in major depressive disorder. *Journal of abnormal psychology.* 2003;112(2):203.
21. Wichers M, Schrijvers D, Geschwind N, et al. Mechanisms of gene–environment interactions in depression: evidence that genes potentiate multiple sources of adversity. *Psychological medicine.* 2009;39(7):1077-1086.
22. Wichers M, Barge-Schaapveld D, Nicolson N, et al. Reduced stress-sensitivity or increased reward experience: the psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology.* 2009;34(4):923.
23. Wichers M, Myin-Germeys I, Jacobs N, et al. Genetic risk of depression and stress-induced negative affect in daily life. *The British Journal of Psychiatry.* 2007;191(3):218-223.
24. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evidence-based mental health.* 2016;19(3):86-89.
25. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds III CF. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. *Focus.* 2014;12(3):347-358.

26. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological medicine.* 2010;40(12):1943-1957.

27. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. *Journal of medical Internet research.* 2016;18(6):e135.

28. Kramer I, Simons CJ, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry.* 2014;13(1):68-77.

29. Sterling P. Homeostasis vs allostasis: implications for brain function and mental disorders. *JAMA psychiatry.* 2014;71(10):1192-1193.

30. Lardinois M, Myin-Germeys I, Bak M, Mengelers R, Van Os J, Delespaul PA. The dynamics of symptomatic and non-symptomatic coping with psychotic symptoms in the flow of daily life. *Acta Psychiatrica Scandinavica.* 2007;116(1):71-75.

31. Depp CA, Ceglowski J, Wang VC, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. *Journal of affective disorders.* 2015;174:23-30.

32. Batink T, Bakker J, Vaessen T, et al. Acceptance and commitment therapy in daily life training: A feasibility study of an mHealth intervention. *JMIR mHealth and uHealth.* 2016;4(3):e103.

33. Barge-Schaapveld DQ, Nicolson NA. Effects of antidepressant treatment on the quality of daily life: an experience sampling study. *The Journal of clinical psychiatry.* 2002.

34. Lataster J, Myin-Germeys I, Wichers M, Delespaul PA, van Os J, Bak M. Psychotic exacerbation and emotional dampening in the daily life of patients with schizophrenia switched to aripiprazole therapy: a collection of standardized case reports. *Therapeutic advances in psychopharmacology.* 2011;1(5):145-151.

35. Lataster J, Thewissen V, Bak M, et al. Emotional experience and estimates of D2 receptor occupancy in psychotic patients treated with haloperidol, risperidone, or olanzapine: an experience sampling study. *The Journal of clinical psychiatry.* 2011;72(10):1397-1404.

36. Murray RM, Quattrone D, Natesan S, et al. Should psychiatrists be more cautious about the long-term prophylactic use of antipsychotics? *The British Journal of Psychiatry.* 2016;209(5):361-365.

37. Barge-Schaapveld DQ, Nicolson NA, van der Hoop RG, DeVries MW. Changes in daily life experience associated with clinical improvement in depression. *Journal of Affective Disorders.* 1995;34(2):139-154.

38. van Os J, Delespaul P, Barge D, Bakker RP. Testing an mHealth momentary assessment Routine Outcome Monitoring application: a focus on restoration of daily life positive mood states. *PLoS One.* 2014;9(12):e115254.

39. Broen MP, Marsman VA, Kuijf ML, Van Oostenbrugge RJ, van Os J, Leentjens AF. Unraveling the relationship between motor symptoms, affective states and contextual factors in Parkinson's disease: A feasibility study of the experience sampling method. *PloS one.* 2016;11(3):e0151195.

40. Bruehl S, Liu X, Burns JW, Chont M, Jamison RN. Associations between daily chronic pain intensity, daily anger expression, and trait anger expressiveness: An ecological momentary assessment study. *PAIN®.* 2012;153(12):2352-2358.

41. Fischer S, Doerr JM, Strahler J, Mewes R, Thieme K, Nater UM. Stress exacerbates pain in the everyday lives of women with fibromyalgia syndrome—the role of cortisol and alpha-amylase. *Psychoneuroendocrinology.* 2016;63:68-77.

42. Leue C, Kruimel J, Vrijens D, Masclee A, Van Os J, Van Koeveringe G. Functional urological disorders: a sensitized defence response in the bladder–gut–brain axis. *Nature Reviews Urology.* 2017;14(3):153.

43. Mujagic Z, Leue C, Vork L, et al. IBS and over-reporting of abdominal pain in retrospective questionnaires: advantages of experience sampling method as new digital tool in symptom measurement: OP004. *Neurogastroenterology & Motility.* 2014;26:2-3.

44. Mujagic Z, Leue C, Vork L, et al. The Experience Sampling Method-a new digital tool for momentary symptom assessment in IBS: an exploratory study. *Neurogastroenterology & Motility.* 2015;27(9):1295-1302.

45. Van Knippenberg R, De Vugt M, Ponds R, Myin-Germeys I, van Twillert B, Verhey F. Dealing with daily challenges in dementia (deal-id study): an experience sampling study to assess caregiver functioning in the flow of daily life. *International journal of geriatric psychiatry.* 2017;32(9):949-958.

46. Vrijens D, Drossaerts J, van Koeveringe G, Van Kerrebroeck P, van Os J, Leue C. Affective symptoms and the overactive bladder—a systematic review. *Journal of psychosomatic research.* 2015;78(2):95-108.

47. Silk JS, Tan PZ, Ladouceur CD, et al. A randomized clinical trial comparing individual cognitive behavioral therapy and child-centered therapy for child anxiety disorders. *Journal of Clinical Child & Adolescent Psychology.* 2018;47(4):542-554.

48. Geschwind N, Nicolson NA, Peeters F, van Os J, Barge-Schaapveld D, Wichers M. Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. *European Neuropsychopharmacology.* 2011;21(3):241-247.

49. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of consulting and clinical psychology.* 2010;78(2):169.

50. Kuyken W, Hayes R, Barrett B, et al. Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): a randomised controlled trial. *The Lancet.* 2015;386(9988):63-73.

51. Geschwind N, Peeters F, Drukker M, van Os J, Wichers M. Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: a randomized controlled trial. *Journal of consulting and clinical psychology.* 2011;79(5):618.

52. Wichers M, Peeters F, Rutten BP, et al. A time-lagged momentary assessment study on daily life physical activity and affect. *Health Psychology.* 2012;31(2):135.

53. Steptoe A, Wardle J. Positive affect measured using ecological momentary assessment and survival in older men and women. *Proceedings of the National Academy of Sciences.* 2011;108(45):18244-18248.

54. Tugade MM, Fredrickson BL, Feldman Barrett L. Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of personality.* 2004;72(6):1161-1190.

55. Strauss J, Peguero AM, Hirst G. Machine learning methods for clinical forms analysis in mental health. *MedInfo.* 2013;192:1024.

4

Demonstrating the reliability of transdiagnostic mHealth routine outcome monitoring in mental health services using experience sampling technology

S. Verhagen, J. Berben, C. Leue, A. Marsman, P. Delespaul, J. van Os, R. Lousberg

Abstract

Background: Routine Outcome Monitoring (ROM) should provide a dynamic, within-treatment forward feedback loop to guide individual treatment decisions across diagnostic categories. It has been suggested that the Experience Sampling Method (ESM), capturing the film of daily life adaptive processes, offers a flexible, personalised and transdiagnostic feedback system for monitoring and adapting treatment strategies. This is the first study that uses an ESM application (the PsyMate™) as a routine mobile-ROM (mROM) tool in an ambulatory mental health setting.

Objective: To demonstrate adequate psychometric properties of the PsyMate™ app assessing both symptom severity levels as well as daily life functioning.

Method: In a transdiagnostic sample of 64 outpatients, an mROM protocol (ESM for 6 days, at 10 semi-random moments a day) and a standard ROM instrument (HADS) were administered at baseline and at three-month follow-up. We measured positive affect (PA), negative affect (NA), quality of sleep, positive social interaction, activity-related stress, tiredness, and feeling unwell.

Results: Subjects completed 53% of the measurements at baseline (N = 64) and 48% at follow-up (N = 29). Factor analysis and subsequent reliability analysis of PA and NA confirmed the two constructs. Significant and meaningful correlations were found between PA, NA and HADS scores (ranging from r =.4 to r =.7). Multilevel analyses yielded significant change scores for all measures.

Conclusion: The ESM-based, transdiagnostic mROM tool can be used reliably in clinical settings: it shows adequate psychometric properties, as well as concurrent validity and sensitivity to change over time with respect to relevant ROM constructs. Person-tailored items can be added. In addition, mROM offers added value over standard symptom-based ROM, as it provides information on adaptive functioning in the daily environment of patients.

Introduction

Routine Outcome Monitoring (ROM) in mental health care refers to the process of collecting clinical data, which can serve as feedback to guide informed treatment decisions[1]. ROM should be transdiagnostic, flexible and follow the flow of the clinical process. Furthermore, it must be adjusted to the needs and wishes of individual patients[2]. Given these requirements, there is considerable debate as to what degree clinical ROM measures in mental health care can serve as input to benchmark outcomes across clinicians and institutions – a process referred to as Benchmark-ROM[3]. Benchmark-ROM requires rigid standardization, both in terms of instruments and data collection, to allow cross-patient and cross-centre comparisons. The implementation of rigid benchmark-ROM negatively influences clinical ROM, which requires a more flexible approach[4-6]. Clinical ROM requires repeated assessments at specific time points over successive diagnostic and treatment phases. Although there is no agreed measure for ROM in mental health care, many ROM scales have been developed for use in clinical practice[7-9]. In the literature, ROM is often referred to as patient-reported outcome measures (PROM), stressing the need for unbiased assessments of outcome collected by the patient rather than the treating clinician[10,11]. ROM instruments should be sensitive to change over time. Selected scales can be generic and, therefore, relevant across populations, or, in contrast, target specific patient groups. Moreover, ROM instruments should map several outcome domains, including psychopathology, daily life functioning, personal recovery, and quality of life[3,12]. Unfortunately, current ROM implementations mainly focus on psychopathology[13], disregarding the other outcome domains. In addition, measures of psychopathology, expressed as total scores from traditional technical scales, do not reflect outcomes that are important to patients[14,15]. Given that ROM is part of clinical routine, instruments should be time efficient. Furthermore, ROM should be set up in such a way that it allows the patient to become the empowered co-owner of the process of diagnosis and treatment. Therefore, ROM should be constructed as an informed forward feedback loop: the iterative process that includes clinical decision-making, outcome evaluation, treatment adjustment, and further outcome evaluation.

A wide variety of ROM instruments exists, often developed to assess psychopathology in specific patient populations (e.g., mood disorders and psychosis) using traditional rating scales that lack relevance to the daily life of patients. Comprehensive assessment batteries that incorporate all different outcome domains are cumbersome and time consuming. Inclusion of clinicians (or trained personnel) as assessors dramatically adds to the complexity and cost of data collection logistics and potentially introduces bias. In many cases, a trade-off has to be made between brevity and completeness[7]. Finally, ROM assessments often rely

on retrospective assessments, which are vulnerable to recall bias and limit sensitive assessment of change[16]. Most individuals are poor assessors when asked retrospectively to assess emotional experiences[17]. The most essential problem with traditional ROM assessments is that they only marginally inform adequate treatment selection as they do not inform on personal goals in real life settings and are not set up in such a way that a constructive forward feedback loop arises. For customized treatment selection, symptom variability and contextual variation is relevant. Mental health problems and related disabilities occur in the flow of daily life. Effective treatment should address the issue that patients are most vulnerable and consider the coping strategies available within their daily routines. Traditional rating scales, completed in a clinical setting, are often not representative of daily life functional adaption. A successful recovery process assumes functional adaptation and increasing resilience against mental illness. Resilience factors, such as experience of positive mood states and the strength of social connections, are increasingly being recognized and used in mental health treatment[18].

Thus, ROM measurements should target contextual factors and positive mood states in addition to negative mood states and symptomatology, rather than being exclusively focused on negative symptomatic states. Experience sampling techniques take into account all of these factors. The *Experience Sampling Method* (ESM), also referred to as *Ecological Momentary Assessment* (EMA), is an umbrella term for several ambulatory assessment strategies that randomly sample mental state in the context, following the flow of daily life. ESM is a structured diary technique, in which patients are asked to complete short questionnaires in response to auditory cues (beep signals), at semi-random moments during the day, for a number of consecutive days. Patient assessments include momentary emotional experiences (e.g., positive and negative mood), symptoms, cognition, context information (e.g., activity, company and location), and an appraisal of the context. Because ESM is a repeated assessment strategy, additional constructs, such as stress-sensitivity or coping strategies, can be quantified[18]. ESM was proposed as a comprehensive, generic ROM instrument and has several advantages over typical ROM assessments[3,18,19]. First, ESM increases ecological validity, because reports are provided in the patient's natural environment. Second, memory bias is minimized because patients report in the moment. Third, ESM is contextualized, allowing for a better understanding of the person-environment interactions that give rise to psychopathology. Being aware of environmental reactivity and pattern changes over time will aid the understanding of treatment progress[18,20]. Fourth, ESM is highly relevant to the individual and is person-tailored when used as a clinical tool, thus creating transparency and facilitating patient engagement when discussing the data with the

clinician. Last, ESM facilitates and enhances shared decision making processes, because patients are actively participating in the data collection, and data can be used to improve treatment efficiency. Study results show that ESM during treatment can enhance feelings of empowerment and the ability for self-management[21]. Across the psychopathology spectrum, ESM has been proven feasible and can be successfully applied, in combination with feedback, as a treatment method for depression[22]. Moreover, ESM has been found valuable for a range of uses, including ROM in clinical practice[23].

Van Os and colleagues studied the use of ESM as a mobile ROM (mROM) tool. They assumed that ESM data was sensitive to change over time, using several clinically relevant parameters. Observational analyses were performed using data from the treatment arm of a randomized control trial in patients diagnosed with major depressive disorder[24-26]. The results confirmed the assumption, with the greatest sensitivity to change found for measures of positive adjustment, positive affect, and increases of positive affect due to natural rewards in daily life. These results show that resilience factors, such as positive affect, are informative for outcome assessment. Positive mood states are relevant for treatment outcomes across mental disorders[27,28] and can be seen as a generic trans-diagnostic indicator of underlying resilience that can be captured in ROM. Research showed that positive affect and negative affect are related but different dimensions, i.e., they are not the extremes of a single continuum. Positive affect is non-heritable and an indicator of resilience, whereas negative affect is highly heritable and an indicator of vulnerability[29]. As both dimensions are important in clinical context, both should be captured in ROM. With the advent of modern mHealth applications for use on mobile devices, such as smartphones, ESM can now be carried out cheaply and routinely[23]. We refer to mHealth ESM for the purpose of ROM as *mROM*. There are multiple ESM mobile applications available (e.g., https://www.lifedatacorp.com/, http://experiencesampler.com/, https://pielsurvey.org/). Here, we present the first study that uses a mobile device to implement mROM in an ambulatory mental health setting, using an open trial design with PsyMate™, one of these ESM mobile applications (www.psymate.eu).

Objective
The aims of this study were twofold. First, to demonstrate transdiagnostically adequate psychometric properties of routine mROM with a PsyMate™ moderated mROM application and second, to investigate sensitivity to change during treatment. Based on previous work, it was expected that (i) the psychometric properties of the PsyMate™ method would be adequate and suitable for therapeutic monitoring; and (ii) change over time could be demonstrated for a series of variables (e.g., mood, activity-related stress, and sleep quality)

indexing both psychopathology and daily life functioning. In this regard, we expected to find a weak to moderate association between a traditional rating scale (the Hospital Anxiety and Depression Scale[30]) and the mROM mood measures, leaving sufficient room for the added value of mROM.

Methods

Sample

Patients were recruited at the outpatient mental health service of the Maastricht University Medical Centre (MUMC+). In total, 115 consecutively attending patients were asked to participate in routine mROM, of whom 75 provided their consent. All patients were 18 years or older and capacity for consent was established by the psychiatrist providing the care to the patient. Exclusion criteria for inclusion in the analysis were not being able to read Dutch or not being able to handle a mobile device with the PsyMate™ app.

Procedures

Analysis Design

mROM with the PsyMate™ app is applied routinely in MUMC+. There were two measurement periods, one at baseline (t = 0), the second at follow-up approximately three months later (t = 1). Due to logistical reasons, the second visit did not always take place exactly 3 months later (M = 111.6 days later, SD = 27.3, range 80-189 days). Each measurement period consisted of six consecutive ESM sampling days and the administration of a traditional ROM questionnaire at the beginning of the ESM sampling days.

Briefing

During the baseline session, patients were helped while downloading the Psymate™ application on their smartphone via the App Store or Google play store. A specific code was required to enter the study protocol. However, a free demo app is available in different languages, for those wishing to implement the app in routine clinical practice. In case a patient could not use his own device, an iPod was provided for the duration of the mROM period. A clinician explained the PsyMate™ procedure during a briefing session and coached patients through the PsyMate™ items. Patients were instructed to continue their normal routine during the Psymate™ data collection period. After the briefing, a sampling period of six consecutive ESM days started, not including the briefing day on which the PsyMate™ protocol was activated.

The standing medical ethical committee approved the anonymous use of routine clinical data, if patients provided informed consent. Thus, all patients provided informed consent to use routine clinical data for the purpose of scientific investigation.

mROM

For the purpose of mROM, the Psymate™ application was programmed to emit 10 random beeps each day. Signals notify when a short questionnaire has to be completed (lasting approximately 1 minute). Beeps were semi-randomised in ten blocks of 90 minutes, between 7:30 AM and 10:30 PM. The questionnaire consisted of 13 mood items, 5 context items (what, where, and with whom the patient was spending time and whether they were enjoying it), 5 items about important events happening since the last beep, 1 item questioning specific somatic complaints, 5 items assessing patients' physical condition, and 1 item assessing levels of beep disturbance. In addition to the beep questionnaire, patients filled in some extra questions at the beginning and the end of each day. The morning questions were related to the quality of sleep of the night before, while the questions in the evening required the patient to give an estimate of their average mood and somatic complaints over the past day. Most items were presented on a 7-point Likert scale, ranging from 1 (not at all) to 7 (very).

In order to assess change over time, items were clustered to form measurable constructs. Previous work showed positive and negative mood items reliably form a Positive Affect (PA) and a Negative Affect (NA) cluster[31]. In this analysis, PA consisted of the items: I feel cheerful, satisfied, relaxed and globally feeling well. The remaining nine mood items (I feel lonely, guilty, worried, down, threatened, insecure, irritated, frightened, and suspicious) formed the NA scale. The quality of sleep was assessed using the (ordinal-coded) items: (1) time needed to fall asleep, (2) number of times the person woke up during the night, (3) the time lying awake before getting up, (4) whether the person felt rested, and (5) how the person globally felt about their sleep quality last night. Approximately three months after the initial visit, patients were asked to participate again in the second part of the PsyMate™ mROM procedure.

Debriefing

After each sampling period, a debriefing session was scheduled. Patients were asked whether the past week was representative for their daily life and whether the Psymate™ interfered with their thoughts, feelings, activities, or social contacts. Furthermore, patients were asked whether unusual incidents occurred, and to what degree the use of Psymate™ was bothersome.

HADS

The Hospital Anxiety and Depression Scale (HADS)[30] is a reliable and validated rating scale[32,33] which is often used for mental health ROM purposes worldwide, particularly in somatic hospital settings[34]. The HADS consists of fourteen items (scaled from 0 to 3) assessing aspects of anxiety and depression experienced during the last week. The HADS was administered twice digitally (via the Psymate™) at the end of each briefing session.

Statistical analyses

Analyses were performed to investigate the structure of the 13 mood items. First, a principal component analysis (PCA) was conducted on these items with orthogonal rotation (varimax), requiring two factors to be extracted. Second, a reliability analysis (i.e., Cronbach's alpha) was performed on each factor to determine the internal consistency of each scale.

The data collected with ESM have a multilevel structure - successive beeps (level 1) are nested within patients (level 2). Multilevel regression analyses took the variability of both levels into account. With respect to the analyses of change, there were 7 *a priori* dependent variables: NA, PA, quality of sleep, preferring to do something else (activity-related stress), enjoying company (social stress), feeling unwell, and feeling tired. The following items were incorporated as covariates in the multilevel models: age, sex, a dichotomous variable indicating whether a patient participated in one or both measurement periods, and DSM-IV-TR diagnosis prior to participation (i.e., a dichotomous variable indicating whether or not a patient had been previously diagnosed with a mental disorder (e.g., depressive disorder, panic disorder, somatoform disorder, no diagnosis, etc.)). Additionally, when examining the variable *feeling unwell*, a dichotomous variable describing the presence or absence of somatic complaints was taken into account and added as a covariate in the statistical model. The predictor variable of main interest was measurement period, indicating a possible change over time.

Analyses were carried out using SPSS Statistics version 23.0. Two-sided p-values < .05 were considered significant. All dependent variables were checked for approximately normal distribution before performing further analyses.

Results

Sample

In the period from February 2015 until May 2016, 75 patients at the MUMC outpatient mental health service fulfilled criteria for inclusion in the analysis. In case a patient completed less than ten beeps during a measurement period (either baseline or follow-up), the measurement period was excluded from further analysis. There were 11 patients who did not complete sufficient beeps during both measurement periods, leading to a final dataset of 64 patients. Thirty-six patients completed one measurement period, 28 patients completed both measurement periods. A logistic regression model examined whether these two groups differed with respect to age, sex and being diagnosed (yes or no). This was not the case for either the omnibus model ($-2 LL = 78.9$; $df = 3$; $p = .18$) or for the univariate tests (all p-values > .075).

The average age of the $N = 64$ analysable group was 48.7 years ($SD = 13.9$, range 18-73). There were 39 women and 25 men. All but 11 patients received a clinical diagnosis by the treating psychiatrist, based on DSM-IV-TR criteria. The 11 non-diagnosed patients did experience a degree of mental discomfort and were therefore included in the analysis. Diagnoses included depression ($N = 25$), panic disorder ($N = 18$), somatoform disorder ($N = 4$), posttraumatic stress disorder ($N = 2$), bipolar disorder ($N = 1$), anxiety disorder ($N = 1$), dysthymia ($N = 1$) and psychotic disorder ($N = 1$).

App statistics

The complete protocol presented 120 beeps per patient (6 days x 10 beeps x 2 measurement periods). At baseline, the response percentage was 52.7% of 3780 presented beeps, comparable to the 48.2% of 1740 presented beeps at follow-up.

In October 2015, an update of the PsyMate™ app was introduced. Since then, the number of errors concerning the Internet connection considerably reduced. The proportion of valid beeps with no missing items within the beep increased substantially after this update: from 71.1% to 86.2%.

Compliance

There was no suggestion of relevant differences in response rates between the days of the week at either baseline or three-month follow-up. However, the data suggest that there is a difference in response rate over consecutive days of the ROM protocol (Figure 1). During the first day and last day of the protocol, patients tend to complete less beeps than on the remaining days. At both baseline and follow-up, a reduction in response rate over time was

apparent. At follow-up, the reduction started two days earlier (day 3), compared to baseline. Regarding the hours of the day, there was a slight increase in response over time, as illustrated in Figure 2. The lowest response was in the early hours of the day (7:30 am – 10:30 am).

Factor structure of PA and NA

The principal component analysis of the 13 PA and NA items resulted in two factors, each having a value larger than 1 (7.1 and 1.5 respectively), with a cumulative percentage of explained variance of 66.1%. An item was assigned to a factor when the factor loading exceeded .45 with a difference larger than .1 for both factor loadings. Using these criteria, all items could be assigned unambiguously to either the PA or the NA factor. As a next step, a reliability analysis was performed. Cronbach's alphas for factor 1 and 2 (NA and PA) were .91 and .92 respectively.

Consequently, the PA subscale was constructed by computing the sum of the 4 positive affect items, the NA scale by computing the sum of the 9 negative affect items. Sub-scores containing missing items were computed by inserting the mean value of the available items. A maximum of two missing items for PA and 4 items for NA was allowed. PA and NA could not be computed for 41 beeps, leaving an analysable number of beeps (records) of 2832.
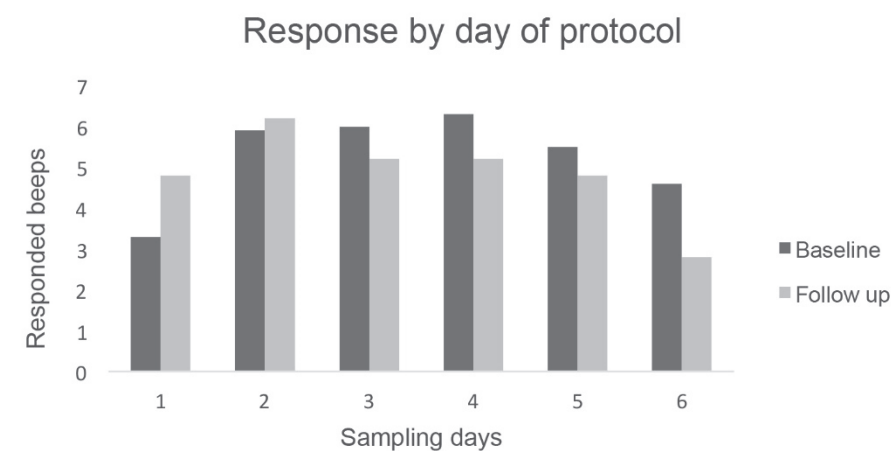
Figure 1

This figure shows the average number of beeps that were completed per day per subject, both for the baseline assessment and the three-month follow-up assessment, over the 6-day ESM period.

Validation analyses

A correlation matrix with the patient-level means of the PA and NA scores (beep-level data) and the patient-level means of HADS total, HADS anxiety, and HADS depression was computed (Table 1). The results confirmed the a priori hypothesized relationships between these five variables. Next, an analysis was performed to investigate whether a change in PA and NA over time correlated with a change on the HADS total score. Significant correlations between change scores (baseline versus 3 months) were found: ΔPA with ΔHADS -0.634 p = 0.002, ΔNA with ΔHADS 0.603 p = 0.004.
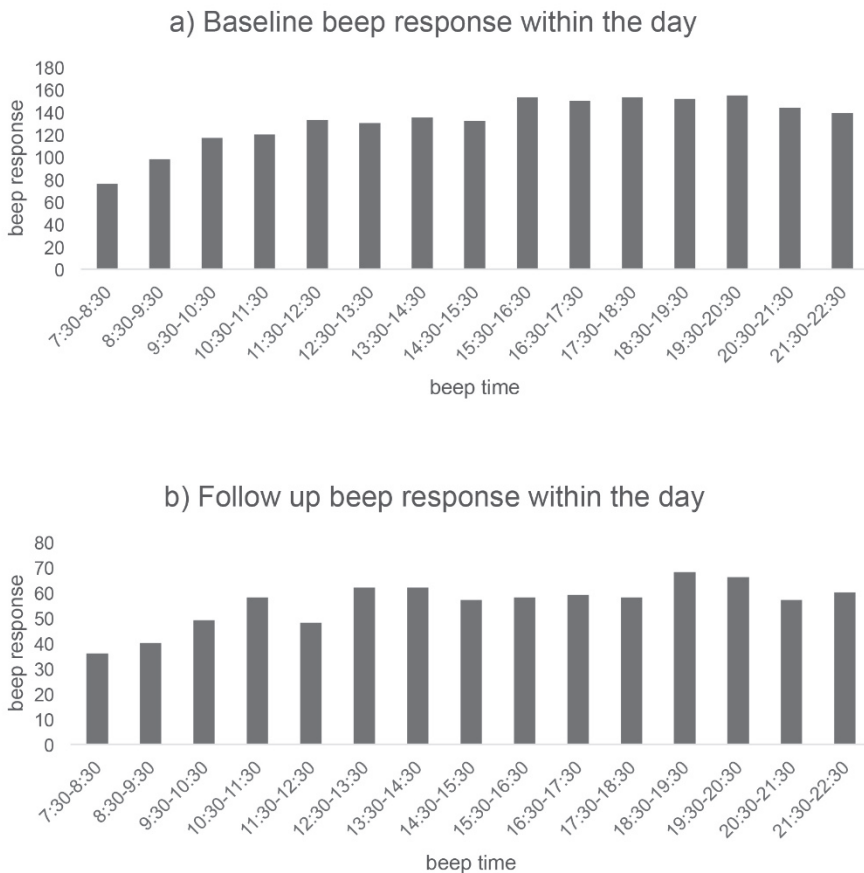
Figure 2

This figure shows the average number of responded beeps, per time window, within a day.

(A) Number of beep responses within the day, derived from baseline ESM data.

(B) Number of beep responses within the day, derived from the three-month follow-up ESM data.

Table 1

Pearson correlates between positive affect scores, negative affect scores and Hamilton anxiety and depression scale scores

Variables	1	2	3	4	5
1. NA mean	-				
2. PA mean	-.71***	-			
3. HADS anxiety	.71***	-.51***	-		
4. HADS depression	.56***	-.45***	.66***	-	
5. HADS total	.66***	-.51***	.89***	.92***	-

Note. Correlations between 1 and 2 were performed over *N* = 63. All others were performed over *N* = 59. NA = Negative affect, PA = Positive affect, HADS = Hamilton Anxiety and Depression Scale. NA and PA represent mean scores over all subjects for one measurement period. HADS anxiety, HADS depression, and HADS total represent subject scores for one measurement period. *p < .05. **p < .01. ***p < .001.

Sensitivity to change

Analyses were performed with multilevel regression to examine whether the positive change that was *a priori* expected could be demonstrated (Table 2). Each model used the same set of predictor variables (i.e., a dichotomous variable for the assessment moment, and age, sex, being diagnosed yes or no, and participation in both measurement periods as covariates). As reported in table 2, the mROM Psymate™ was able to detect significant changes over time in all the variables.

Table 2

Multilevel regression model estimates for the effects of measurement period on several variables

Dependent Variables	B (*SE)*	t-values	p-values
Positive affect	.65 (.19)	3.45	.002
Negative affect	-.37 (.14)	-2.61	.015
Quality of sleep	1.89 (.85)	2.23	.033
Positive social interaction	.44 (.15)	3.69	.001
Activity-related stress	-.37 (.18)	-2.10	.046
Feeling tired	-.54 (.18)	-2.94	.007
Feeling unwell	-.67 (.21)	-3.17	.003

Note. The analyses are based on 2874 beeps nested within 64 persons. SE = standard error.

Discussion

This study examined the applicability of a routine mobile-ROM tool within an ambulatory mental health setting. The PsyMate™ application was used to implement experience sampling methodology (ESM). The first aim was to demonstrate adequate psychometric properties of mROM, by investigating the reliability and validity of the mROM application in PsyMate™. The second aim was to investigate sensitivity to treatment change using PsyMate™ data.

Results demonstrated adequate psychometric properties of the PsyMate™ app when used with the present mROM protocol. The reliability of the method was examined by exploring user characteristics of the PsyMate™. Compliance rates during both the baseline and follow-up periods (a period consisted of 60 possible beeps per person; 10 per day, for 6 days), were around 50 percent, in terms of beeps completed. Although ESM compliance may seem lower compared to previous ESM findings, which showed compliance rates around 80 percent[35], sufficient beeps were completed for reliable data analyses. There are a number of possible reasons for the reduced compliance. Even conscientious patients miss some beeps due to daily life demands. Furthermore, the compliance rates found in earlier studies were based on traditional sampling techniques, using booklets combined with wristwatches or Personal Digital Assistants. Therefore, a direct comparison to the PsyMate™ app is misleading. In addition, there were technical issues with the first release of the app, leading to data loss when the Wi-Fi communication was unstable. After a system update, these problems were reduced. Hence, the compliance with the PsyMate™ app is considered acceptable. Different strategies may be used to improve compliance. The briefing session is important. It should create transparency about the reasons to collect mROM data: to optimize treatment, participate in clinical research and/or to comply with administrative requirements. The briefing session is the start of building an alliance in which clinicians try to understand needs arising in the context of daily life. It helps to discuss ESM sampling situations that shed light on aspects of resilience and vulnerability (with the possibility to customize ESM-questionnaires if necessary). Elements of 'gamification' can be included to motivate patients to fill in the ESM sampling sheets. A crucial factor is the personalized feedback on daily life strengths and vulnerabilities. The PsyMate™ includes an automated web-based ESM feedback that clinicians and patients can consult. The option should be better advertised because it makes data collection more relevant to individual patients. Looking at the response behaviour at week level, few differences were found across the days of the week, although the first and last day of the ESM protocol were characterised by slightly fewer responses. During the baseline assessment, the first day showed the lowest response rate,

whereas in the follow-up period, the last day had the lowest response rate. A lack of familiarity with the ESM routine is arguably the reason for the low response rate on the first day, while the anticipation of completing the ESM period might account for the low numbers on the final day[36]. Considering data at day level, fewer beeps were completed in the morning hours. Again, this is to be expected, since patients were instructed to follow their own daily pattern and sometimes were still asleep at the early pre-programmed moments. When considering attrition rates, only 28 patients completed both measurement periods, or 37% of the 75 patients initially included. Some patients were excluded because there was not enough data and others did not complete the follow-up measurement. The loss of patients at follow-up could indicate that the method is too burdensome or not experienced as relevant. This seems unlikely, given that previous research has shown that the method was feasible in patients with a wide variety of mental disorders[23] and current findings indicate feasibility within a single measurement period. Furthermore, a substantial proportion was already discharged from the mental health service by the time of the three month follow-up, as can be expected within the dynamics of a hospital outpatient mental health setting. After three months, patients that were seen in the context a diagnostic assessment would already have been referred back. Other patients would have discontinued treatment in the high-attrition risk context of referral from a somatic department to a mental health setting. To a degree, limited follow-up may reflect the natural flow of a general hospital mental health care setting. Future research should further examine this issue. From a clinical perspective, mROM is most relevant at the beginning and during treatment, when data can be used to customize interventions. Nearer the end of treatment, patients motivation may naturally decrease as they start to leave the episode of mental distress behind[37].

The constructs Positive Affect (PA) and Negative Affect (NA) were selected based on the existing ESM literature[31]. Factor analysis of these items confirmed the structure of the two concepts and the subsequent reliability analyses yielded excellent internal consistency coefficients. To assess the concurrent validity of mROM, comparisons where made between the traditional ROM questionnaire used in the hospital (HADS; assessing anxiety and depression) and the PsyMate™ moderated mROM protocol. Overall, significant and clinically relevant correlations were found between the PA and NA constructs on the one hand, and the HADS (total, anxiety, and depression) scores on the other. Although the measures show substantial overlap, the ESM based PsyMate™ moderated mROM protocol offers a unique contribution to clinical routine outcome assessment given that recall bias and contextual biases are controlled for to a great extent. In addition, the ecological measures within the mROM protocol inform on daily life adaptive functioning. Seven measures (PA, NA, positive

social interactions, quality of sleep, activity-related stress, tiredness, and feeling unwell), considered relevant in mental health care, were assessed and compared over baseline and follow-up. All measures changed significantly over time. Thus, the mROM measurements are sensitive to change over time. At follow-up, scores on NA, activity-related stress, tiredness and feeling unwell were reduced, whereas scores on PA, positive social interactions, and quality of sleep were increased. This overall sensitivity to change is in line with previous ESM research[2,21].

The study sample was heterogeneous, although mood and anxiety disorders were most prevalent in this general hospital setting. Despite a degree of heterogeneity, the same Psymate™-moderated mROM protocol was used transdiagnostically for all patients. The ESM questionnaires used in different mental health target populations typically share 80% of the items, in order to generically assess contextualized symptom variability and well-being in daily life[20]. The mROM tool shows initial usefulness as a transdiagnostic instrument. However, the small sample size and limited heterogeneity in our sample warrant further investigation across more diagnostic groups. Moreover, in contrast to traditional ROM tools, the data collection is not restricted to symptom intensity, but also includes assessment of functioning and quality of life. The ESM procedure allows for patient-reported outcome measures that index adaptive daily life functioning, avoiding potentially biased data collection methods based on clinical interviews conducted by the treating therapist. Self-reports of mental states can have their own biases; the ESM sampling procedure is designed to avoid these risks, which may be especially advantageous in general hospital patients with somatic complaints due to underlying mental conditions[38,39].

The logistics of ESM data sampling, and thus the mROM data-collection, were simplified by using modern smartphones. Patients can use their own device to collect data in daily life. The Psymate™ methodology allows customization; beeps can be programmed, questionnaires adapted and person-tailored items can be added. Collected data is immediately sent to the database whenever a connection is available and remains continually available for feedback through a website with understandable graphs and figures (Figure 3). The improved logistics broaden the applicability of the ESM method from limited use in research centres to routine use in clinical care[40].

ROM is often positioned as a management tool, which suggests that ROM-data can be used to benchmark different centres. ESM methodology, applied as an mROM tool, assures that ROM data has a far broader applicability. The high level of patient involvement increases its clinical relevance to users of services[41]. Moreover, clinical and patient relevance is further

impacted by the possibility of accessing and viewing the data independent from services. Thus, in order to make ROM relevant to both the patient and the clinician, ESM may be rolled out routinely.
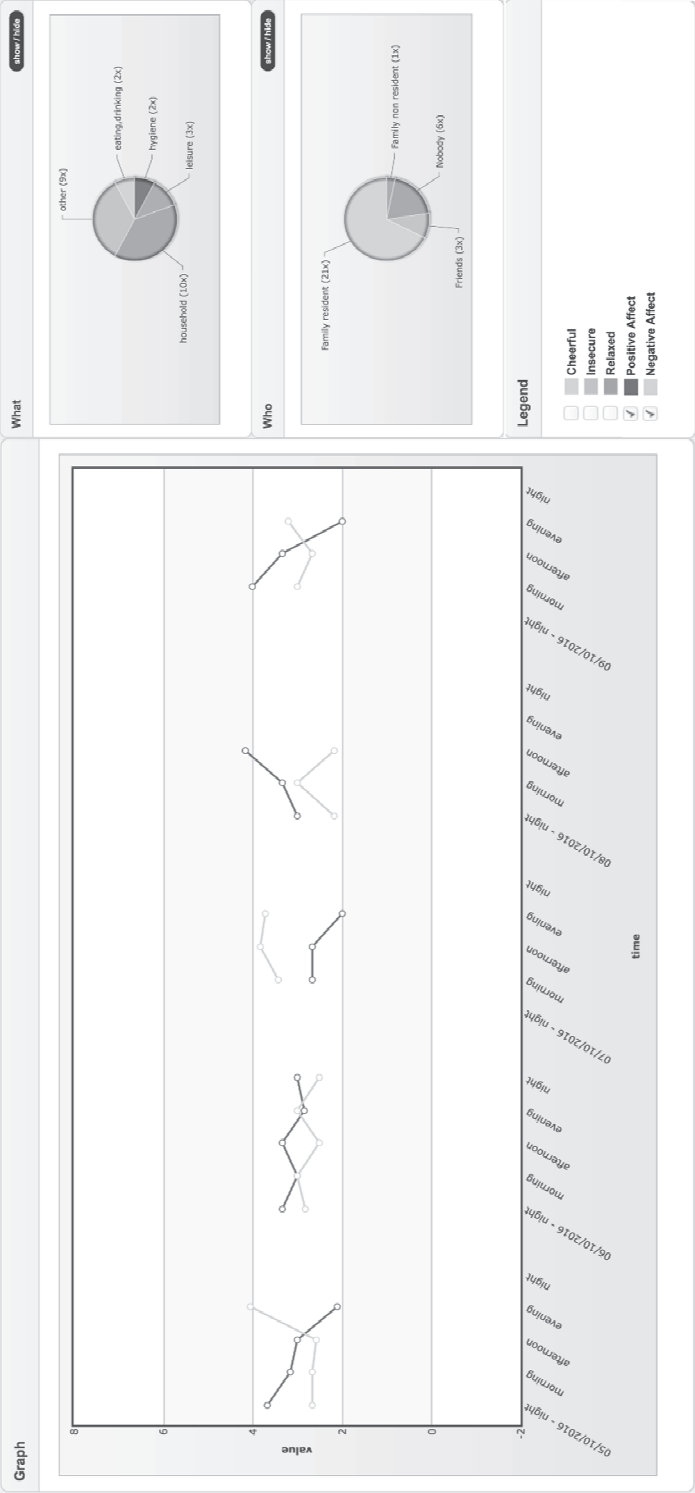
Figure 3

This figure is an example of the PsyMate™ ESM feedback, provided on a dedicated website with choice of 'contextual' or 'functional' analysis. The data (here five consecutive days) are displayed over time (here clustered in dayparts; morning, afternoon ...). Aggregated Likert scores of the scale items (here positive- and negative affect is selected) reflect the mental state at the time points. What and Who pie charts display the time budgets for the sampling period. In an interactive feedback meeting with the patient, the clinician can select options of the pie chart to break down the subject's responses and assess whether mental states are contextualized (subject feels better in one situation, compared to the other) or zoom in on specific moments, that reflect vulnerability (crisis) or resilience (coping).

References

1. Hafkenscheid A, Van Os J. Naar een deugdelijke ROM. *Maandblad Geestelijke Volksgezondheid.* 2014;69(1):20-28.
2. Hafkenscheid A, Van Os J. ROM van geïndividualiseerde behandeldoelen. *Psychoparktijk.* 2014;6(5):29-32.
3. Delespaul PA. Routine outcome measurement in the Netherlands–A focus on benchmarking. *Int Rev Psychiatry.* 2015;27(4):320-328.
4. Cofiel L, Bassi DU, Ray RK, Pietrobon R, Brentani H. Detecting dissonance in clinical and research workflow for translational psychiatric registries. *PloS One.* 2013;8(9):e75167.
5. Hafkenscheid A, van Os J. [Current routine outcome monitoring undermines valid quality measurement]. *Tijdschrift voor psychiatrie.* 2013;55(3):179-181.
6. Hafkenscheid A, van Os J. ROM measurements in mental health care: users need to be aware of the problems and pitfalls. *Tijdschrift voor psychiatrie.* 2016;58(5):388.
7. Roe D, Drake RE, Slade M. Routine outcome monitoring: An international endeavour. *Int Rev Psychiatry.* 2015;27(4):257-260.
8. Wing J, Beevor A, Curtis R, Park S, Hadden S, Burns A. Health of the Nation Outcome Scales (HoNOS). Research and development. *Br J Psychiatry* 1998;172(1):11-18.
9. Evans C, Mellor-Clark J, Margison F, et al. CORE: clinical outcomes in routine evaluation. *J Ment Health.* 2000;9(3):247-255.
10. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. 1998.
11. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ.* 2010;340:c186.
12. Carlier IV, Meuldijk D, Van Vliet IM, Van Fenema E, Van der Wee NJ, Zitman FG. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract.* 2012;18(1):104-110.
13. Ferrero A, Simonelli B, Fassina S, et al. Psychopathological Functioning Levels (PFLs) and their possible relevance in psychiatric treatments: a qualitative research project. *BMC Psychiatry.* 2016;16(1):253.
14. Tylee A. Major depressive disorder (MDD) from the patient's perspective: overcoming barriers to appropriate care. *Int J Psychiatry Clin Pract.* 2001;5(1):37-42.
15. Zimmerman M, Chelminski I, McGlinchey JB, Posternak MA. A clinically useful depression outcome scale. *Compr Psychiatry.* 2008;49(2):131-140.
16. 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(3):425-436.
17. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychol Bull.* 2002;128(6):934.
18. Van Os J, Verhagen S, Marsman A, et al. The Experience Sampling Method as an mHealth Tool to Support Self-monitoring, Self-Insight and Personalised Health Care in Clinical Practice. *Depression and Anxiety.* 2017;in press.
19. van Os J, Delespaul P, Barge D, Bakker RP. Testing an mHealth momentary assessment Routine Outcome Monitoring application: a focus on restoration of daily life positive mood states. *PloS One.* 2014;9(12):e115254.
20. 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(9):1533.
21. Wichers M, Simons C, Kramer I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatr Scand.* 2011;124(4):262-272.
22. Kramer I, Simons CJ, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry.* 2014;13(1):68-77.
23. van Os J, Verhagen SJ, Marsman A, et al. The Experience Sampling Method as an mHealth Tool to Support Self-Monitoring, Self-Insight and Personalised Health Care in clinical Practice. *J Depress Anxiety.* 2017.
24. Barge-Schaapveld DQ, Nicolson NA, van der Hoop RG, DeVries MW. Changes in daily life experience associated with clinical improvement in depression. *J Affect Disord.* 1995;34(2):139-154.
25. Barge-Schaapveld DQ, Nicolson NA, Berkhof J. Quality of life in depression: daily life determinants and variability. *Psychiatry Res.* 1999;88(3):173-189.
26. Barge-Schaapveld DQ, Nicolson NA. Effects of Antidepressant Treatment on the Quality of Daily Life: An Experience Sampling Study. *J Clin Psychiatry.* 2002;63(6):477-485.
27. Garland EL, Fredrickson B, Kring AM, Johnson DP, Meyer PS, Penn DL. Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clin Psychol Rev.* 2010;30(7):849-864.

28. Fredrickson BL, Joiner T. Positive emotions trigger upward spirals toward emotional well-being. *Psychol Sci.* 2002;13(2):172-175.

29. Jacobs N, Menne-Lothmann C, Derom C, Thiery E, van Os J, Wichers M. Deconstructing the familiality of variability in momentary negative and positive affect. *Acta Psychiatrica Scandinavica.* 2013;127(4):318-327.

30. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361-370.

31. Myin-Germeys I, Peeters F, Havermans R, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr Scand.* 2003;107(2):124-131.

32. Spinhoven P, Ormel J, Sloekers P, Kempen G, Speckens A, Van Hemert A. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychol Med* 1997;27(02):363-370.

33. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *J Psychosom Res.* 2002;52(2):69-77.

34. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale-a review of validation data and clinical results. *J Psychosom Res.* 1997;42(1):17-41.

35. Palmier-Claus JE, Myin-Germeys I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatr Scand.* 2011;123(1):12-20.

36. Delespaul PA. *Assessing schizophrenia in daily life: The experience sampling method*, Maastricht university; 1995.

37. de Beurs E, Barendregt M, Warmerdam L, Warmerdam L. *Behandeluitkomsten - Bron voor kwaliteitsbeleid in de ggz.* Boom; 2017.

38. Leue C, Kruimel J, Vrijens D, Masclee A, van Os J, van Koeveringe G. Functional urological disorders: a sensitized defence response in the bladder-gut-brain axis. *Nat Rev Urol.* 2016.

39. Mujagic Z, Leue C, Vork L, et al. The Experience Sampling Method-a new digital tool for momentary symptom assessment in IBS: an exploratory study. *J Neurogastroenterol Motil.* 2015;27(9):1295-1302.

40. Vilardaga R, Bricker J, McDonell M. The promise of mobile technologies and single case designs for the study of individuals in their natural environment. *J Contextual Behav Sci.* 2014;3(2):148-153.

41. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evid Based Ment Health.* 2016;19(3):86-89.

5

Monitoring my journey from doctor, to patient, to doctor with lived experience

J. Lenssen, S. Verhagen

As a medical student, and thereafter as a psychiatry resident, I (JL) spent more than ten years studying diagnosis, treatment and prevention of mental, emotional and behavioural disorders. After several years of clinical experience, I thought I had gained some insight into what mental suffering is about. However, as I found out, nothing out there can prepare you for what it is really like.

Here, I describe my journey towards finding meaning in depression, and how low mood becomes a psychotic prism through which the world is observed - my life the last few years.

In the spring of 2017, my life took an unexpected turn as I suddenly found myself in the other chair: a patient instead of a student in psychiatry. The year before that, I had been using bupropion, an antidepressant, to help me with attention deficit and improving structure in everyday life. At that time, I was working in an outpatient clinic for mood disorders, treating patients and prescribing medications if necessary. I was therefore quick to recognize side effects, like muscle tension and cramps in my jaw and upper legs, associated with my own treatment. Together with my psychiatrist, we decided to discontinue the bupropion, as I was doing well. With the wisdom of hindsight, we should perhaps have considered a risk, given a personal history of depression, a family history of mood disorders and a particularly busy period at work. In the first weeks after stopping, no problems arose. The muscle cramps disappeared and I was relieved not to have to take medication every day.

However, after a while I began feeling tired, had trouble concentrating, and found it increasingly difficult to get up in the morning. At some point, I decided to use light therapy for 20 minutes each morning, hoping it would help me get out and about more easily. It did not work. Sleeping, getting started in the morning, and staying motivated throughout the day was not self-evident anymore, but increasingly felt as hard labour. I also felt less confident at work. Internal struggles arose because I was feeling down and it seemed inappropriate that I would be helping others while secretly struggling myself. People around me started noticing changes and finally my colleagues gently but clearly indicated I should seek help. In April 2017, I called in sick and visited my psychiatrist. He confirmed a major depression and together we started to look for new treatment options.

A difficult period started during which I felt utterly lost. I do not remember much from the first weeks, as everything was blurred. My thoughts and movements went as if in slow-

motion. I could not enjoy even the smallest things and suddenly my future perspective was gone. I experienced psychosis – perceiving a world that was bleak, threatening, lonely and devoid of meaning. The only part of my psychiatry training that – fortunately – continued was the compulsory trajectory of personal psychotherapy. I was also prescribed antidepressants. One evening, when I felt a little better and found myself able to take perspective, I wondered if I could monitor the lived experience of my suffering, making it somehow part of my training and my desire to have a future as a psychiatrist. As my concentration was at a low point, I required a relatively simple way to document my journey.

I send an email with a query to a professor I knew. Although I felt embarrassed immediately afterwards, his reaction was supportive, saying my story was important. It made me feel validated and even generated a tiny bit of confidence. He suggested monitoring the recovery process by using the Experience Sampling Method in the form of a freeware smartphone application called the PsyMate. Experience sampling is a self-monitoring tool for momentary data collection, capturing the evolution of mood, cognition, and behaviour over time. The person is required to complete multiple short questionnaires of 1-minute duration at several moments per day, in response to auditory signals. These data can then be displayed in graphs and figures, making patterns in daily life visible. Experience sampling is used in research and has relevance for clinical practice. Often, for short sampling periods of one week. In the past, I had used the PsyMate during a research internship. Back then, it had no personal significance and soon the routine became annoying. Not so, it turned out, when I was ill.

Together with an ESM expert, I began the process that eventually would result in over 2 years of data collection. The app consists of a standard set of questions, monitoring mood (e.g., "I feel anxious"), current context (e.g., "what am I doing right now") and appraisal of context (e.g., "I like this company") at 10 random moments per day. In addition to these assessments, a morning and evening questionnaire are available to capture sleep quality and overall feelings at the end of the day. This was our starting point. Every week, we sat together to look at my results on the PsyMate website. In the beginning, this process of weekly contact and daily monitoring, thereby contributing to research, provided an anchor of meaning and purpose.

Over the following weeks, we personalized items to include other relevant symptoms and experiences. For example, we added "I feel numbed", an overwhelming feeling at the time.

Important was that we made sure the list remained balanced and contained positive items as well, to prevent a negative influence on my mood. We also decided to monitor medication more closely by adding a category to the question "Since the last beep I used…". In addition, we added an item about meditation, because we were wondering if the classes I started would impact my mood. Not everything could be personalized; beeps were programmed between 8.30AM and 10.30PM, which frequently resulted in missed beeps in the morning.

As the collected data was growing, multiple interesting patterns became visible. Often, we would see changes, which made me conscious of minor shifts that I could not have consciously formulated in advance, but which I then recognized were present. In the beginning, few fluctuations were apparent in my mood and I was feeling down all the time. After some time, my treatment started to have effect and more reactivity became visible during the day. Antidepressants often take some time before taking effect, which was also reflected in the data. PsyMate became a way to help me assess treatment effects and propose changes. Examining time-budgets was also helpful and a little confronting at first. I was spending a lot of time at home and by myself, although we learned, by graphically combining mood with activities or company, that being outdoors and with others resulted in an increase in positive feelings (see Figure 1).
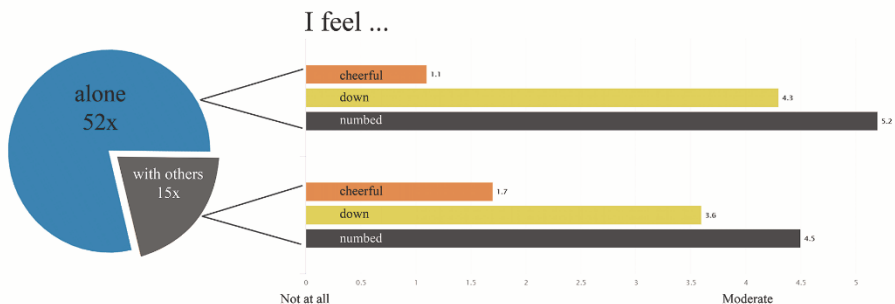
Figure 1

The pie chart (left) displays my answers on the question "Who am I with" during 2 weeks of sampling, indicating that I was alone 52 times and with others 15 times. The bar graph (right) shows how cheerful, down and numbed I felt when I was alone versus when I was with others. I felt more cheerful, less down and less numb when with others.

There were many incredibly discouraging drawbacks in my recovery process. However, because I kept monitoring my mood, we could always inspect the 'bigger picture' showing overall slow progress over the momentary hollows. This helped me realize I was still moving forward, even though it sometimes did not feel remotely as if my mental state had improved (Figure 2).
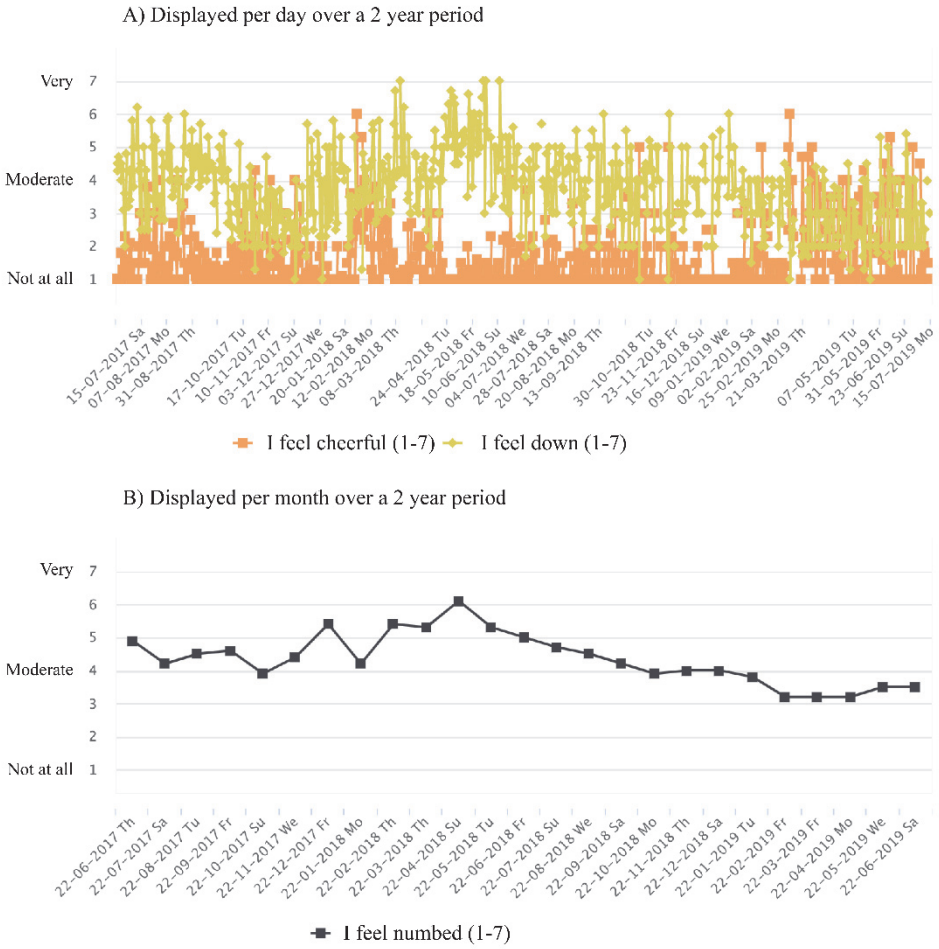
Figure 2

Online feedback graph showing my experience sampling data over a period of 2 years. Part A) displays the mood items cheerful and down over time with answers averaged by day. Fluctuations in my mood can be seen, next to how the feelings relate to each other. Part B) displays the item numbed over time with answers averaged by month. A general decline can be seen in how numb I felt, showing my overall recovery.

After two gruelling years, I am doing better now. Monitoring helped me to appraise my situation objectively and kept me on top of my depression instead of being lived by it. Recently, I restarted my psychiatry residency training. Having experienced the benefits of monitoring during different stages of recovery, I would definitely encourage others to integrate such tools into their therapy. Early on, it provides an easy way to keep track of what is going on without having to think too much. Later on, it provides concrete clues for change, guiding treatment choices. At the end, it displays a story about you, as a person, with all the thousands of literally ups-and-downs as a lived reality. In a strange way, it is both comforting and aiding in a sense of control that one can never completely disappear into the darkness. Having constructed the journey of my lived experience, I am confident that I will be a better doctor. It gives me courage to share my story and inspire healthcare professionals to do the same.

Closing perspective - I (SV) feel privileged to have been able to reflect on the experiences collected by Jaimie. Working together on personalizing PsyMate and finding meaning in a challenging period of life has been a valuable and special experience. As a researcher and psychologist, I learned what is important in experience sampling and what is less relevant. The key is to attempt to reduce complexity to simplicity. Pervasive patterns, once discovered, are turning points in guiding change or finding understanding. I am proud of Jaimie, it takes courage and endurance to open up and use one's own experience and share it for the benefit of others.

Part II

Methodological relevance

6

Constructing a reward-related quality of life statistic in daily life – a proof of concept study using positive affect

S. Verhagen, C. Simons, C. van Zelst, P. Delespaul

Abstract

Background: Mental healthcare needs person-tailored interventions. Experience Sampling Method (ESM) can provide daily life monitoring of personal experiences. This study aims to operationalize and test a measure of momentary reward-related Quality of Life (rQoL). Intuitively, quality of life improves by spending more time on rewarding experiences. ESM clinical interventions can use this information to coach patients to find a realistic, optimal balance of positive experiences (maximize reward) in daily life. rQoL combines the frequency of engaging in a relevant context (a 'behavior setting') with concurrent (positive) affect. High rQoL occurs when the most frequent behavior settings are combined with positive affect or infrequent behavior settings co-occur with low positive affect.

Method: Resampling procedures (Monte Carlo experiments) were applied to assess the reliability of rQoL using various behavior setting definitions under different sampling circumstances, for real or virtual subjects with low-, average- and high contextual variability. Furthermore, resampling was used to assess whether rQoL is a distinct concept from positive affect. Virtual ESM beep datasets were extracted from 1058 valid ESM observations for virtual and real subjects.

Results: Behavior settings defined by Who-What contextual information were most informative. Simulations of at least 100 ESM observations are needed for reliable assessment. Virtual ESM beep datasets of a real subject can be defined by Who-What-Where behavior setting combinations. Large sample sizes are necessary for reliable rQoL assessments, except for subjects with low contextual variability. rQoL is distinct from positive affect.

Conclusion: rQoL is a feasible concept. Monte Carlo experiments should be used to assess the reliable implementation of an ESM statistic. Future research in ESM should asses the behavior of summary statistics under different sampling situations. This exploration is especially relevant in clinical implementation, where often only small datasets are available.

Introduction

Mental health care is becoming more patient-centered. Every person is unique and the current classification systems miss relevant nuances for customized therapeutic interventions[1,2]. Van Os[3] argued to innovate assessment by making it more person-tailored and actively involve patients in the process. Clinicians should focus less on group characteristics and more on the individual's daily adaptation strategies and need for care[3]. Over the years, diagnostic procedures have become more time consuming. Results are often complex latent structures that are not transparent and do not facilitate a collaborative communication between clinician and patient[3,4]. The main purpose of mental health care is to improve functioning as well as quality of life, in an empowering way. Most psychological interventions require motivation and engagement. An alienating communication does not help to engage patients in treatment. The ultimate goal is to assist patients in becoming more resilient, improve autonomy, reduce the impact of mental illness and improve well-being.

Resilient individuals are able to reduce their vulnerability by reducing the impact of symptoms and complaints in daily life. The reference point for therapeutic success is the actual moment-to-moment experience and functioning. The Experience Sampling Method (ESM) is a structured diary technique specially developed to appraise subjects in their daily interactions[5]. The technique makes it possible to study subtle dynamic changes in momentary affective states that are difficult to assess in cross-sectional questionnaires. ESM has high ecological validity[6,7] and allows person-tailored data collection. Since it reflects the subject's own daily life, the data facilitates transparent communication and collaborative care. ESM has been used for numerous research purposes in the general population[8], somatic health care[9] and mental health care[10]. The method proved valuable in monitoring treatment effects[11] and can be used as a treatment intervention[12].

Quality of Life (QoL) is an important outcome indicator. Mental illness often decreases QoL, resulting in lowered subjective well-being and functioning[13,14]. QoL can be defined as "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" [15, p. 1405]. Researchers and clinicians essentially consider QoL as a subjective concept. A broad operationalization combines different domains, such as social contact, physical health and environmental resources[15,16]. Affect, cognition, behavior and physical functioning influence experienced QoL[17]. The World Health Organization Quality of Life Assessment (WHOQOL) was developed to assess this comprehensive, multi-domain view of QoL[15]. Other cross-sectional QoL measures either assess the impact of mental illness[14,18] or monitor successes

in treatment[19-21]. Structured interviews are used in the assessment of patients with a severe mental illness, both for subjective information (e.g., self-assessment) and objective information (e.g., social functioning)[22,23]. Lehman[24] showed that patients with severe mental illness were able to give an account of experienced QoL, similar to subjects from the general population.

Barge-Schaapveld and colleagues[25] used ESM to study subjective well-being in daily life for patients with depression. They assumed that QoL varies from moment-to-moment and assessed momentary QoL (mQoL) repeatedly, using the question "In general, how is it going with you right now?" ESM questionnaires were administered ten times a day for six consecutive days. Results confirmed between-subject and within-subject (temporal) variation in mQoL. Compared to healthy control subjects, depressed subjects reported lower mQoL, less activity, and experienced more negative affect and less positive affect. In addition, the variation of mQoL was higher in the depressed group. Furthermore, situational factors had a large influence on mQoL in both groups[25]. ESM is a feasible method for the assessment of momentary health-related QoL[26]. QoL can be measured in the moment, under different daily life situations and varies between and within persons[25,26]. ESM is a compelling and comprehensive method because it assesses different factors that influence QoL, namely momentary affect and contextual variability.

To date, mQoL was an outcome statistic. Repeated self-assessments are made for a representative sample of moments during a specified period. Individual mQoL assessments are aggregated, yielding a statistic that represents pre or post intervention mQoL for a subject. The clinical relevance is limited, because these aggregated mQoL scores miss the necessary information to inform patients and clinicians on how to improve QoL dynamically over time. This requires an mQoL statistic that directly links treatment aims such as improving well-being with adaptation strategies in the moment and informs individuals and clinicians about choices in daily life.

The mechanism of reward or the process of reward seeking can be relevant to improve mQoL. Subjective well-being is related to reward experiences. Moreover, subjective well-being and reward-related neural activity are related[27]. Reward experiences are the drivers of operant learning[28]. In operant conditioning, people learn from the consequences of their response and use that knowledge to guide future behavior choices[28,29]. Stimulus-response associations are computed internally and updated frequently, allowing people to predict the outcome and choose responses from their repertoire accordingly. This implicit associative learning is driven by reinforcement. According to behavior theory, reinforcement

strengthens or weakens the selection of behavior in a similar situation. Positive reinforcement occurs when a certain response to a new stimulus results in a valuated outcome and is thus rewarded, thereby increasing the likelihood of similar behavior in the future[30].

In mental health research, mechanisms of operant conditioning and reward are repeatedly linked to well-being. Lewinsohn and colleagues[31] hypothesized that depression is a consequence of low levels of response-contingent positive reinforcement. In a sample of college students, they correlated depressed mood with time spent on pleasant activities. Increased time in pleasant activities was viewed as an indicator of positive reinforcement. The results showed a moderately negative association: where pleasant activities decreased, experienced depression increased[31,32]. A causal link between the two variables could not be established[33]. Other studies have demonstrated positive affective experiences and reward experiences in relation to resilience in depressive subjects[34,35]. A randomized control trial (RCT) conducted by Geschwind, Peeters, Drukker, van Os and Wichers[36] showed that mindfulness-based cognitive therapy aimed at increasing positive affect and the enjoyment of reward experiences during daily life was associated with a reduction in experienced depression.

Kramer and colleagues[37] used an ESM-based intervention (ESM-I) to examine whether self-monitoring of positive affect (PA) is beneficial for depressive patients in addition to treatment as usual. ESM-derived feedback was used as a therapeutic tool to gain insight in implicit dynamic patterns that arise over time. Through person-tailored feedback sessions, hidden patterns were made explicit using visualization in graphs and figures[37]. Weekly ESM-I feedback sessions influenced the treatment of depression positively, with an effect still present at six-month follow-up[37]. Contrary to these long-term effects, they found no significant impact of the ESM-I on daily experienced PA during the intervention or shortly after[38]. Another RCT in young adults with depression also provides evidence that ESM-I may have positive impact on pleasure and PA by providing personalized lifestyle advice[39]. Wichers and colleagues[40] used ESM as a tool to prospectively observe implicit learning processes for reward-seeking and punishment-avoidant behavior in the context of daily life. They hypothesized that current behavior could be predicted by the experience of related behavior at previous time points. Results confirmed that affect moderates this association over time, both at beep and day level[40].

ESM is rooted in ecological psychology where contextual embeddedness is widely recognized[41]. Barker introduced the term 'behavior settings', to reflect the mutual relation

between human behavior and the environment[42]. Behavior settings are eco-behavioral entities which exist independent of persons and form self-regulating systems[43]. They represent stable and identifiable constructs with both spatial and temporal indices and can provide opportunities or constrain the actions of persons[42,44]. The term behavior setting puts emphasis on processes and structures that often go unnoticed in the daily life of individuals[44].

Imagine yourself on a market. Being there with friends will likely provide a different experience than being alone. In addition, activities influence the experience. You could be working, buying a last ingredient for dinner or simply enjoying yourself. Other factors play a role as well, such as the location of the market, its attributes, the weather and time of day. All ingredients form the behavior setting – the rich and meaningful context.

Future research should emphasize the linkage between a person's affective experience and behavior setting characteristics, including the beliefs and know-how of this behavior setting[44]. With ESM, information on momentary affect can be gathered in the context of daily life[5]. Behavior settings are important because they provide insight in the contextual variability of positive or negative affect. Clinicians can coach patients to engage in contexts (i.e., future behavior selection) to maximize individual patterns of positive affect and avoid negative affect, leading to more experienced QoL.

In line with the literature above, subjects can improve QoL by engaging more often and for longer time in affectively rewarding situations. Some may argue that extremely rewarding, low frequency behaviors can boost QoL (e.g., a holiday travel compensates for a boring job). However, exceptional situations do not rule moment-to-moment experiences, while frequent minor events that occur naturally in the flow of daily life have an impact on mental well-being[45]. The pursuit of reward cannot be a unidimensional focus. QoL does not increase linearly (more is better), but optimizes by balancing challenges of daily life. Even the most enjoyable job will be perceived differently when there is no time to relax anymore. Maximizing reward experiences in daily life, means spending most of our time in rewarding situations (eliciting high PA) and avoiding situations with low PA. Depressive patients, for instance, are out of balance and spend insufficient time on pleasant activities[31,45-47]. Their time budgets are not optimized and too much time is spent in behavior settings with low PA.

The reward-related QoL function (rQoL) reflects the momentary fit between context and optimal affective experience. ESM can monitor this process in treatment[37]. rQoL can be used in shared decision-making. Feeling good in some situations and bad in others is a

transparent communication and most patients understand this intuitively. The collaborative ESM feedback sessions between patients and clinicians are the setting to discuss reward optimization. In a shared decision process, patients and clinicians select rewarding situations and explore how to increase the occurrence of well-being in daily life.

Can these situations be detected and personal profiles computed? When applied in clinical practice, does this improve the subject's overall well-being? This paper describes the development of a reward-related QoL function (rQoL) and the proof of concept of its applicability. A reward statistic is defined. It reflects individual daily life moment-to-moment variation in reward efficiency, by combining the actual (positive) affect with the occurrence frequency of the actual behavior setting. Data is collected using ESM. To our knowledge, this is the first study that uses mechanisms of reward to design a momentary quality of life measure.

Method

Sample

To assess the feasibility of the rQoL concept, ESM data from an existing dataset — the D-STIGMI study[48 p. 103] was used. The D-STIGMI study evaluates a psycho-educational coping skill training in people with severe mental illness using a RCT. The aim of the training was to increase resilience against stigmatization. The ESM data collection was an optional add-on to explore innovative outcome parameters. The Medical Ethics Committee of Maastricht University Medical Centre approved the study protocol under the number of NL3179406810. In the current study, only baseline ESM assessment data were used as seeds for the random simulated sets. ESM data were available for 27 participants.

Measurements

Experience sampling method (ESM). ESM is a structured diary technique to assess moment-to-moment mental state changes in relation to situations in the daily life of individuals[5,49]. The data of the D-STIGMI study was collected with the PsyMate™ device, a palm-top assessment tool developed for ESM data collection (http://www.psymate.eu/). The PsyMate™ was programmed to emit ten beeps each day for six consecutive days. Beeps were generated at semi-random moments, within 90-minute blocks, between 7.30 and 22.30 hours. An auditory and visual signal indicated the availability of a short questionnaire. Responding lasted less than a minute. The questionnaires remained available for 15 minutes and subjects were instructed to promptly reply. Daily life experiences are captured in items assessing current affect, activities and context. Most items were rated on a 7-point Likert scale (ranging

from 1 = 'not at all' to 7 = 'very'). A bipolar scale (-3 = negative, 0 = neutral, to 3 = positive) was used to assess stressful events. Context (activity, location and person present) and the use of substances were assessed with multi-optional checklists. The item 'Overall, I feel well at the moment' was added to assess mQoL. Beeps are considered valid when the whole questionnaire is completed. Each subject could respond to a maximum of 60 beeps. In line with ESM guidelines, subjects were included in the analyses when they completed at least a third of the beeps (20 valid beeps)[5].

Assessment of rQoL

Momentary reward-related quality of life (rQoL) is defined as the fit between frequency of situations (behavior setting) and actual mental state. rQoL can be computed for each moment and provides feedback to clinicians and patients to collaboratively select intervention strategies that optimize the mental state in daily life and lead to more overall well-being. The related rQoL statistic uses the subject's own data to assess if he/she optimizes the selection of contexts to maximize positive mental states, meaning that rewarding situations should occur often and less rewarding situations should be avoided.

The ingredients of the function are:

Mental states. The ESM questionnaire contains mood items that assess PA: I feel 'cheerful', 'satisfied', 'relaxed' and 'enthusiastic'. Momentary PA was normalized by subject ($zPA_{(ij)} = PA_{(ij)} - PA_{(i.)}$ – with 'i' for subjects and 'j' for moments). $zPA_{(ij)}$ yields positive scores for better than average mental states and negative scores for below average mental states. Better than average mental states are assumed to be rewarding.

Behavior setting. A meaningful situation is a behavior setting. Time, place, persons and activities define them. The ESM definition of a behavior setting uses the context information available in the beep-level questionnaires: the time of the day (morning, afternoon, evening), persons present ('with whom am I': 'partner', 'resident family', 'family living away from home', 'friends', 'colleagues', 'acquaintances', 'strangers/others' and 'nobody'), activity ('what am I doing': 'resting', 'work/school', 'household/groceries', 'hygiene', 'eating/drinking', 'relaxation', 'doing something else' and 'nothing') and location ('where am I': 'home', 'someone else's home', 'work/school', 'public place', 'on the go' and 'somewhere else'). This results in 3 x 8 x 8 x 6 = 1152 potential combinations, of which many infrequently occur or never occur and result in empty cells for individual subjects. The detailed number of situations does not allow the selection of high and low frequency behavior settings for each subject. Therefore, the time of the day was omitted and the number of options for who, what and where limited to

six each. For persons present, 'partner' was included into 'resident family' and 'colleagues' into the 'acquaintances' category. For activity, we combined 'resting' with 'doing nothing' and 'household/groceries' with 'hygiene'. The options for location remained unchanged. The occurrence (as a proportion) of the 216 resulting behavior settings was computed for each individual subject. The cumulative proportion was computed with a break at 0.5 to differentiate the large set of infrequent situations and the much smaller set of frequent situations.

Reward function. The momentary rQoL statistic combines the frequency of the momentary behavior setting with the actual mental state. Specifically, the function uses the normalized positive affect score by individual ($zPA_{(ij)}$) and weights it with the frequency of occurrence of the individuals' behavior setting ($cp(BS_{ij})_i$). Reward efficiency occurs when high frequency situations yield positive moods. High rQoL occurs when high frequency situations are combined with positive mental states or low frequency is limited to negative mental states. Low rQoL situations combine poor mental states with high frequency situations or elevated mental states with infrequent behavior settings. These characteristics are reflected in the formula:

$$rQoL(PA)_{ij} = zPA_{ij} \times \left(cp(BS_{ij})_i - 0.5 \right)$$

in which:

$rQoL(PA)_{ij}$ is reward-based QoL computed on PA for subject i on moment j; zPA_{ij} is the standardized PA score for subject i at moment j; and $cp(BS_{ij})_i$ is the cumulative proportion for subject i of occurrence of the current behavior setting for that subject i on moment j.

-0.5 is used to generate a cut-off score for high and low proportions. Other cut-off scores were explored (and are further explained in the analyses section).

Using this formula, a specific rQoL score was generated for each ESM moment. Negative scores (e.g., -0.1) represent low rQoL, whereas positive scores (e.g., 0.9) represent high rQoL. An alternative reward function can be defined with negative affect (NA) but PA was selected in line with other scholars in the field[35,40]. The rQoL statistic was computed using a small program written in a Stata™ (v13.0) script.

Analyses
To assess the feasibility of the rQoL function, we assessed the impact of different choices using resampling methods (Monte Carlo experiments) to generate ESM data for virtual

subjects. The sample size (N) is the number of ESM observations or beeps drawn from D-STIGMI dataset used as seeds. Different selections of contextual domains and different cut-off scores were explored to generate rQoL. Simulations were run using virtual subjects with data extracted from all beeps from the D-STIGMI seeding database. Because real subjects have more specific frequency distributions in behavior settings, virtual datasets were extracted for subjects with low, average and high situational diversity.

Which sample size do we need to reliably assess behavior setting in individuals? A Monte Carlo experiment was executed to study the effect of the number of available beeps on the average number of unique context combinations (behavior settings). Three different definitions for behavior setting were explored: a What only definition (BS_W – 6 situations), a What and Who -based definition (BS_WW – 6 x 6 = 36 situations) and a What, Who and Where definition (BS_WWW – 6 x 6 x 6 = 216 combinations). Resampling made it possible to explore the alternative behavior setting definitions for different sampling sizes (20, 40, 60, 80, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 8000, 10000 'valid' observations by virtual subject).

Which sample size do we need for optimal variation in reward (rQoL)? A second Monte Carlo experiment was executed to study the effect of available observations on the average scores and variation of rQoL, independently for BS_W, BS_WW and BS_WWW. This was done to provide insight in the number of observations needed to reliably generate rQoL. Initially, different cut-off scores (.50, .40, .30, .20) were explored to differentiate low and high frequency situations. Only the cut-off score .50 proved sensitive enough for meaningful rQoL detection and was therefore used in further analyses. Sample sizes that were explored, are 20, 40, 60, 80, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 8000 and 10000 observations for the virtual subjects in the simulation.

What is the effect of the actual contextual variation in real subjects? In contrast to virtual subjects who live in the contexts of the group of individuals, actual subjects live in environments that are more restricted. We selected subjects with different levels of behavior setting differentiation (the number of non-empty BS_WWW categories) and ranked these individuals to compute percentiles. A Monte Carlo experiment was executed to simulate the rQoL functions for subjects with low (5th percentile), average (50th percentile) and high (95th percentile) variability in contextual domains. Sample sizes that were explored are 20, 40, 60, 80 and 100 'valid' observations by virtual subject.

Is reward-related rQoL something different than positive affect? A final Monte Carlo experiment was executed to assess whether momentary PA and rQoL were separate concepts. Pearson's product-moment correlations between PA scores and rQoL scores were computed for each resampled set of momentary data, using varying sampling sizes. This was done separately for BS_WWW, BS_WW and BS_W in the overall sample.

The seeding data for the Monte Carlo experiments were selected using ESM observations from real subjects combined together as the sampling domain. When not enough unique empirical data were available (simulated samples exceeding available observations), we sampled with replacement. For each simulation, 1000 samples were drawn. Analyses were performed in Stata™ (v13.0). The do-file is added a supplementary material.

Results

Subject characteristics

Twenty-seven patients with a severe mental illness were included in the ESM baseline measurement of the D-STIGMI study. Four patients had insufficient valid beeps (<1/3), thus 23 patients were included as seeds for Monte Carlo experiments. The sampling set includes 1058 valid beeps, at average 48 per subject (*SD* = 9.03, range 22 - 63). No significant differences were found between the original sample and the seed-sample on age (*p* = .97) and sex (*p* = .53). Demographic information is summarized in table 1.

Which sample size do we need to assess behavior settings in individuals reliably?

Figure 1 shows the number of unique behavior settings generated by the resampling procedures for different sample sizes. The theoretical number of behavior settings was 216 but the beeps actually only contained 113 options (the empirical ceiling). The graph has three phases: from $N = 20$ (*mean* = 13; *SD* = 1.75) to $N = 100$ (*mean* = 33; *SD* = 3.31), from $N = 100$ to $N = 1000$ (*mean* = 92; *SD* = 3.6) and from $N = 1000$ to $N = 10,000$ (*mean* = 113; *SD* = 0.16). Around $N = 8000$ (*mean* = 113; *SD* = 0.16), saturation is reached. Standard errors are low in small samples (*SEM* = .39; $N = 20$) and increase over the second phase (max *SEM* = 4.08; $N = 500$) and finally to reduce again (*SEM* = .002; $N = 10,000$). Figure 1b and 1c reflect simulations for the BS_WW (theoretical 6 x 6 = 36 options, sample 36 options) and BS_W behavior settings (theoretical 6 options, sample 6 options). The same pattern replicates but the ceiling is reached with smaller samples (around 500 observations in the BS_WW and 100 in the BS_W situation).

Table 1

Demographics and characteristics of the 23 participants seed-sample

Characteristics	N (%)	Mean	SD	Range	Details
Age (years)		46	9.95	29-62	
Gender					
Female	60%				
Male	40%				
DSM-IV diagnostics					
Single diagnosis	61%				
Multiple diagnoses	35%				
Missing diagnostics	4%				
Clinical classification					
Primary psychotic vulnerability	68%				(292.12/295.30/295.70 /298.9/296.40)
Primary non-psychotic vulnerability	32%				(296.31/299.80/301.83 /301.83/309.81)
GAF symptoms		46.45	14.27	28-70	
GAF handicap		45.00	9.83	31-70	
Beep characteristics (N = 1058)					
Positive affect (overall beeps)		4.00	1.41	1-7	
Positive affect (by subject)		4.00	0.99	1.3-6.3	
Number of context combinations					
Unique BS_WWW combinations	113				Possible = 216
Unique BS_WW combinations	36				Possible = 36
Unique BS_W combinations	6				Possible = 6
Context combinations per subject					
Low BS variability subject	5				
Average BS variability subject	17				
High BS variability subject	28				

Note. 292.12 = Induced psychotic disorder, with hallucinations. 295.3 = Schizophrenia: Paranoid type. 295.7 = Schizoaffective disorder. 298.9 = Psychotic disorder NOS. 296.4 = Bipolar I disorder, most recent episode hypomanic. 296.31 = Major depressive disorder, mild. 299.8 = Rett's disorder; Asperger's disorder; Pervasive developmental disorder NOS. 301.83 = Borderline personality disorder. 309.81 = posttraumatic stress disorder. GAF = Global Assessment of Functioning. BS = Behavior Setting.
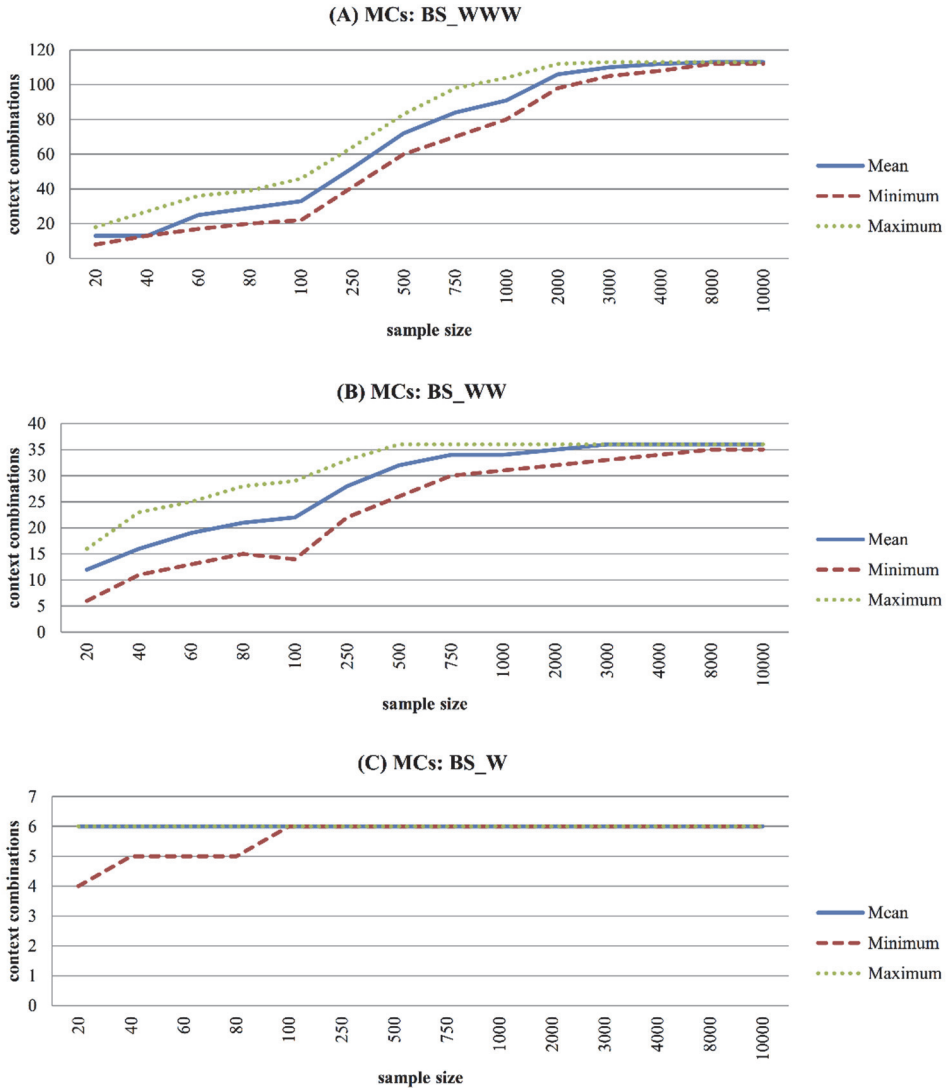
Figure 1

Figure 1a: Monte Carlo Experiment (MCs) to Explore Average-, Minimum- and Maximum Number of Unique Context Combinations with Increased Sample Size for Behavior Settings Including Who, What, Where Information (BS_WWW)

Figure 1b: Monte Carlo Experiment (MCs) to Explore Average-, Minimum- and Maximum Number of Unique Context Combinations with Increased Sample Size for Behavior Settings Including Who, What Information (BS_WW).

Figure 1c: Monte Carlo Experiment (MCs) to Explore Average-, Minimum- and Maximum Number of Unique Context Combinations with Increased Sample Size for Behavior Settings Including What Information (BS_W).

Which sample size do we need for optimal variation in reward quality of life (rQoL)?

For different sampling sizes, the rQoL was computed at each moment of the simulated virtual subject. As expected, all samples had an average of 0.00, the neutral point of the rQoL function. Standard errors of the mean were low and are considered negligible (*SEM* = .06 to *SEM* = .003). For BS_WWW and BS_WW (figure 2a and b), the range of rQoL scores increased up to sample sizes of 500 beeps. For BS_W the range was more restricted but reached its maximum for sampling sizes of 60.

What is the effect of the actual contextual variation in real subjects?

The selected subjects (low-, average-, high variability in behavior setting) responded reliably to respectively 31, 45 and 55 beeps, with respectively 5, 17 and 28 different behavior settings (using the BS_WWW combination). Results are presented in Figure 3. Part 3a shows the increase in average rQoL scores for the subject with low behavior variability (p5; 5th percentile), average behavior setting variability (p50; 50th percentile) and high behavior setting variability (p95; 95th percentile). Increases in sampling size did not affect the scores for subjects with low behavior variation (1.1 mean difference), but did for subjects with high variation (12.6 mean difference). Part 3b shows the range of minimum and maximum scores. This confirms the previous observation. Smaller sampling sizes are possible in subjects living restricted lives.

Is reward-related QoL something different than positive affect?

The Pearson's product-moment correlations of the Monte Carlo experiments on sampled sets of PA scores and rQoL are summarized in figure 4. Looking at the average correlation scores, weak associations were found between PA and rQoL for BS_WWW, BS_WW and BS_W (range $r(18) = .11$, $p < .001$ to $r(9998) = .005$, $p = <.001$). The range between minimum and maximum correlational scores in all three variations of behavior setting is large in smaller sample sizes (highest difference from *min* = -.91 to *max* = .90) and decreases with larger sample sizes (lowest difference from *min* = -.02 to *max* = .04).
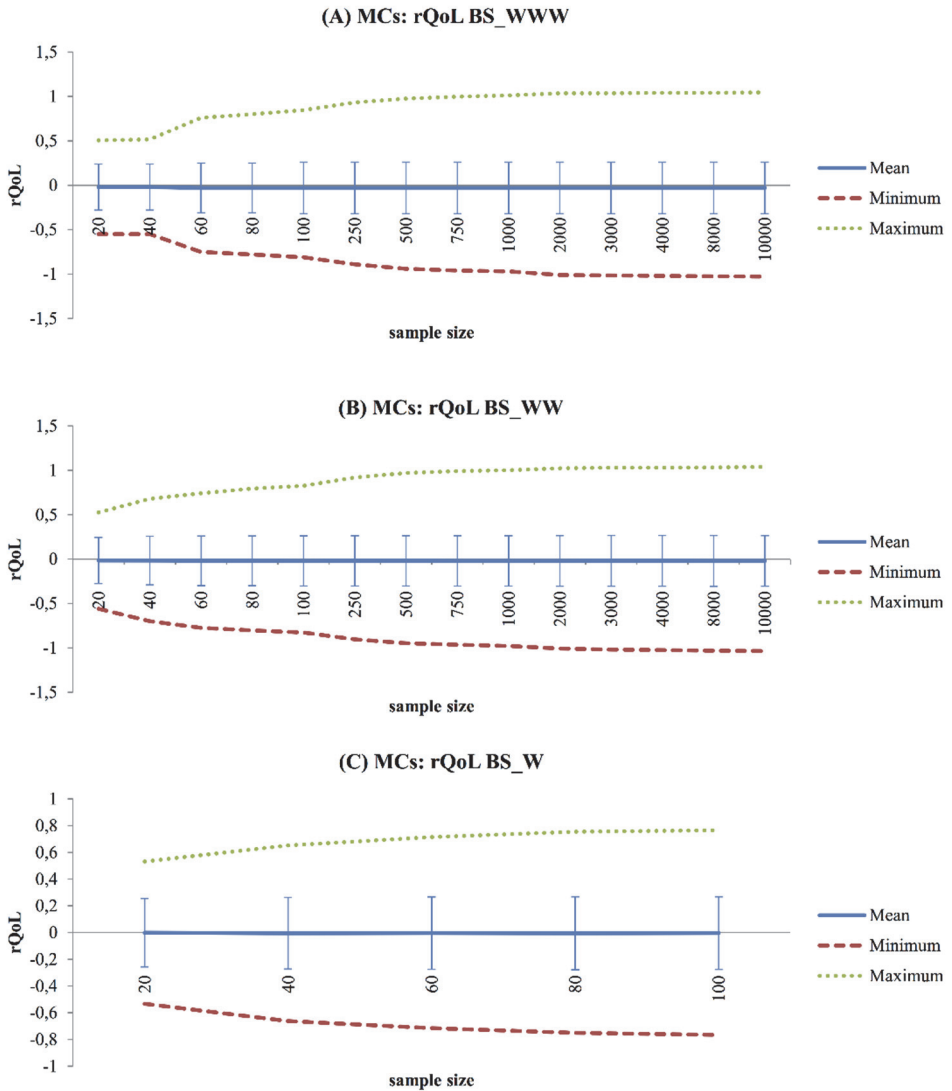
Figure 2

Figure 2a: Monte Carlo Experiment (MCs) to Explore Average-, Minimum-, Maximum- and Standard Deviation Scores of rQoL(PA) with Increased Sample Size for Behavior Settings Including Who, What, Where Information (BS_WWW).

Figure 2b: Monte Carlo Experiment (MCs) to Explore Average-, Minimum-, Maximum- and Standard Deviation Scores of rQoL(PA) with Increased Sample Size for Behavior Settings Including Who, What Information (BS_WW).

Figure 2c: Monte Carlo Experiment (MCs) to Explore Average-, Minimum-, Maximum- and Standard Deviation Scores of rQoL(PA) with Increased Sample Size for Behavior Settings Including What Information (BS_W).
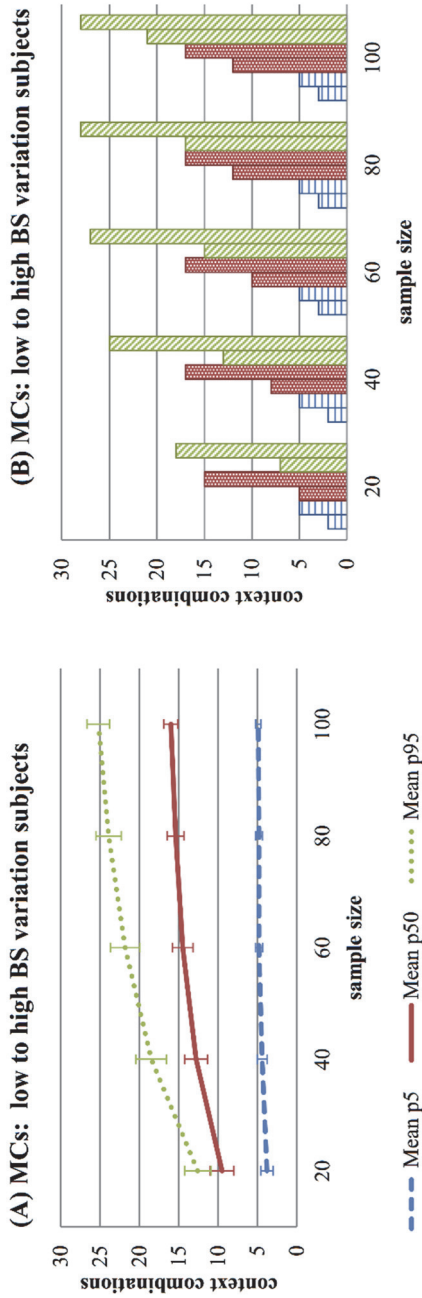
Figure 3

Figure 3a (left): *Monte Carlo Experiment (MCs) to Explore Average- and Standard Deviations Scores of Unique Context Combinations for Subjects with Low-, Medium- and High Diversity in Behavioral Setting, with Increased Sample Size for Behavior Settings Including Who, What, Where Information (BS_WWW).*

Figure 3b (right): *Monte Carlo Experiment (MCs) to Explore Minimum- and Maximum Number of Unique Context Combinations for Subjects with Low-, Medium- and High Diversity in Behavioral Setting, with Increased Sample Size for Behavior Settings Including Who, What, Where Information (BS_WWW).*
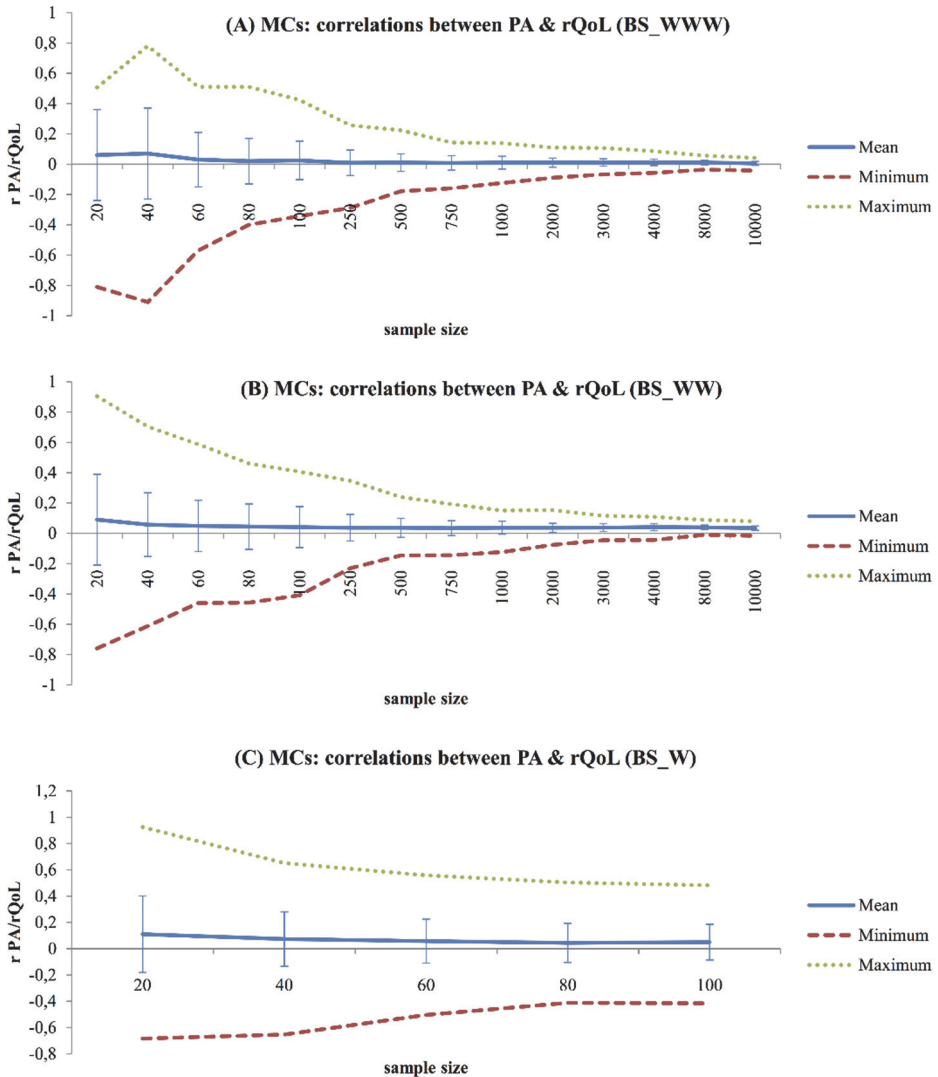
Figure 4

Figure 4a: Monte Carlo Experiment (MCs) to Explore the Relationship between PA Scores and rQoL Scores with Increased Sample Size for Behavior Settings Including Who, What, Where Information (BS_WWW), Using Pearson's Product-Moment Correlations.

Figure 4b: Monte Carlo Experiment (MCs) to Explore the Relationship between PA Scores and rQoL Scores with Increased Sample Size for Behavior Settings Including Who, What Information (BS_WW), Using Pearson's Product-Moment Correlations.

Figure 4c: Monte Carlo Experiment (MCs) to Explore the Relationship between PA Scores and rQoL Scores with Increased Sample Size for Behavior Settings Including What Information (BS_W), Using Pearson's Product-Moment Correlations.

Discussion

General conclusion

The purpose of the current proof-of-concept study is two-fold. First, a momentary rQoL statistic was defined. Second, Monte Carlo experiments were performed with samples of beep-level data for virtual subjects to check the feasibility and initial validity of the statistic. Momentary rQoL integrates affective experience (positive affect) and situations (behavior setting). A cut-off score of .50 of the cumulative proportion was chosen to separate low and high frequency situations. This proved to be the best choice to detect relevant contexts for reward efficiency in most subjects. Results show that the rQoL statistic is feasible. The rQoL statistic is defined at the moment level, allowing assessment of change over time. At a specific moment in time, a positive rQoL score indicates good reward efficiency, whereas a negative score indicates bad reward efficiency.

Persons present (Who), activity (What) and location (Where) were used to define different conceptualizations of a behavior setting, namely Who-What-Where (BS_WWW), Who-What (BS_WW) and What (BS_W). For Who-What-Where combinations, not all theoretical possibilities were available in the reference beep dataset. This dataset included 113 unique combinations out of the 216 theoretical options. Some situations had low frequencies (e.g., working with your partner or doing household/groceries at work) or simply did not occur in the group of patients with a severe mental illness. First, Monte Carlo experiments were performed to check which sample size is needed for reliable behavior setting calculations in individuals. Overall, large sample sizes were needed to generate realistic frequency distributions (time budgets) in virtual subjects (saturation was reached at $N = 8000$ for BS_WWW, $N = 500$ for BS_WW and $N = 100$ for BS_W). Next, Monte Carlo experiments were performed to explore which sample size is needed to detect optimal rQoL variation. The range of rQoL scores increased up to generated samples of 500 beeps for BS_WWW and BS_WW, whereas the limit was reached at 60 beeps for BS_W. Therefore, a behavior setting defined by Who-What proved most useful: it provides sufficient behavior setting variation to generate a reliable frequency distribution at acceptable sample sizes. An optimal spread in rQoL variation is obtained after collecting 500 observations, meaning that all possible behavior-setting combinations are present in the sampling period (see figure 2B). However, a minimum of 100 beeps is required to reliably calculate the rQoL statistic. The sample sizes needed for BS_WWW are unrealistic in ESM. These results suggest that an extended sampling period is needed before the rQoL statistic can be integrated as an active part of treatment.

Further, Monte Carlo experiments were used to explore actual situational variations in real subjects with low, average and high behavior setting variation with combinations of persons present, activity and location as behavior setting. This behavior setting definition was chosen because it theoretically provides the largest chance of finding a decent spread in unique situations in this patient group. Results show that small sample sizes are possible in subjects with low behavior setting variations ($N > 40$), whereas larger sample sizes ($N > 100$) are needed for subjects with average and high behavior setting variations. It is difficult to assess which definition of behavior setting is sufficient for individual subjects. For now, behavior settings defined by Who-What-Where combinations seem the best option because the restricted living environments of individual subjects results in less overall situational variation.

Finally, Monte Carlo experiments were run to see whether momentary rQoL is distinct from PA. Only weak correlations were found and results confirmed that the concepts assess different aspects of daily life mental states. Additionally, the momentary rQoL scores were correlated with 'In general, how is it going with you right now' (mQoL). For this, the overall beep sample of 1058 valid observations was used. Results show a moderate positive correlation ($r(1056) = .33$, $p < .001$), with mQoL explaining 11 percent of the variation in rQoL.

Strengths

To our knowledge, this is the first study that combines affective experience with behavior settings in an operationalization of rQoL in daily life. A main strength is the use of ESM data collected in the flow of daily life, making the rQoL statistic highly specific to the situation of the subject. Monte Carlo experiments are especially suited for exploration of properties and sampling characteristics of specific combined statistics[50]. Researchers compute different parameters, but insufficiently realize the relation and biases due to sampling characteristics. Knowing how a statistic responds to different sampling characteristics is particularly important when applied in treatment.

Limitations

The choices made to operationalize momentary rQoL can seem arbitrary. Other options could be explored. For example, the rQoL statistic can be computed with negative affect. Additional cut-off scores could be explored to differentiate low and high frequency situations. The behavior settings are defined on persons present, activity and location, although more aspects of the environment could be relevant. Behavior settings are complex entities which include a number of contextual variables[42]. Here we excluded, for example, temporal indices such as the time of the day and limited the categorical options so that a critical mass of

workable data remains. With advances in technology, other factors such as heart rate or weather reports could be more easily combined with ESM data, thereby increasing the accuracy of the behavior setting. In the future, it is possible to harvest big data, such as GPS location, sensor data, or geo-political events to enrich the situational information without increasing subject burden. However, the main purpose of this study was to explore a first operationalization based on available ESM data and test the behavior of the momentary rQoL statistic using Monte Carlo experiments.

Furthermore, the generated samples using real subject data (for the analysis of subjects with low-, average- and high behavior setting variability) were oversaturated and the same records were used repeatedly (due to replacement). The used sample included insufficient subjects with large beep datasets (>1000 observations). It would be interesting to replicate these analyses when longer series become available.

Another limitation is the use of a specific sample of patients with severe mental illness as seeds for the Monte Carlo experiments. These subjects often lead restricted lives with limited variation in daily life activities[31,51]. It would be interesting to see how the frequencies in behavior setting differ between these patients and the general population. This first operationalization was made from a clinical perspective, to explore the possibility of an rQoL statistic that is meaningful to patients with severe mental illness. However, the basic components of the rQoL statistic could be relevant across clinical populations. For example, patients with depression or anxiety could also use the statistic within treatment to optimize their balance in rQoL. Future research should use Monte Carlo simulations on ESM data collected in other populations; to see how the frequency in behavior setting is distributed, to calculate which sample size would work best, and to see in what situation the rQoL statistic can be meaningfully calculated. It is conceivable that situational variability differs between populations (and between individuals).

Implications and further research
This proof-of-concept study indicates that momentary rQoL is a feasible statistic. Monte Carlo experiments provide valuable insight in the behavior of the statistic under different sampling restraints. The methodology can be used to further improve rQoL. Monte Carlo experiments should be used more frequently in ESM studies. Several suggestions were made for future research. The question remains whether reward-related optimized well-being is actually quality of life; maybe a better description is more adequate. The link between the rQoL statistic and mQoL could further be explored, as well as the relation of rQoL with other

(cross-sectional) measures of QoL. Furthermore, the statistic should be explored in other populations that engage in more diversified behavior settings.

It is interesting to explore whether the rQoL statistic improves targeted communication with patients. The statistic could be used to identify situations that result in low rQoL, so that changes can be made in daily life and progress can be monitored (see the supplementary material for a hypothetical case example). A previous ESM-based feedback intervention[37] improved therapeutic outcome. The question remains if well-being can be improved by a person-tailored rQoL feedback intervention that monitors reward experiences in daily life. Shared decision-making is facilitated when clinicians and patients share the same information. ESM data, disclosed by smart feedback, can provide this context. By integrating the clinician's expertise with the goals and knowledge of patients and relatives, and by looking at environmental daily life challenges and opportunities, more suggestions that are realistic can be made for optimizing reward in daily life, possibly leading to improved well-being and QoL.

References

1. McGorry P, van Os J. Redeeming diagnosis in psychiatry: timing versus specificity. *Lancet.* 2013;381(9863):343-345.

2. Evans SC, Reed GM, Roberts MC, et al. Psychologists' perspectives on the diagnostic classification of mental disorders: Results from the WHO-IUPsyS Global Survey. *Int J Psychol.* 2013;48(3):177-193.

3. van Os J. *De DSM-5 voorbij! Persoonlijke diagnostiek in een nieuwe GGZ.* . Leusden: Diagnosis Uitgevers; 2014.

4. van Staden WC. Diagnostic understanding and diagnostic explanation in psychiatry. *Curr Opin Psychiatry.* 2003;16(6):667-671.

5. Delespaul P. *Assessing schizophrenia in daily life: The experience sampling method.* Maastricht: Universitaire Pers Maastricht; 1995.

6. 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(9):1533-1547.

7. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. *Psychol Assess.* 2009;21(4):457-462.

8. Jacobs N, Menne-Lothmann C, Derom C, Thiery E, van Os J, Wichers M. Deconstructing the familiality of variability in momentary negative and positive affect. *Acta Psychiatr Scand.* 2013;127(4):318-327.

9. Parati G, Omboni S, Bilo G. Why is out-of-office blood pressure measurement needed? *Hypertension.* 2009;54(2):181-187.

10. Walz LC, Nauta MH, aan het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: A systematic review. *J Anxiety Disord.* 2014;28(8):925-937.

11. Munsch S, Meyer AH, Milenkovic N, Schlup B, Margraf J, Wilhelm FH. Ecological Momentary Assessment to Evaluate Cognitive-Behavioral Treatment for Binge Eating Disorder. *Int J Eat Disord.* 2009;42(7):648-657.

12. Wichers M, Hartmann J, Kramer I, et al. Translating assessments of the film of daily life into person-tailored feedback interventions in depression. *Acta Psychiatr Scand.* 2011;123(5):402-403.

13. Williams E, Sands N, Elsom S, Prematunga RK. Mental health consumers' perceptions of quality of life and mental health care. *Nurs Health Sci.* 2015;17(3):299-306.

14. Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes.* John Wiley & Sons; 2013.

15. the WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Social science & medicine.* 1995;41(10):1403-1409.

16. Saxena S, Orley J, Group W. Quality of life assessment: the World Health Organization perspective. *European psychiatry.* 1997;12:263s-266s.

17. Spilker B. *Quality of life assessments in clinical trials.* New York: Raven Press; 1990.

18. Auquier P, Simeoni MC, Sapin C, et al. Development and validation of a patient-based health-related quality of life questionnaire in schizophrenia: the S-QoL. *Schizophr Res.* 2003;63(1):137-149.

19. Fleury MJ, Grenier G, Bamvita JM, Tremblay J, Schmitz N, Caron J. Predictors of quality of life in a longitudinal study of users with severe mental disorders. *Health Qual Life Outcomes.* 2013;11(1):92-92.

20. Ruggeri M, Nose M, Bonetto C, et al. Changes and predictors of change in objective and subjective quality of life: Multiwave follow-up study in community psychiatric practice. *Br J Psychiatry.* 2005;187(2):121-130.

21. Yamauchi K, Aki H, Tomotake M, et al. Predictors of subjective and objective quality of life in outpatients with schizophrenia. *Psychiatry Clin Neurosci.* 2008;62(4):404-411.

22. Oliver J, Huxley P, Priebe S, Kaiser W. Measuring the quality of life of severely mentally ill people using the Lancashire Quality of Life Profile. *Social Psychiatry and Psychiatric Epidemiology.* 1997;32(2):76-83.

23. Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester Short Assessment of Quality of Life (MANSA). *International journal of social psychiatry.* 1999;45(1):7-12.

24. Lehman AF. The well-being of chronic mental patients: Assessing their quality of life. *Arch Gen Psychiatry.* 1983;40(4):369-373.

25. Barge-Schaapveld DQ, Nicolson NA, Berkhof J. Quality of life in depression: daily life determinants and variability. *Psychiatry Res.* 1999;88(3):173-189.

26. Maes IH, Delespaul PA, Peters ML, et al. Measuring health-related quality of life by experiences: the experience sampling method. *Value Health.* 2015;18(1):44-51.

27. Gilleen J, Shergill SS, Kapur S. Impaired subjective well-being in schizophrenia is associated with reduced anterior cingulate activity during reward processing. *Psychol Med.* 2015;45(3):589-600.

28. Skinner BF. Two types of conditioned reflex: A reply to Konorski and Miller. *J Gen Psychol.* 1937;16(1):272-279.

29. Staddon JER, Cerutti DT. Operant conditioning. *Annu Rev Psychol.* 2003;54(1):115-144.

30. Flora SR. *The power of reinforcement.* Albany: State University of New York Press; 2004.

31. Lewinsohn PM. A behavioral approach to depression. *Essential papers on depression.* 1974:150-172.

32. Lewinsohn PM, Graf M. Pleasant activities and depression. *J Consult Clin Psychol.* 1973;41(2):261.

33. Sweeney PD, Shaeffer DE, Golin S. Pleasant events, unpleasant events, and depression. *J Pers Soc Psychol.* 1982;43(1):136-144.

34. Wichers Mc, Myin-Germeys I, Jacobs N, et al. Evidence that moment-to-moment variation in positive emotions buffer genetic risk for depression: a momentary assessment twin study. *Acta Psychiatr Scand.* 2007;115(6):451-457.

35. Peeters F, Wichers MC, Barge-Schaapveld DQCM, et al. Reduced Stress-Sensitivity or Increased Reward Experience: The Psychological Mechanism of Response to Antidepressant Medication. *Neuropsychopharmacology.* 2009;34(4):923-931.

36. Geschwind N, Peeters F, Drukker M, van Os J, Wichers M. Mindfulness Training Increases Momentary Positive Emotions and Reward Experience in Adults Vulnerable to Depression: A Randomized Controlled Trial. *J Consult Clin Psychol.* 2011;79(5):618-628.

37. Kramer I, Simons CJP, Hartmann JA, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry.* 2014;13(1):68-77.

38. Hartmann JA, Wichers M, Menne-Lothmann C, et al. Experience sampling-based personalized feedback and positive affect: a randomized controlled trial in depressed patients. *PloS One.* 2015;10(6):e0128095.

39. Van Roekel E, Vrijen C, Heininga VE, Masselink M, Bos EH, Oldehinkel AJ. An Exploratory Randomized Controlled Trial of Personalized Lifestyle Advice and Tandem Skydives as a Means to Reduce Anhedonia. *Behav Ther.* 2017;48(1):76-96.

40. Wichers M, Kasanova Z, Bakker J, et al. From Affective Experience to Motivated Action: Tracking Reward-Seeking and Punishment-Avoidant Behaviour in Real-Life. *PloS One.* 2015;10(6):e0129722.

41. Heft H. An ecological approach to psychology. *Rev Gen Psychol.* 2013;17(2):162.

42. Barker RG. Explorations in ecological psychology. *Am Psychol.* 1965;20(1):1.

43. Barker RG. *Ecological psychology: Concepts and methods for studying the environment of human behavior.* Stanford University Press; 1968.

44. Wicker AW. Perspectives on Behavior Settings: With Illustrations From Allison's Ethnography of a Japanese Hostess Club. *Environ Behav.* 2012;44(4):474-492.

45. Peeters F, Nicolson NA, Berkhof J, Delespaul P, deVries M. Effects of daily events on mood states in major depressive disorder. *J Abnorm Psychol.* 2003;112(2):203.

46. Thompson RJ, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Gotlib IH. The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *J Abnorm Psychol.* 2012;121(4):819.

47. Roekel E, Bennik EC, Bastiaansen JA, et al. Depressive symptoms and the experience of pleasure in daily life: An exploration of associations in early and late adolescence. *J Abnorm Child Psychol.* 2016;44(5):999-1009.

48. van Zelst CJAM. *Inside out: on stereotype awareness, childhood trauma and stigma in psychosis* [Doctoral dissertation]. Maastricht, Maastricht University; 2014.

49. Jacobs N, Van Os J, Derom C, Thiery E, Delespaul P, Wichers M. Neuroticism explained? From a non‐informative vulnerability marker to informative person–context interactions in the realm of daily life. *British Journal of Clinical Psychology.* 2011;50(1):19-32.

50. Mooney CZ. *Monte carlo simulation.* Vol 116: Sage Publications; 1997.

51. Holloway F, Carson J. Quality of life in severe mental illness. *International Review of Psychiatry.* 2002;14(3):175-184.

Internet site: www.psymate.eu

Supplementary material

List of abbreviations

ESM	Experience Sampling Method (or EMA: Ecological Momentary Assessment)
rQoL	reward-related Quality of Life
mQoL	Momentary Quality of Life
MCs	Monte Carlo simulations
BS	Behavior Setting (context or situation)
BS_W	BS: What = situation defined by current activity
BS_WW	BS: What Who = situation defined by combining current activity and company
BS_WWW	BS: What Who Where = situation defined by combining current activity, company, location
PA	Positive Affect
NA	Negative Affect
N	Number of observations or so-called beeps

Clinical implications rQoL statistic –hypothetical case example-

Patient X has used ESM for several weeks to monitor her progress towards recovery. She struggles with her daily routine and does not feel satisfied with the way her life is going at the moment. However, she does not really know what to change in her life to feel satisfied again. Together with her clinician, ESM data are discussed using an online reporting page that explores the subject's mental state and context using interactive graphs and figures. The reporting page is intended as a tool for collaboration and can be used to generate functional analyses. For each beep moment, a rQoL score is presented on the screen, ranging from negative to positive, with 0 as a neutral point. A positive score represents feeling good in a situation that also occurs frequently (plus-plus) or feeling less good in a situation that is infrequent (minus-minus). Negative scores reflect an imbalance in daily life reward experiences (feeling good in infrequent (plus-minus) or bad in frequent situations (minus-plus). Series of rQoL scores are displayed over time, to identify time trends and outliers. At each time point, the rQoL score can be explored to zoom into the behavior setting of that moment. The exploration assesses both strengths and weaknesses of the individual. Together with the clinician, patient X decides on action points that could lead to small changes in daily life that are deemed realistic. Continued monitoring provides the necessary feedback to see if progress can be made towards improved reward-efficiency and ultimate to well-being.

STATA™ (v13.0) do file for Monte Carlo experiments

```
*ssc install egenmore
clear all
program define mc_reward, rclass
file open Total_OK using "C:\...\result_file_of  simulations.csv",  write
replace
* writing header
        file write Total_OK "Iteration"
        file write Total_OK _tab
        file write Total_OK "cnt catheg"
        file write Total_OK _tab
        file write Total_OK "Mean rPA"
        file write Total_OK _tab
        file write Total_OK "sd rPA"
        file write Total_OK _tab
        file write Total_OK "min rPA"
        file write Total_OK _tab
        file write Total_OK "max rPA"
        file write Total_OK _tab
        file write Total_OK "corr PArPA"
        file write Total_OK  _n
set more off
* iterations
forvalues xi=1(1)1000 {
        clear
        use "C:\...\reward_datafile_of_ESM_seeds.dta"
        bsample 100
        egen cnt_bs=nvals(behav_setting)
        egen mPA=mean(sub_pa)
        egen sdPA=sd(sub_pa)
        gen zPA=(sub_pa-mPA)/sdPA
        so behav_setting
        by behav_setting:egen cfreq=count(behav_setting)
        so cfreq behav_setting
        egen crfreq=rank(cfreq)
        generate crprop=crfreq/_N
        gen rPA=zPA*(crprop-.50)
        summarize cnt_bs
        file write Total_OK "Iteration"
        file write Total_OK (`xi')
        file write Total_OK _tab
        file write Total_OK (string(r(mean)))
        summarize rPA
        file write Total_OK _tab
        file write Total_OK (string(r(mean)))
        file write Total_OK _tab
        file write Total_OK (string(r(sd)))
        file write Total_OK _tab
        file write Total_OK (string(r(min)))
        file write Total_OK _tab
        file write Total_OK (string(r(max)))
        correlate sub_pa rPA
        file write Total_OK _tab
        file write Total_OK (string(r(rho)))
        file write Total_OK  _n
}
file close Total_OK
clear
import delimited "C:\...\result_file_of simulations.csv"
su
histogram cnt catheg
histogram mean rpa
```

Overview Monte Carlo experiments
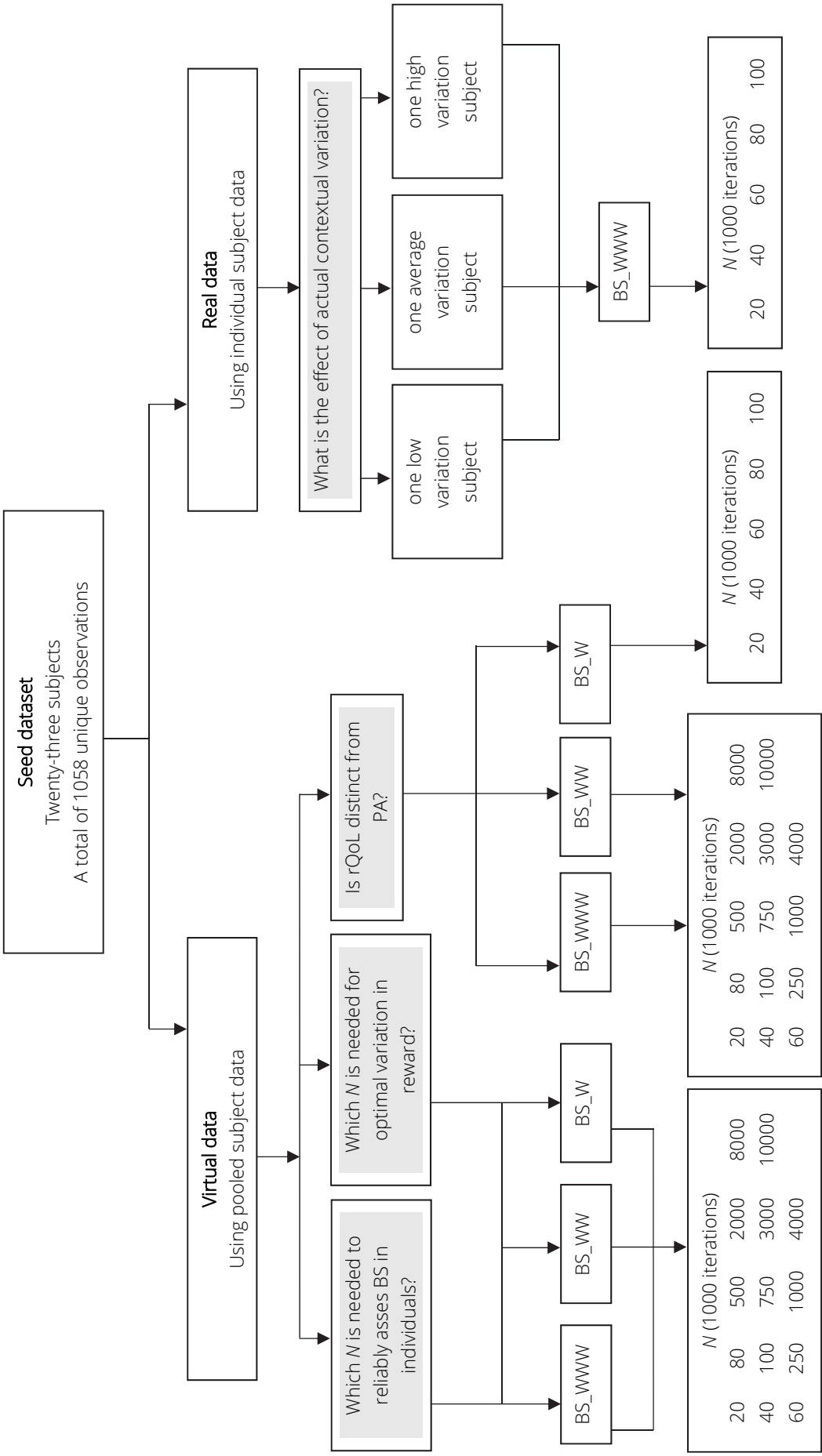
Part III

Cognitive relevance

7

Investigating mood and cognition in the daily lives of patients with a diagnosis of bipolar disorder – a pilot study

S. Verhagen, F. Peeters, B. Arts, T. Lataster, J. van Os, P. Delespaul, N. Jabben
This thesis.

Abstract

Background: Neurocognitive alterations are common in mental disorders and often persist after symptom remission, influencing aspects of daily interpersonal functioning. Few studies have focused on the dynamics of cognition in relation to mood and symptom changes that span episodes and within-day fluctuations. In bipolar disorder, mood states change, and cognition is affected, offering an attractive paradigm for investigating cognition related changes in daily life.

Objective: Investigating the feasibility of two momentary cognition tasks as add-on to experience sampling in patients with bipolar disorder in remission, and exploring associations between mood, context, and cognition.

Method: A symbol-task measuring processing speed and a quiz-task measuring liberal acceptance bias were implemented as a pilot in the Experience Sampling Method (ESM). Patients completed ESM ten times daily, for six days and were randomized to either a symbol-task (n=5) or a quiz-task (n=5) protocol. Personal ESM experiences were collected afterwards. Positive affect, negative affect, and cognition outcomes (speed; accuracy; liberal acceptance bias) were calculated.

Results: ESM completion was 82% (including the symbol-task), and 59% for the quiz-task. Patient experiences were sufficiently represented, with few uncertainties, and a minimal assessment burden. Participants completed on average 14 trials within 30-seconds (speed), with 98% correct (accuracy) on the symbol-task. A liberal acceptance bias occurred in 3% of the assessments and was only present in two patients. Power problems limited further explorations.

Conclusion: ESM-cognition implementations are feasible and especially the symbol-task warrants further research. Both tasks need adjustments for a fluent daily life integration. More data is needed to examine cognitive variation in relation to affective fluctuations and contextual factors. Future studies should assess general- and patient populations to further test and improve feasibility and to examine clinical information about relevant micro-dynamics in daily life.

Introduction

People with mental disorders often experience cognitive alterations and sometimes do not regain baseline cognitive abilities for long periods after symptom remission[1,2]. Cognitive alterations have been postulated as endophenotypes or underlying vulnerabilities, being also present – albeit to a lesser extent- in the healthy first degree relatives[3,4]. In patients with bipolar disorder, variability in cognitive alteration was found during shifting states of depression and mania[5]. Furthermore, neurocognitive alterations cluster in domains of executive functioning, verbal learning, memory, attention, and processing speed[3,6-8]. Cognitive alterations can impact functioning, sometimes resulting in dissociation between symptomatic and functional recovery. Persisting neurocognitive alterations are thought to represent important determinants of psychosocial disability[9-11] and are therefore a relevant research and treatment target.

The exact origin of the neurocognitive problems in psychopathology is unknown and appears to be multi-factorial. Neurocognitive complaints differ among individual patients within a similar diagnostic category and within individuals over time[12,13]. These heterogeneities combined with the evidence for cognitive alterations as endophenotypes, suggest a neurodevelopmental origin. However, clinical factors such as social marginalization due to chronicity and despair caused by a history of psychosis or manic episodes increase the likelihood of neurocognitive alterations[14,15].

Social cognition is an important domain in psychopathology research because it is assumed relevant for daily interpersonal functioning[16]. Social cognition alterations are less evident in patients with bipolar disorder, compared to patients with a psychotic disorder[17,18]. A meta-analysis of social cognition in euthymic bipolar patients found alterations in emotion processing and theory of mind. However, there was much heterogeneity and it was suggested that the results were caused by a subgroup of patients[19]. Furthermore, social cognitive functioning in bipolar disorder seems state dependent and varies with symptoms[18].

It may be productive to study the bipolar phenotype for research linking cognition to mood states. However, longitudinal studies of (social) cognitive functioning in relation to clinical symptoms are lacking. It remains unclear whether patients only experience social cognitive alterations during mood episodes and how mood influences social performance. Sporadically, mood is experimentally induced in labs and changes in cognition assessed[20]. However, most research to date has used an episodic timeframe. Scholars argue that cognition also varies with mood changes in reaction to environmental cues[21,22]. The timeframe to assess the relation of mood and cognition then shifts from illness episodes

(months) to daily life (hours). This provides a powerful rationale for investigating these processes in context.

The Experience Sampling Method (ESM), also called Ecological Momentary Assessment (EMA), is an assessment strategy to study symptoms in daily life[23,24]. ESM uses a structured diary technique and provides ecologically valid real-time and real-world monitoring of mood, behavior, and symptomatology in relation to contextual factors[21]. Repeated assessments are performed in normal daily life. Consequently, ESM typically provides information of symptoms and situational characteristics over hours and days. Bipolar disorder research may benefit from measuring within-day mood shifts, as symptoms fluctuate dynamically over time[25]. Prospective ESM assessments avoid recall bias and are better suited to record micro level processes within the individual[21]. Adding objective cognitive assessments to ESM would provide a relevant extension that could help shed light on underlying mechanisms and the daily life functioning of patients.

Two momentary cognitive tasks were developed for the assessment of cognition in daily life, as part of a large ESM study in patients with severe mental illness (EU-GEI HEALTH-F2-2009-231909)[26]. The current pilot study investigated the feasibility of both objective momentary tasks as an add-on cognition assessment to experience sampling in patients with remitted bipolar disorder. Furthermore, the within-subject longitudinal design of ESM was used to explore the relation between cognition, mood, and symptomatology and untangle the determinants of their fluctuations. Knowledge in this regard could aid the future development of meaningful cognitive assessments for both research as well as clinical purposes.

Method
Participants
Ten patients with either bipolar I disorder or bipolar II disorder in remission, were recruited from outpatient mental health clinics in the Netherlands. Inclusion was based on a lifetime presence of a diagnosis of bipolar I or bipolar II disorder according to DSM-IV criteria. All participants were between 18 and 65 years old, with sufficient knowledge of the Dutch language. Exclusion criteria were not being able to provide informed consent (as judged by the clinician), current alcohol abuse (i.e., within the last 3 months), frequent drug use (i.e., > 2 times a week), mental retardation (as defined by IQ scores below 70), a history of neurological illness or traumatic brain injury with loss of consciousness, and

electroconvulsive treatment in the past year. The local medical ethics committee approved the study (NL5168.068.15) and all participants provided a written informed consent.

Measurements
Questionnaires

Participants were interviewed with the Dutch adaption of the Questionnaire for Bipolar Illness (QBP-NL, part A and B)[27]. The QBP-NL is a well-validated questionnaire, which takes approximately 60 minutes to complete. Demographic characteristics, illness history, and medication and drug use were recorded. Severity of depressive and manic symptomatology was assessed using the Quick Inventory of Depression Symptomatology – short version (QIDS-C). This depression rating scale consists of 16 items and is completed by the clinician[28]. The Young Mania Rating Scale (YMRS) consists of 11 items and is also completed by the clinician, and was used to assess current manic symptoms[29].

Experience sampling protocol

Experience-sampling data was collected using the palm-top PsyMate™ device (http://www.psymate.eu/). Sampling occurred ten times a day, for six consecutive days (Thursday to Tuesday) using a short questionnaire lasting approximately two minutes. These self-report questionnaires were programmed at semi-random intervals of 90 minute blocks, between 7.30 AM and 10.30 PM. Beeps signaled when a questionnaire had to be completed and patients were instructed to respond immediately. The beep questionnaire consisted of ten mood items, eight psychopathology items, three context items (activity (what), location (where), company (with whom)), and an item assessing beep disturbance. See supplementary materials table 1 for an overview. Most items were rated on a seven-point Likert scale (1=not at all, 4=moderate, 7=very much), with the exception of the item "How (un)pleasant was the event?" which was rated on a seven-point bipolar scale (-3=very unpleasant, 0=neutral, 3=very pleasant). The context items were pre-coded in categories. After the sampling week, patients completed two short questionnaires to evaluate the ESM period. Items included both seven-point Likert scale ratings as well as open-ended questions, assessing possible difficulties with the device and its use in daily life.

Cognition in daily life

Two momentary cognition tasks were implemented as add-on to the ESM procedure, focusing on assessments of processing speed (neuro-cognition) and a liberal acceptance bias (related to social cognition).

The neuro-cognition task is a momentary Digit Symbol Substitution task - or symbol task in short, partly derived from the Symbol Substitution-Coding subtest of the Wechsler Adult Intelligence Scale[30]. The original task measures both information processing speed and short-term working memory. The modified version for ESM primarily assesses information processing speed. After the regular beep questionnaire, patients saw a row of digits (1 to 9) and directly under these, nine corresponding symbols (encoding information) at the top of the screen. At the bottom, a row of the same nine symbols was presented, but in a different random order. Each symbol at the bottom was an active response button and in the middle of the screen, one digit was presented. Patients were instructed to select the button representing the symbol in the bottom row that matches the digit according to the top row code and to complete as many trials as possible within 30 seconds, while making as few mistakes as possible. See figure 1 for a visual representation of the task. There were sixty sessions (similar to the beep total), with a maximum of 45 trials per beep questionnaire. To minimize learning, ten unique encoding combinations with corresponding answer keys were presented at random. The outcome measures were number of trials within a 30-second timeframe (speed) and the percentage of correct trials (accuracy).
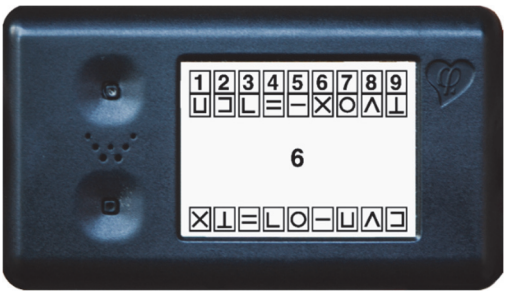
Figure 1
Momentary Digit Symbol Substitution task on the PsyMate™ device.

The social cognition task was derived from an earlier laboratory study that used a forced choice reasoning task with quiz questions to assess liberal acceptance bias[31]. This task is the same as the one used by Reininghaus and colleagues[32]. The mobile Probabilistic Quiz task (previously called experimental ESM task) – or quiz task in short - used sixty moderately difficult multiple choice quiz questions. They were randomized and one was presented for each beep. Answer options were selected that had a similar face-value likelihood of being correct (e.g., "To which country belongs Greenland?" "A: Iceland, B: The Netherlands, C: Denmark, D: Canada").

After viewing the quiz question, patients were required to provide probability estimates for all four answer options to the question: "*What is the chance that A, B, C, or D is correct?*". Ratings were presented in 10%-step categories (i.e., 0%, 1-9%, …, 90-99%, 100%). After selecting a probability answer for each of the multiple-choice options, patients were asked if they wanted to endorse a specific answer (yes or no). If the answer was yes, patients were required to indicate which one (A, B, C, or D). The next question was whether they wanted to reject a specific answer (yes or no), followed by which one (A, B, C, or D).

According to Reinighausen and colleagues[32], a liberal acceptance bias or reasoning bias is apparent if patients select an answer they had given a low subjective probability estimate. This suggests a low decision threshold compared to a rational expectation. One out of four answers is always correct, so the likelihood of guessing the right answer at random is 25%. Assuming that patients are unfamiliar with the right answer, endorsing a multiple-choice option with 20-29% or lower personally rated likelihood is illogical, so that a liberal acceptance bias or reasoning bias is assumed apparent. Another indicator of a (more pronounced) bias is an incorrect decision to endorse an answer (as being correct), that participants rated having a low probability estimate (<= 20-29%). Thus, the two outcome measures are a liberal acceptance bias 1) based on any decision and 2) based on an endorsed answer being incorrect.

Procedure

Participants were provided with the PsyMate™ device on which the ESM protocol was activated. They received either ESM followed by the symbol task or ESM followed by the quiz task, depending on randomization. During an hour-long briefing, the ESM procedure was explained and test trials were completed to familiarize participants with the device, the questions, and the cognition task. The participants were instructed to practice on the briefing day, and that the six-day ESM period would start on the following day. On the second day of data collection, participants were contacted by phone to check whether everything was working well and to ask any remaining questions. After the ESM period, a debriefing session took place. Participants handed in the device and completed the debriefing questionnaire.

Analyses

Descriptive statistics (averages and frequencies) were used to assess patient characteristics, feasibility, and acceptability. The completion rate was defined as the overall percentage completed beeps, calculated from the total number of possible beeps per individual, averaged over individuals. Principal Component Analysis (PCA) with varimax (orthogonal) rotation was performed over the averaged mood item scores per participant to assess the

structure of the 10 ESM mood items. Thereafter, reliability analyses (i.e., Cronbach's alpha) were run to determine the internal consistency of each factor. Retrospectively, the ESM data was combined with either the symbol task or the quiz task outcome measures. Pearson pairwise correlations were used to examine the relation among the mood variables and cognition outcomes. To see whether cognitive performance varied over time, to assess possible learning effects (within-day and between-day time effects), and to explore relations with ESM mood items, psychopathology items, and context, Pearson's pairwise correlations were calculated and exploratory multilevel-regression analyses were run. The multilevel structure is recommended because ESM data consists of beep assessments (level 1) nested within participants (level 2). Two dummy variables were created for location (at home vs. somewhere else) and company (alone vs. with others). However, these explorations yielded limited information due to power constrictions (5 participants per cognition task) and were therefore excluded from the results. A two-sided significance level of .05 was used and all analyses were performed using STATA, version 13.1[33].

Results

Patient sample
In the period from March until September 2016, ten patients were included. All participants were Dutch and of white ethnic group. The average age was 48 years (SD=14.42, range 23-64) and 70% were women. Six patients were diagnosed with a Bipolar II disorder and four patients with a Bipolar I disorder according to the DSM-IV-TR criteria, with three patients in partial remission and seven in full remission. Compliance rates with the ESM protocol were good, with no participant reporting less than 20 beeps (i.e., the cut-off score for reliable ESM analyses[34]). In total, 486 beep moments were collected with a mean of 49 completed beeps per patient (SD=4.58, range=38-55). On average, participants completed 82% of the total possible responses.

Subjective ESM evaluation
The ESM items represented participant experiences relatively well (M=5.2, SD=.88). Some difficulties were encountered with the distinction between mood items in the same conceptual domain (e.g., angry and agitated). The categorical questions related to context were not always inclusive and participants sometimes struggled to find a good fit for their activities. The ESM procedure had little influence on mood (M=2.4, SD=1.90), activities (M=2.2, SD=1.62), and social contacts (M=2.7, SD=1.95). Sometimes, participants felt that they made mistakes when navigating the device (M=2.9, SD=.99). These mistakes were often the result of going too quickly, and correcting the mistake by going back to a previous item was not

possible. The PsyMate™ was easy to use (reverse coded: *M*=1.5, *SD*=.97), with few unclear questions (*M*=3.2, *SD*=1.40). Problems were reported with ambiguous mood ratings. Two participants specifically mentioned the item "*It feels as if everything goes by itself*" as unclear. The ESM procedure was experienced as minimally burdensome with regard to the amount of beeps (*M*=3.6, *SD*=2.20), the duration of an assessment (*M*=2.8, *SD*=1.87), and the sound of the beep (*M*=1.9, *SD*=1.3). One participant remarked that the ESM procedure was feasible for a short period, but not longer. She advised to make the questionnaire less redundant, to avoid annoyance in repeated assessments. Another advice was to implement the PsyMate™ in a smartphone app: a more convenient tool and less easy to forget. The participants who completed the symbol task remarked that the screen sometimes responded too fast or too slow (three times). Some participants who completed the quiz task experienced the questions as difficult and therefore frustrating (two times).

Factor structure ESM mood items

The exploratory within-subject PCA of the ten mood items resulted in three factors, each having a value larger than 1 (4.03 for *cheerful*, *satisfied*, *relaxed* and *enthusiastic*; 1.25 for *insecure*, *lonely*, *anxious* and *down*; 1.04 for *agitated* and *guilty*) with a cumulative explained variance of 63%. Range item loading; factor 1 (*N*=487, range .70-.80), factor 2 (*N*=487, range .51-.85), factor 3 (*N*=487, range .72-.82). Next, reliability analyses were performed, with a Cronbach's alpha of .84 for factor 1 (positive mood), and .74 for factor 2 and 3 together (negative mood). Consequently, we calculated a positive affect subscale by computing the sum of the four positive mood items (*M*=4.28, *SD*=.80, range 3-5.45), and a negative affect subscale form the six negative mood items (*M*=1.84, *SD*=.84, range 1.04-3.79). A Pearson's pairwise correlation between the positive and negative affect scales yielded a negative correlation (*r*=-.35, *p*<.001).

Momentary digit symbol substitution task

Five patients were presented the additional symbol task following the beep questionnaires. In total 247 (82% of the total beeps) symbol task beep moments were collected, with a mean of 50 completed beeps per participant (*SD*=3.99, range = 43-55). Each task consisted of 45 trials, of which patients completed on average 14.19 trials (*SD*=2.56, range 5-22) within the 30 second timeframe. In total, 98% of the trials were correct (*SD*=3.75, range 75%-100%). Pearson's pairwise correlation yielded a positive correlation of .14 (*p*=.03) between both symbol task outcome measures.

Pearson's pairwise correlations were also used to correlate the cognition outcomes with relevant ESM items (see supplementary materials table 2 for person-centered averages).

Number of trials was positively associated with PA ($r=.34$, $p<.001$). Furthermore, the speed outcome was positively associated with feeling energetic ($r=.45$, $p<.001$), 'as if everything goes by itself' ($r=.42$, $p<.001$), harried ($r=.19$, $p=.003$), talkative ($r=.43$, $p<.001$) and impatient ($r=.35$, $p<.001$). Number of trials correct (accuracy) was positively associated with feeling as if everything goes by itself ($r=.03$, $p=.03$). No correlations were found between speed and respectively NA ($r=.04$, $p=.57$), the item busy in my head ($r=.03$, $p=.63$), feeling suspicious ($r=-.09$, $p=.15$), worrying ($r=-.04$, $p=.53$), company ($r=.11$, $p=.09$), location ($r=.01$, $p=.83$) and beep disruption ($r=-.10$, $p=.11$). Accuracy did not correlate with PA ($r=-.001$, $p=.98$), NA ($r=.04$, $p=.58$), the item busy in my head ($r=.11$, $p=.09$), worrying ($r=.06$, $p=.39$), feeling energetic ($r=.001$, $p=.99$), harried ($r=-.03$, $p=.61$), suspicious ($r=-.07$, $p=.31$), impatient ($r=.17$, $p=.09$), and talkative ($r=-.04$, $p=.51$). Furthermore, accuracy did not correlate with company ($r=-.04$, $p=.56$), location ($r=-.08$, $p=.23$), and beep disruption ($r=-.05$, $p=.41$)

Mobile probabilistic quiz task
Data from one patient was excluded because that patient reported searching for the correct answers on the internet. In total, four patients provided a sample consisting of 178 quiz moments (59% of the total beeps), with a mean of 48 completed beeps per participant ($SD=4.97$, range 38-52). In 51% of the time, a decision to endorse an alternative was made (91 out of 178). Based on the criteria of endorsing an answer combined with the probability estimate they gave for that specific answer, a liberal acceptance bias was present in 2.8% of possible instances (5 out of 178). The same script was run with the difference of looking at correct or incorrect items as criteria, instead of just looking at the decision yes or no. Here, a liberal acceptance bias was also present in 2.8% (5 out of 178): providing an incorrect answer with a low probability estimate. In 36% of instances, an incorrect decision was made as opposed to a correct or no decision (64 out of 178). Looking at individual data, only two patients showed a liberal acceptance bias. Individual results are summarized in Table 1.

Table 1

Individual outcomes on the mobile probabilistic quiz task

		Participant			
		1	2	3	4
Decision to endorse	*n*(%)	25 (86%)	9 (18%)	20 (44%)	33 (63%)
LA bias (any decision)	*n*(%)	0	2 (4%)	0	3 (6%)
Incorrect decision	*n*(%)	19 (66%)	6 (12%)	15 (33%)	24 (46%)
LA bias (incorrect decision)	*n*(%)	0	2 (4%)	0	3 (6%)
Probability estimate	*M*(*SD*; range)	8.7(2.2; 5-12)	7.4(3.5; 5-12)	10.1(2.0; 2-11)	9.3(3.9; 1-12)

Note. LA = Liberal Acceptance Bias.

Discussion

The aim of this pilot study was to assess the feasibility of an add-on experimental cognition task to experience sampling in patients with explicit mood shifts (bipolar disorder). It is important to understand the micro-level processes that are involved with cognition and mood states in the context of daily life.

General ESM findings

Adding a cognition task to the ESM protocol did not alter ESM completion as compared to other ESM studies[35-37]. Furthermore, patients reported both tasks as feasible and acceptable in addition to the ESM questionnaire. Participant feedback suggests that some adjustments may be helpful. In order to assess mood states intuitively, the mood items need to be distinct and vary across time[35]. Items that remain vague are frustrating and increase assessment burden; an example is the item "*It feels like everything goes by itself*". Some patients showed variation on this item, whereas others did not. Non-varying items can be informative, but become redundant when assessed repeatedly[38]. The items agitated and guilty seem to have a slightly different connotation from the other negative affect items (insecure, lonely, anxious, and down). However, reliability was still sufficient to calculate a separate PA and NA score for further cognition explorations. These aggregations are in line with other ESM studies in the mental health field that focus on dynamic mood[21,39]. Although helpful, large between-individual differences exist between the single mood items, next to subtle valence

differences. In future studies, it might be necessary to consider single mood items in relation to cognitive performance.

Symbol task findings

Participant ratings and task outcomes indicate the momentary digit-symbol substitution task to be acceptable and feasible for repeated assessments in everyday life. Patients reported to have made mistakes due to the (fast or slow) responsiveness of the device. This needs to be addressed in future task adaptions. However, the subjective mistakes were incongruent with task outcome, showing a high ceiling effect (98% correct) with few actual mistakes. The original processing speed measure also proves highly accurate, whereas speed is more sensitive to cognitive performance variability[30]. In the symbol task, patients completed on average 14 trials and although the standard deviation (*SD*=3) and range (5-22) seem to indicate variability, firm conclusions cannot be drawn. Tentative evidence towards meaningful cognitive variation in relation to mood and context was found when exploring overall ESM correlations. PA, feeling energetic, harried, impatient, and talkative were associated with speed, and "*feeling like everything goes by itself*" with both speed and accuracy. All items seem to imply a high energy level and a relation with speed is therefore unsurprising. Fatigue as the inverse of high energy has been associated with decrements in information processing performance[40]. Exploratory analyses on learning effects (examining within-day time effects and between-day time effects) showed a trend towards variability over time for speed. Initial multilevel-regression explorations also yielded tentative evidence for a negative association between distraction, as well as being in company of others, with speed. Detrimental performance while distracted or in busy environments is intuitively logic and fits with previous findings[41,42]. However, a power problem limited the statistical exploration with multilevel regression analyses. Future studies should explore these findings further.

Quiz findings

The mobile probabilistic quiz task was generally experienced as acceptable and feasible, although patients reported frustration due to task difficulty. This subjective feeling seems reflected in the relatively low response rate (59%), especially when compared to the symbol task responses (82%). The lower response rate suggests that patients stopped prematurely, either to avoid quiz completion or because they were unable to finish in their current environment. Probability estimates judgements were required for each answer option, which might be too cumbersome for high frequency sampling. For an intuitive daily life integration, the assessment burden should be kept to a minimum[38]. Finding out what led to lowered

response rate is therefore important. Future studies could consider a debriefing questionnaire that is dedicated to a subjective evaluation of a new cognition task.

Task conceptualization was complex and warrants adjustments. One patient confessed to have looked up the answers, making the data invalid to assess liberal acceptance bias. Reliance on participant compliance and prior knowledge makes it important to control for these factors. While implementing momentary probabilistic tasks, intellectual abilities and years of education need to be taken into account[32]. However, fully controlling for someone's level of general knowledge remains difficult. The task is less suited for longer assessment periods with a similar sampling frequency, because quiz options would run out and participants already may be frustrated. A solution to minimize look-up times is to limit the response time to a question. Timestamps or response times were not available and should be considered in future adaptations. Only two patients showed some evidence towards a liberal acceptance bias (3% of the total), suggesting low task sensitivity, or a limited presence of the bias in the current patient group. A lowered decision threshold has previously been associated with psychotic symptoms and mechanisms of aberrant salience[43,44]. A meta-analysis showed that alterations in cognitive functioning are overall stronger in patients suffering from psychosis, compared to bipolar disorder[3]. It is possible that the task is less relevant for bipolar disorder, especially when manic symptoms have remitted. Previously, Reininghaus and colleagues have applied the task successfully[32]. The liberal acceptance bias outcome was able to distinguish between patients with psychotic experiences and healthy controls. Good concurrent validity with psychotic experiences and clinical status was furthermore shown. Interestingly, affective disturbance was not associated with the presence of a reasoning bias[32]. In the current sample, further exploration of the associations between mood, symptoms, context, and cognition were omitted due to the low rate of the bias. It may be necessary to increase the number of trials within a beep, so that the likelihood of capturing a liberal acceptance bias is increased and the possible influence of coincidences or distractions minimized.

The task was meant to assess social cognition, but taps only indirectly into the cognitive domain. Social cognitive performance is by definition ecological and embedded in daily life [45]. A more direct measure would be clinically informative, but is difficult to implement without asking people to reflect on current social interactions. Future studies should reflect on the clinical usefulness and think about the design of a good objective social cognition task, used for repeated daily assessments.

Conclusion

This pilot study provided feasibility and initial task-performance information for remitted patients with a bipolar disorder. Task adjustments are recommended, with the symbol task showing the most promise as a broadly applicable and well-suited task for repeated assessments. More research is needed to establish the usability, feasibility and validity of a combined ESM and cognition paradigm in varying populations with and without cognitive complaints. Larger sample sizes are needed to investigate possible learning effects, time-varying factors, and associations among relevant ESM outcomes and cognitive performance. Comparisons with related cross-sectional cognition measures could establish convergent validity. Technological advances have made it possible to use smartphone applications, instead of PDA's for ESM implementation. The use of smartphones greatly improve data logistics and seems a promising tool for future ESM research and clinical practice[46]. Mobile cognition measures show promise and insight into dynamic cognitive performance can enhance clinical meaningful information for researchers, clinicians and patients alike.

References

1. Hasselbalch BJ, Knorr U, Kessing LV. Cognitive impairment in the remitted state of unipolar depressive disorder: a systematic review. *Journal of affective disorders.* 2011;134(1-3):20-31.

2. Millan MJ, Agid Y, Brüne M, et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature reviews Drug discovery.* 2012;11(2):141.

3. Arts B, Jabben N, Krabbendam L, Van Os J. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychological medicine.* 2008;38(6):771-785.

4. Snitz BE, MacDonald III AW, Carter CS. Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. 2005.

5. Martínez-Arán A, Vieta E, Reinares M, et al. Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder. *American Journal of Psychiatry.* 2004;161(2):262-270.

6. Bora E, Yücel M, Pantelis C, Berk M. Meta-analytic review of neurocognition in bipolar II disorder. *Acta Psychiatrica Scandinavica.* 2011;123(3):165-174.

7. Robinson LJ, Thompson JM, Gallagher P, et al. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of affective disorders.* 2006;93(1):105-115.

8. Torres I, Boudreau V, Yatham L. Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. *Acta Psychiatrica Scandinavica.* 2007;116(s434):17-26.

9. Depp CA, Mausbach BT, Harmell AL, et al. Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder. *Bipolar disorders.* 2012;14(3):217-226.

10. Martinez-Aran A, Vieta E, Torrent C, et al. Functional outcome in bipolar disorder: the role of clinical and cognitive factors. *Bipolar disorders.* 2007;9(1-2):103-113.

11. Wingo AP, Harvey PD, Baldessarini RJ. Neurocognitive impairment in bipolar disorder patients: functional implications. *Bipolar disorders.* 2009;11(2):113-125.

12. Martino DJ, Strejilevich SA, Scápola M, et al. Heterogeneity in cognitive functioning among patients with bipolar disorder. *Journal of affective disorders.* 2008;109(1-2):149-156.

13. Joyce EM, Roiser JP. Cognitive heterogeneity in schizophrenia. *Current opinion in psychiatry.* 2007;20(3):268.

14. Lewandowski K, Cohen B, Öngur D. Evolution of neuropsychological dysfunction during the course of schizophrenia and bipolar disorder. *Psychological medicine.* 2011;41(2):225-241.

15. López-Jaramillo C, Lopera-Vásquez J, Gallo A, et al. Effects of recurrence on the cognitive performance of patients with bipolar I disorder: implications for relapse prevention and treatment adherence. *Bipolar disorders.* 2010;12(5):557-567.

16. Fett A-KJ, Viechtbauer W, Penn DL, van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neuroscience & Biobehavioral Reviews.* 2011;35(3):573-588.

17. Burdick K, Goldberg J, Harrow M. Neurocognitive dysfunction and psychosocial outcome in patients with bipolar I disorder at 15-year follow-up. *Acta Psychiatrica Scandinavica.* 2010;122(6):499-506.

18. Lee J, Altshuler L, Glahn DC, Miklowitz DJ, Ochsner K, Green MF. Social and nonsocial cognition in bipolar disorder and schizophrenia: relative levels of impairment. *American Journal of Psychiatry.* 2013;170(3):334-341.

19. Samamé C, Martino D, Strejilevich S. Social cognition in euthymic bipolar disorder: systematic review and meta-analytic approach. *Acta Psychiatrica Scandinavica.* 2012;125(4):266-280.

20. Chepenik LG, Cornew LA, Farah MJ. The influence of sad mood on cognition. *Emotion.* 2007;7(4):802.

21. 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. *Psychological medicine.* 2009;39(9):1533-1547.

22. van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature.* 2010;468(7321):203.

23. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. *The Journal of nervous and mental disease.* 1987;175(9):526.

24. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1-32.

25. Nica EI, Links PS. Affective instability in borderline personality disorder: Experience sampling findings. *Current psychiatry reports.* 2009;11(1):74-81.

26. Korver N, Quee PJ, Boos H, Simons CJ, Haan L. Genetic Risk and Outcome of Psychosis (GROUP), a multi site longitudinal cohort study focused on gene–environment interaction: objectives, sample characteristics, recruitment and assessment methods. *International journal of methods in psychiatric research.* 2012;21(3):205-221.

27. Leverich GS, Nolen WA, Rush AJ, et al. The stanley foundation bipolar treatment outcome network: I. Longitudinal methodology. *Journal of affective disorders.* 2001;67(1):33-44.

28. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The inventory of depressive symptomatology (IDS): psychometric properties. *Psychological medicine.* 1996;26(3):477-486.

29. Aprahamian I, Ladeira RB, Diniz BS, Forlenza OV, Nunes PV. Cognitive impairment in euthymic older adults with bipolar disorder: a controlled study using cognitive screening tests. *The American Journal of Geriatric Psychiatry.* 2014;22(4):389-397.

30. Wechsler D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson.* 2008;22:498.

31. Moritz S, Woodward TS, Hausmann D. Incautious reasoning as a pathogenetic factor for the development of psychotic symptoms in schizophrenia. *Schizophrenia Bulletin.* 2005;32(2):327-331.

32. Reininghaus U, Oorschot M, Moritz S, et al. Liberal acceptance bias, momentary aberrant salience, and psychosis: an experimental experience sampling study. *Schizophrenia bulletin.* 2018;45(4):871-882.

33. StataCorp L. Stata multilevel mixed-effects reference manual. *College Station, TX: StataCorp LP.* 2013.

34. Delespaul PA. *Assessing schizophrenia in daily life: The experience sampling method.* Maastricht university; 1995.

35. Palmier-Claus JE, Myin-Germeys I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatrica Scandinavica.* 2011;123(1):12-20.

36. Havermans R, Nicolson NA. Daily hassles, uplifts, and time use in individuals with bipolar disorder in remission. *The Journal of nervous and mental disease.* 2007;195(9):745-751.

37. Kimhy D, Delespaul P, Corcoran C, Ahn H, Yale S, Malaspina D. Computerized experience sampling method (ESMc): assessing feasibility and validity among individuals with schizophrenia. *Journal of psychiatric research.* 2006;40(3):221-230.

38. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evidence-based mental health.* 2016;19(3):86-89.

39. Brans K, Koval P, Verduyn P, Lim YL, Kuppens P. The regulation of negative and positive affect in daily life. *Emotion.* 2013;13(5):926.

40. Johansson B, Berglund P, Rönnbäck L. Mental fatigue and impaired information processing after mild and moderate traumatic brain injury. *Brain injury.* 2009;23(13-14):1027-1040.

41. Rodrigues PF, Pandeirada JN. Attention and working memory in elderly: the influence of a distracting environment. *Cognitive processing.* 2015;16(1):97-109.

42. Lavie N. Attention, distraction, and cognitive control under load. *Current directions in psychological science.* 2010;19(3):143-148.

43. Moritz S, Pfuhl G, Lüdtke T, Menon M, Balzan RP, Andreou C. A two-stage cognitive theory of the positive symptoms of psychosis. Highlighting the role of lowered decision thresholds. *Journal of behavior therapy and experimental psychiatry.* 2017;56:12-20.

44. Menon M, Mizrahi R, Kapur S. 'Jumping to conclusions' and delusions in psychosis: relationship and response to treatment. *Schizophrenia research.* 2008;98(1-3):225-231.

45. McArthur LZ, Baron RM. Toward an ecological theory of social perception. *Psychological review.* 1983;90(3):215.

46. van Os J, Verhagen S, Marsman A, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and anxiety.* 2017;34(6):481-493.

Supplementary materials

Table 1

Experience sampling method questions – beep level ten times a day

Negative affect	I feel … (1=not at all, 4= moderate, 7= very much)
	Insecure
	Lonely
	Anxious
	Agitated
	Down
	Guilty
Positive affect	I feel … (1=not at all, 4= moderate, 7= very much)
	Cheerful
	Relaxed
	Enthusiastic
	Satisfied
Psychopathology	I feel … (1=not at all, 4= moderate, 7= very much)
	Energetic
	Suspicious
	Impatient
	Talkative
	Harried
	As if everything goes by itself
	It is busy in my head
	I worry
Event stress	Since the last beep the most important things that has happened is …
	How (un)pleasant was this event? (-3=very unpleasant, 0=neutral, 3=pleasant)
Context	
Who am I whit	partner, residents, family non-resident, friends, colleagues/classmates, healthcare professional, acquaintances, strangers/others, nobody
Where am I	at home, at family\friends, work or school, healthcare setting, public place, in transit, somewhere else
What am I doing	work/school, domestic tasks, groceries, eating/drinking, personal hygiene, smoking, active leisure, passive leisure, gaming/internet use, sports, hanging out, resting, something else, nothing
	This beep disturbed me (1=not at all, 4=moderate, 7=very much)

Table 2

Person centered averages of experience sampling outcomes (N=5 symbol-task subsample)

	M	SD	min	max
Positive affect	4.75	.74	3.53	5.45
Negative affect	1.41	.34	1.04	1.80
Energetic	4.31	1.12	2.47	5.17
Suspicious	2.30	.63	1.56	3.10
Impatient	1.66	.77	1.04	2.67
Talkative	3.06	1.89	1	4.98
Harried	1.40	.62	1	2.48
As if everything goes by itself	3.39	1.74	1.08	4.95
It is busy in my head	2.09	1.28	1	4.07
I worry	2.82	.86	1.10	2.95
Beep disruption	2.31	.62	1.56	3.10

8

Measuring within-day cognitive performance using the experience sampling method – a pilot study in a healthy population

S. Verhagen / N. Daniëls, S. Bartels, S. Tans, K. Borkelmans, M. de Vugt, P. Delespaul

Abstract

Background: People with depression, anxiety, or psychosis often complain of confusion, problems concentrating or difficulties cognitively appraising contextual cues. The same applies to people with neurodegenerative diseases or brain damage such as dementia or stroke. Assessments of those cognitive difficulties often occurs in cross-sectional and controlled clinical settings. Information on daily moment-to-moment cognitive fluctuations and its relation to affect and context is lacking. The development and evaluation of a digital cognition task is presented. It enables the fine-grained mapping of cognition and its relation to mood, intrapersonal factors and context.

Method: The momentary Digit Symbol Substitution Task is a modified digital version of the original paper-and-pencil task, with a duration of 30 seconds and implemented in an experience sampling protocol (8 semi-random assessments a day on 6 consecutive days). It was tested in the healthy population (N=40). Descriptive statistics and multilevel regression analyses were used to determine initial feasibility and assess cognitive patterns in everyday life. Cognition outcome measures were the number of trials within the 30-second sessions and the percentage of correct trials.

Results: Subjects reported the task to be easy, pleasant and do-able. On average, participants completed 11 trials with 97% accuracy per 30-second session. Cognitive variation was related to mood, with an interaction between positive and negative affect for accuracy (% correct) (p=.001) and an association between positive affect and speed (number of trials) (p=.01). Specifically, cheerful, irritated and anxious seem to covary with cognition. Distraction and location are relevant contextual factors. The number of trials showed a learning effect (p<.001) and was sensitive to age (p<.001).

Conclusion: Implementing a digital cognition task within an experience-sampling paradigm shows promise. Fine-tuning in further research and in clinical samples is needed. Gaining insight into cognitive functioning could help patients navigate and adjust the demands of daily life.

Introduction

Various patient populations experience confusion, difficulties to concentrate or problems to cognitively grasp contextual cues[1-5]. To assess an individual's ability to function and cope in everyday life, neuropsychological tests are crucial. The information generated can be included in evaluating whether someone is, for example, capable of independent living or self-care[6]. Given these far-reaching consequences, it is important that the performance measured with a neuropsychological test accurately reflects performance in daily life. A review by Chaytor and Schmitter-Edgecombe (2003) suggests, however, that, when the relationship between tests and measures of daily functioning is considered, neuropsychological tests might only have moderate ecological validity for predicting everyday cognitive functioning[7].

While the general use of neuropsychological tests has gained importance in recent years, the tests themselves as well as the standardized context of administration remained largely the same[8]. Often, a battery of cognition tests (e.g., CANTAB) are used to determine someone's cognitive potential on a range of domains[9]. Individual tests often take several minutes to administer. Furthermore, they are performed in the presence of a professional in minimum distraction environments. The goal is to determine a stable cognition factor that provides insight into the individual's general strengths and vulnerabilities[8]. However, the clinical test conditions sharply contrast with everyday environments. Everyday life is comprised of multi-sensory elements such as distracting sounds, smells, lights, or tactile stimuli. Furthermore, daily stressors and mental states can influence an individual's cognitive ability[10,11]. Mood, for example, follows a dynamic pattern in everyday life[12] and its effect on cognition from one moment to the next is seldom considered. Moreover, cognition is known to fluctuate over the day, depending on factors such as the level of alertness or food intake[13,14]. To improve the understanding of cognition in everyday life, the assessments need to take place in natural daily environments. Ideally, other domains such as mood and behavior are monitored simultaneously so that underlying associations can be learned. Insight into these implicit patterns would enrich treatment for cognitive complaints and provide additional clues for recovery and rehabilitation, next to opportunities to tailor interventions to the individual[15]. By providing cognitive assessments within the Experience Sampling Method (ESM) this strategy becomes possible.

ESM, also called Ecological Momentary Assessments (EMA), is a (digital) structured self-assessment diary technique that allows insight into the everyday life of an individual[16]. At several (semi-) random times during the day, eHealth technologies such as Personal Digital

Assistants or smartphone apps give signals (beeps) to prompt the collection of momentary experiences. At those moments, participants are asked to reflect on their current mood, environmental context, and activities and report their real-time information to the eHealth technology used. ESM is characterized by a high ecological validity as it collects experiential and contextual data in situ[17]. In-the-moment reflections reduce the recall bias that troubles retrospective self-reports[18]. Furthermore, repeated ESM measures allow a better understanding of between- and within-person variability in psychopathology and beyond[12]. As ESM can be experienced as time-consuming, the questionnaires need to be kept short and the design transparent to avoid overburdening[17].

The initial feasibility and acceptability of cognition tasks in an ESM paradigm are supported by a small number of studies , including domains such as working memory, attention, or processing speed[19]. The feasibility of a digital trail making test assessing processing speed in everyday life, for example, was found to be feasible in Chinese patients with depression[20]. Another study investigated the reliability and validity of three ambulatory cognition tasks measuring different cognitive domains (i.e., Symbol search, Dot memory, and an N-back task)[21]. Results indicated that all three tasks are feasible within an ESM paradigm and show excellent between-person reliability, reliable within-person variability, and construct validity with cross-sectional cognitive assessments[21]. In young adults, a digital processing speed task was not only feasible, but also sensitive to blood alcohol concentration[22].

Notably, most studies on daily life cognition focus only on a limited number of contextual factors in relation to cognitive performance. As everyday life is extremely complex, more research is needed to contextualize daily cognition with extensive intrapersonal (e.g., mood, age, fatigue) and contextual factors (e.g., location, company). Additionally, cognition tasks in everyday environments that take multiple minutes to perform[23] might, on one hand, provide valuable information on daily cognitive functioning. On the other hand, the length of the task can result in a relatively low sampling frequency to not overburden the participant and thus limit the exploration of cognitive fluctuations over the course of the day. In order to learn which factors influence cognitive variation over time, a higher sampling rate is required with shorter beep durations to minimize burden. This strategy would enable to study the influence of different daily situations on cognition. Ultimately, the test results could be reported back to patients and discussed together with a clinician in relation to other relevant health domains.

The present study aims to build an objective cognition task with a short duration for repeated assessments and to implement this task into a daily life setting. Accordingly, a modified digital version of the Digital Symbol Substitution Task was used within the ESM-based PsyMate™ application on an iPod for six consecutive days by healthy individuals. This digital cognition task is called momentary Digital Symbol Sustitition Task (mDSST).

First, the utility and feasibility of the mDSST was determined through the participants' compliance rate and retrospective subjective experience. Second, the focus lay on validation via comprehensive contextualization of daily cognitive performance. The relationship between intrapersonal as well as contextual factors and the mDSST performance was investigated using high frequency ESM sampling (eight times a day).

Prospectively, digital cognition tasks in everyday life may be relevant for improved prevention, treatment, and rehabilitation of psychopathology.

Methods

Participants

Individuals from the general population were recruited via poster advertisement at Maastricht University and through social media as seeds for snowball sampling[24]. Sample size was based on recommendations for pilot studies and other exploratory ESM studies[25-27]. In total, 45 participants provided written informed consent. All individuals were 18 years or older, had sufficient command of the Dutch language, and were able to handle an iPod with the PsyMate™ app. Exclusion criteria were medication use that influences cognitive performance and current treatment for mental illnesses or cognitive complaints. Ethical approval was obtained by the standing ethical committee of the Faculty of Psychology and Neuroscience, Maastricht University (ref.no.183_02_09_2017).

Measurements

PsyMate™

The PsyMate™ is a web-based platform for moment-to-moment assessment of mood and behavior in daily life. It includes an App (iOS and Android), cloud-based data storage, and reporting tool. The PsyMate™ was developed by Maastricht University and Maastricht UMC+ (www.psymate.eu) and programmed to prompt participants using auditory signals eight times a day to complete a self-report questionnaire (approximately two minutes). Signals (beeps) were provided between 7.30 AM and 10.30 PM in semi-random time blocks of 112,5 minutes. The self-report questionnaire assessed mood, physical status (i.e., fatigue, hunger),

and context (i.e., location, activity, and persons present). The mood items were combined in two independent constructs: Positive Affect (PA) by averaging 'cheerful', 'energetic', 'relaxed', 'enthusiastic', and 'satisfied', and Negative Affect (NA) using 'insecure', 'down', 'irritated', 'lonely', 'anxious', and 'guilty'. The mood and physical status items were rated on a 7-point Likert scale (1=not at all, 4=moderate, 7=very) and the context items were assessed categorically. The complete item list is included as supporting information (see supplementary materials S1). In addition to the self-report questionnaire on the beep level, participants were asked to complete a morning and an evening questionnaire. These additional questionnaires consisted of self-report items that assessed respectively sleep duration and sleep quality, and general appraisal of the day. Most items of the morning questionnaire were assessed categorically, whereas all the items of the evening questionnaire were rated on a 7-point Likert scale (1=not at all, 4=moderate, 7=very). Participants were included in the analyses if they completed a minimum of sixteen valid beep moments (1/3 of total), conform with ESM guidelines[28]. All participants were provided with an iPod on which the PsyMate™ app (version 2.0.0.) was installed to standardize the administration of the momentary Digit Symbol Substitution Task (mDSST). To evaluate the PsyMate™ procedure, debriefing questionnaires were provided after the ESM completion.

PsyMate™ mDSST
The mDSST is based on the Digit Symbol Substitution Task from the Wechsler Adult Intelligence Scale (WAIS)[29]. It measures information processing speed and short-term working memory. The modified mDSST primarily assesses information processing speed, but not short-term working memory due to design choices (e.g., short duration, one-by-one presentation) that are part of the ESM set-up. The task was selected after consultations with psychiatric and neuropsychological healthcare professionals and scholars of daily life assessment. The constraints were that the digital cognition task could be performed multiple times per day and therefore had to be short, sensitive to cognitive fluctuations, and show no or only a small learning effect. The mDSST is thought to fulfill these criteria.

The mDSST started after the standard ESM beep questionnaire. Participants viewed an instruction screen including a button to start the task. The item screen displays the numbers 1 to 9 with a corresponding symbol at the top of the screen (encoding information). For each trial, a number was presented one-by-one in the middle of the screen. Participants had to select the corresponding symbol at the bottom of the screen (see Fig 1). Symbols were kept similar to the original Digit Symbol Substitution Task. The task duration was 30 seconds and participants were instructed to complete as many trials as possible while also being as

accurate as possible. Five unique combinations of numbers and symbols with corresponding answer keys were programmed beforehand and presented in random order over the course of the 48 beeps. Outcome measures of the PsyMate™ mDSST are the number of trials (how many one-by-one trials are completed within 30-second sessions) and the percentage of correct trials (the number of correctly answered trials divided by the total number of trials).
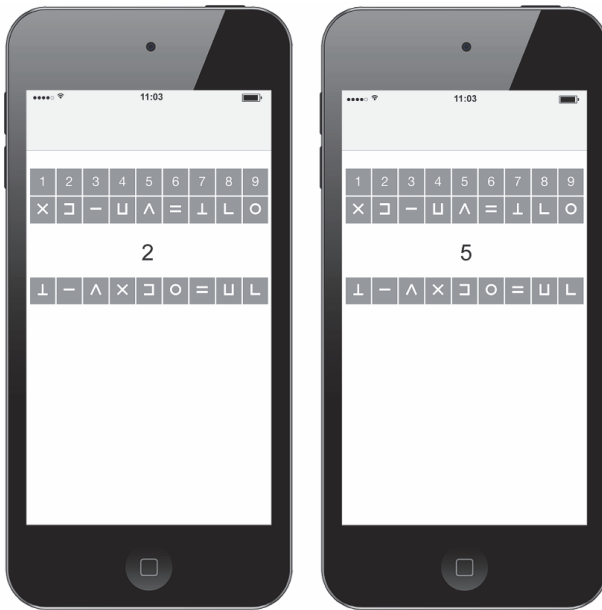
Figure 1
Momentary Digit Symbol Substitution Task in the PsyMate™ application.

Debriefing questionnaire

Participants received a debriefing questionnaire with three parts using open-ended and 7-point Likert scale questions: 1) to assess the general experience of participants throughout the week (e.g., was this a normal week, did participation influence your mood, social contact or activities); 2) to evaluate the usability of the PsyMate™ in general (e.g., was the PsyMate™ difficult to use , was the number of questionnaires burdensome, were there any technical issues); and 3) to assess the experiences with the mDSST (e.g., how well do you think you performed on the task, was the task difficult, was the task enjoyable).

Procedure

After participants provided written informed consent, a briefing session of one hour took place. Participants provided sociodemographic information including gender, age, living situation, education level, current occupation, and ethnicity. Additionally, current medication use and treatment for mental illnesses and cognitive complaints were assessed through self-reports. Furthermore, participants received an iPod (5th generation) with the PsyMate™ (v2.0.0.) preinstalled. They were instructed how to use the PsyMate™ and performed a test trial to familiarize themselves with the ESM procedure. Then, the participants used the PsyMate™ for six consecutive days, starting on the day after the briefing session. On the second day of the ESM period, participants were contacted by telephone to assist with potential problems or answer questions. After the ESM period, a debriefing session of one hour took place in which participants completed the debriefing questionnaire and returned the iPod.

Statistical analyses

Descriptive statistics were used to assess participant characteristics, initial feasibility, and acceptability (frequencies) of the ESM protocol. The completion rate was calculated by comparing the mean percentage of valid beep moments to the total number of beep moments. The data collected with the PsyMate™ have a multilevel structure; beeps (level 1) were nested in participants (level 2). Average scores of the variables of interest were person-mean centered to take into account the within-person effect. In order to look at contextualized variation, dummy variables were created for location (at home versus somewhere else), company (alone versus with others), and coffee use since the last beep (yes or no). Furthermore, activity-related stress was conceptualized as an average of the items 'I would rather be doing something else', 'This is difficult for me', and 'I can do this well' (reverse coded). In order to look at learning effects, a log transformation of the replication (sequence number of responded beeps within subjects ranging from 1 (first beep) to 48 (last beep)) was calculated as a proxy measure of time across the six day period. Additionally, within-day time effect was explored using hour of the day and its quadratic function. To assess cognitive variation over time and to check for learning effects, multilevel regression analyses were run with the number of trials within the 30-seconds interval and the percentage of correct trials (for each assessment moment) on the mDSST as dependent variables and respectively time (i.e., log transformation of replication), hour, squared hour, and a log transformation of day number (from day 1 to day 6) as independent variables. Furthermore, multilevel analyses were run to assess the association between positive affect, negative affect, its interaction, and various other contextual factors (e.g., fatigue, distraction)

as independent variables and both cognition outcomes as dependent variables. Additionally, multilevel stepwise regression procedures were used to explore the effect of individual mood items on cognition. Both forward and backward strategies were applied. The individual mood items and various other contextual factors were seen as independent variables and cognition as dependent variable. Quadratic function of age, gender, possible learning effects, and within-day effects were considered as covariates in all multilevel models. Analyses were carried out using Stata version 13.0[30]. A two-sided significance level of .05 was used.

Results

Participants

Forty-five participants were included in the ESM protocol, resulting in 1330 valid beep records. Two participants were unable to finish the ESM protocol due to problems with the iPod device (loss of 11 records, 0.83%), one participant was excluded because the iPod was stolen (loss of 9 records, 0.68%), and two participants did not reach the criteria of at least 16 valid beeps due to various reasons (loss of 20 records, 1.50%). The complete dataset consisted of 40 participants with 1293 valid beep records. The participants' age ranged from 21 to 72 years of age with a mean of 30.4 (*SD*=14.79, *Mdn*=23.0). On average, participants completed 33 beeps (*SD*=4.9, range 21-43) of the 48 scheduled beeps. ESM completion rate was 69%. See Table 1 for descriptive statistics of the healthy population sample.

Table 1
Descriptive statistics for the healthy population sample (N=40)

	N (%)
Gender (women)	29 (72.5%)
Education level	
Secondary vocational education	5 (13%)
Bachelor degree	20 (50%)
Master degree	15 (37%)
Occupation	
Students	25 (62%)
Fulltime work	10 (25%)
Part-time work	3 (8%)
Voluntary work	1 (2,5%)
No occupation	1 (2,5%)

Feasibility

In order to assess feasibility, the available data from the debriefing questionnaire was used. One participant, whose iPod was stolen, did not complete this evaluation questionnaire, leaving 44 participants in the sample. All other analyses based on ESM/PsyMate™ data were performed with a sample size of forty participants.

Evaluation PsyMate™ procedure

Participants reported that the ESM items were a good representation of their experience (M=5.1, SD=1.26). They had no difficulty using the PsyMate™ (M=1.59, SD=1.06) and the verbal and written instructions were clear (verbal: M=6.64, SD=.53; written: M=6.43, SD=.70). Furthermore, completing the items had little influence on their mood (M=2.07, SD=1.26), activities (M=1.89, SD=1.5), and social contact (M=1.55, SD=.93). Participating in ESM did not hinder their daily activities (M=2.16, SD=1.31). With regard to the burden, participants reported that the number of beeps a day (M=3.23, SD=1.46), the duration of beep completion (M=2.32, SD=1.29), and the beep sound (M=3.18, SD=1.97) had low impact.

Evaluation of the mDSST

Participants were motivated to perform well on the mDSST (M=5.70, SD=.93), the mDSST was moderately pleasant to perform (M=4.43, SD=1.37), and participants would recommend the task to others (M=5.48, SD=1.17). Overall, the task was experienced as easy (M=1.80, SD=1.15). However, when participants had to assess their own performance retrospectively, they indicated to have performed moderately on the mDSST (M=4.55, SD=1.19). Also in retrospect, they reported to be moderately distracted during the task (M=3.51, SD=1.39).

Variation in cognition

Participants completed on average 11.39 trials within 30-second sessions (SD=1.32, range 3-15), with an average percentage correct of 97.11 (SD=2.01, range 28.6-100). The number of trials was positively associated with time (B=.36, SE=.033, p<.001, 95% CI=.30, .43), with a positive within-day effect for hour of the day (B=.03, SE=.007, p<.001, 95% CI=.01, .04), and a positive between-day effect for day number (B=.45, SE=.05, p<.001, 95% CI=.35, .54). The percentage of correct trials was not associated with time (B=-.26, SE=.19, p=.17, 95% CI=-.63, .11), with no within-day (B=-.04, SE=.04, p=.36, 95% CI=-.11, .04) or between-day effect (B=-.21, SE=.27, p=.45, 95% CI=-.74, .32).

Mood, contextual factors, and cognition

Participants experienced high positive affect (M=4.82, SD=.77, range 2.68-6.48) and low negative affect (M=1.65, SD=.47, range 1.01–2.98) throughout the study. They were a little worried (M=2.52, SD=1.00, range 1.00–4.78) and felt moderately fatigued (M=3.69, SD=1.03,

range 1.55–5.77). Furthermore, they experienced low activity-related stress (*M*=2.68, *SD*=.61, range 1.37–3.73) and were moderately focused on their current activities (*M*=4.87, *SD*=.73, range 3.39–6.63). On the mDSST, they reported a low to moderate level of distraction during this task (*M*=2.88, *SD*=.88, range 1.15-4.67).

Only the main significant aggregated findings from the multilevel regression analyses are reported. Single-item analyses are included in the supplementary material (see S1 Table). Participants performed more trials (*B*=.08, *p*=.04) and made less mistakes (*B*=.62, *p*=.001) when experiencing high positive affect. They made more mistakes when experiencing high negative affect (*B*=-1.41, *p*<.001). With regard to the contextual factors, participants performed less trials when being at a different location then home (*B*=-.20, *p*=.002) and when reporting to be distracted (*B*=-.17, *p*<.001). They also made more mistakes when distracted (*B*=-.46, *p*<.001). Fatigue, activity-related stress, worrying, current company, coffee use, and being able to focus were unrelated to both cognition outcome measures. With regard to possible covariates, less trials were performed with higher age (*B*=-.001, *p*<.001), and when being male. A positive association was found between the time measures (i.e., the log-transformed replication variable as time measure, hour, squared hour) and the number of trials (e.g., the log-transformed replication variable as time measure; *B*=.36, *p*<.001). The variables with an association with the cognitive outcome measures were included in further multilevel regression models.

In the final model of the number of trials, participants again performed more trials when experiencing high positive affect (*B*=.20, *p*=.01). In addition, a positive learning effect was present with more trials completed over time (*B*=.38, *p*<.001). Moreover, participants completed less trials when distracted (*B*=-.19, *p*<.001) and at an older age (*B*=-.0008, *p*<.001). The results of this analysis indicated that the six predictors explained 36% of the overall variance (16% within-subject variance and 47% between-subject variance).

In the final model of the percentage of correct trials, a positive interaction effect was found between positive affect and negative affect for the percentage of correct trials. In other words, the influence of negative affect on correctness is limited when positive affect is high, but stronger when positive affect is low (*B*=.71, *p*=.001). Additionally, participants made more mistakes when distracted (*B*=-.46, *p*<.001). The results of this analysis indicated that the four predictors explained 3% of the overall variance (5% within-subject variance and 0.1% between-subject variance). The results of the final models are presented in Table 2 (the number of trials) and Table 3 (the percentage of correct trials).

Table 2

Multilevel regression analyses of mood, distraction, time, and age during the mDSST on the number of trials

	B	SE	p	95% CI	
			Number of trials		
Model 1			<.001*		
Positive affect	.20	.08	.01*	.04,	.36
Negative affect	.27	.17	.12	-.07,	.60
Interaction between positive and negative affect	-.04	.04	.34	-.12,	.04
Distracted	-.19	.02	<.001*	-.22,	-.15
Time$^\$$	.38	.03	<.001*	.31,	.44
Age^2	-.0008	.0001	<.001*	-.001,	-.0005

Note. CI = Confidence Interval, Time$^\$$ = log-transformed replication score, Age^2 = squared age. *$p < .05$.

Table 3

Multilevel regression analyses of mood and distraction during the mDSST on the percentage of correct trials

	B	SE	p	95% CI	
			Percentage of correct trials		
Model 1			<.001*		
Positive affect	-.89	.43	.04*	-1.73,	-.05
Negative affect	-4.10	.97	<.001*	-5.99,	-2.21
Interaction between positive and negative affect	.71	.22	.001*	.28,	1.15
Distracted	-.46	.11	<.001*	-.67,	-.26

Note. CI = Confidence Interval, *$p < .05$.

Exploratory analyses on individual mood items

The pairwise correlation of individual mood items ranged from .42 to .74 for positive affect items and from .30 to .54 for negative affect items. These correlations disregard the nested within-subject variance. When substracting by subject means to assess within-subject

variance only, the correlations were considerably lower (from .24 to .63 for positive affect, and from .18 to .40 for negative affect). Results are presented in the supporting materials (see S2 Table).

Exploratory multilevel regression analyses of individual mood items on cognition were computed, using mood items as independent variables and cognitive outcome measures as dependent variables (see S3 Table for an overview). Only the items cheerful and energetic were positively associated with the number of trials (respectively $B=.12, p<.001$; $B=.06, p=.02$). The positive affect items cheerful ($B=.54, p<.001$), relaxed ($B=.51, p<.001$), and satisfied ($B=.53, p=.001$) were positively associated with the percentage of correct trials. All negative affect items were negatively associated with percentage of correct trials.

In order to weigh item covariation, both forward and backward stepwise strategies were applied. These results are also presented in the supporting materials (see S3 Table). In the backward-approach, cheerful remained the most prominent positive mood variable associated with the number of trials ($B=.13\ p<.001$) and the percentage of correct trials ($B=.36\ p=.03$). For the negative affect items, irritated showed a positive association with the number of trials ($B=.07\ p=.01$), whereas anxious was negatively associated with the percentage of correct trials ($B=-.69\ p=.01$).

Discussion

A novel digital cognition task, the mDSST, was evaluated for use within a daily life ESM protocol. The first aim was to assess the utility and initial feasibility of the mDSST. The second aim was to study the preliminary internal validation of measuring cognition in daily life, both as varying over time and in relation to contextual and intrapersonal factors.

Feasibility and Utility of the PsyMate™ mDSST

ESM data from three participants were removed due to circumstances outside our control and two participants did not reach the minimum beep requirements, leaving 40 participants with analyzable data. Participants completed on average 33 beeps within a 48-beep protocol, resulting in a completion rate of 69%. The participants' overall experience was positive; ESM completion did not hinder daily life and the burden was reported as acceptable. This result is satisfactory and similar to other ESM research with and without a cognition task[19,23,31,32]. The cognition task was evaluated as easy and pleasant to perform. Task motivation was high and participants felt competitive towards the task, although several participants indicated that this competitiveness faded towards the end of the six day assessment period. This is an

indication that the task is less suited for longer datacollection periods, as is relevant in clinical practice. Solutions in this context should alternate the task with another cognition measure or provide cognitive assessments in a subset of beep-moments each day.

Contextualization of the PsyMate mDSST

Information processing speed was measured with a modified momentary version of the Digit Symbol Subtitution Task that yielded two outcome measures: the number of trials within 30 seconds and the percentage of correct trials[29]. On average, participants completed 11 trials within 30-second sessions (speed) and answered 97% correct (accuracy). This high correctness score indicates that the task is easy, something that is also reflected in the participants' retrospective evaluation. The choice for a DSST-based task was deliberate because it proved sensitive to detect cognitive complaints and changes in cognitive functioning in clinical samples[33,34]. As this is a cognitive healthy sample, it is unsurprising that participants made little mistakes. Generally, cognitive performance can be viewed as a trade-off between accuracy and speed. Here, accuracy showed a ceiling effect (with reduced variability) while speed is a more sensitive measure. Only the number of trials showed a learning effect over time, with a slight increase of trials during the first half of the ESM period followed by a stabilization. Additionally, more trials were completed towards the end of the day.

The relationship between mood and the accuracy outcome reflected a positive interaction effect between positive affect, negative affect, and the percentage of correct trials. In situations were negative affect is high, participants also tend to make more mistakes, an effect that is strongest when positive affect is low. Zooming in on individual mood items, only cheerful and anxious seemed to be associated with the accuracy outcome. Therefore, it has merit to unpack the positive and negative mood aggregations to get relevant information and clues for clinical practice. A possible explanation could be that people are less able to focus on a task when they feel anxious. This negative influence of mood on cognitive performance is observed in clinically depressed patients and might be caused by distractions due to ruminations[35,36]. Here, participants who got distracted during the task also made more mistakes. As distraction was assessed after task completion, it is possible that participants who noticed that they made mistakes, consequently scored higher on distraction. Overall, the explained variance for accuracy in relation to mood and contextual factors is neglible (3%) and combined with a ceiling effect it seems to be an irrelevant chance finding in a population without cognitive complaints.

A small positive association was found between mood (positive affect and more specifically cheerful) and the speed outcome. Participant's who were more cheerful also completed more trials irrespective of learning effects. With regard to contextual and intrapersonal factors, a small negative association was found between age and speed, indicating that older participants overall completed less trials. The original Digit Symbol Substitution task is known to be sensitive in identifying age-related performance and processing speed often explains a large part of the variance in these studies[37]. Our modified digital version of the task was also age-sensitive. With regard to gender, males seemed to perform slower compared to females, an effect that disappeared in the final model. In the original Digit Substitution tests, men also seem to perform less well when averaged[38,39]. In this convenience sample however, females were overrepresented (73%) and further research is needed.

Similar to the accuracy outcome, higher distraction was associated with fewer completed trials within a 30-second session. Here, the overall explained variance is clearly higher (36%). There is more variation over time with only a small learning effect. Indicating that the speed outcome is more suited to assess cognition in the current sample.

Several daily life factors were explored. Only distraction was associated with cognition, whereas other factors such as activity-related stress, company, and being able to focus were not. One other study looked at situational cues in relation to cognitive performance within an ESM paradigm. They found that working memory performance did not differ for people at work versus at home, but that short-term memory improved during worktime[40]. Possibly, processing speed is less sensitive to contextual changes.

Notably, fatigue did not vary significantly over time and had no effect on cognition. This was surprising, since other studies with a young population show a negative impact of tiredness on mental processing and increased difficulties with focusing on a task[41-43]. However, the mDSST was only 30 seconds while a standard cognitive assessment is longer (often 2 minutes). It is likely that the association of cognition with fatigue only occurs in longer or more demanding tasks, which are not suited to the ESM paradigm.

Strenghts and limitations
The PsyMate™ app with the mDSST can be used on an individual's own smartphone and is not restricted to the provided iPod. The use of cognition tasks on smartphones is feasible[44,45]. By using iPod devices across participants, the device specifications during the initial validity were standardized. In the early stages of task development, uncertainty about test characteristics, design choices, and device specifications exists. The use of the same device,

the iPod, reduced the uncertainties about factors that might influence outcome across the study sample. In later stages, the influence of different devices (i.e., own smartphones) will become less problematic as the goal shifts towards an evaluation of within-person variability for clinical purposes.

Additionally, the mDSST was developed in an inter-professional context. Researchers (both in mental health and somatic care), physicians, neuropsychologists, clinicians, and software developers worked together to accomplish a tool that can prospectively be used across disciplines and in daily practice.

Although the study has several advantages, limitations need to be kept in mind. First, our sample was mainly restricted to female students (70% women, 61% students, median age was 23). The study, however, was intended as a pilot study using convenience sampling to assess initial feasibility and validity. The mDSST has shown merit for daily life assessment and age sensitivity of the mDSST could already be indicated. Nonetheless, using a more heterogeneous population, a broader age range (through stratification), as well as populations with cognitive impairments, will increase knowledge about task sensitivity as well as a more diverse examination of between- and within-person variance in task performance.

Second, technical problems have influenced the study outcomes. The beep questionnaire was only abailable for ten minutes. When participants initiated the questionnaire within the ten-minute boundary, the software should allow them to finish the task. However, the PsyMate™ app stopped after 10 minutes sharp, which resulted in 15 unfinished and interrupted tasks. The number of trials statistic was unreliable in these cases. Furthermore, the first participants indicated not hearing the beep sound (leading to eighteen missed beeps). This problem was resolved by a system update that enabled a louder and more intrusive beep sound. The technological issues concerning the mDSST seem unlikely to have influenced the performance outcome; the proportion of correct answers was high. Nevertheless, participants experienced those issues as unpleasant and in the future a more reliable technology should be used.

Finally, while reflecting on the task, two participants reported making mistakes by accidentally pressing the wrong symbol since the buttons were too small. In addition, sixteen participants reported that they made mistakes due to the slow processing of the iPod. The mDSST could be improved by using smartphones with a larger screen so that the size of the buttons is increased. Another option would be to rotate the screen into landscape mode.

Future direction

In light of the current study results, several questions still remain. Valuable, but limited information on the psychometric properties of the 30-second mDDST is gathered. It would be interesting to examine if the time interval can be further decreased (e.g., to 15 seconds) and still yields reliable data. A shorter duration could increase the feasibility and decrease the influence of distractions. The outcome measures of the task can be extended to include response time (milliseconds) to get an idea about the influence of distractions on task performance. Future research should investigate construct validity by comparing performance to the paper-and-pencil version of the DSST. This study is in progress.

Sleep quality was assessed using the morning questionnaire, but not taken into account here due to power problems. Poor sleep quality can negatively influence cognitive performance during the day[13]. More attention needs to be paid to the influence of sleep quality and fatigue on cognitive performance in daily life. Smartwatches exist that can accurately track sleep patterns. It would be interesting to link objectively gathered sleep data to ESM cognition and fatigue outcomes.

The mDSST predominantly focuses on processing speed, but other tasks measuring additional cognitive constructs could be designed for use in an experience-sampling paradigm. This would allow to compute the discriminant validity, as was done by Sliwinski and colleagues[21]. However, it is unclear whether a battery of mobile cognitive tasks is necessary for clinical purposes. Insight into daily cognitive fluctuations may be possible with a nonspecific cognition task. Repeated cognitive testing using ESM technology do not allow for a conclusive assessment across cognitive constructs, cross-sectional test batteries are more suited for this purpose. Gaining a general sense of cognitive functioning in relation to other domains can provide concrete ideas on how to deal with cognitive deficits that are individually relevant during everyday life. Although in this study, the various contextual factors did not show an effect on cognitive performance it still seems valuable to examine possible links more closely. All these factors arguably influence daily cognitive functioning and should further be explored in the context of the rehabilitation process.

Clinical implications

This study is moving away from a classic cross-sectional assessment of cognition to an ecological assessment of cognitive variation. The combination of the mDSST with experience sampling allows for an examination of the link between cognition and contextual and intrapersonal information. ESM is used in clinical assessments and to implement in situ interventions in various populations. Using this method helps to raise awareness for

variability patterns in everyday life and it is used to support self-management and improve well-being[15]. Thus, making ESM a valuable tool to supplement assessments of behavior and mood, with the monitoring of cognitive abilities and its daily fluctuations.

Cognitive impairments are known to influence recovery and self-care behavior in various populations. In schizophrenia and depression, there is evidence that cognitive deficits contribute to poor psychosocial functioning[46,47], while in bipolar disorders there is an association between cognitive dysfunction and the course and length of the illness[48]. A study by Cameron et al. (2010) showed that in patients with heart failure, cognitive problems hindered decision-making[49]. Individuals with diabetes, who experienced greater cognitive difficulties, were less likely to remain adherent to exercise or diet[50]. Teaching individuals self-management techniques is generally recommended for rehabilitation purposes, for example after a stroke[51].

Understanding oneself and one's (cognitive) abilities is important for self-management. By monitoring cognition with ESM and by examining the results afterwards, knowledge can be gained about previously nontransparent patterns between behavior, mood, and cognition, facilitating this understanding[15]. Learning when difficulties arise and under which circumstances, could help patients to adjust their tasks accordingly. Individuals might thus plan their days according to their cognitive abilities and, for example, schedule resting moments when cognitive exhaustion occurs. Keeping track of minor changes towards recovery motivates patients and helps clinicians to adapt treatment plans. Cognition tasks like the mDSST can be helpful in supporting future treatment, prevention, and rehabilitation.

Conclusions

Adding a digital cognition task to an experience-sampling paradigm proved to be feasible in healthy individuals. The mDSST is promising and sensitive to detect cognitive variability in relation to mood, intrapersonal, and contextual factors. Although the task seems promising, further exploration is needed in more diverse age samples and in clinical populations with cognitive complaints. The implementation could be improved by providing some minor changes to the task (e.g., larger buttons or screen for visibility). It is clinically relevant to grasp how cognition fluctuates over time and relates to daily life functioning. By providing patients and clinicians with feedback on this data, cognitive rehabilitation and self-management can be improved.

References

1. Planton M, Peiffer S, Albucher J, et al. Neuropsychological outcome after a first symptomatic ischaemic stroke with 'good recovery'. *European Journal of Neurology.* 2012;19(2):212-219.

2. Wright SL, Persad C. Distinguishing between depression and dementia in older persons: neuropsychological and neuropathological correlates. *Journal of Geriatric Psychiatry and Neurology.* 2007;20(4):189-198.

3. Bouchard RW. Diagnostic criteria of dementia. *Canadian journal of neurological sciences.* 2007;34(S1):S11-SS18.

4. Castaneda AE, Tuulio-Henriksson A, Marttunen M, Suvisaari J, Lönnqvist J. A review on cognitive impairments in depressive and anxiety disorders with a focus on young adults. *Journal of affective disorders.* 2008;106(1-2):1-27.

5. Reichenberg A, Harvey PD, Bowie CR, et al. Neuropsychological function and dysfunction in schizophrenia and psychotic affective disorders. *Schizophrenia bulletin.* 2008;35(5):1022-1029.

6. Heaton RK, Pendleton MG. Use of Neuropsychological tests to predict adult patients' everyday functioning. *Journal of consulting and clinical psychology.* 1981;49(6):807.

7. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review.* 2003;13(4):181-197.

8. Casaletto KB, Heaton RK. Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society.* 2017;23(9-10):778-790.

9. Sweeney JA, Kmiec JA, Kupfer DJ. Neuropsychologic impairments in bipolar and unipolar mood disorders on the CANTAB neurocognitive battery. *Biological psychiatry.* 2000;48(7):674-684.

10. Mitchell RL, Phillips LH. The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia.* 2007;45(4):617-629.

11. Gray JR. Emotional modulation of cognitive control: Approach–withdrawal states double-dissociate spatial from verbal two-back task performance. *Journal of Experimental Psychology: General.* 2001;130(3):436.

12. 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. *Psychological medicine.* 2009;39(9):1533-1547.

13. Dijk DJ, Duffy JF, Czeisler CA. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *Journal of sleep research.* 1992;1(2):112-117.

14. Gómez-Pinilla F. Brain foods: the effects of nutrients on brain function. *Nature reviews neuroscience.* 2008;9(7):568.

15. van Os J, Verhagen S, Marsman A, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and anxiety.* 2017;34(6):481-493.

16. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. In: *Flow and the foundations of positive psychology.* Springer; 2014:35-54.

17. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evidence-based mental health.* 2016;19(3):86-89.

18. Scollon CN, Prieto C-K, Diener E. Experience sampling: promises and pitfalls, strength and weaknesses. In: *Assessing well-being.* Springer; 2009:157-180.

19. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: A systematic review. *International journal of methods in psychiatric research.* 2017;26(4):e1562.

20. Hung S, Li M-S, Chen Y-L, Chiang J-H, Chen Y-Y, Hung GC-L. Smartphone-based ecological momentary assessment for Chinese patients with depression: An exploratory study in Taiwan. *Asian journal of psychiatry.* 2016;23:131-136.

21. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. *Assessment.* 2018;25(1):14-30.

22. Suffoletto B, Goyal A, Puyana JC, Chung T. Can an app help identify psychomotor function impairments during drinking occasions in the real world? A mixed-method pilot study. *Substance abuse.* 2017;38(4):438-449.

23. Bouvard A, Dupuy M, Schweitzer P, et al. Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls. *The American journal on addictions.* 2018;27(7):553-556.

24. Goodman LA. Snowball sampling. *The annals of mathematical statistics.* 1961:148-170.

25. Johanson GA, Brooks GP. Initial scale development: sample size for pilot studies. *Educational and Psychological Measurement.* 2010;70(3):394-400.

26. Verhagen SJ, Berben JA, Leue C, et al. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. *PloS one.* 2017;12(10):e0186294.

27. van Knippenberg R, De Vugt M, Ponds R, Myin-Germeys I, van Twillert B, Verhey F. Dealing with daily challenges in dementia (deal-id study): an experience sampling study to assess caregiver functioning in the flow of daily life. *International journal of geriatric psychiatry.* 2017;32(9):949-958.

28. Delespaul PAEG. *Assessing schizophrenia in daily life: The experience sampling method.* Maastricht, Maastricht University; 1995.

29. Wechsler D. *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV).* San Antonio, TX: The Psychological Corporation; 2008.

30. StataCorp L. Stata multilevel mixed-effects reference manual. *College Station, TX: StataCorp LP.* 2013.

31. Schuster RM, Mermelstein RJ, Hedeker D. Acceptability and feasibility of a visual working memory task in an ecological momentary assessment paradigm. *Psychological assessment.* 2015;27(4):1463.

32. Waters AJ, Li Y. Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. *Psychopharmacology.* 2008;197(1):25-35.

33. Jaeger J. Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in Neuropsychological Testing. *Journal of clinical psychopharmacology.* 2018;38(5):513-519.

34. Simons CJ, Bartels-Velthuis AA, Pijnenborg GH, Investigators OoP. Cognitive Performance and Long-Term Social Functioning in Psychotic Disorder: A Three-Year Follow-Up Study. *PloS one.* 2016;11(4):e0151299.

35. McDermott LM, Ebmeier KP. A meta-analysis of depression severity and cognitive function. *Journal of affective disorders.* 2009;119(1-3):1-8.

36. Van Vugt MK, van der Velde M, Investigators EM. How does rumination impact cognition? A first mechanistic model. *Topics in cognitive science.* 2018;10(1):175-191.

37. Hoyer WJ, Stawski RS, Wasylyshyn C, Verhaeghen P. Adult age and digit symbol substitution performance: a meta-analysis. *Psychology and aging.* 2004;19(1):211.

38. Majeres RL. Sex differences in symbol-digit substitution and speeded matching. *Intelligence.* 1983;7(4):313-327.

39. Van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J. The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24–81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. *Journal of clinical and experimental neuropsychology.* 2006;28(6):998-1009.

40. von Stumm S. Feeling low, thinking slow? Associations between situational cues, mood and cognitive function. *Cognition and Emotion.* 2018;32(8):1545-1558.

41. Van der Linden D, Eling P. Mental fatigue disturbs local processing more than global processing. *Psychological research.* 2006;70(5):395-402.

42. Van der Linden D, Frese M, Meijman TF. Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica.* 2003;113(1):45-65.

43. Lorist MM, Klein M, Nieuwenhuis S, De Jong R, Mulder G, Meijman TF. Mental fatigue and task control: planning and preparation. *Psychophysiology.* 2000;37(5):614-625.

44. Brouillette RM, Foil H, Fontenot S, et al. Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PloS one.* 2013;8(6):e65925.

45. Schweitzer P, Husky M, Allard M, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *International journal of methods in psychiatric research.* 2017;26(3):e1521.

46. Green MF. Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. *The Journal of clinical psychiatry.* 2006;67:3-8; discussion 36-42.

47. Rock P, Roiser J, Riedel W, Blackwell A. Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological medicine.* 2014;44(10):2029-2040.

48. Robinson LJ, Nicol Ferrier I. Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence. *Bipolar disorders.* 2006;8(2):103-116.

49. Cameron J, Worrall-Carter L, Page K, Riegel B, Lo SK, Stewart S. Does cognitive impairment predict poor self-care in patients with heart failure? *European Journal of Heart Failure.* 2010;12(5):508-515.

50. Feil DG, Zhu CW, Sultzer DL. The relationship between cognitive impairment and diabetes self-management in a population-based community sample of older adults with Type 2 diabetes. *Journal of behavioral medicine.* 2012;35(2):190-199.

51. Jones F. Strategies to enhance chronic disease self-management: how can we apply this to stroke? *Disability and rehabilitation.* 2006;28(13-14):841-847.

Supplementary materials
S1
Experience sampling protocol: beep questionnaire

	Item	7-point Likert scale or categorical options
1	I feel cheerful	1 = not at all 4 = moderate 7 = very much
2	I feel energetic	1 = not at all 4 = moderate 7 = very much
3	I feel insecure	1 = not at all 4 = moderate 7 = very much
4	I feel relaxed	1 = not at all 4 = moderate 7 = very much
5	I feel down	1 = not at all 4 = moderate 7 = very much
6	I feel irritated	1 = not at all 4 = moderate 7 = very much
7	I feel satisfied	1 = not at all 4 = moderate 7 = very much
8	I feel lonely	1 = not at all 4 = moderate 7 = very much
9	I feel enthusiastic	1 = not at all 4 = moderate 7 = very much
10	I feel anxious	1 = not at all 4 = moderate 7 = very much
11	I feel guilty	1 = not at all 4 = moderate 7 = very much
12	I'm worrying about things	1 = not at all 4 = moderate 7 = very much
13	I generally feel well at the moment	1 = not at all 4 = moderate 7 = very much
14	What am I doing	work, school/ housekeeping/ self-care/ relaxing/ sport/ eating, drinking/ traveling, on the road/ having a conversation/ something else/nothing
15	I can do this well	1 = not at all 4 = moderate 7 = very much
16	This is difficult for me	1 = not at all 4 = moderate 7 = very much
17	I would rather be doing something else	1 = not at all 4 = moderate 7 = very much
18	I am focused	1 = not at all 4 = moderate 7 = very much
19	Where am I	at home/ at someone else's home/ work, school/ public space/ on the road/ somewhere else
20	Who am I with	partner/ family/ housemates/ friends/ colleagues/ acquaintances/ strangers, others / nobody
21a	Company: I like this company	1 = not at all 4 = moderate 7 = very much
22a	Company: I would rather be alone	1 = not at all 4 = moderate 7 = very much
21b	Alone: I like being alone	1 = not at all 4 = moderate 7 = very much
22b	Alone: I would rather be in company	1 = not at all 4 = moderate 7 = very much
23	I don't feel well	1 = not at all 4 = moderate 7 = very much
24	I am tired	1 = not at all 4 = moderate 7 = very much
25	Since the last beep I have used	alcohol/ medication/ coffee, caffeine/ smoking, nicotine/ cannabis/ other drugs/ nothing
26	mDSST instruction screen	
	mDSST	30 seconds task duration
27	I got distracted during the task	1 = not at all 4 = moderate 7 = very much
28	This beep disturbed me	1 = not at all 4 = moderate 7 = very much
29	Thanks!	

S1 Table. Individual multilevel regression analyses

Multilevel regression analyses of the explored ESM items separate for number of trials and percentage of correct trials

	Number of trials					Percentage of correct trials				
	B	SE	p	95% CI		B	SE	p	95% CI	
Positive affect	.08	.04	.04*	.005,	.16	.62	.19	.001*	.24,	.99
Cheerful	.12	.03	<.001*	.06,	.18	.54	.15	<.001*	.24,	.85
Energetic	.06	.03	.02*	.01,	.12	.14	.14	.32	-.14,	.42
Relaxed	.01	.03	.78	-.05,	.06	.51	.14	<.001*	.23,	.80
Satisfied	.02	.03	.51	-.04,	.08	.53	.16	.001*	.22,	.83
Enthusiastic	.02	.03	.53	-.04,	.07	.13	.14	.34	-.14,	.39
Negative affect	-.03	.06	.58	-.14,	.08	-1.41	.28	.000*	-1.96,	-.86
Down	-.04	.03	.24	-.10,	.03	-.40	.17	.02*	-.74,	-.06
Insecure	-.01	.03	.80	-.08,	.06	-.62	.18	<.001*	-.96,	-.27
Irritated	-.004	.03	.89	-.06,	.05	-.54	.14	<.001*	-.81,	-.27
Lonely	.02	.04	.68	-.06,	.09	-.56	.19	.003*	-.94,	-.19
Anxious	-.005	.05	.92	-.10,	.09	-.90	.25	<.001*	-1.39,	-.41
Guilty	-.03	.04	.41	-.12,	.05	-.65	.21	.002*	-1.07,	-.23
Fatigue	-.01	.02	.52	-.06,	.03	-.06	.11	.59	-.29,	.16
Worrying	.02	.03	.54	-.04,	.07	-.26	.14	.07	-.54,	.02
Focused	.01	.03	.70	-.04,	.06	.10	.13	.44	-.16,	.36
Distracted	-.17	.02	<.001*	-.21,	-.13	-.46	.11	<.001*	-.67,	-.25
Act. stress	.02	.03	.44	-.03,	.07	-.11	.14	.44	-.38,	.17
Location	-.20	.06	.002*	-.32,	.07	.26	.34	.45	-.41,	.93
Company	-.13	.07	.05	-.26,	.001	-.02	.35	.95	-.72,	.67
Coffee use	-.08	.09	.37	-.25,	.09	-.37	.46	.42	-1.28,	.53
Age^2	-.001	.0001	<.001*	-.001,	-.0005	-.0002	.0003	.50	-.0007,	.0004
Gender	-1.09	.44	.01*	-1.95,	-.22	.04	.72	.96	-1.38,	1.46
Time$^\$$	.36	.03	<.001*	.30,	.43	-.26	.19	.17	-.63,	.11
Hour	.03	.01	<.001*	.01,	.04	-.03	.04	.36	-.11,	.04
Hour^2	.001	.0002	.001*	.0003,	.001	-.001	.001	.30	-.004,	.001

Note. CI = Confidence Interval, PA = Positive Affect, NA = Negative Affect, Act. Stress = Activity-related Stress. Location = dummy variable of being at home versus somewhere else. Company = dummy variable of being alone versus with others. Coffee use = dummy variable of coffee use since the last beep versus no coffee use. Age^2 = quadratic function of age. Time$^\$$ = log-transformed replication score. Hour = hours within a day. Hour^2 = quadratic function of hour. *p < .05.

S2 Correlations between mood items
Table A
Pearson correlates for single positive affect items

Variables	1	2	3	4	5
Overall variance					
1. Cheerful	-				
2. Energetic	.74*	-			
3. Relaxed	.47*	.42*	-		
4. Satisfied	.59*	.55*	.57*	-	
5. Enthusiastic	.67*	.66*	.48*	.62*	-
Within-subject variance					
1. Cheerful	-				
2. Energetic	.63*	-			
3. Relaxed	.29*	.24*	-		
4. Satisfied	.43*	.39*	.41*	-	
5. Enthusiastic	.50*	.50*	.30*	.43*	-

Note. All *p*< .001.

Table B
Pearson correlates for single negative affect Items

Variables	1	2	3	4	5	6
Overall variance						
1. Down	-					
2. Insecure	.45*	-				
3. Irritated	.44*	.27*	-			
4. Lonely	.49*	.40*	.30*	-		
5. Anxious	.45*	.54*	.32*	.46*	-	
6. Guilty	.33*	.34*	.31*	.30*	.48*	-
Within-subject variance						
1. Down	-					
2. Insecure	.33*	-				
3. Irritated	.40*	.18*	-			
4. Lonely	.35*	.25*	.20*	-		
5. Anxious	.31*	.37*	.22*	.28*	-	
6. Guilty	.27*	.20*	.21*	.18*	.29*	-

Note. All *p*< .001.

S3 Multilevel stepwise regression analyses
Table A

Multilevel stepwise forward and backward regression analyses of individual ESM items for number of trials

	B	SE	p	95% CI	
				Number of trials	
Forward (overall)			<. 001*		
• Cheerful	.10	.03	<. 001*	.05,	.16
• Location	-.003	.06	.96	-.13,	.12
• Distraction	-.19	.02	<. 001*	-.23,	-.15
• Time$	.35	.03	<. 001*	.28,	.41
• Hour	.13	.05	.01	.03,	.23
• Hour^2	-.004	.002	.02	-.007	-.0006
• Age^2	-.0008	.0001	<. 001*	-.001	-.0006
Backward (overall)			<. 001*		
• Cheerful	.13	.03	<.001*	.07,	.19
• Irritated	.07	.03	.01*	.02,	.12
• Age^2	-.0008	.0001	<.001*	-.001,	-.0006
• Time$	.35	.03	<.001*	.29,	.42
• Hour	.12	.05	.02*	.02,	.22
• Hour^2	-.004	.002	.03*	-.009,	-.0004

Note. CI = Confidence Interval. Age^2 = quadratic function of age. Time$ = log-transformed replication score. Hour = hours within a day. Hour^2 = quadratic function of hour. *$p<$ .05.

Table B

Multilevel stepwise forward and backward regression analyses of individual ESM items for percentage of correct trials

	B	SE	p	95% CI	
				Percentage of correct trials	
Forward (overall)			<.001*		
• Cheerful	.33	.17	.06	-.007,	.66
• Relaxed	.21	.16	.20	-.11,	.52
• Insecure	-.36	.18	.05	-.72,	.002
• Irritated	-.27	.15	.07	-.57,	.02
• Distracted	-.43	.11	<.001*	-.63,	-.22
Backward (overall)			<.001*		
• Cheerful	.36	.17	.03*	.04,	.69
• Relaxed	.31	.16	.05	.003,	.62
• Anxious	-.69	.26	.01*	-1.19,	-.19
• Location	.75	.35	.04*	.05,	1.44
• Distracted	-.53	.11	<.001*	-.74,	-.32

Note. CI = Confidence Interval. *$p <$ .05.

9

Digital assessment of working memory and processing speed in everyday life: feasibility, validation and lessons-learned

N. Daniëls / S. Bartels, S. Verhagen, R. van Knippenberg, M. de Vugt, P. Delespaul
Pending revisions, *Internet Interventions.*

Abstract

Background: Experience-sampling tools are used in research and clinical practice to study and support individuals in everyday life. One key aspect of human functioning is cognition. To address this element in momentary assessments, an existing experience-sampling smartphone app was upgraded. The objectives of this study were to test the feasibility and validity through contextualization of two momentary cognition tasks.

Method: The momentary Visuospatial Working Memory Task (mVSWMT) and the momentary Digit Symbol Substitution Task (mDSST) were developed as add-on to an experience-sampling app. Healthy adults (n=49) between 19 and 73 years of age performed the tasks within an experience sampling method (ESM) questionnaire eight times a day, over six consecutive days. Feasibility was assessed through the completion rate and participant experience. Validity was assessed through the contextualization of cognitive performance within a broad range of intrapersonal and situational factors in everyday life.

Results: Participants experienced the tasks as pleasant, felt motivated, and the ESM completion rate was high (71%). Social context, age, and distraction had an influence on cognitive performance in everyday life. The mDSST speed outcome seems the most sensitive outcome measure to capture between- and within-person variance. The mVSWMT was too difficult (37% of recalls correct), and needs further adjustments (i.e., more trials per moment).

Conclusion: Short momentary cognition tasks for repeated assessment are feasible and hold promise for future studies. Recommendations for prospective development include expanding the approach into other cognitive domains and clinical populations. The involvement of a multidisciplinary team and flexible tailoring of the set-up to the individual's needs and abilities to maintain motivation and prevent frustration are furthermore recommended. Finally, a balance between the enjoyment and competitiveness of the tasks and seriousness of the ESM items should be preserved.

Introduction

Memory functions, processing speed, and other cognitive abilities are usually assessed in clinical or laboratory settings rather than natural environments. This approach may affect the ecological validity of neuropsychological test results[1]. An empirically validated daily diary method, known as the Experience Sampling Method (ESM) or Ecological Momentary Assessment (EMA), allows to collect real-world information[2-4]. Digital ESM technologies using smartphone apps (e.g., PsyMate™ or Smart-T) prompt participants repeatedly over the day to reflect on their own behavior, affect, and contextual factors[5,6]. Over the last decade, interest has increased to not only depict affect and activities with experience sampling, but also to include the area of cognitive functioning. It is relevant to bridge the lab-life gap and observe cognition closely in everyday life[7]. For ecologically valid cognitive assessments, digital diary methods delivered via smartphone apps offer unique opportunities.

The status of cognitive assessments in everyday life

Cognitive assessments in everyday life through technology are relatively new. A recent review by Moore, Swendsen, Depp [8] has identified 12 studies that use self-administered digital cognitive assessments. A brief literature search on PubMed identified 13 additional studies (see Figure 1 for a visualized summary).

Various cognitive domains are considered for the assessments in everyday life. Often, more than one task is used and multiple domains are evaluated. Working memory and attention/reaction time are most prevalent, which may be explained by the fact that these domains are generally relevant for various patient populations as they are affected in neurological conditions[9,10] as well as mental illnesses[11,12]. Participants' age ranged from adolescents to older adults. While some studies focused on healthy individuals, others included patient populations. Normative data remains unavailable and, therefore, the validation of mobile cognitive assessments in healthy individuals is still relevant.

To describe the internal validity of the momentary tasks, the contextualization of the momentary performance is a key element. Therefore, previous studies took intrapersonal factors such as age, mood, and drug use into account. Furthermore, psychometric properties of the mobile cognition tasks, including between-person reliability, within-person reliability, and construct validity should be elaborated[8].

Visualized Summary of Study Diversity

Publication
1 Allard et al. (2014)
2 Bouvard et al. (2018)
3 Brouillette et al. (2013)
4 Dirk & Schmiedek (2016)
5 Frings et al. (2008)
6 Hung et al. (2008)
7 Hyun et al. (2019)
8 Keenan et al. (2014)
9 Kennedy et al. (2011)
10 Lee et al. (2018)
11 Marhe et al. (2013)
12 Moore et al. (2016) (Review)
13 Riediger et al. (2014)
14 Schulze et al. (2018)
15 Schuster et al. (2015)
16 Schuster et al. (2016)
17 Schweltzer et al. (2017)
18 Sliwinski et al. (2018)
19 Suffoletto et al. (2017)
20 Thomson et al. (2009)
21 Timmers et al. (2014)
22 Tiplady et al. (2009)
23 Vaeset et al. (2015)
24 Von Stumm et al. (2018)
25 Waters & Li (2008)
26 Waters et al. (2013)

Figure 1

Visualized summary of studies focusing on cognitive assessments in everyday life by cognitive domains, populations, and the intrapersonal factors set in relation with outcome(s) (see Publication); this summary does not aim to be exhaustive, but provides a first overview of the field). NP = Neuropsychological [8,13-37].

The Present Study

The present study aims to evaluate the feasibility and validity through contextualization of two newly developed, short momentary cognition tasks implemented in the PsyMate™ platform with a high sampling frequency. The PsyMate™ is an ESM tool broadly used in research as well as in clinical settings[38,39] and has the advantage of being available for both Android and iOS users (www.psymate.eu). Therefore, this ESM tool is an important target for future cognition task development. A healthy sample was recruited to perform a visuospatial working memory task and a processing speed task within a momentary assessment. Additionally, the ESM items assess a wide range of momentary intrapersonal and situational factors such as mood and social context.

To determine the feasibility, completion rate and participant satisfaction were used. Validity was assessed through an exploration of the contextual variation of cognitive performance. Cognitive performance measured with the momentary cognition tasks was evaluated by relating cognition outcomes to each other as well as to relevant ESM-measures such as mood, fatigue, and current company. The results were discussed with regard to lessons-learned and future implications. Prospectively, researchers and clinicians who are already familiar with the PsyMate™ app can include cognitive measures in everyday life alongside affect and context to understand and support various patient populations.

Methods

Participants

Recruitment from the general population was performed via snowball sampling, using media advertisements and the personal network as seeds. Sample size was based on previous PsyMate™ feasibility studies and aimed for a minimum of 30 participants[40,41]. Fifty-one participants provided written informed consent. Participants were at least 18 years old, had a full understanding of the Dutch language, and were able to handle a smartphone device (Android) with a beta version of the PsyMate™ app (version 213–253). Participants who could not use their own smartphone device were provided with a 5th generation iPod on which the same version of the PsyMate™ app was installed. Individuals were excluded based on medication use that could influence cognitive performance and current treatment for cognitive or mental health complaints. The standing ethical committee of the Faculty of Psychology and Neuroscience, Maastricht University (ref.no.183_02_09_2017) granted ethical approval and the study was carried out in accordance with the Declaration of Helsinki.

Measurements

Experience Sampling Method (ESM)

The ESM was administered using the PsyMate™ Suite; a smartphone app and a cloud-based platform developed by Maastricht University and Maastricht UMC+ (www.psymate.eu). PsyMate™ is a parametrized and flexible tool for repeated assessments in everyday life. The application was programmed to emit an auditory and visual prompt (beep signal) eight times a day for six consecutive days, signaling the availability of a self-report questionnaire. These beep questionnaires were provided at semi-random time blocks of 112.5 minutes, between 7.30 AM and 10.30 PM and remained available for response during 15 minutes. Beep questionnaires included mood (i.e., positive and negative affect), physical status (i.e., hunger, fatigue, and pain), and context (i.e., location, activity, and social company) items as well as the two cognition tasks. Positive affect (PA) included the items 'cheerful', 'energetic', 'relaxed', 'enthusiastic', and 'satisfied', while negative affect (NA) was composed of the items 'insecure', 'down', 'irritated', 'lonely', 'anxious', and 'guilty'. In line with ESM guidelines[3], a minimum of 16 valid beep questionnaires (1/3 of total) per participant had to be completed to be included in the analyses.

In addition to the beep questionnaires, participants completed a morning assessment and an evening assessment every day, each consisting of seven self-report items. The morning questionnaire focused on self-reported sleep quality. The evening questionnaire focused on a global appraisal of the day. The majority of the items were answered on a seven-point Likert scale (1=not at all, 4=moderate, 7=very). Some items contained categories (e.g. 0, 1, 2, 3, 4, 5, or more than 5 times awake during the night). The full list of ESM items is included in supplementary materials Table A.1.

PsyMate™ momentary Digit Symbol Substitution Task (mDSST)

The PsyMate™ mDSST was inspired by the paper-pencil version of the Digit Symbol Substitution Task (DSST) of the Wechsler Adult Intelligence Scale (WAIS). The original WAIS task measures information processing speed and short-term working memory[42]. The PsyMate™ version aims to measure information processing speed. The development team included psychiatric and neuropsychological healthcare professionals as well as ESM researchers. This team defined the following requirements for the mDSST: participants should be able to perform the mobile cognition task several times a day, the task needs to be short, sensitive to cognitive variation, and demonstrates no or a small learning effect.

At the end of the regular ESM beep questionnaire, an instruction screen appeared with a start button to be pressed when ready. At the top of the task screen, the numbers 1 to 9

with a corresponding symbol (similar to the WAIS DSST) were displayed for encoding. In the middle of the screen, different numbers were displayed one-by-one for each trial. At the bottom of the screen, participants had to select the symbol that corresponds to the number presented in the middle of the screen (see Figure 2). Within a 30-seconds timeframe, participants had to accurately complete as many trials as possible. Ten unique encoding combinations with corresponding answer keys were used at random. Two mDSST outcome measures were computed: the number of trials (speed) and the percentage of correct trials (accuracy).
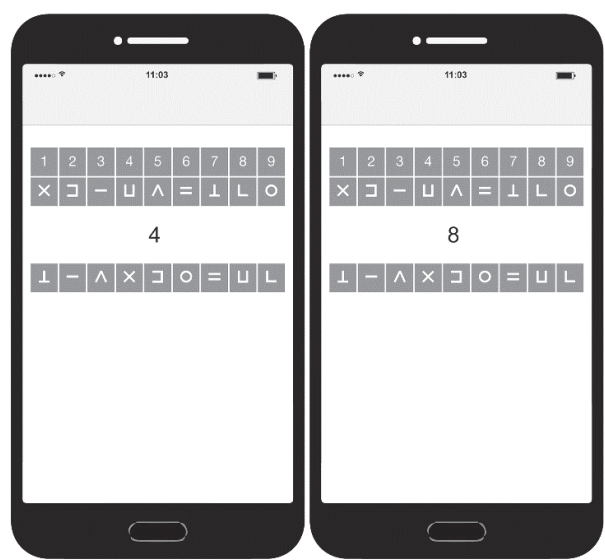
Figure 2
PsyMate™ momentary Digit Symbol Substitution Task.

PsyMate™ momentary Visuospatial Working Memory Task (mVSWMT)

The concept of the PsyMate™ mVSWMT is based on the popular card game 'Memory', also known as 'Concentration' or 'Match Match', where players turn cards to find matching pairs. 'Memory' has been used to study concentration and memory functions in various age groups[43,44]. The mVSWMT fulfils the momentary task requirements (performable several times a day; short; sensitive to cognitive variation; no/small learning effect) and the same team was consulted during task development. The mVSWMT aims to measure concentration and visuospatial working memory (i.e., encoding, maintaining, and retrieving visual information).

The participants were instructed that they would see nine icons to remember. After the participant pressed the start button, icons were presented in a three-by-three grid for eight seconds (encoding phase; see left part of Figure 3 a). Next, participants answered two interference questions on a seven-point Likert scale: 'I think I remembered it all', and 'Generally, I feel well at the moment'. During the recall phase (see right part of Figure 3 a), participants were presented with a three-by-three grid of blanc squares with one icon from the original nine above. An instruction stated to select the square of the original location of the presented icon. The selected square revealed the icons underneath to provide feedback. There was only one trial per beep moment. The grids were filled at random from a selection of 122 unique icons (see Figure 3 b) and a random icon cue was selected from the grid. The outcome measure was correct/incorrect (correct =1) during recall. The icons presented in the mVSWMT were chosen as they represent well-known objects of everyday life and are easily recognizable.

a)
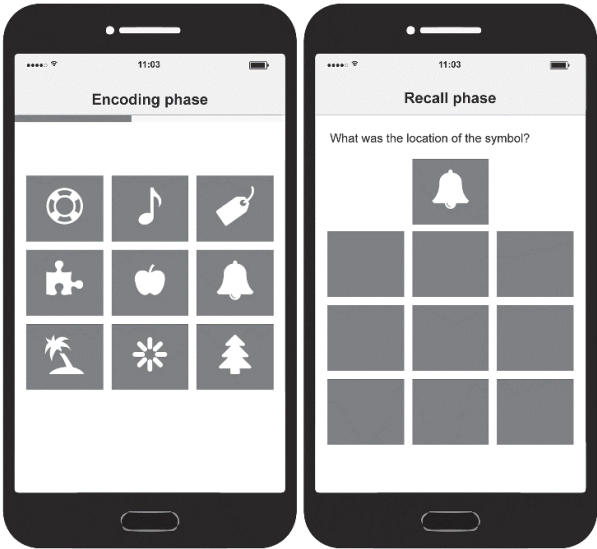
b)
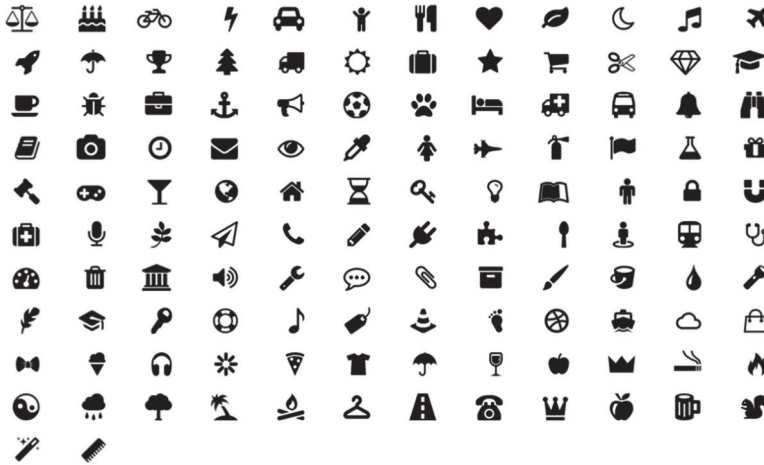
Figure 3

a) PsyMate™ momentary Visuospatial Working Memory Task (mVSWMT) encoding and recall phase.

b) Summary of icons presented in the PsyMate™ mVSWMT.

Debriefing questionnaire

Participants received a debriefing questionnaire that focused on their general experiences during the ESM week (e.g., 'Was this a normal week?', 'Did the PsyMate™ use influence your daily activities?'), the usability of the PsyMate™ (e.g., 'Were the PsyMate™ instructions clear?', 'Was using the PsyMate™ stressful?'), and their experiences with the PsyMate™ mDSST and mVSWMT (e.g., 'To what extent was the task pleasant to perform?', 'Did you experience any technical difficulties?'). Both seven-point Likert scale questions and open-ended questions were used.

Procedure

Following informed consent, a one-hour briefing session took place at Maastricht University or at the participant's home. Sociodemographic information was gathered (e.g., gender, age, living situation, education level, current occupation, and ethnicity) and additional information was asked on current medication use, earlier treatment for mental illnesses, and cognitive complaints. After these general assessments, either the PsyMate™ was installed on the participant's smartphone device or the participant received an iPod with the latest PsyMate™ version installed. After checking the device settings for battery saving options and allowing push notifications, participants were instructed on how to use the PsyMate™ and the

cognition tasks. Test trials were completed to become acquainted with the ESM procedure. Participants started with their six-day ESM period on the following day. During the second day, participants were called to check whether the application was working properly and to clarify potential questions. After the ESM period, a one-hour debriefing session took place during which participants completed the debriefing questionnaire and provided specific feedback with regard to the two cognition tasks.

Statistical analyses

Participant characteristics, feasibility, and acceptability of both cognition tasks were assessed by means of descriptive statistics (frequencies). The completion rate was calculated by comparing the mean percentage of valid beep moments to the total number of beep moments. The data collected with the PsyMate™ have a multilevel structure; beeps (level 1) were nested within participants (level 2). Multilevel regression analyses were used to assess cognitive variation over time and to check for learning effects. Replication score was used as a proxy measure of time and consists of a sequence of beeps within subjects, ranging from 1 (first beep) up to 48 (last beep). Learning effects were examined by using hours to assess a within-day time effect and study day (day 1 to 6) to assess a between-day time effect. Correct/Incorrect (mVSWMT), the number of trials within the 30-seconds time interval (speed), and the percentage of correct trials (accuracy) (mDSST) were used as dependent variables and a log transformation of replication (time), hour and its quadratic function, and a log transformation of study day as independent variables. To assess contextualization, dummy variables were created for location (at home vs. somewhere else) and company (alone vs. with others). Activity-related stress was conceptualized as an average of the items 'I would rather be doing something else', 'This is difficult for me', and 'I can do this well' (reverse coded). To assess the association between PA, NA, fatigue, activity-related stress, distraction, worrying, focusing, location, company, and sleep quality as independent variables and the cognition outcomes of the mVSWMT and mDSST as dependent variables, multilevel regression analyses were computed. Covariates in these multilevel models were quadratic age, gender, and possible learning effects as measured with the time variables. In order to investigate age effects, subgroup multilevel regression analyses were performed, splitting participants into a young group (<45 years) and an old group (45 years or older). Furthermore, Fischer-z transformations of by-subject Pearson's pairwise correlations were calculated between the cognitive outcome measures. Analyses were carried out using Stata version 13.0[45]. A two-sided significance level of .05 was used throughout.

Results

Participants

Seventy-one individuals expressed their interest in the study, of which 66 met the eligibility criteria. From these individuals, seven were not allowed to use mobile phones during work, four individuals did not have enough time, one could not participate because no device was available, and three individuals did not reply after receiving the information. In total, 51 individuals consented to be included, resulting in a 66% recruitment rate.

From the 51 participants who provided informed consent, two participants were excluded because of current treatment for mental health problems. This left data from 49 participants and 1499 beep records. Data of five participants could not be used due to technical problems (beta release of the software: loss of 70 records (4.67%), n=3 transmission problems, n=2 broken devices), leaving a final dataset with 44 participants and 1429 valid beep records. On average, participants completed 34 out of 48 beeps (SD=7.03, range 17-47), resulting in a completion rate of 71%. The age of the n=44 participants ranged from 19 to 73 years with a median of 36 years (M=40, SD=14.82). Sixty-six per cent were women. Highest education level was skewed, with 6% having finished low education, 18% middle education, and 76% high education. Most participants had a fulltime job (61%), others worked part time (23%), studied (7%), took care of their own household (5%), or were retired (4%).

Feasibility

Evaluation of the PsyMate™ procedure

Participants reported that the items represented their experiences well (M=5.70, SD=1.53), that the PsyMate™ was easy to use (reverse coded, M=1.25, SD=.78), and that the verbal and written instructions were clear (respectively M=6.86, SD=.41; M=6.84, SD=.48). The PsyMate™ did not influence their mood (M=2.11, SD=1.53), their activities (M=1.91, SD=1.01), or their social contact (M=1.91, SD=1.18). However, six people indicated that the ESM procedure could be unpleasant or annoying during work, meetings, or conversations. The number of beeps, the duration of a beep, and the sound had a low impact on the burden (respectively M=3.2, SD=1.79; M=2.61, SD=1.86; M=2.32, SD=1.62). Three people found the length of the questionnaire too long.

Evaluation of the mVSWMT

Participants reported that the mVSWMT was pleasant to use (M=5.05, SD=1.57), but rather difficult (M=4.61, SD=1.79). Six participants indicated that the interference questions between encoding and recall made the task difficult. Participants did not get distracted

during the task (*M*=2.93, *SD*=1.53) and were highly motivated to perform well (*M*=5.84, *SD*=1.16). They indicated that they made few inaccuracies (*M*=1.75, *SD*=1.28) and would recommend the task to others (*M*=5.25, *SD*=1.62). They provided some suggestions for further improvement, namely a longer encoding phase and a timer. Participants reported strategies to recall the icons: reading aloud (7 times), creating a story or a mnemonic (6 times), remembering the icons and the location of the icons (5 times), and remembering the first, the middle, or the last row (6 times).

Evaluation of the mDSST
Participants reported that the mDSST was pleasant (*M*=5.66, *SD*=1.22) and easy to use (reverse coded, *M*=1.86, *SD*=.90). They were not distracted during the task (*M*=3.00, *SD*=1.43) and highly motivated to perform well (*M*=5.93, *SD*=1.13). Participants reported that they made few inaccuracies (e.g., tapping symbol X instead of symbol Y) (*M*=3.00, *SD*=1.54). Fourteen people commented that the size of the response buttons was too small, potentially leading to inaccuracies. Participants would recommend the task to others (*M*=5.82, *SD*=1.11). They provided some suggestions for further improvement: to increase the symbol and number size or rotate the screen horizontally. This was especially an issue for iPod users since the screen was smaller.

Contextual factors
Participants experienced high PA (*M*=5.08, *SD*=.69, range 3.35–6.66) and low NA (*M*=1.49, *SD*=.55, range 1.01–3.20). Furthermore, they felt moderately fatigued (*M*=3.01, *SD*=1.19, range 1–5.64), were a little worried (*M*=2.23, *SD*=1.15, range 1–5.43), and experienced low activity-related stress (*M*=2.44, *SD*=.56, range 1.61–3.69). Overall, participants reported a high level of focus during an activity (*M*=4.78, *SD*=.77, range 3.11–6.45) and experienced low to moderate distraction during the mDSST (*M*=2.79, *SD*=.88, range 1.05–4.53). Participants were alone in 29% of the time and in company 71% of the time. Furthermore, they spend 56% of the time at home and 44% somewhere else. According to the morning questionnaire, participants fell asleep after 5 to 15 minutes (40%) and woke up once during the night (34%). Participants slept well (*M*=5.25, *SD*=.80, range 3.46–7) and felt well rested at the start of the day (*M*=4.62, *SD*=1.03, range 2.67–6.65).

Cognition in relation to contextual factors
mVSWMT
Overall, participants were correct in 37% of the mVSWMT assessments (*SD*=.16, range .07–.74). There was no association between time (replication score) and the mVSWMT outcome, (*B*=.01, *SE*=.01, *p*=.32, 95% CI =-.01, .04), showing no within-day time effect (*B*=-.004, *SE*=.003,

p=.17, 95% CI=-.01, .002), nor between-day time effect (*B*=.04, *SE*=.03, *p*=.17, 95% CI=-.02, .09), indicating no learning-effect.

Participants made more mistakes when experiencing high PA (*B*=-.03, *p*=.04), when in company (vs. being alone; *B*=-.10, *p*<.001), and when being distracted (*B*=-.03, *p*=.001). Being able to focus during an activity resulted in more correct answers (*B*=.04, *p*<.001). NA, fatigue, activity-related stress, location, worrying, and sleep quality (morning questionnaire) were not associated with the mVSWMT outcome. More mistakes were made with higher age (*B*=-.00005, *p*=.002), whereas gender was not associated. For all results, see supplementary materials B.

To build the final multilevel regression model, a basis of PA, NA, and its interaction effect was extended with variables that were associated with the cognition outcome measure. No interaction effect was found between PA, NA, and the mVSWMT outcome. The effect of PA disappeared and participants again made more mistakes when in company (*B* =-.09, *p*=.002), when distracted (*B*=-.02, *p*=.03), and with older age (*B*=-.00006, *p*<.001). Being able to focus during an activity was associated with more correct answers (*B*=.05, *p*<.001). The results of this analysis indicated that the seven predictors explained 6% of the overall variance (3% within-subject variance and 28% between-subject variance). Results of the final model for correct/incorrect are presented in Table 1.

Table 1
Final model of the mVSWMT outcome correct/Incorrect

			Correct/incorrect		
	B	SE	p	95% CI	
Model			< .000*		
PA	-.06	.03	.08	-.13,	.01
NA	-.06	.09	.48	-.25,	.12
PA x NA	.01	.02	.33	-.03,	.05
Focus	.05	.01	< .001*	.03,	.07
Company$^\$$	-.09	.03	.002*	-.14,	-.03
Distraction	-.02	.01	.03*	-.03,	-.001
Age^2	-.00006	.00002	< .001*	-.00009,	-.00002

Note. PA = positive affect. NA = negative affect. PA x NA = interaction between positive and negative affect. Company$^\$$ = dummy variable of being alone versus with others. Age^2 = quadratic function of age. CI = Confidence Interval. **p* < .05.

mDSST

Due to technical problems, 82 times (5.7%) the mDSST did not follow the ESM questionnaire. These records were removed, leaving a sample of 42 participants with 1347 beep records. Participants completed on average 12 trials (*SD*=2.57, min=7.21, max=16.83) within the 30-second timeframe, with on average 97% accuracy (*SD*=1.81, min=92.29, max=100). Participants completed more trials over time (beep 1 vs. beep 48; *B*=.32, *SE*=.04, *p*<.001, 95% CI=.24, .40), showed no within-day time effect (*B*=.01, *SE*=.01, *p*=.12, 95% CI =-.003, .03), but more completed trials at later study days (*B*=.61, *SE*=.08, *p*<.001, 95% CI=.45, .76). Accuracy was not associated with time (*B*=.08, *SE*=.20, *p*=.71, 95% CI=-.32, .47), showing no within-day time effect (*B*=.05, *SE*=.04, *p*=.25, 95% CI=-.03, .13) nor between-day time effect (*B*=-.02, *SE*=.36, *p*=.96, 95% CI=-.72, .68).

NA, activity-related stress, location, sleep quality, and gender were not associated with either cognitive outcome measure. Looking at speed, participants completed less trials when in company (*B*=-.29, *p*=.001), when being distracted (*B*=-.15, *p*<.001), and with older age (*B*=-.001, *p*<.001). Furthermore, participants completed more trials when worrying (*B*=.09, *p*=.02). PA and being able to focus were not associated with speed. Looking at accuracy, participants made more mistakes when being tired (*B*=-.34, *p*=.007) and with more distraction (*B*=-.54, *p*<.001). Participants made less mistakes when they experienced more PA (*B*=.45, *p*=.03) and when they could focus better (*B*=.36, *p*=.01). Company was not associated with accuracy. For all results, see supplementary materials C.

Again, the basic model was extended with variables that were associated with the mDSST outcome measures. In the final model of speed (see Table 2), mood showed no effect, and time and worry effects disappeared. Participants completed fewer trials when in company (*B*=-.18, *p*=.04), when being distracted (*B*=-.15, *p*<.001), and with older age (*B*=-.002 *p*<.001). The results of this analysis indicated that the nine predictors explained 48% of the overall variance (8% within-subject variance and 56% between-subject variance).

In the final model of accuracy (see Table 3), the effect of fatigue and PA disappeared and no other mood effects were found. Participants made more mistakes when being distracted (*B*=-.51, *p*<.001). The results of this analysis indicated that the six predictors explained 3% of the overall variance (2% within-subject variance and 20% between-subject variance).

Table 2

Final model of the mDSST speed outcome

	B	SE	p	95% CI	
			Speed		
Model			< .001*		
PA	.04	.11	.74	-.18,	.26
NA	.01	.30	.97	-.57,	.59
PA x NA	-.02	.06	.81	-.14,	.11
Worry	.06	.04	.14	-.02,	.14
Company$	-.18	.09	.04*	-.34,	-.01
Distraction	-.15	.02	< .001*	-.20,	-.10
Age^2	-.002	.0002	< .001*	-.002,	-.001
Time$	.17	.10	.08	-.02,	.37
Study day$	.34	.18	.06	-.02,	.70

Note. PA = positive affect. NA = negative affect. PA x NA = interaction between positive and negative affect. Company$ = dummy variable of being alone versus with others. Age^2 = quadratic function of age. Time$ = log-transformed replication score. Study day$ = log-transformed day of study score. CI = Confidence Interval. *p < .05.

Table 3

Final model of the mDSST accuracy outcome

	B	SE	p	95% CI	
			Accuracy		
Model			< .001*		
PA	-.22	.50	.67	-1.19,	.76
NA	-1.15	1.28	.37	-3.66,	1.36
PA x NA	.25	.28	.37	-.30,	.81
Fatigue	-.22	.14	.13	-.50,	.06
Focus	.14	.15	.34	-.15,	.44
Distraction	-.51	.11	< .001*	-.72,	-.30

Note. PA = positive affect. NA = negative affect. PA x NA = interaction between positive and negative affect. CI = Confidence Interval. *p < .05.

Young vs. older age and cognition

In order to gain more insight into age effects, exploratory subgroup analyses were performed for both cognition tasks. Splitting age groups for the mVSWMT data resulted in n=26 (59%) in the young group (<45 years) and n=18 (41%) in the older age group (45 years or older). In both groups, being able to focus resulted in more correct answers (young age: B=.05, p<.001; older age: B=.04, p=.01). However, in the older group, participants made more mistakes when being in company (B=-.13, p=.001) and with increasing age (B=-.00007, p=.047).

With the mDSST data, data from two participants were excluded due to technical problems (see section 3.4.2), leaving n=26 in the young group (62%) and n=16 in the older age group (38%). Results for speed remained largely the same in both groups, showing a main effect of distraction with higher distraction resulting in more mistakes. A similar result was found in both groups for accuracy, where higher distraction resulted in fewer trials. The difference was that the younger group completed more trials over time (B=.30, p=.047). The older group did not show a time effect on number of beeps (log-transformed replication score), but completed more trials over study days (B=.47, p=.02) and when PA was higher (B=.33, p=.04). In addition, fewer trials were completed when being in company (B=-.22, p=.02), and with increasing age (B=-.001, p<.001) within the older group but not the younger group. Results of the subgroup analyses are presented in supplementary materials D.

Correlations between mVSWMT and mDSST

Fisher-z transformations of by subject pairwise correlations were used between the mVSWMT outcome correct/incorrect, and the mDSST outcomes speed and accuracy. Over subject averages, there were no significant correlations between correct/incorrect and speed (z=1.36, p=.09), nor accuracy (z=-0.35, p=.64).

Discussion

Feasibility of the momentary cognition tasks

This study confirms the feasibility of two newly developed momentary cognition tasks within the PsyMate™ app in healthy individuals. The completion rate was high (71%) and is in line with other ESM studies with and without cognition tasks[46,47]. Furthermore, participants overall experienced the cognition tasks as pleasant and were motivated to perform well.

Although entertaining, the mVSWMT was experienced as difficult. In only 37% of the mVSWMT assessments the icon location was remembered correctly, with a range from 7% to 74% between participants. Differences may reflect the use of strategies, such as thinking

of an 'icon-story' or trying to group the icons per row. In other momentary visuospatial working memory studies, participants identified the correct location(s) in 90% of their responses[26,27,29]. In these studies, neutral circles in a grid were presented while meaningful icons were used in the mVSWMT (see Figure 3 b). Remembering the location combined with the meaning of the icon requires a higher cognitive demand. The choice for the here-used icons was made to test a greater working memory capacity[48,49]. Furthermore, the meaningful icons were expected to motivate the participant, which was confirmed by the positive feedback (see section 3.2.2). Participants responded randomly on the interference items to perform better on the mVSWMT. Prospectively a different interference[29] should be considered. For example, a fixed timer could be applied to standardize the interference between encoding and recall.

In contrast, participants experienced the mDSST as easy, which resulted in a ceiling effect for accuracy (97% correct). This high accuracy is in line with another digital processing speed task named the Colour-Shape Test, where participants answered correctly in 97% of the attempts[15]. In the paper-pencil DSST version, participants also make little mistakes reflected by a high accuracy, while the speed outcome proves more sensitive to cognitive variations[42]. The mDSST speed performance varied between and within subjects, which indicates this outcome to be suitable in detecting momentary cognitive fluctuations.

These feasibility results confirm that both tasks are appropriate, but need fine-tuning. For instance, the font size could be increased or the screen rotated to further improve the mDSST. Nevertheless, some limitation need to be acknowledged. In the mVSWMT analyses, the position of the icons could not be taken into account due to technical limitations. Descriptive background analysis revealed that a slight primacy and recency effect appeared, as participants remembered the first or last icons slightly more often. The discrepancy between identified location and actual location may be an interesting aspect of a momentary working memory task in the future. Additionally, the sample was healthy and highly educated resulting in limited generalizability of the feasibility results. Next steps may include testing the adjusted mDSST and mVSWMT in a more diverse healthy and clinical populations.

Validity of momentary cognition tasks through contextualization
The validity was demonstrated by relating cognitive performance with intrapersonal factors (e.g., mood, fatigue, stress, sleep-related outcomes) and contextual factors (e.g., being in company of others) to evaluate and understand momentary cognitive performance. Surprisingly, mood and fatigue had no effect on cognitive performance. One explanation could be that the participants were healthy, well rested, and overall in a positive mood (see

section 3.3). Timmers, Maeghs, Vestjens, Bonnemayer, Hamers, Blokland [32] also found no indication for an effect of fatigue on cognition (i.e., short-term memory) in healthy young adults. Previous findings on the relation between mood and cognition in daily life are inconclusive. While one study found no association between changes in mood and cognitive functioning[35], another study reported that higher positive mood resulted in less interference during an emotional Stroop task[36]. Stronger associations may appear in clinical populations[18,25].

Individual performance on the mVSWMT and mDSST was diminished when distracted and in social company. Logically and confirmed by experiments, distraction has a negative influence on cognitive performance[50]. The negative influence of being in company was also previously reported: von Stumm [35] argues that being alone may help to focus one's attention and thus improve cognitive performance. In contrast, it cannot be assumed that social context truly lowers a person's cognitive ability. This result may rather be related to variations in situational demands[35]. In the present study, participants were in company in 71% of the time. Hence, future studies may take the potentially mediating factors of company and distraction into account when analyzing momentary cognitive fluctuations.

Age sensitivity was found for both tasks with associations between age and visuospatial working memory and processing speed in everyday life. As expected[51], younger adults performed better. It is important to disentangle the cognitive decline in performance from an assessment bias. Compared to younger adults, older adults tend to experience technologies as less easy to use[52]. Additionally, older adults may have impaired hearing or vision, potentially affecting the usability and thus outcomes of technology-based assessments. Previous studies, however, confirmed the feasibility, reliability, and validity of digital assessments in the elderly[13,15,28]. All participants became better over time, but the learning curve was steeper for younger adults than for older adults. Learning effects were reported in previous mobile cognition tasks and may not affect sensitivity negatively[28]. Descriptive background analysis revealed that learning stabilized after 10 to 15 beeps and a steady state is reached within the first days of the study.

The non-significant correlation between the two momentary tasks suggests that different cognitive domains are measured, namely processing speed (mDSST) and visuospatial working memory (mVSWMT). This finding is in line with previous research that also found no correlation between momentary working memory and processing speed, possibly due to the unreliability of one task[35]. Strategy use was different between the two momentary tasks in this study. The lack of correlation could be due to a different approach in both tasks,

hindering a comparison of cognitive performance per se. It is expected that the current tasks will correlate after adjustment are made and when tested in a clinical sample with cognitive complaints. If no correlation shows, it might be that strategy use is moderating cognitive performance.

The present validation study did not focus on correlations with traditional neuropsychological tests. However, an in-house (unpublished) trial with 50 healthy participants showed that outcomes of a two-minute mDSST and paper-pencil DSST correlated (partial r=.50, p<.001). Participants reported that the digital version was slightly more difficult, as learning the digit-symbol combinations was challenging and the next number could not be anticipated. This preliminary finding provides initial evidence that the 30-second mDSST measures processing speed, but needs extension to confirm construct validity.

Overall, these validity findings can be seen as a proof of concept for the contextualization of momentary cognitive performance. Future research is needed to disentangle the complex interaction between mood, context, and cognition further. In addition, limitations include not taking the education level and the relation with traditional neuropsychological tests into account as part of the validation. Furthermore, the developed tasks are still artificial[7], in the sense that individuals normally do not perform these tasks, but actually search for their keys or process information to plan their days.

Next steps are the final adjustments of the app and testing both tasks in populations with different cognitive profiles. Clinical populations may include people with schizophrenia, major depression, or brain damage. Furthermore, situations that influence the cognitive performance such as tiredness, alcohol use, or medication intake may provide relevant insight into the validity of the tasks. Cognitive assessments in everyday life, in combination with momentary sampling of mood and context, may prospectively give individuals more insight into their functioning and thus support self-management and planning. This study is an important step in the anticipated personalization of holistic mobile health[6].

Suggestions for future development and use of momentary cognition tasks
Reflecting on the overall development and use of the mDSST and mVSWMT, a number of lessons were learned and can be implemented in future studies (see Table 4). Most momentary cognition tasks are inspired by traditional neuropsychological tests. A direct translation can be problematic due to the variability of everyday life that needs repeated assessments in complex environments. Smartphone assessments can offer new possibilities

for task development and use. Furthermore, input from a multi-disciplinary team involving neuropsychological healthcare professionals and ESM experts should guide the development. Other stakeholders, including clinicians and patients, should be consulted during development and evaluation. The participant's self-report and observations of the technology use[53] can provide insights into their perspective and experience when measuring momentary cognition within an ESM paradigm.

Gamification is a strategy to increase motivation. An example of gamification to measure cognition is the Sea Hero Quest smartphone app (www.seaheroquest.com). Participants orient themselves in a virtual sea world and get rewards when performing well. This quest may be a valid method to asses navigation skills in a fun way[54], however, it may be less suited for clinical practice were repeated assessments of cognitive performance are interwoven with assessments of mood and context. It is important to strike a balance between a thoughtful completion of the ESM items and the competitiveness and enjoyment of a cognitive performance task.

Testing the cognition task in a healthy sample is a useful way to test feasibility and validity[26,32]. Adjustments can be made before introducing the task to a more vulnerable clinical population. A benefit of ESM is that individuals can be their own controls and performance can be compared within one dataset[8,55].

Another aspect to consider is the beep frequency. While a high intensity may reveal more fluctuations, beep length and time investment need to be considered. The beep length should not exceed a couple of minutes to prevent interference with everyday life. Potentially, tasks can alternate at random across assessments to minimize fatigue effects. In general, participant's experience should be explored and if necessary, guide task adjustments when feelings of over- or under stimulation appear. Further, task difficulty may be tailored to the individual's ability. Ideally, momentary assessment promotes a flow in everyday life. Tasks should remain challenging when users reach a level of experienced achievement that relates to the individual's overall cognitive ability (e.g., being correct in 70% of the cases). The right difficulty level can prevent a loss of motivation and simultaneously preclude frustration or even resentment[56,57]. Prospectively, studies may consider automatically adjusting the level based on an individual's performance.

When implementing momentary cognition tasks, the sampling duration needs to be tailored to the research or clinical question. Examples include single case assessments running over months or years, where continued experience sampling can illustrate useful insight into the

course of a disease[58,59]. In the assessment process at memory clinics, weeklong momentary cognition tasks can supplement traditional neuropsychological test batteries and provide information on cognitive fluctuations in everyday life[13]. Side effects of new medications may furthermore be detected when using momentary cognition tasks parallel to the dose adjustments of drugs[17]. Finally, in rehabilitation centers, e.g., after brain damage, momentary cognition tasks may determine the effectiveness of the treatment when applied before and after a program. Insight into momentary cognitive fluctuations in context can be used to provide individuals with feedback and guidance to deal with cognitive complaints in daily life.

Table 4

Suggestions for future task development and use

Suggestions for momentary cognition tasks	
Task development	Task use
• Involve a multi-disciplinary team • Orientate concepts on traditional neuropsychological tests • Balance enjoyment/gamification with context information (experiences and physical context) • Ideal outcomes need to show clinically relevant within- and between-subject variance, be age-sensitive, and show no ceiling-effect • Use comparison data to determine between-subject variance; within-subject data serves as its own control	• Tailor beep frequency and sampling duration to the research/clinical question • Balance length and number of tasks, and additional momentary items • Limit assessment time (e.g., 2 minutes) • Adjust difficulty levels to individual abilities to prevent frustration and maintain motivation • Consider momentary context during interpretation (e.g. distraction) • Consider learning-effects (particularly in early trials)

Conclusions

Momentary cognition tasks aim to depict fluctuation of cognitive performance in everyday life and hold promise for future research and clinical use. Prospectively, the task application needs to be extended, for example into different cognitive domains or patient populations. Furthermore, the interaction with other intrapersonal factors requires further disentanglement. Next steps can be guided by the suggestions resulting from this study such as involving a multi-disciplinary team, tailoring the set-up to the individual, and balancing the level of enjoyment and seriousness.

References

1. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review.* 2003;13(4):181-197.

2. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience Sampling Method *J Nverv Ment Dis.* 1987;175(9):526-536.

3. Delespaul PAEG. *Assessing schizophrenia in daily life: The experience sampling method*. Maastricht, Maastricht University; 1995.

4. Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment *The Annual Review of Clinical Psychology* 2008;4(1):1-32.

5. Hébert ET, Stevens EM, Frank SG, et al. An ecological momentary intervention for smoking cessation: the associations of just-in-time, tailored messages with lapse risk factors. *Addictive behaviors.* 2018;78:30-35.

6. van Os J, Verhagen S, Marsman A, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and anxiety.* 2017;34(6):481-493.

7. Bielak AA, Hatt CR, Diehl M. Cognitive Performance in Adults' Daily Lives: Is There a Lab-Life Gap? *Research in Human Development.* 2017;14(3):219-233.

8. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: A systematic review. *International journal of methods in psychiatric research.* 2017;26(4):e1562.

9. Huntley J, Howard R. Working memory in early Alzheimer's disease: a neuropsychological review. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences.* 2010;25(2):121-132.

10. Silveri MC, Reali G, Jenner C, Puopolo M. Attention and memory in the preclinical stage of dementia. *Journal of Geriatric Psychiatry and Neurology.* 2007;20(2):67-75.

11. Luck SJ, Gold JM. The construct of attention in schizophrenia. *Biological psychiatry.* 2008;64(1):34-39.

12. Roitman SEL, Mitropoulou V, Keefe RS, et al. Visuospatial working memory in schizotypal personality disorder patients. *Schizophrenia Research.* 2000;41(3):447-455.

13. Allard M, Husky M, Catheline G, et al. Mobile technologies in the early detection of cognitive decline. *PLoS One.* 2014;9(12):e112197.

14. Bouvard A, Dupuy M, Schweitzer P, et al. Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls. *The American journal on addictions.* 2018;27(7):553-556.

15. Brouillette RM, Foil H, Fontenot S, et al. Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PloS one.* 2013;8(6):e65925.

16. Dirk J, Schmiedek F. Fluctuations in elementary school children's working memory performance in the school context. *Journal of Educational Psychology.* 2016;108(5):722.

17. Frings L, Wagner K, Maiwald T, et al. Early detection of behavioral side effects of antiepileptic treatment using handheld computers. *Epilepsy & Behavior.* 2008;13(2):402-406.

18. Hung S, Li M-S, Chen Y-L, Chiang J-H, Chen Y-Y, Hung GC-L. Smartphone-based ecological momentary assessment for Chinese patients with depression: An exploratory study in Taiwan. *Asian journal of psychiatry.* 2016;23:131-136.

19. Hyun J, Sliwinski MJ, Smyth JM. Waking up on the wrong side of the bed: The effects of stress anticipation on working memory in daily life. *The Journals of Gerontology: Series B.* 2018;74(1):38-46.

20. Keenan EK, Tiplady B, Priestley CM, Rogers PJ. Naturalistic effects of five days of bedtime caffeine use on sleep, next-day cognitive performance, and mood. *Journal of caffeine research.* 2014;4(1):13-20.

21. Kennedy DO, Veasey RC, Watson AW, et al. Vitamins and psychological functioning: a mobile phone assessment of the effects of a B vitamin complex, vitamin C and minerals on cognitive performance and subjective mood and energy. *Human Psychopharmacology: Clinical and Experimental.* 2011;26(4-5):338-347.

22. Lee W, Williams DR, Evans A. Feasibility of Smartphone-Based Testing of Interference in Parkinson's Disease. *Neurodegenerative Diseases.* 2018;18:133-142.

23. Marhe R, Waters AJ, van de Wetering BJ, Franken IH. Implicit and explicit drug-related cognitions during detoxification treatment are associated with drug relapse: an ecological momentary assessment study. *Journal of consulting and clinical psychology.* 2013;81(1):1.

24. Riediger M, Wrzus C, Klipker K, Müller V, Schmiedek F, Wagner GG. Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and aging.* 2014;29(1):103.

25. Schulze L, Bürkner P-C, Bohländer J, Zetsche U. Cognitive control and daily affect regulation in major depression and borderline personality disorder: protocol for an experimental ambulatory assessment study in Berlin, Germany. *BMJ open.* 2018;8(10):e022694.

26. Schuster RM, Mermelstein RJ, Hedeker D. Acceptability and feasibility of a visual working memory task in an ecological momentary assessment paradigm. *Psychological assessment.* 2015;27(4):1463.

27. Schuster RM, Mermelstein RJ, Hedeker D. Ecological momentary assessment of working memory under conditions of simultaneous marijuana and tobacco use. *Addiction.* 2016;111(8):1466-1476.

28. Schweitzer P, Husky M, Allard M, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *International journal of methods in psychiatric research.* 2017;26(3):e1521.

29. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. *Assessment.* 2018;25(1): 14-30.

30. Suffoletto B, Goyal A, Puyana JC, Chung T. Can an app help identify psychomotor function impairments during drinking occasions in the real world? A mixed-method pilot study. *Substance abuse.* 2017;38(4):438-449.

31. Thomson A, Nimmo A, Tiplady B, Glen J. Evaluation of a new method of assessing depth of sedation using two-choice visual reaction time testing on a mobile phone. *Anaesthesia.* 2009;64(1):32-38.

32. Timmers C, Maeghs A, Vestjens M, Bonnemayer C, Hamers H, Blokland A. Ambulant cognitive assessment using a smartphone. *Applied Neuropsychology: Adult.* 2014;21(2):136-142.

33. Tiplady B, Oshinowo B, Thomson J, Drummond GB. Alcohol and cognitive function: assessment in everyday life and laboratory settings using mobile phones. *Alcoholism: Clinical and Experimental Research.* 2009;33(12):2094-2102.

34. Veasey R, Haskell-Ramsay C, Kennedy D, Tiplady B, Stevenson E. The effect of breakfast prior to morning exercise on cognitive performance, mood and appetite later in the day in habitually active women. *Nutrients.* 2015;7(7):5712-5732.

35. von Stumm S. Feeling low, thinking slow? Associations between situational cues, mood and cognitive function. *Cognition and Emotion.* 2018;32(8):1545-1558.

36. Waters AJ, Li Y. Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. *Psychopharmacology.* 2008;197(1):25-35.

37. Waters AJ, Szeto EH, Wetter DW, Cinciripini PM, Robinson JD, Li Y. Cognition and craving during smoking cessation: an ecological momentary assessment study. *nicotine & tobacco research.* 2013;16(Suppl_2):S111-S118.

38. Jongeneel A, Scheffers D, Tromp N, et al. Reducing distress and improving social functioning in daily life in people with auditory verbal hallucinations: study protocol for the 'Temstem'randomised controlled trial. *BMJ open.* 2018;8(3):e020537.

39. Verhagen SJ, Berben JA, Leue C, et al. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. *PloS one.* 2017;12(10):e0186294.

40. Edwards CJ, Cella M, Tarrier N, Wykes T. The optimisation of experience sampling protocols in people with schizophrenia. *Psychiatry research.* 2016;244:289-293.

41. van Knippenberg R, De Vugt M, Ponds R, Myin-Germeys I, van Twillert B, Verhey F. Dealing with daily challenges in dementia (deal-id study): an experience sampling study to assess caregiver functioning in the flow of daily life. *International journal of geriatric psychiatry.* 2017;32(9):949-958.

42. Wechsler D. *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV).* San Antonio, TX: The Psychological Corporation; 2008.

43. Lavenex PB, Lecci S, Prêtre V, et al. As the world turns: short-term human spatial memory in egocentric and allocentric coordinates. *Behavioural brain research.* 2011;219(1):132-141.

44. Schumann-Hengsteler R. Children's and adults' visuospatial memory: The game concentration. *The Journal of genetic psychology.* 1996;157(1):77-92.

45. StataCorp L. Stata multilevel mixed-effects reference manual. *College Station, TX: StataCorp LP.* 2013.

46. Lenaert B, Colombi M, van Heugten C, Rasquin S, Kasanova Z, Ponds R. Exploring the feasibility and usability of the experience sampling method to examine the daily lives of patients with acquired brain injury. *Neuropsychological rehabilitation.* 2019;29(5):754-766.

47. Swendsen J, Schweitzer P, Moore RC. Mobile cognitive testing using experience sampling. In: Palmier-Claus J, Haddock G, Varese F, eds. *Experience Sampling in Mental Health Research.* New York: Routledge; 2019:142-155.

48. Brady TF, Störmer VS, Alvarez GA. Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences.* 2016;113(27):7459-7464.

49. Chen X, Li B, Liu Y. The impact of object complexity on visual working memory capacity. *Psychology.* 2017;8(06):929.

50. Lustig C, Hasher L, Tonev ST. Distraction as a determinant of processing speed. *Psychonomic bulletin & review.* 2006;13(4):619-625.
51. Zimprich D, Kurtz T. Individual differences and predictors of forgetting in old age: the role of processing speed and working memory. *Aging, Neuropsychology, and Cognition.* 2013;20(2):195-219.
52. Hauk N, Hüffmeier J, Krumm S. Ready to be a silver surfer? A meta-analysis on the relationship between chronological age and technology acceptance. *Computers in Human Behavior.* 2018;84:304-319.
53. Bartels S, Assander S, Patomella A-H, Jamnadas-Khoda J, Malinowsky C. Do you observe what I perceive? The relationship between two perspectives on the ability of people with cognitive impairments to use everyday technology. *Aging & mental health.* 2019:1-11.
54. Coutrot A, Schmidt S, Pittman J, et al. Virtual navigation tested on a mobile app (Sea Hero Quest) is predictive of real-world navigation performance: preliminary data. *bioRxiv.* 2018:305433.
55. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evidence-based mental health.* 2016;19(3):86-89.
56. Csikszentmihalyi M. *Finding flow: The psychology of engagement with everyday life.* Basic Books; 1997.
57. Engeser S, Rheinberg F. Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion.* 2008;32(3):158-172.
58. Lenssen J, Verhagen S. Monitoring my journey from doctor, to patient, to doctor with lived experience. *Schizophrenia Bulletin.* In Press.
59. Wichers M, Groot PC, Psychosystems E, Group E. Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and psychosomatics.* 2016;85(2):114-116.

Supplementary materials
Table A.1
Experience sampling protocol: Beep questionnaire

	Item	7-point Likert scale or categorical options
1	I feel cheerful	1 = not at all 4 = moderate 7 = very much
2	I feel energetic	1 = not at all 4 = moderate 7 = very much
3	I feel insecure	1 = not at all 4 = moderate 7 = very much
4	I feel relaxed	1 = not at all 4 = moderate 7 = very much
5	I feel down	1 = not at all 4 = moderate 7 = very much
6	I feel irritated	1 = not at all 4 = moderate 7 = very much
7	I feel satisfied	1 = not at all 4 = moderate 7 = very much
8	I feel lonely	1 = not at all 4 = moderate 7 = very much
9	I feel enthusiastic	1 = not at all 4 = moderate 7 = very much
10	I feel anxious	1 = not at all 4 = moderate 7 = very much
11	I feel guilty	1 = not at all 4 = moderate 7 = very much
12	I'm worrying about things	1 = not at all 4 = moderate 7 = very much
13	mVSWMT instruction screen	
	mVSWMT part 1: encoding	
14	I think I remembered it all	1 = not at all 4 = moderate 7 = very much
15	I generally feel well at the moment	1 = not at all 4 = moderate 7 = very much
16	mVSWMT instruction screen	
	mVSWMT part 2: recall	
17	What am I doing (right before the beep)	work, school/ housekeeping/ self-care/ relaxing/ sport, movement/ eating, drinking/ traveling, on the road/ having a conversation/something else/nothing
18	I can do this well	1 = not at all 4 = moderate 7 = very much
19	This is difficult for me	1 = not at all 4 = moderate 7 = very much
20	I would rather be doing something else	1 = not at all 4 = moderate 7 = very much
21	I am focused	1 = not at all 4 = moderate 7 = very much
22	Where am I (just before the beep)	at home/at someone else's home/work, school/ public space/ on the road/ somewhere else
23	Who am I with (just before the beep)	partner/ family/ housemates/ friends/ colleagues/ acquaintances/ strangers, others/ nobody
24a	Company: I like this company	1 = not at all 4 = moderate 7 = very much
25a	Company: I would rather be alone	1 = not at all 4 = moderate 7 = very much
24b	Alone: I like being alone	1 = not at all 4 = moderate 7 = very much
25b	Alone: I would rather be in company	1 = not at all 4 = moderate 7 = very much
26	I don't feel well	1 = not at all 4 = moderate 7 = very much
27	I am tired	1 = not at all 4 = moderate 7 = very much

28	Since the last beep I have used	alcohol/ medication/ coffee, caffeine/ smoking, nicotine/ cannabis/ other drugs/ nothing
29	mDSST instruction screen	
	mDSST	30-seconds timeframe
30	I got distracted during the task	1 = not at all 4 = moderate 7 = very much
31	This beep disturbed me	1 = not at all 4 = moderate 7 = very much
31	Thanks!	

Table A.2

Experience sampling protocol: Morning questionnaire

	Item	7-point Likert scale or categorical options
1	I generally felt well today	1 = not at all 4 = moderate 7 = very much
2	I generally felt tired today	1 = not at all 4 = moderate 7 = very much
3	I generally felt tense today	1 = not at all 4 = moderate 7 = very much
4	I generally worried a lot today	1 = not at all 4 = moderate 7 = very much
5	I generally could concentrate well today	1 = not at all 4 = moderate 7 = very much
6	I generally felt forgetful today	1 = not at all 4 = moderate 7 = very much
7	Goodnight!	1 = not at all 4 = moderate 7 = very much

Table A.3

Experience sampling protocol: Evening questionnaire

	Item	7-point Likert scale or categorical options
1	How long did it take before I fell asleep last night?	0-5 minutes/ 5-15 minutes/ 15-30 minutes/ 30-45 minutes/ 45 minutes-1 hour/ 1-2 hours/ 2-4 hours/ >4 hours
2	How often did I wake up last night?	0/1/2/3/4/5/>5
3	How long did I lie awake this morning before getting up?	0-5 minutes/ 5-15 minutes/ 15-30 minutes/ 30-45 minutes/ 45 minutes-1 hour/ 1-2 hours/ 2-4 hours/ >4 hours
4	I slept well	1 = not at all 4 = moderate 7 = very much
5	I feel well rested	1 = not at all 4 = moderate 7 = very much
6	I am looking forward to this day	1 = not at all 4 = moderate 7 = very much
7	Thanks!	1 = not at all 4 = moderate 7 = very much

Table B

Individual multilevel regression analyses of mood, context, and sleep quality on the mVSWMT outcome correct/incorrect

	Correct/Incorrect			95% CI	
	B	SE	p		
Positive affect	-.03	.02	.04*	-.06,	-.001
Negative affect	-.0003	.03	.99	-.05,	.05
Worry	.002	.01	.89	-.02,	.02
Fatigue	-.01	.01	.54	-.02,	.01
Focus	.04	.01	<.001*	.02,	.06
Activity-related stress	-.01	.01	.49	-.03,	.02
Company[$]	-.10	.03	<.001*	-.16,	-.05
Location[$]	-.02	.03	.38	-.07,	.03
Distraction	-.03	.01	.001*	-.04,	-.01
Time until sleep	.01	.01	.23	-.01,	.04
Number of wake-ups	-.005	.01	.67	-.03,	.02
Slept well	.01	.01	.47	-.01,	.03
Well rested	.005	.01	.65	-.02,	.03
Age^2	-.00005	.00002	.002*	-.00009,	-.00002
Gender	-.04	.05	.47	-.14,	.06

Note. Company[$] = dummy variable of being alone versus with others. Location[$] = dummy variable of being at home versus somewhere else. Age^2 = quadratic function of age. CI = Confidence Interval. *$p < .05$.

Table C

Individual multilevel regression analyses of mood, context and sleep quality on the mDSST outcomes speed and accuracy

	Speed			95% CI		Accuracy			95% CI	
	B	SE	p			B	SE	p		
PA	-.01	.05	.90	-.11,	.10	.45	.21	.03*	.03,	.87
NA	-.05	.10	.61	-.24,	.14	-.39	.34	.25	-1.04,	.27
Worry	.09	.04	.02*	.01,	.17	-.20	.15	.17	-.50,	.09
Fatigue	-.02	.03	.46	-.09,	.04	-.34	.12	.007*	-.58,	-.09
Focus	.02	.03	.50	-.04,	.08	.36	.14	.008*	.09,	.63
Act-stress[$]	.05	.04	.18	-.03,	.13	-.14	.18	.44	-.48,	.21
Company[$]	-.29	.09	.001*	-.46,	-.12	-.003	.39	.99	-.77,	.76
Location[$]	-.09	.08	.26	-.25,	.07	.34	.36	.34	-.36,	1.05
Distraction	-.15	.02	<.001*	-.20,	-.10	-.54	.41	<.001*	-.74,	-.33
Time-sleep[$]	-.04	.04	.32	-.11,	.04	-.07	.17	.67	-.40,	.25
Wake-ups[$]	-.01	.04	.82	-.08,	.07	-.08	.16	.63	-.40,	.24
Slept well	-.01	.04	.84	-.08,	.07	-.16	.17	.35	-.49,	.17
Well rested	.01	.04	.86	-.07,	.08	.08	.16	.62	-.24,	.40
Age^2	-.001	.0002	<.001*	-.002,	-.001	.0002	.0002	.47	-.0003,	.0006
Gender	.47	.85	.58	-1.19,	2.14	.71	.58	.23	-.44,	1.85

Note. PA = positive affect. NA = negative affect. Act-stress[$] = activity-related stress. Company[$] = dummy variable of being alone versus with others. Location[$] = dummy variable of being at home versus somewhere else. Time-sleep[$] = time until sleep. Wake-ups[$] = number of wake-ups at night. Age^2 = quadratic function of age. CI = Confidence Interval. *$p <$ .05.

Table D.1

Subgroup analyses on age for the mVSWMT outcome correct/incorrect

	< 45 years of age					45 years or older				
	B	SE	p	95% CI		B	SE	p	95% CI	
Model			< .001*					< .001*		
PA	-.08	.05	.08	-.18,	.009	-.01	.06	.88	-.14,	.12
NA	-.10	.11	.36	-.32,	.12	.11	.22	.62	-.32,	.54
PA x NA	.02	.03	.56	-.04,	.07	-.03	.04	.55	-.11,	.06
Focus	.05	.01	< .001*	.02,	.08	.04	.02	.01*	.008,	.07
Company$	-.05	.04	.21	-.13,	.03	-.13	.04	.001*	-.21,	-.05
Distraction	-.02	.01	.11	-.04,	.004	-.01	.01	.22	-.04,	.009
Age^2	-.00005	.00009	.55	-.0002,	.0001	-.00007	.00004	.047*	-.0001,	-1.13e-06

Note. PA = positive affect. NA = negative affect. PA x NA = interaction between positive and negative affect. Company$ = dummy variable of being alone versus with others. Age^2 = quadratic function of age. CI = Confidence Interval. *p < .05.

Table D.2

Subgroup analyses on age for the mDSST speed outcome

	< 45 years of age					45 years or older				
	B	SE	p	95% CI		B	SE	p	95% CI	
Model			<.001*					<.001*		
PA	-.05	.17	.77	-.37,	.28	.33	.16	.04*	.02,	.63
NA	-.27	.39	.49	-1.04,	.49	1.01	.53	.06	-.02,	2.05
PA x NA	.03	.09	.74	-.14,	.20	-.17	.11	.10	-.38,	.04
Worry	.10	.06	.10	-.02,	.21	-.01	.05	.78	-.12,	.09
Company$	-.17	.13	.20	-.43,	.09	-.22	.09	.02*	-.40,	-.03
Distraction	-.19	.04	<.001*	-.26,	-.13	-.08	.03	.004*	-.13,	-.02
Age^2	-.0007	.001	.57	-.003,	.002	-.001	.0003	<.001*	-.002,	-.0008
Time$	.30	.15	.047*	.004,	.59	.009	.11	.94	-.21,	.23
Study day$	.24	.28	.40	-.31,	.78	.47	.21	.02	.07,	.88

Note. PA = positive affect, NA = negative affect. PA x NA = interaction between positive and negative affect. Location = dummy variable of being at home versus somewhere else. Company$ = dummy variable of being alone versus with others. Age^2 = quadratic function of age, Time$ = log-transformed replication score. Study day$ **=** log-transformed day of study score. CI = Confidence Interval. *p < .05.

Table D.3

Subgroup analyses on age for the mDSST accuracy outcome

	< 45 years of age			95% CI		45 years or older			95% CI	
	B	SE	p			B	SE	p		
Model			.004*					.003*		
PA	-.86	.68	.21	-.2.18,	.47	.88	.91	.34	-.91,	2.67
NA	-1.92	1.57	.22	-5.004,	1.15	1.11	3.11	.72	-4.98,	7.20
PA x NA	.37	.36	.30	-.34,	1.08	-.15	.61	.80	-1.36,	1.05
Fatigue	-.34	.20	.10	-.74,	.06	-.08	.20	.70	-.47,	.31
Focus	.25	.21	.23	-.15,	.65	-.02	.22	.93	-.45,	.41
Distraction	-.44	.15	.003*	-.73,	-.15	-.57	.16	<.001*	-.88,	-.27

Note. PA = positive affect. NA = negative affect. PA x NA = interaction between positive and negative affect. CI = Confidence Interval. *p < .05.

10

General discussion

General discussion

Variability is an important factor to understand what is going on in someone's life. People continuously navigate in changing environments and are subject to a multitude of impressions over short amounts of time. Not only the outside world changes. Internal dynamics, involving mood and cognition, are equally relevant. The experience sampling method (ESM) is a powerful assessment tool that can help to grasp the individual variability of internal mental states and external context factors. It has a long tradition in academia and in scientific research[1,2]. The central aim of this thesis is to explore the use of ESM in clinical practice and expand its scope to mental health services. This discussion focusses on the main findings and their implications, the lessons-learned and the future steps towards new horizons of good clinical practice.

Individual landscapes

PsyMate™ is an online ESM platform including an app, a data cloud-based data solution, and a reporting tool (https://www.psymate.eu). It allows persons to collect data and access their data online, using a secure website. Reviewing their own data allows people to make sense of mental state variability, a process that becomes meaningful when more ESM assessments are collected. Feeling cheerful most likely varies somewhat during a day. The reporting module presents the mental state at each moment, as a dot on a line. Sometimes, patterns start to appear. Activities, places, and current company may also vary across time. Feedback on these patterns (depicted as pie charts and called time budgets) can increase insight. For instance, it is possible to see in what activities a person engages in most. Clinically, combining mental state variability with context (time, place, activity, and people) is most relevant. Imagine that someone wonders if their effort to go outside also results in higher positive mood. The scores on feeling cheerful can then be paired with the time budget of location. Bar charts indicate the level of cheerfulness when being outside, and compare these with other situations. An illustration of reporting module is presented in **chapter 5**. A high average score of cheerfulness when being outdoors can yield the clinically relevant hypothesis that this is a preferred context. Such hypotheses can be explored during therapy, for instance by instructing the subject to spend more time outdoors. Iteratively, this process can improve global well-being.

The above-mentioned description of the PsyMate™ reporting page illustrates how a personal landscape can be explored. A well-designed user interface is crucial to facilitate the exploration process and allow users to make sense of the data[3]. There are different ways to present the data and researchers continually collect feedback from the field. With contracted

designers, the reporting tool is periodically updated. The actual tool allows exploring different mood states. To contrast simple situations is a powerful method for creating awareness and insight, as demonstrated in the papers presented in **part I** of the thesis. The best illustration is the personal story presented in **chapter 5**, were the PsyMate™ is used for two years during the recovery of depression. These clinical explorations have let me to realize how important ESM can be for clinical practice. While collecting repeated assessments can be intrusive and burdensome, much is acceptable and motivation increases when the data is meaningful.

ESM feedback as hypothesis generating

Being able to see one's own mood patterns in relation to one's environment, without having to reconstruct these patterns from the mind is a great advantage. However, there is a risk that patients and clinicians misinterpret the data. As researchers and developers of these tools, we are responsible for the way we present data and confusion should be avoided. Patients might not be aware that differences lack significance and consequently over-interpret the data. The question rises if enough data is available to warrant conclusions with enough power. Patterns in sets of only ten assessments are more uncertain than patterns in fifty assessments. Often, even more observations are needed to reliably assess computed indicators of health and the actual number needed is related to the living circumstances of subjects. In **chapter 6** (**part II**), Monte Carlo simulations were used to find out how much assessments are needed to calculate a stable reward-related quality of life statistic. The study demonstrates that often more assessments are needed (> 100) than are available in a standard high frequency sampling week (< 60). Differences were found between persons. Individuals who spend more time at home and alone need fewer assessments than persons with more varying situational patterns. This statistical exploration reminds us that it is difficult and challenging to explore daily life. When patients (but also clinicians) explore and interpret the patient's ESM results, additional coaching is necessary.

It is important to realize that information is only partial information, especially in (predicting) human behavior. People live in a dynamic world and are in constant interaction with their environment. There are many unknowns when anticipating future mental states. This does not mean that data collection is useless and efforts towards a global, holistic understanding of the ecological reality in which individuals operate are fruitless. However, some caution is justified. Users should be conscious of the correlational nature of the data. Observed patterns are probabilistic, making deterministic interpretations tempting. Users should judge results in an open-minded manner and consider them as hypotheses ripe for testing. This hermeneutic exploration is a repeating, iterative process. In research, this exploration is best

done using a multimethod strategy, sometimes including experimental manipulation[4]. In clinical practice, a good hypothesis is often sufficient to act because the impact can be evaluated and adjustments made during treatment. ESM feedback can be viewed as a hypothesis-generating assessment instrument. It invites the individual to explore emerging patterns and provides clues for improving the personal health[2,5]. Clinical practice is enriched, because clinicians and patients gain a better understanding of what is going on. This contrasts with the classic static approach of diagnostics that tries to understand why complaints emerge[5]. This classical approach however, does not result in real etiological knowledge. The attempt is categorical at best and clinically rather futile (see the introduction).

Meaningfulness of ESM feedback

In an effort to enrich the clinical meaningfulness of ESM data, I have visited the Center for Self-Report Science at the University of Southern California in Los Angeles. This research group chaired by Arthur Stone – a pioneer in ESM research, develops expertise to improve accurate and reliable self-reports. They perform valuable work in the field of public health, exploring alternative ways of calculating outcome measures (for example in pain research) to increase their clinical value[6-8]. We worked together on a longitudinal single case ESM dataset, collected in a naturalistic way as part of a recovery process from depression (see **chapter 5**). The goal during the work visit was to better understand the approaches used in Los Angeles to process data. We focused on how to recognize patterns from data and to assess whether these patterns were clinically meaningful. Possibly, the PsyMate™ reporting page (used to view ESM data) could be enriched with new algorithms that help visualizing state changes during treatment.

The case data contained 868 unique assessments, collected over 209 days. Over that period, different interventions occurred: repeated medication switches, psychotherapy, and self-guided treatment such as daily exercises and vitamin use. When ESM is used for a long time, multiple patterns emerge in the data. Showing and understanding patterns of increased positive mood in different situations was relatively straightforward. It was much harder to pinpoint when a different state, episode, or recovery phase starts. The changes do not always relate to interventions and effects shift in time. State differences are ambiguous and relate to treatment-independent events in someone's life[9]. Observed differences might be triggered by changed circumstances or accumulated stress and are not the consequence of treatment choices per se. These complexities make it a challenge to preprocess the data in an online reporting module, to facilitate clinical decision-making.

In this case study, medication switched four times at recorded moments. This allows the mapping of medication windows over the data to assess mood changes or context changes (outcome). We chose to focus on positive and negative mood and calculated the mean and standard deviation over time intervals. We explored whether different ways to characterize the data provided other clinical meaningful information for the assessment of medication effects. Additionally, a synchronicity score between positive and negative mood (based on correlations) and a measure of mixed-emotions (using the lowest mood value) was calculated. Epochs of one week yielded sufficient observations for a reliable momentary assessment and assessments of change in the 200+ day monitoring period.

The results of this exercise showed high weekly variability among all indices and differences emerged between medication types. Furthermore, each of the indices provided unique information. However, the results were not very transparent and interpretation of its clinical meaning remained challenging. In such a long period of data collection, chaos emerges and formulating a clear message that is understandable for the average person is difficult. It might be that conclusions would be easier if the data stabilized in the end (as was not the case here). However, in clinical practice a true stable state is rare[9]. Maybe the above-mentioned indices are the wrong outcome measures for good clinical communication. A focus on contextual changes might yield better insights. For now, the strength of ESM feedback lies in the exploration of shorter periods, with a linear evolution. Episodic patterns remain challenging.

Daily cognitive assessments

Cognition, mood, perceptions and behavior are central elements in psychology. Traditional ESM assessment focusses on behavior and mood[10,11]. A narrow focus on mood within psychopathology can lead to bias[12]. ESM has been widely used for mood assessments in relation to behavior, but is less used to assess cognitions or perceptions (except for hallucinations in psychosis research). Assessment of cognition, a relevant domain in mental health, remains unexplored (or underexplored). This thesis intends to broaden the clinical usefulness of ESM by adding cognition measures. Multi-domain assessments fit well into the new health definition and the new healthcare movement (**chapter 1**).

Cognition is a relevant health domain that deserves attention in daily assessments. In mental healthcare, cognitive functioning plays a central role, in part because cognitive skills can vary over time[13,14]. Psychopathological complaints often co-occur with or lead to compromises in cognitive performance[15]. In psychosis-spectrum disorders or mood disorders, lowered cognitive functioning is even regarded as a core symptom[16,17]. In addition, cognitive

complaints often remain after symptomatic remission[18]. These lingering cognitive complaints can challenge recovery.

Traditionally, cognition is measured with large test-batteries, at specific moments in time and in controlled and standardized settings[19]. These assessments provide valuable information on state performance across cognitive domains, but fail to provide insight into cognitive fluctuations. In clinical practice, it is helpful to know when someone struggles to perform a task or when performance is actually going well. Insight in these dynamics with attention to the role of mood and environments can be helpful for patients, allowing them to adjust their daily life accordingly. **Part III** focused on enriching ESM data with momentary cognitive assessments and evaluating its benefit for future use in clinical practice.

Several short mobile cognitive tasks were conceptualized, developed, and tested for repeated assessments in the flow of daily life. Feasibility and initial validity were critically evaluated and the objective cognitive measurements seem promising in this regard. However, we are still in the initial stage of development and implementation. Correlational outcomes are found between mood, context and cognitive performance[20]. For example, people seem less able to perform well when in company or when tired. Knowledge on the causality of the correlations is still lacking. What do we measure when we find these outcomes? Is being tired actually leading to a decrease in cognitive performance or are people just less inclined to engage in the cognitive assessment? Choice of strategy was an important determinant in a momentary visuospatial memory task (**chapter 9**). This could potentially confound a conclusion on cognitive performance. Generalization needs to be determined and further research is needed to disentangle the different elements underlying outcomes.

The presented studies do not directly examine concurrent validity with traditional neuropsychological assessments. This step is important in order to understand the cognitive domains better. For daily cognitive assessments, do we need a cognitive measure for each cognitive domain or is it enough to have a general measure that can inform on cognitive performance fluctuations? Alternatively, maybe it is enough to have one good task that taps into neurocognitive functioning and one that taps into social cognitive skills, as is presented in **chapter 7**. There seems to be merit in these developments, but more research is needed before ESM is ready to be extended with objective cognition measures in clinical practice. It is highly recommended, however, to add subjective cognition items to an ESM questionnaire. Assessing fluctuations in concentration or memory already provide valuable insight for patients and are of clinical use[21].

Where do we stand with ESM in clinical practice?

ESM in its current form is particularly useful to monitor different health-related domains in daily life[5]. Assessments are possible on mobile phones and can be person tailored[2]. Moreover, the research in this thesis helped to promote ESM as an mHealth tool for regular clinical practice. Further development remains necessary. The representation of the data to patients and clinicians in the reporting tool should be simplified and more transparent. However, ESM gets momentum and more people are getting interested in using experience sampling and mHealth technology for health purposes. Time is ripe for change.

Commercial companies are noticing these developments[22,23]. They develop 'fancy' apps with many interesting features. Consumers are getting used to intuitive and attractive interfaces and expect a degree of user-friendliness in their tools. The focus of an academic center is not to develop market-ready apps, but to explore the aspects and working mechanisms of mHealth instrument experimentally. It is difficult to keep up with the fast changing design fashions of the market. This puts researchers under pressure. The development of a modern looking app requires large funds and multidisciplinary collaborations with app developers and design bureaus. Moreover, hardware and software of consumer devices (e.g., Android) continually change, challenging developers to keep the ESM app operational. Meanwhile, the field formulates new specs and innovative clinical applications are the most demanding[23]. Moreover, the European privacy laws (GDPR) redefine the role and responsibilities of eHealth service providers[24]. On the one hand, research is needed to strengthen the experience-sampling methodology for clinical practice. On the other hand, investments are needed to increase the data security, the user-friendliness, and intuitiveness of ESM as a mobile tool. Alternatives to mobile sampling such as passive tracking or smartwatch use might be necessary in the future.

In Maastricht, the department of Psychiatry and Neuropsychology accepted the challenge to implement this promising technology in routine clinical practice. This requires that patients and clinicians can access the technology, preferably by making it freely available to the public. Initial steps in this direction are being made and colleagues are working on the implementation of ESM at the general practitioners office and as a self-help tool[25]. Healthcare professionals are enthusiastic about the possibilities that ESM offers and are open to try the method in practice. Patients in turn become more empowered and actively engage in the process[26]. A lesson-learned is that implementation in clinical practice requires involvement from different stakeholders. It is important to involve users from the start and include patients, close-relatives, clinicians, healthcare providers, and healthcare managers[25].

Example of a clinical service innovation

In Dialectical Behavior Therapy (DBT) for borderline personality disorder, daily diaries are used to track suicidal ideations and dysfunctional behavior[27]. Reports are reviewed at the start of each therapy session. Clinicians experienced some problems with these diaries: end of day reports suffer from memory biases and lack contextualized information. Patients often struggle to reconstruct the circumstances of dysfunctional behavior from memory. Reporting is often forgotten and diaries are sometimes completed in the parking lot before a therapy session. ESM can be useful to support in situ data collection and supplement the information of the diaries with context descriptions of problematic moments, but also input on resilience (moments of optimal experience). A more dynamic picture can emerge. Due to the repeated momentary assessments, memory load is reduced, herewith facilitating data collection[28].

In this context, an ESM-DBT protocol was developed for use during individual therapy sessions at a mental health facility in the Netherlands (Mondriaan Mental Health Trust). The team members were a supervisor in cognitive behavioral therapy (also in training as a clinical psychologist) acting as project leader, DBT therapists and an ESM expert. The first step was to motivate the mental health facility and the DBT team to become involved. The process was a mutual exchange. ESM experts needed to learn about the DBT treatment and assess the needs, and DBT therapist had to explore the possibilities of ESM. Initial ideas for the development of a customized PsyMate™ supported DBT app were theory driven and contained straightforward implementations of existing assessment instruments. However, a direct translation to within-day assessments proved difficult. Therapeutic theory came in conflict with ESM momentary assessments needs, which have to be intuitive, short, and require minimal effort[1]. At that time, the PsyMate™ app was mainly developed for classic ESM research protocols and adjustments were needed to get the app technology and DBT framework ready for integration in clinical use. The collaborative development process took a year and resulted in modular applications for three different phases in the DBT treatment. One patient and her clinician started to use the app, providing valuable information on feasibility and its usefulness in clinical practice.

In the first module, ESM was used as a functional analyses tool that replaced the daily diaries. Patients download the app on their own smartphone and receive a questionnaire at eight random moments a day. To reduce reactivity and burden, three out of the eight within-day assessment moments assessed suicidal ideations and dysfunctional behavior. After a week, the patient and therapist examined the data together by using the online report platform. This data exploration formed the start of the treatment session. The second module was

initiated during the treatment phase aimed at skill acquisition. It includes elements of mindfulness and uses focusing exercises on internal processes (*kernoplettendheid* in Dutch). Later, the module was extended to include an exercise on attentiveness to help patients integrate the therapy skill into their daily life. An algorithm was used that helped patients to direct their attention to exercises that are appropriate considering their current state of mind. After three weeks, the last ESM module was activated. This module did not include exercises but provided momentary in-app feedback based on mood states. An algorithm was created that used the answer score on negative affect and combined this with the responses on the tendency towards dysfunctional behavior (e.g., impulsivity or self-harm). At the end of each questionnaire, a remark was made on their current mood states, followed by advice on actions that could be taken to lower negative tension or to stay in a healthy mode.

Several lessons were learned from this bottom-up development and implementation attempt. Implementation takes time and only works if there is support from the board of the institute. They should free clinicians from regular (production) duties and allow them to invest time during co-development. Starting from the theoretical background of DBT and within the limitations of technology resulted in a very interesting but rather complicated PsyMate™ mediated mHealth protocol. Trying to incorporate a multitude of elements leads to a multitude of problems. For instance, we found that the security regulations on internet traffic within the clinical setting hindered accessing the online reporting page. It took a long time before the data could pass correctly through the firewall, frustrating the clinical adoption time. In the end, the protocol was simplified to fit the current technology limitations and to make the procedure more intuitive for clinicians. Furthermore, institutional politics led to repeated drastic changes in the DBT team and to stagnation of the project. The therapist who co-developed the protocol switched jobs, leaving an insufficient work force for the ESM implementation of the DBT modules.

With these barriers in mind, it is worth mentioning that both the patient and therapist were enthusiastic. Exploring the data opened up the conversation to a more positive approach of the patient's weekly struggles in her daily life. Moments when suicidal ideations were low provided clues for resilience. The repeated assessments helped to reflect on the changes in emotional state in daily life, herewith increasing awareness. Repeated exercises and momentary feedback helped to strengthen the application of learned skills outside the therapy session. The patient asked if she could still use the app when her treatment ended, a promising sign for future implementation.

New horizons

There are still new horizons for ESM in clinical practice. The rich sampling data leaves room for further exploration to increase its merit. The in-app feedback that was used in the implementation example seems promising and provides an elegant way to strengthen the therapeutic effect. Promising is that in the future it will become easier to personalize the experience sampling methodology. Daily life assessments do not require the same nuances as traditional cross-sectional assessments need. While engaging in ESM research and guiding people in their clinical use of the method I became convinced that a basic, limited set of ESM items is universally relevant across disorders. Sometimes an item does not vary within an individual. To lower the burden of ESM assessments, these nonresponsive items could be disabled. Machine learning techniques might be useful to detect non-responsive items and disable them to customize the questionnaires. For instance, the item *I feel irritated* may be relevant for one person, but not for another person. Such an item could be disabled. Personalization also means a close fit to someone's perception of health. One solution is to add several personal items to a standard ESM questionnaire. This allows the monitoring of additional aspects that are relevant for the individual, but not part of the standard ESM questionnaire. The need for these personal aspects is an important point raised by clinicians and patients in the implementation project at the general practitioner office[25]. Personalization also includes the adjustment of sampling frequency. It is important to customize sampling schemes to fit the waking hours of an individual. However, we also learned that giving subjects too many options is frustrating and an inefficient way to collect data.

A good balance is essential between freedom of choice to increase motivation and limitations due to scientific knowledge of experts. People will intuitively ask for less frequent assessments, but experience learned that a continuous monitoring helps to increase flow and ultimately lowers the cognitive burden of repeated assessments. Explaining the background or concept of ESM is still required. A good mHealth tool for public use needs to be self-explaining and intuitive[22]. New horizons will include the development of mHealth tools that can be used without the briefing of an ESM expert. Gamification elements can aid in the motivation for continued ESM use[29]. Furthermore, there is a promising future in the use of passive monitoring to complement the active ESM elements[30]. Using GPS to map environmental aspects or activity trackers to map activation can provide additional information with low effort from participants[31].

The future horizon looks bright. Increasingly, we are moving forward and with every mile, we come closer to a better mental health for the majority of people. Research is an important facilitator that should go hand in hand with practical implications and societal implementations.

References

1. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. *Evidence-based mental health.* 2016;19(3):86-89.

2. van Os J, Verhagen S, Marsman A, et al. The experience sampling method as an mHealth tool to support self‑monitoring, self‑insight, and personalized health care in clinical practice. *Depression and anxiety.* 2017;34(6):481-493.

3. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. *Journal of medical Internet research.* 2007;9(2):e15.

4. Brewer J, Hunter A. *Multimethod research: A synthesis of styles.* Sage Publications, Inc; 1989.

5. van Os J, Delespaul P, Wigman J, Myin-Germeys I, Wichers M. Beyond DSM and ICD: introducing "precision diagnosis" for psychiatry using momentary assessment technology. *World Psychiatry.* 2013;12(2):113.

6. Schneider S, Stone AA, Schwartz JE, Broderick JE. Peak and end effects in patients' daily recall of pain and fatigue: a within-subjects analysis. *The Journal of Pain.* 2011;12(2):228-235.

7. Stone AA, Broderick JE, Schneider S, Schwartz JE. Expanding options for developing outcome measures from momentary assessment data. *Psychosomatic medicine.* 2012;74(4):387-397.

8. Schneider S, Junghaenel DU, Ono M, Stone AA. Temporal dynamics of pain: an application of regime-switching models to ecological momentary assessments in patients with rheumatic diseases. *Pain.* 2018;159(7):1346-1358.

9. Nelson B, McGorry PD, Wichers M, Wigman JT, Hartmann JA. Moving from static to dynamic models of the onset of mental disorder: a review. *JAMA psychiatry.* 2017;74(5):528-534.

10. Larson R, Csikszentmihalyi M. The experience sampling method. In: *Flow and the foundations of positive psychology.* Springer; 2014:21-34.

11. 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. *Psychological medicine.* 2009;39(9):1533-1547.

12. van Os J, Delespaul P, Barge D, Bakker RP. Testing an mHealth momentary assessment Routine Outcome Monitoring application: a focus on restoration of daily life positive mood states. *PLoS One.* 2014;9(12):e115254.

13. Lewandowski K, Cohen B, Öngur D. Evolution of neuropsychological dysfunction during the course of schizophrenia and bipolar disorder. *Psychological medicine.* 2011;41(2):225-241.

14. Joyce EM, Roiser JP. Cognitive heterogeneity in schizophrenia. *Current opinion in psychiatry.* 2007;20(3):268.

15. Millan MJ, Agid Y, Brüne M, et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature reviews Drug discovery.* 2012;11(2):141.

16. Fusar-Poli P, Deste G, Smieskova R, et al. Cognitive functioning in prodromal psychosis: a meta-analysis. *Archives of general psychiatry.* 2012;69(6):562-571.

17. Van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature.* 2010;468(7321):203.

18. Hasselbalch BJ, Knorr U, Kessing LV. Cognitive impairment in the remitted state of unipolar depressive disorder: a systematic review. *Journal of affective disorders.* 2011;134(1-3):20-31.

19. Casaletto KB, Heaton RK. Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society.* 2017;23(9-10):778-790.

20. Swendsen J, Schweitzer P, Moore RC. Mobile cognitive testing using experience sampling. In: Palmier-Claus J, Haddock G, Varese F, eds. *Experience Sampling in Mental Health Research.* New York: Routledge; 2019:142-155.

21. van Knippenberg RJ, de Vugt ME, Ponds RW, Verhey FR, Myin-Germeys I. Emotional reactivity to daily life stress in spousal caregivers of people with dementia: An experience sampling study. *PloS one.* 2018;13(4):e0194118.

22. Marzano L, Bardill A, Fields B, et al. The application of mHealth to mental health: opportunities and challenges. *The Lancet Psychiatry.* 2015;2(10):942-948.

23. Torous J, Andersson G, Bertagnoli A, et al. Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry.* 2019;18(1):97.

24. Voigt P, Von dem Bussche A. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed, Cham: Springer International Publishing.* 2017.

25. Daniëls NEM, Hochstenbach LMJ, Bokhoven MAv, Beurskens AJHM, Delespaul PAEG. Implementing Experience Sampling technology for functional analysis in family medicine - a design thinking approach. *Submitted for publication.* 2019.

26. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of Patients and Clinicians on the Promise of the Experience Sampling Method for Psychiatric Care. *Psychiatric Services.* 2019:appi. ps. 201900050.

28. Scollon CN, Prieto C-K, Diener E. Experience sampling: promises and pitfalls, strength and weaknesses. In: *Assessing well-being.* Springer; 2009:157-180.

29. Van Berkel N, Goncalves J, Hosio S, Kostakos V. Gamification of mobile experience sampling improves data quality and quantity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.* 2017;1(3):107.

30. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal.* 2015;38(3):218.

31. Doherty ST, Lemieux CJ, Canally C. Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. *Social Science & Medicine.* 2014;106:83-92.

Addendum

Addendum

Summary

Individual landscapes

The work presented in this thesis aims at exploring and expanding the possibilities of the Experience Sampling Method (ESM) for use in clinical practice. **Chapter 1** describes the current positive view on mental health care that includes a broad range of recovery domains and goes beyond symptom reduction. ESM as an mHealth tool can enhance care by making the individual's everyday life, where problems occur, accessible for assessment and treatment. After this introduction, the thesis has three separate parts. **Part I** describes the clinical relevance of ESM with a focus on the individual patient. **Part II** describes methodological considerations that are relevant to determine data reliability for innovative parameters developed for use in clinical care. **Part III** explores the benefits of complementing ESM with momentary cognition tests.

Clinical relevance

In **Chapter 2**, a general overview of ESM is provided, including advantages and disadvantages of the method for use in clinical trial research. ESM is a valuable tool to assess the effectiveness of an intervention as it provides fine-grained information on relevant outcomes in everyday life. Researchers are advised on how to design, analyze and administer the method. Applications include monitoring and intervening for group comparison and for single-case clinical trials. **Chapter 3** uses research findings to illustrate the benefit of ESM as a regular mobile tool for individual support. ESM as an mHealth tool can support self-monitoring, self-insight and personalized health care in routine clinical practice, if the data are readily accessible for users. The method helps to highlight vulnerability as well as resilience and provides insight into implicit patterns between mood, behavior and context. In treatment, ESM facilitates shared decision-making and empowerment. Blended care, with ESM complementing treatment as usual, may reduce costs because ESM can be used outside the context of professional care. **Chapter 4** assesses the reliability of ESM as a momentary Routine Outcome Monitoring (mROM) instrument in the general hospital. ESM as an mROM tool is reliable and has the advantage of informing individuals on their daily adaptive functioning. The method shows adequate psychometric properties, concurrent validity and sensitivity to change over time on relevant recovery outcomes (i.e., positive affect, negative affect, quality of sleep, positive social interactions, activity-related stress, feeling tired, and feeling unwell). **Chapter 5** concludes with a personal experience story. A user explains how ESM supported her recovery from depression. Monitoring had a different meaning during consecutive recovery phases, first by providing insight in determinants of

vulnerability and later by confirming patterns of general recovery. This long case study shows that prolonged monitoring (> 2 years) is possible and meaningful.

Methodological work

ESM researchers often develop innovative parameters and data presentation scheme's to implement in regular clinical decision-making. **Chapter 6** explores the clinical utility of such an ESM summary statistic. Subjective momentary well-being is important for good mental health. Intuitively, quality of life improves by increasing the involvement in rewarding experiences (situations that feel good). Quality of life does not increase linearly (more is better), but can be optimized by balancing challenges in daily life. A reward-related quality of life statistic was operationalized by combining situational frequencies (relevant activities, locations and company) with concurrent positive affect. It generates a score (representing good to bad reward efficiency) for each ESM observation. Reliability of the statistic was tested with resampling methods, using various context definitions under different sampling circumstances, for real or virtual patients with low-, average- or high contextual variation. Overall, reliability was achieved after 100 or more observations were collected. Extra observations may be needed in different individual situations. Small sample sizes were only feasible in patients with limited contextual variation. Future studies should consider resampling procedures to assess the behavior of summary statistics under different sampling situations and across populations. This exploration is relevant for clinical practice, where often few observations are available.

Cognition in daily life

Cognitive problems are common in mental disorders and often linger even when other symptoms diminish. These lingering effects influence everyday life adaptation. We need our cognitive skills to function properly. Traditionally, cognition tests are administered at specific moments and in controlled environments. These assessments provide insight in state abilities, but lack information on cognitive variation in real-world environments. To compensate for this lack, brief cognition tasks are designed specifically for ESM. **Chapter 7** describes the initial feasibility of two cognition tasks in a sample of 10 patients with bipolar disorder. A symbol task measuring processing speed had good completion rates (82%), whereas a quiz task measuring liberal acceptance bias resulted in fewer completions (52%). The most sensitive cognition parameter on the symbol task – warranting further research – was processing speed (reaction time). In **Chapter 8**, the same task was adapted for smartphones and tested in a general population sample (*N*=40). The protocol included eight ESM questionnaires (beeps) with the symbol task, repeated for six days. Cognition outcomes

were the number of trials completed within a 30-seconds time interval (speed) and the percentage of correct trials (accuracy). The task was perceived as easy, pleasant and do-able. These subjective feelings were reflected in a high task accuracy (97%). Cognition varied over time and was related to mood, specifically feeling cheerful, irritated and anxious. Context mattered, with distraction and location influencing speed. Moreover, speed showed a learning effect (faster towards the end of the six-day period) and an age effect (slower with older age). The present study thus suggests that implementation of a short digital cognition task for repeated measurement in daily life is feasible. In **Chapter 9**, the same protocol was used in a new general population sample (N=49). An additional cognition task assessing visuospatial working memory was developed and added to the protocol. Results show positive experiences with both cognition tasks, with a sufficient completion rate (71%). Social context, age and feeling distracted influence cognitive performance in everyday life. The most sensitive outcome measure for cognition was speed on the symbol task. The memory task yielded few correct responses (being too difficult), resulting in little variation in scores. Thus, it requires further adjustments. More research is needed to establish validity and the usefulness of ESM-based brief objective cognition measures for patients in clinical practice.

New Horizons

Chapter 10 discusses the main findings, the lessons learned, and future implications. Assessing variability is key in understanding what is happening with relevant recovery domains on a daily basis. Resilience is related to the efficacy of daily life adaptation strategies. Standard instruments often lack insight into contextual variability and ESM can fill this gap. Due to innovations in computer technology, the method can be applied clinically at low cost. The method can transfer thus from a research instrument to a useful clinical instrument and be disseminated in mental health care. Patients monitor their individual landscapes of mood, behavior and environment and learn from it. The risk of misinterpreting the data remains. Therefore, users need to realize that information is partial. Researchers need to assess the behavior of ESM parameters and invest in new ways to guide data interpretation. The power of individual data lies in hypotheses generation and the opportunity to test them in daily life. In the future, ESM feedback possibilities will likely expand (e.g., by using artificial intelligence). In essence, however, its simplicity is its strength. The first steps towards a widespread clinical implementation are taken. Now, the path needs to be continued. Research is a good facilitator for this, next to hands-on experience in the field. Let us continue the path towards a bright new horizon for ESM in clinical practice.

Addendum

Samenvatting
Individuele landschappen

Dit proefschrift richt zich op het verkennen en uitbreiden van de Experience Sampling Method (ESM), een op wetenschap gebaseerde dagboekmethode, voor de geestelijke gezondheidszorg (GGZ). **Hoofdstuk 1** beschrijft de GGZ op dit moment, waar aandacht is voor positieve gezondheid en herstel. ESM als app biedt de mogelijkheid om behandeling naar het dagelijks leven te brengen op een manier die toegankelijk is en bekrachtigend werkt. Na deze introductie volgen drie delen. **Deel I** beschrijft de zorgrelevantie van ESM voor individuele patiënten. **Deel II** beschrijft de methodologische overwegingen die relevant zijn bij het beoordelen van databetrouwbaarheid. **Deel III** verkent de meerwaarde van het objectief meten van cognitie met ESM.

Klinische relevantie

Hoofdstuk 2 geeft een algemeen overzicht, waarbij de voor- en nadelen van ESM als meetinstrument voor onderzoek naar de effectiviteit van behandelingen worden beschreven. ESM is hiervoor geschikt, omdat het informatie verzamelt over uitkomsten in het dagelijks leven. Onderzoekers krijgen advies over hoe ze een ESM-studie kunnen opzetten, analyseren en toepassen. Dit kan bijvoorbeeld door ESM te gebruiken als monitoringsinstrument of als behandelinterventie, maar de methode is ook geschikt voor het uitvoeren van individuele behandelstudies. **Hoofdstuk 3** illustreert met het gebruik van verschillende onderzoeksresultaten hoe ESM werkt als reguliere app binnen de zorg. De voordelen voor individuele ondersteuning worden besproken. Naast een persoonlijke aanpak, kan de app zelfmonitoring bevorderen en zelfinzicht bieden, als de data toegankelijk zijn voor patiënten. Deze data kunnen helpen om weerbaarheid te ontdekken en verborgen patronen tussen stemming, gedrag en omgeving zichtbaar te maken. Daarnaast faciliteert ESM een gedeelde besluitvorming binnen een behandeling en zet het patiënten in hun kracht. Zorg waarbij traditionele hulpverlening wordt aangevuld met ESM als digitaal hulpmiddel kan behandeluitkomsten verbeteren en kosten besparen. **Hoofdstuk 4** presenteert een betrouwbaarheidsstudie in het ziekenhuis, waarbij ESM is getest als vast meetinstrument voor behandelingen in het ziekenhuis. De uitkomsten lieten zien dat de methode betrouwbaar is, met voldoende psychometrische kwaliteiten. ESM meet wat het hoort te meten en is in staat om subtiele veranderingen over tijd op te pikken die relevant zijn voor herstel (den aan positief affect, negatief affect, slaapkwaliteit, activiteit-gerelateerde stress, moeheid en niet goed voelen). De methode verlegt de focus naar het moment en informeert personen over hun dagelijks aanpassingsvermogen. **Hoofdstuk 5** beschrijft in een

ervaringsverhaal hoe ESM kan bijdragen aan het herstel van depressie. Het ervaringsverhaal laat zien dat langdurig monitoren (> 2 jaar) mogelijk en betekenisvol is. ESM biedt houvast gedurende verschillende herstelfases, waarbij de betekenis verandert over tijd. Dit kan variëren van een doel op de dag tot een algemeen overzicht van herstel.

Methodologisch werk

Hoofdstuk 6 verkent de bruikbaarheid van ESM-data voor het bijhouden van herstel. Subjectief welzijn is belangrijk voor een goede mentale gezondheid. Intuïtief gezien verbetert de kwaliteit van leven door meer tijd te spenderen aan belonende ervaringen (situaties die goed voelen). Een scenario waarbij alle ervaringen positief zijn is niet realistisch. Het doel is om een goede balans te vinden tussen fijne ervaringen en vervelende momenten. Om dit te meten, hebben we een op beloning gebaseerde welzijnsmaat geoperationaliseerd. Deze maat combineert hoe vaak situaties voorkomen (bestaande uit activiteiten, locaties en gezelschap) met de mate van bijbehorende positieve gevoelens. Per ESM-moment ontstaat er een positieve score (iemand is efficiënt bezig en spendeert de meeste tijd in situaties die goed voelen) of een negatieve score (iemand is minder efficiënt bezig). De betrouwbaarheid van deze maat is getest met resampling-procedures waarbij verschillende definities van situaties, onder verschillende steekproefomstandigheden zijn getest, zowel voor echte als virtuele patiënten met lage-, gemiddelde-, of hoge variatie in dagelijkse situaties. Gemiddeld 100 momenten waren nodig voor een betrouwbare meting, met verschillende aantallen tussen personen. Kleinere aantallen (bijv. 40 momenten) waren alleen haalbaar voor iemand met weinig situationele variatie op een dag (bijv. iemand die veel thuis is). Resampling mag vaker gebruikt worden om het gedrag van uitkomstmaten te onderzoeken. Deze verkenning is relevant voor de praktijk, omdat minder ESM-momenten vaak de norm zijn.

Cognitie in het dagelijks leven

Cognitieve problemen (aandacht, geheugen en concentratieklachten) komen veel voor binnen de GGZ. Klachten blijven vaak lang aanwezig, ook wanneer andere symptomen (bijvoorbeeld stemmingsklachten) afnemen. Cognitieve vaardigheden zijn nodig om goed te functioneren en beïnvloeden ons dagelijks leven. Om iemands stabiele capaciteiten te meten, worden traditionele cognitie testen afgenomen in gecontroleerde omgevingen en op specifieke momenten. Wat hierbij mist is informatie over hoe cognitie varieert over tijd en hoe het cognitief functioneren samenhangt met andere factoren. Om hier inzicht in te krijgen hebben we korte cognitieve taken ontwikkeld, speciaal voor ESM. Hoofdstuk 7 beschrijft een eerste haalbaarheidsstudie waarbij twee cognitietaken zijn getest bij 10 patiënten met een stemmingsstoornis, een symbooltaak en een quiztaak. De symbooltaak (voor

verwerkingssnelheid) werd goed ingevuld door deelnemers (82% van de taken), terwijl de quiztaak (voor een verhoogde aannamebias) minder goed werd ingevuld (52%). Het meest gevoelig voor verandering bleek de snelheid op de symbooltaak, deze rechtvaardigt verder onderzoek. In **Hoofdstuk 8** is de symbooltaak als app getest bij 40 gezonde personen. Deelnemers ontvingen gedurende zes dagen elke dag acht ESM-vragenlijsten met de symbooltaak. De uitkomstmaten voor cognitie waren snelheid en accuraatheid op de taak. De taak werd ervaren als gemakkelijk, plezierig en haalbaar en deelnemers waren in 97% van de gevallen correct (hoge accuraatheid). Resultaten laten zien dat cognitie varieert over tijd en samenhangt met stemming (opgewektheid, irritatie en angst). De context bleek belangrijk, waarbij afleiding en locatie invloed hadden op de snelheid. Ouderen waren gemiddeld langzamer dan jongeren en mensen werden sneller naarmate de testweek vorderde (leereffect). In **hoofdstuk 9** is hetzelfde protocol uitgevoerd, maar nu bij 49 gezonde personen en een extra geheugentaak (voor het meten van visueel ruimtelijk werkgeheugen). De ervaringen waren positief en de vragenlijsten zijn voldoende ingevuld (71%). Sociale omgeving, leeftijd en afleiding hadden invloed op cognitieve prestaties. Snelheid bij de symbooltaak blijkt een sensitieve uitkomstmaat voor cognitie, terwijl de geheugentaak te moeilijk is en aanpassingen vergt. Vervolgonderzoek is nodig om de validiteit vast te stellen en te onderzoeken of de taken geschikt en waardevol zijn binnen de zorg.

Nieuwe horizon

Hoofdstuk 10 gaat in op de hoofdbevindingen, de geleerde lessen en de toekomst van ESM in de praktijk. Inzicht in variatie op relevante herstelgebieden is essentieel om te begrijpen wat er dagelijks gebeurd. ESM als zorginstrument kan dit bieden. Patiënten monitoren hun individuele landschappen van stemming, gedrag en omgeving en leren ervan. Voorzichtigheid is hierbij geboden, omdat het risico bestaat dat er verkeerde conclusies worden getrokken. Gebruikers moeten zich realiseren dat de verzamelde informatie slechts deelinformatie is. Het is aan onderzoekers om het gedrag van parameters te evalueren en te investeren in een betrouwbare dataweergave. De kracht ligt bij het formuleren van hypothesen en de kans om deze in het dagelijks leven te toetsen. ESM-feedback zal zich in de toekomst uitbreiden, maar voor nu werkt eenvoud het beste. De eerste stappen richting de zorg zijn gemaakt, nu moet het pad verder gevolgd worden. Onderzoek faciliteert dit, naast praktijkervaring in het veld. Laten we blijven bewegen naar een heldere nieuwe horizon voor ESM binnen de GGZ.

Addendum

Knowledge valorization

The general aim of this thesis was to explore and expand the utility of the experience sampling method for use in mental health care. This means moving away from a singular research focus towards creating an additional value for the public domain.

Societal relevance

Psychological problems are common and societal costs to deal with these problems are high. Research shows that every year, a quarter of the Dutch population is in need of mental health support[1]. This is a problem, because only 7.3 percent of the Dutch population actually receive professional mental healthcare[2]. The finances for expanding the mental health sector are lacking. So not everyone will receive accurate treatment[1]. A continuous effort is required to improve the effectiveness of available care, reduce the costs involved, and prioritize care better. What could help is to provide tools that allow people to work on their mental health without the presence of a professional.

In light of these efforts, a transition is ongoing within the mental health care system in the Netherlands. The focus shifts from illness to positive health[3]. This is important, because a strong focus on illness, both in somatic care as well as mental health care, has iatrogenic effects and keeps patients ill. For sustainable well-being and the prevention of suffering, it is crucial to build resilience and increase autonomy[1]. The professional health care field has neglected these aspects in the past, but is now adopting a more holistic care approach. This includes paying attention to someone's social network, his or her environment, work situation and general health. Resilience, in this respect, is very much a concept of everyday life. However, the professional field still misses the tools to assess these aspects.

The experience sampling method (ESM) facilitates monitoring in a holistic way; it can assess vulnerability as well as resilience, and fits within the positive health concept. Since it works on smartphones, it is accessible during the day and to the majority of people. By providing an evidence-based mobile application in the App store or Google Play store and by making sure the collected data are secure and freely accessible to users, a costs-effective alternative is available to complement treatment. The added benefit is that patients become empowered, because ESM allows them to gather information and actively test personal hypotheses outside treatment hours. Building resilience should be possible without using professional resources. People who are unwell and want to explore what helps them to feel better are also able to use ESM. Thereby providing the chance to help people early on and prevent suffering.

Innovation

The PsyMate™ as an ESM tool was used throughout this thesis. One of the benefits of this app is that it is accessible in different languages and freely available to the public. The studies in this thesis contributed to the development of this app. The knowledge gained by the naturalistic single-case trials have helped design a basic set of ESM questions that are universal across complaints and relevant during recovery. Up to five additional personal questions can be added to the free questionnaire, allowing the monitoring of specific wishes during recovery (e.g., I feel suspicious). Currently, efforts are made to provide intuitive guidelines, so that people are not dependent on an elaborate information session before they use the method. The app and the web-based reporting module (to view ESM results) could benefit from information screens that explain the method. The positive results presented in this thesis warrant a continuous investment in ESM for clinical practice. This means working towards increased user-friendliness (e.g., making the app available for people who have difficulty reading), improving the reliability of ESM feedback, and investigating the validity of ESM-based cognition tasks. Furthermore, insight is needed into the facilitators and barriers while implementing ESM in clinical practice. Colleagues at the department of Psychiatry and Neuropsychology in Maastricht are currently working on the above-mentioned projects.

Target audience

The finding described in this thesis are relevant for people suffering from mental health complaints, health care providers, researchers, health insurance companies and policy makers.

Everyone, at some point in their life, experiences mental health or somatic complaints. It could help to explore when and why complaints occur. Often, symptomatic variability is influenced by context. Knowing when pain occurs, or equally so, when someone is able to relax and become unburdened, is valuable in treatment and care. Knowing what happens is the first step towards exploring alternatives towards feeling better. ESM can provide insight into mood and context for people with somatic complaints, psychological complaints, or for anyone who is interested to increase their general well-being. The audience includes youth, elderly people, and people who suffer from severe health problems or mental illness. In the future, ESM will likely be suited to provide insight into daily cognitive functioning. This is especially relevant for people suffering from neurological conditions, such as dementia or mild cognitive impairment.

Health care professionals benefit from viewing ESM data together with their patient. It can help move away from assumptions based on group findings and center the therapy session on the individual situation. Patients sometimes struggle to express themselves. Looking at their momentary mood ratings over the past week, might open up the conversation and provide insight for both parties. Therapy then becomes a collaborative practice were both parties can equally engage. Current practices often require the use of end-of-the-day diary methods that suffer from memory bias. ESM could replace them and provide additional fine-grained information on situational processes. A well-balanced ESM questionnaire focuses on several relevant health aspects, helping clinicians to focus beyond symptomatic recovery. Important is that the user is also the owner of the data. They can choose to share the data with their care provider, not the other way around.

Our findings could be of interest to insurance companies and policy makers, given that ESM as an mHealth tool has proven merits in routine clinical practice. Next to being a useful and more accurate assessment instrument during treatment, ESM is also a relevant prevention strategy. Supporting people to increase their resilience is a useful strategy to prevent mental suffering in the future.

Implementation

To disseminate our findings among health care professionals, we give presentations and workshops on how to use ESM as a clinical tool and work with ESM feedback. Currently, a workshop is provided for psychological assistants at the general practitioners office. They are enrolled in a pilot were ESM is adjusted for and implemented in their current care practice. The clinical value of the method for both clinicians as well as patients is evaluated and adjustments are made accordingly. Most presentations are held at health care institutes in the South of the Netherlands, but efforts are made to reach interested parties across the country too. In a recent presentation at the Psychose café in Roermond, patients with a psychosis background, family members, and interested people were informed on how to use the PsyMate™ during recovery. These outreaching activities are very informative and help to guide research towards designing a practical, valuable and useful tool for the public. Future projects should continue along these lines and make sure the method is implemented in health care and freely available to the public. Ultimately, the clinician becomes an additional coach in the recovery process where the patient remains fundamentally in charge.

References

1. Delespaul P, Milo M, Schalken F, Boevink W, Van Os J. *Goede GGZ!: Nieuwe concepten, aangepaste taal en betere organisatie.* Springer; 2017.
2. Van Hoof F, Knispel A, Hulschbosch L, et al. Landelijke monitor ambulantisering en hervorming langdurige GGZ. *Utrecht: Trimbos Instituut.* 2016.
3. Huber M, Knottnerus JA, Green L, et al. How should we define health? *Bmj.* 2011;343:d4163.

Addendum

About the author

Simone Verhagen was born on 13 September 1991 in Heerlen, the Netherlands. After graduating from secondary school (Gymnasium) at College Rolduc Kerkrade in 2010, Simone moved to Maastricht to study Psychology at Maastricht University. During her study, she enrolled in the MaRBLe program, were she studied treatment integrity to group schema-focused therapy for borderline personality disorder. At the program, she rated Dutch, German and Australian therapy sessions on protocol adherence using a scale she co-authored. Additionally, she followed the minor Mental Health and graduated in 2013. Subsequently, Simone was enrolled in the one-year Master's program Neuropsychology. She graduated in 2014 and was then enrolled in the Master's program Mental Health, were she obtained the certificate Psychodiagnostic assessment while working as a clinical intern.

Immediately after, Simone started working as a research assistant at the department of Psychiatry and Neuropsychology of Maastricht University. In 2016, this work transgressed into a PhD project. During this trajectory, she was highly involved in several projects that aimed to bridge the gap between research and clinical use of the Experience Sampling Method and broaden its scope for mental health care. During her PhD, Simone spent four months at the Center for Self-Report Science, University of Southern California, Los Angeles. She is currently working as a psychologist at Mondriaan, next to working as a postdoc research fellow in the same department of Maastricht University.

Addendum

Publications and presentations

International publications in peer-reviewed journals

2019

Verhagen, S. J. W. / Daniëls, N. E. M., Bartels, S. L., Tans, S., Borkelmans, K. W. H., de Vugt, M. E., & Delespaul, Ph. A. E. G. (2019). Measuring within-day cognitive performance using the experience sampling method. – A pilot study in a healthy population. *PloS ONE.*

Lenssen, J., & **Verhagen**, S. J. W. (2019). Monitoring my journey from doctor, to patient, to doctor with lived experience. *Schizophrenia bulletin*, sbz097.

Michielse, S., Lange, I., Bakker, J., Goossens, L., **Verhagen**, S., Papalini, S., ... & van Os, J. (2019). Reward anticipation in individuals with subclinical psychotic experiences: A functional MRI approach. *European Neuropsychopharmacology*.

Michielse, S., Lange, I., Bakker, J., Goossens, L., **Verhagen**, S., Wichers, M., ... & Marcelis, M. (2019). White matter microstructure and network-connectivity in emerging adults with subclinical psychotic experiences. *Brain imaging and behavior*, 1-13.

Stamate, D., Katrinecz, A., Stahl, D., **Verhagen**, S. J. W., Delespaul, P. A. E. G., van Os, J., & Guloksuz, S. (2019). Identifying psychosis spectrum disorder from experience sampling data using machine learning approaches. *Schizophrenia research*, *209*, 156-163

2018

Bastick, E., Bot, S., **Verhagen**, S. J. W., Zarbock, G., Farrell, J., Brand-de Wilde, O., ... & Lee, C. W. (2018). The development and psychometric evaluation of the group schema therapy rating scale – Revised. *Behavioural and cognitive psychotherapy*, *46*(5), 601-618.

2017

Verhagen, S. J. W., Berben, J. A., Leue, C., Marsman, A., Delespaul, P. A. E. G., van Os, J., & Lousberg, R. (2017). Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. *PloS ONE*, *12*(10), e0186294.

Verhagen, S. J. W., Simons, C. J. P., van Zelst, C., & Delespaul, P. A. E. G. (2017). Constructing a reward-related quality of life statistic in daily life - a proof of concept study using positive affect. *Frontiers in psychology*, *8*, 1917.

van Os, J., **Verhagen**, S., Marsman, A., Peeters, F., Bak, M., Marcelis, M., ... & Simons, C. (2017). The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and anxiety*, *34*(6), 481-493.

Lange, I., Goossens, L., Michielse, S., Bakker, J., Lissek, S., Papalini, S., ... & Lieverse, R. (2017). Behavioral pattern separation and its link to the neural mechanisms of fear generalization. *Social cognitive and affective neuroscience*, *12*(11), 1720-1729.

2016
Verhagen, S. J. W., Hasmi, L., Drukker, M., van Os, J., & Delespaul, P. A. E. G. (2016). Use of the experience sampling method in the context of clinical trials. *Evidence-based mental health*, *19*(3), 86-89.

2014
Zarbock, G., Farrell, J. M., Schikowski, A., Heimann, A., Shaw, I., Reiss, N., ... & Bastick, E. (2014). Group Schema Therapy Rating Scale–Revised (GSTRS-R). *Available online at: http://www. schematherapysociety.org/Resources/Documents/GSTRS-R-fin-2014-07-07.pdf*.

Submitted and in progress articles

Verhagen, S. J. W., Peeters, F. P. M. L., Arts, B. M. G., Lataster, T., van Os, J., Delespaul, P. A. E. G. (2019; this thesis). Investigating mood and cognition in the daily lives of patients with a diagnosis of bipolar disorder – a pilot study.

Daniëls, N. E. M. / Bartels, S. L., **Verhagen**, S. J.W., van Knippenberg, R. J. M., de Vugt, M. E., & Delespaul, Ph, A. E. G. (pending revisions). Digital assessment of working memory and processing speed in everyday life: feasibility, validation, and lessons-learned.

Daniëls, D., **Verhagen**, S. J. W., Delespaul, P. A. E. G. (accepted). mHealth binnen herstelnetwerken. In N. Mulders, P. Delespaul, J. van Weeghel et al. (Red.), *Handboek Netwerk Psychiatrie*. Utrecht, the Netherlands: De Tijdstroom.

Scientific communications related to this thesis

2019

Oral presentation at the fall conference of the Dutch association for Cognitive Behavioral Therapy (VGCt) (Veldhoven, the Netherlands)

Oral presentation at the 6th annual Conference of the Society of Ambulatory Assessment (Syracuse, USA)

Oral presentation at the 13th European Network for Mental Health Service Evaluation (ENMESH) conference (Lisbon, Portugal)

Oral Presentation at the 47th spring conference of the Dutch Association of Psychiatry (NVvP) (Maastricht, the Netherlands)

Oral presentation at the Institute of Psychology seminar series, UCLouvain (Louvain-La-Neuve, Belgium)

Oral presentation at the Dutch brain awareness week, MUMC+ (Maastricht, the Netherlands)

2018

Oral and poster presentation at the University of Southern California, Mobile Connected Health symposium (Los Angeles, USA)

2017

Workshop Phrenos Psychose congress 'levensloop' (Zwolle, the Netherlands)

Oral presentation at the Phrenos Masterclass 'RCT ja of nee' (Utrecht, the Netherlands)

Oral presentation at the fall conference of the Dutch association for Cognitive Behavioral Therapy (Veldhoven, the Netherlands)

Poster presentation at the 12th European Network for Mental Health Service Evaluation (ENMESH) conference (Groningen, the Netherlands)

Oral presentation the 5th annual Conference of the Society of Ambulatory Assessment (Luxembourg city, Luxembourg)

Oral Presentation at the 45th spring conference of the Dutch Association of Psychiatry (NVvP) (Maastricht, the Netherlands)

2016

Oral presentation at the Phrenos Masterclass 'the science to service gap' (Utrecht, the Netherlands)

Oral presentation at the Better data event (London, United Kingdom)

Dankwoord / Acknowledgements

Het pad is bewandeld en de reis is ten einde. Het was een ongelofelijke tijd. Vol bijzondere momenten, leerzame momenten, soms stressvolle momenten, maar vooral ook een tijd met veel plezier! Ik wil dan ook **iedereen** die betrokken is geweest in de afgelopen jaren van harte bedanken. Dit proefschrift was er nooit geweest zonder de bijdrage en steun van zoveel mensen om mij heen. Lieve **vrienden, familie en collega's**, jullie hebben me de energie en inspiratie gegeven om af te ronden en nu mag ik zeggen: het is klaar!

Een aantal mensen wil ik in het bijzonder bedanken, te beginnen bij mijn promotieteam. **Philippe**, vanuit mijn stages ben ik bij jou terecht gekomen. Wat een geluk! Hoewel ik moest leren om soms aan te geven dat ik het niet snap (mijn woordenschat is inmiddels uitgebreid) en een interventie nodig had in nee zeggen, heb ik enorm veel kunnen leren en mogen groeien als onderzoekster. Er was ruimte voor ontwikkeling en bijscholing, maar vooral ook om te sparren over het leven (al dan niet binnen de GGZ). Onze meetings zijn voor mij zeer waardevol, een dikke merci! **Jim**, toen ik nog werkte als onderzoeksassistent kwam je me vragen of ik het contract kon tekenen. "Welk contract vroeg ik? Waarop jij antwoorde: van jouw PhD, je begint in maart. Had ik dat nog niet gezegd??" Altijd vol inspiratie en altijd gereed om een positieve boost te geven. Het was enorm fijn om met je te werken en ik hoop dat onze paden zich in de toekomst nog vaak mogen kruisen. **Claudia**, bedankt voor al je steun, feedback en gezelligheid! Ook van jou heb ik veel geleerd. Ik hoop dat we nog vaak samen op congres mogen (en naar huis kunnen met o.a. extra zuurstof, servies en een collectie drinkflessen). Inmiddels is er minder presentatie ondersteuning nodig, dus als we een hotelkamer delen dan is er beslist meer tijd voor ontspanning. **Team**, jullie hebben mij begeleid en ervoor gezorgd dat ik op tijd én met een goed gevoel heb kunnen afronden.

Ook wil ik graag alle leden van de **beoordelingscommissie** bedanken voor het beoordelen van mijn proefschrift en de **co-auteurs** voor de prettige samenwerking en bijdrage aan de artikelen. In het bijzonder wil ik **Thérèse** bedanken voor het opnemen van de voorzittersrol.

Onderzoek doe je nooit alleen. Op de **afdeling Psychiatrie en Neuropsychologie** werken (of werkten) geweldige mensen die ondersteunen, begeleiden en motiveren. Aan iedereen ook heel erg en van harte bedankt! Het liefst noem ik iedereen bij naam, maar gezien de hoeveelheid tekst ga ik me een beetje inhouden. Er is op de afdeling een ijzersterk **onderzoeksassistenten-** en **coördinatorenteam**! Ik heb jullie advies, hulp, gesprekjes en gezelligheid altijd erg kunnen waarderen. **Truda** en **Nele V**, naast werkbegeleiding en projectondersteuning, heb ik ook persoonlijk veel steun bij jullie mogen vinden. **Karel**, ik geloof dat het mailverkeer tussen ons niet te tellen is. Naast een geweldige app-assistentie,

heb je ook menig afbeelding in dit proefschrift verzorgd (thanks!). Dit proefschrift was sowieso nergens geweest zonder de steun van het **PsyMate ontwikkelteam**, allen dank! Ondersteuning bestaat ook uit data invoer, hulp met praktische (onderwijs)zaken en agendavoering. Bedankt onder meer aan **Ine**, **Jolanda**, **Trees**, **Dafne**, **Inge** en **Dominique**, maar ook aan **Ron** en **Jo** voor alle digitale ondersteuning. Met methodologische of statistische vragen kon ik altijd goed terecht bij **Marjan**, **Wolfgang** of **Richel**. Bedankt voor jullie geduld en advies. **Catherine**, jouw begeleiding tijdens mijn stage heeft mij het vertrouwen gegeven voor het uitvoeren van onderzoek en uiteindelijk het PhD traject. **Maarten**, fijn dat ik jouw reisschema's mocht uitwerken ;) Bedankt voor alle sparmomenten, ik kijk uit naar onze samenwerking! **Nienke**, bedankt om mij te betrekken bij het toffe DaLiBi project. **Frenk**, bedankt voor de inspiratie en congres-gezelligheid. To the **English-speaking colleagues**, thanks for making the office such a nice environment and for all the fun conversations. **Sinan**, thanks for the additional ESM inspiration, I enjoyed working on the machine learning paper!

De **voormalige PhD's**: Annelie, Yori, Hennie, Thomas, Silvia, Zuzana, Esther, Claudia, Steven, Iris, Stijn, Jindra: het was superfijn om met jullie te kunnen sparren en ontspannen onder het genot van pizza, Weerwolven-spelletjes en vers gemaakte cappuccino's! **PhD'ers van nu**: Kim, Maarten, Laila, Lotta, Mary Rose, Naomi, Christian, Sophie, Boris, Ozan, Stella, Samantha, Nikita, Maud, Emma, Anne, Coline, Suzanne, and Evita: thanks for making the department such a nice place! I enjoy chatting with you, discussing about our research and going for a walk to stretch our legs!

Naast mijn eigen onderzoek, heb ik mogen werken aan een aantal andere projecten waaronder de **Geestkrant** (of Hoofdzaken) redactie. Het blijft leuk om met jullie te mogen brainstormen over welke interessante onderzoeken we kunnen uitlichten. Ook mocht ik werken als **PhD representative**; zonder dit had ik een aantal mensen van de andere divisies niet leren kennen! Mede (ex)**PhD-reps**, het was echt leuk om met jullie PhD evenementen te organiseren. Het is altijd fijn om met **collega's van buitenaf** te kunnen kletsen en sparren. Bedankt allemaal :) **Lotte** en **Niels**, jullie proefschriftadvies was erg welkom! **Christine**, de koffiemomentjes hebben goed gedaan. **Sara**, it was such a pleasure to work with you on developing the cognition tasks! I am happy we have met and hope for more opportunities to work together (or hang out) in the future. Gerelateerd aan cognitie: **Rudolf** en **Bert**, hartelijk dank voor het delen van jullie visie en bereidheid om mee te denken. **Klara**, bedankt voor alle DGT-inspiratie, een supertof project!

To the **ESM-expert group**: Thanks so much for all the ESM enthusiasm, inspiration and advice in these last years! The work in this thesis has benefited from your expertise and I am happy that I can remain a member for a while longer.

In het afgelopen jaar heb ik mogen starten als psycholoog bij de Nieuwe GGZ in Hoensbroek. Lieve **Mondriaan collega's**, het is superleuk om met jullie samen te werken. Binnen het team kan ik groeien en de opgedane ervaringen zijn daarnaast een fijne inspiratie voor toekomstig onderzoek. Een combinatie van beide werkvelden voelt als een verrijking!

I got the chance to travel to Los Angeles and work with some great researchers at the **Center for Self-Report Science** at USC. Dear **Arthur**, thank you for making this visit possible. You and your **team** have challenged me to take a different perspective and inspired me to be critical when examining EMA data. **Stefan**, it was such fun to discuss statistical possibilities to model changes! **Bob**, thanks for connecting research and magic and providing me with new perspectives :) **Marta** and **Filip**, I am so happy that I met you both! Taking me to the beach right after my arrival made all the difference in the world. Your enthusiasm to explore is contagious and without your guidance, I would have missed such cool things!

In de afgelopen periode heb ik meerdere malen bij **mijn netwerk** aangeklopt voor hulp met het testen van de PsyMate. Soms wel tot twee keer toe omdat iets niet meteen wilde werken. Maar de resultaten zijn wel terug te vinden in dit proefschrift. Nogmaals enorm bedankt voor alle bereidheid en jullie deelname! Hiernaast wil ik ook alle **stagiaires** bedanken die zich hebben ingezet voor het PsyMate onderzoek!

Jaimie, **Christoffel**, **Michelle**, jullie wil ik in het bijzonder bedanken. Ik heb zo enorm veel geleerd van het samen zoeken naar een betekenisvolle en passende monitoring, het leren van de rapportagepagina en de verzamelde gegevens. Mijn respect voor jullie is groot en het was een eer om de weg naar herstel met jullie te mogen ontdekken.

Enorm veel liefde en dank voor mijn kamergenootjes, met wie ik stressvolle momenten, onduidelijkheden maar ook successen heb mogen delen en wie mij enorm veel energie en plezier in werk hebben gegeven. **Suzanne**, nog nooit heb ik zo'n enthousiast persoon ontmoet als jij. Een PhD is fantastisch, toch?! Onze kletsmomenten bleven wel beperkt tot koffiepauzes (je hebt mijn koffie-verslaving mede op gang geholpen), want er werd natuurlijk hard gewerkt. Al werden die hyper-focus momenten wel af en toe onderbroken door wat rek- en strekoefeningen op kantoor, want we mogen niet vastroesten ;). **Naomi**, wat was het tof dat er nog een PhD student begon, met ESM ging werken en ook nog eens bij mij op kantoor terecht kwam. Enthousiast nam jij het gezamenlijk afwas-ritueel van Suzanne en mij

over en konden we goed sparren over allerlei ESM en niet ESM-gerelateerde onderwerpen. Erg bedankt voor jouw opvang tijdens stressmomenten. Fijn dat we elkaar leren om af en toe aan paarden te denken en ons perfectionisme los te laten. **Anne**, ook jou wil ik bedanken als kamergenote, voor de fijne gesprekken en de toffe versiering op kantoor (*waar hoop is, is perspectief* en *doe eens even niet normaal!*). Ik heb veel respect voor alles wat je doet om de wereld net wat vriendelijker te maken en al dat stigma rondom psychische klachten te doorbreken.

Vrienden zijn ook hier onmisbaar. Zonder alle gezellige, niet werk gerelateerde feestmomenten was de balans flink zoek geweest! De spellenavonden (of treasure hunts), etentjes, bakmomentjes, filmavondjes, festivals, theatershows, schaats- en pretparkfestijnen, carnavalsfeestjes en diepgaande gesprekken (al dan niet onder het genot van bier) worden erg gewaardeerd! Laten we snel wat nieuws plannen, nu is er tijd ☺.

Alexandra, een betere schoonmama kan ik mij niet wensen. Bedankt voor alle steun en bereidheid om te helpen. Een gezellige wandeling of lunch, het poetsen van ons huisje, mee-testen (plus alle collega's zover krijgen), of door gewoon trots te zijn. **Marthè**, ik weet zeker dat jij ook trots bent en op afstand het proces hebt gevolgd. **Rebecca**, al een lange tijd zusjes, maar binnenkort officieel (oeh yeah!). Bedankt voor al je steun en gezelligheid. Snel weer wat concertjes plannen! **Luc**, ook jij hebt me goed gesteund. Samen verliezen met Knowledge is Power is toch veel gezelliger. Tegen die tweeling kunnen wij natuurlijk niet op!

Pap, **Mam**, **Dennis**, **Michelle**. Zonder jullie steun had ik dit helemaal niet gekund! Altijd trots, ook als het tegenzit. **Pap**, ik heb toch een aantal karaktertrekken overgenomen geloof ik (iets met werkmieren en mestkevers). **Mam**, weet je nog toen we beide slechts nieuws kregen? Moet je zien waar we nu staan! Jullie hebben me de kansen gegeven, waarvoor ik heel dankbaar ben. **Dennis**, als ik professor word, dan word jij luitenant! Bedankt voor je onvoorwaardelijke vertrouwen. **Michelle**, het is supergezellig met jou in de familie erbij. Bedankt voor alle hulp met dit proefschrift!

Het mooiste plaatje hier is toch wel het familiewapen (thanks pap), nu officieel in print.

D'r letste is d'r betste. **Joshua**, my neutron star collision. Jij bent degene die alle dalen en pieken heeft meegemaakt, van A tot Z! Ook voor jou is het een hele reis geweest. Toen ik twijfelde of ik wel zo lang weg kon gaan, was jij diegene die zei dat ik het moest doen. Jouw steun is onvoorwaardelijk. Ik kan niet wachten om ons avontuur voort te zetten en deze zomer te trouwen!