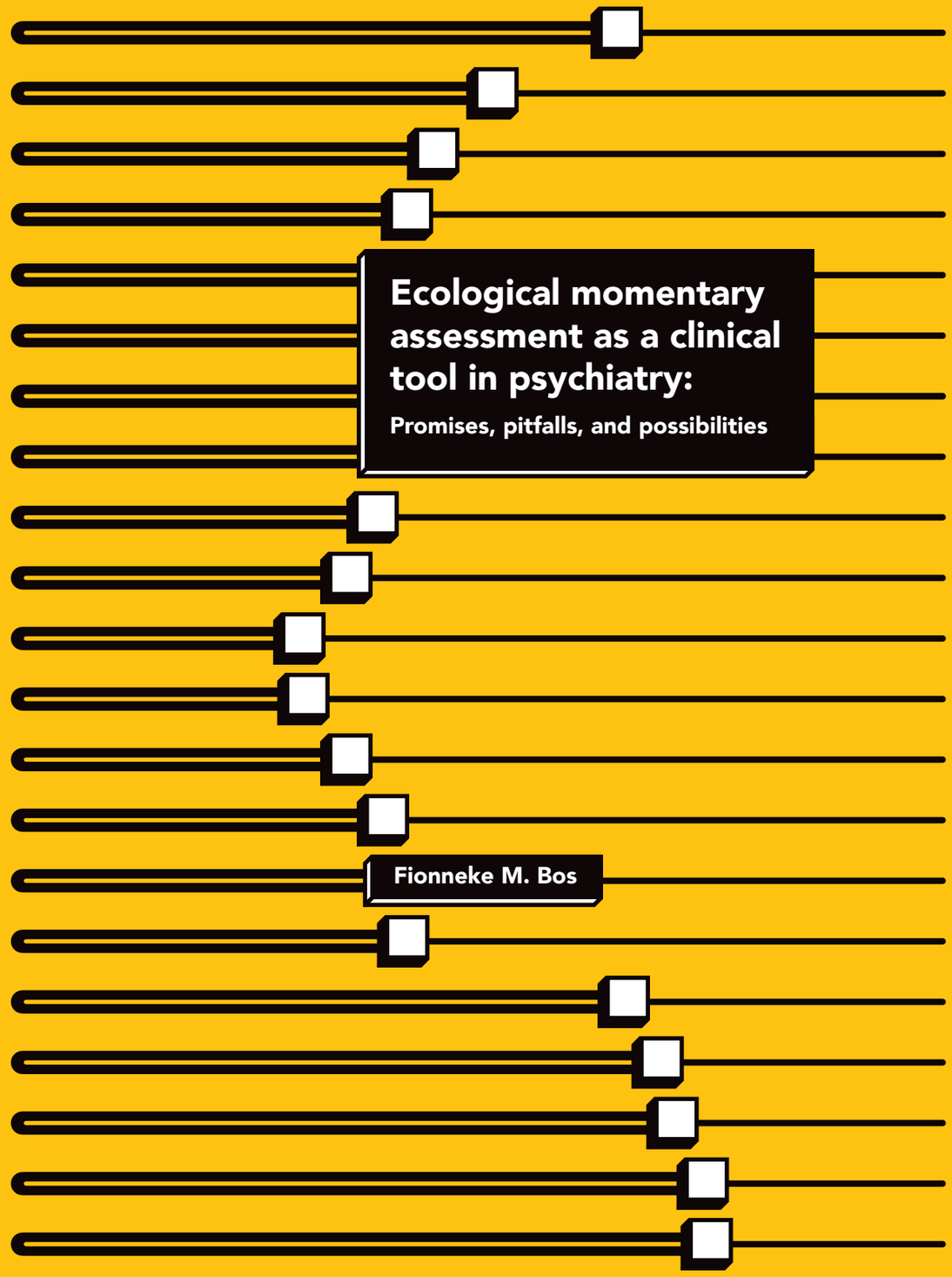
Ecological momentary assessment as a clinical tool in psychiatry: Promises, pitfalls, and possibilities

Fionneke M. Bos



Ecological momentary assessment as a clinical tool in psychiatry:

Promises, pitfalls, and possibilities

Fionneke M. Bos

Ecological momentary assessment as a clinical tool in psychiatry

Promises, pitfalls, and possibilities

Fionneke M. Bos

Cover design | Rik Speel
Layout | Wendy Bour
Printing | Ipskamp

This thesis was financially supported by the Rob Giel Research Center (RGOc).

rijksuniversiteit
groningen

Ecological momentary assessment as a clinical tool in psychiatry: Promises, pitfalls, and possibilities

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 8 september 2021 om 14.30 uur

door

Fiona Marijke Bos

geboren op 2 mei 1992
te Groningen

Promotores

Prof. dr. M.C. Wichers
Prof. dr. R. Bruggeman

Copromotores

Dr. J.A.J. van der Krieke
Dr. E. Snippe

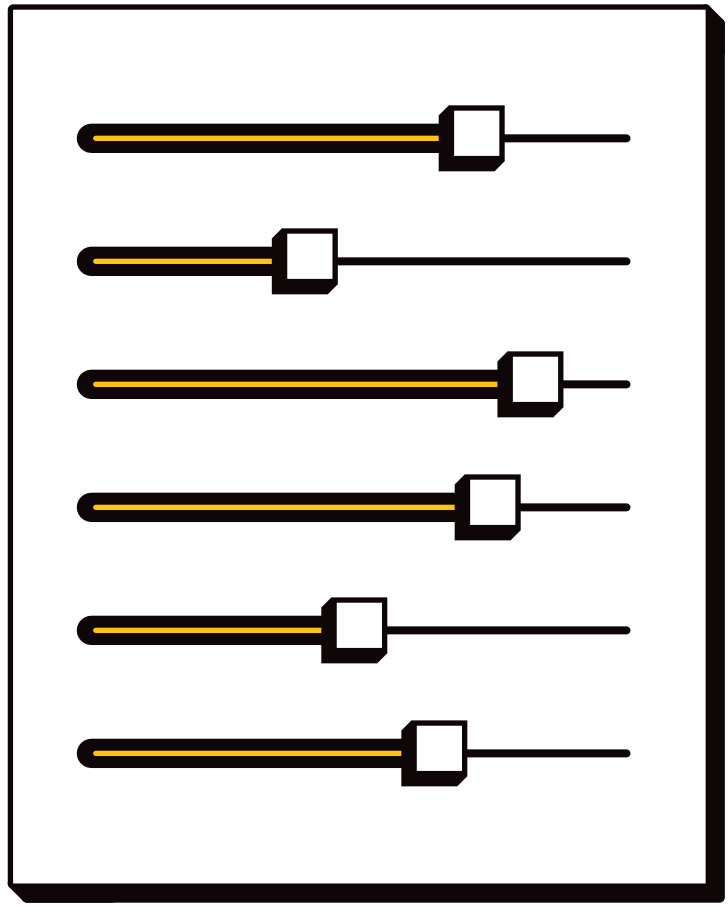
Beoordelingscommissie

Prof. dr. J.G.M. Rosmalen
Prof. dr. I.Y.R. Myin-Germeys
Prof. dr. P.A.E.G. Delespaul

Stellingen

1. Het grootste voordeel van dagboekmetingen is dat ze patiënten meer grip geven op hun welzijn.
2. Dagboekmetingen zijn vanwege de belasting en focus op klachten niet voor iedere patiënt en zorgfase geschikt.
3. Dagboekmetingen hebben het meeste nut voor de psychiatrische zorg als ze gepersonaliseerd kunnen worden.
4. Om bruikbaar te zijn in de praktijk moeten er soms op wetenschappelijk gebied concessies worden gedaan rondom dagboekmetingen.
5. Gepersonaliseerde netwerken zijn nog onvoldoende onderbouwd voor gebruik in de praktijk.
6. De belofte dat complexe analyses, zoals netwerkanalyse en *early warning signals*, patiënten en behandelaren nieuwe inzichten opleveren, is vooralsnog niet waargemaakt.
7. Patiënten en behandelaren moeten veel meer en beter betrokken worden bij wetenschappelijk onderzoek en de ontwikkeling van e-health tools.
8. *Team work makes the science work*: het mooiste wetenschappelijke onderzoek ontstaat in samenwerking.
9. Mentale gezondheid is nog te vaak een ondergeschoven kindje in de samenleving.
10. *'t Is nait aal doage kovvie mit kouke, moar vandoag wel ja!*

Chapter overview



Chapter 1

General introduction

Preface

Approximately 30% of people suffer from a mental disorder at some point in their lives[1,2]. The experience of psychopathology is one of the leading causes of disability in the Western world and is associated with tremendous burden for those suffering from it, their family-members, and society[3,4]. Effective treatment is therefore crucial. Unfortunately, even after decades of academic progress, we still have limited insight in the mechanisms through which mental disorders develop and maintain over time[5]. Since most clinical research has so far focused on group-level averages, most findings have limited utility for individual patients[6,7]. This is because the experience of psychopathology is highly heterogeneous[8], meaning that two persons with the same diagnosis can starkly differ in their experience of symptoms. As such, the field is in need of new methodologies that appropriately address this heterogeneity and are able to map, for each person individually, the mechanisms through which symptoms are developed and maintained. Such knowledge will help us to develop more targeted and effective interventions for people with mental disorders.

One such a novel approach has quickly gained popularity in the recent years: zooming in on the moment-to-moment experiences in daily life. This approach postulates that we should abandon our focus on diagnostic categories and rather focus on the investigation of the smallest building blocks underlying psychopathology in daily life[9,10]. This micro-level approach to understanding psychopathology can be studied using ecological momentary assessment (EMA)[11], also known as the experience sampling method (ESM)[12], which are methods with which participants record their life 'as it is lived'. Although EMA started out as a research methodology, its clinical utility was suggested early on by researchers and scientist-practitioners[13]. The promise of EMA as a clinical tool is two-fold. First, self-monitoring through EMA has been suggested to improve patient self-management and empowerment. Second, statistical analysis of the gathered EMA data could offer patients and their clinicians new insights into the psychopathological mechanisms of this specific individual. The clinical potential of EMA is thus closely tied to the development of personalized statistical models that use EMA data to attempt to illuminate individual mechanisms of psychopathology. Two such analytical frameworks thus far have

shown promise in that regard: the network theory of psychopathology, which investigates the relationships between symptoms and momentary states[9,14], and complex dynamical systems theory, early warning signals in particular, which could help foresee imminent transitions in psychopathology using EMA data.

The present thesis has explored the utility of EMA as a clinical application. In this general introduction, I will first outline the basic characteristics of EMA as a self-monitoring method, after which I will turn to exploring the possibilities for analyzing EMA data using network theory and complex dynamical systems theory. Next, I will introduce the promise of EMA as a clinical tool. Finally, I will describe the outline of this thesis.

Ecological momentary assessment (EMA)

The term ESM was first coined in 1977 by Csikszentmihalyi, Larson, and Prescott, who introduced ESM as a research method to study daily life[12], and was later reintroduced as EMA by Shiffman, Stone, and Hufford[11]. Although both ESM and EMA generally refer to the same methodology, there are some historical differences: ESM has traditionally focused more on affective states and associated contextual information, whereas EMA has concentrated on actual behavior, also including physiological measurements such as heart rate[15]. In addition, EMA is more often used in psychological fields, whereas ESM is mostly used in medicine[15]. However, since most researchers now use ESM and EMA interchangeably, henceforth I will refer to both methodologies as EMA.

With EMA, participants complete assessments on their symptoms, affect, thoughts, activities, and (social) contexts, several times a day for a given period[16] (see Figure 1.1). Typically, EMA diary items refer to momentary experiences and are answered on a visual analogue scale (VAS) ranging from 0 to 100 (“not at all” to “very much”) or on a 7-point Likert scale. In psychiatry, such momentary experiences often pertain to symptoms, such as feeling sad, anxious, agitated, or the hearing of voices. Although older studies had to make due with paper-pencil forms or pagers to administer the EMA assessments, the widespread adoption of smartphones by the general population has enabled easy access to smartphone-based EMA[17].

Figure 1.1 A visual representation of an EMA diary and possible feedback based on EMA data.

As in any assessment method, the specifications of the EMA diary depend on the research question and the burden that is placed upon the participants completing the diary. In addition to the diary items, important specifications include the number of assessments per day, the duration of the EMA period, and the timing of assessments. Most EMA studies thus far utilized 3-10 assessments per day for 1-2 weeks[18,19]. Assessments can occur either at fixed time points (e.g., at 12:00h) or at random points in predefined intervals (e.g., somewhere between 10:30-12:00h), termed fixed or semi-random EMA designs, respectively. Whereas fixed designs are believed to be less burdensome for participants, semi-random designs supposedly provide a more representative overview of daily life experiences[16].

EMA has four important advantages when compared to more traditional retrospective questionnaires. First, EMA assesses experiences in the present moment rather than retrospectively, and pertains to diverse micro-level experiences

rather than a syndrome as a whole. EMA is therefore considered to be less subject to memory biases and more ecologically valid than traditional retrospective questionnaires[20]. EMA should thus yield more reliable data[21,22] and should be easier to complete for participants. Second, EMA enables the study of the impact of context (activities, social interactions, events) on affect and symptoms, which likely plays an important role in the development and maintenance of mental disorders. In traditional questionnaires, context is often overlooked. Third, as it is increasingly recognized that psychopathology evolves dynamically over time, the frequent and repeated nature of EMA is ideally suited to study how mental disorders develop and are maintained over time in daily life[9,23].

A final important advantage is the possibility of a more idiographic approach to psychiatric research[24]. EMA studies result in a rich dataset of experiences specific to the individual, enabling the study of person-specific processes relevant to psychopathology. Until recently, most psychiatric research involved nomothetic studies, in which groups of individuals were compared against one another (for example, people with depression versus healthy individuals). This approach has received criticism because such findings often do not generalize to individual patients: there is no 'average' individual, treatments that may work for some patients often do not work for others[25]. Likewise, studies have shown that the expression of psychopathology is highly heterogeneous[8,26]. Both research and practice are thus in need of more personalized models that explain psychopathological mechanisms in individual patients[27]. EMA can meet that demand.

Because of these advantages, EMA has become increasingly popular, both as a research methodology and as a clinical self-monitoring method. In research, EMA has already been applied to improve our understanding of symptoms, affective variability, contextual influences on psychopathology, and treatment effects[28], in a diverse range of mental disorders[15,29-31]. Although EMA was initially feared to be too burdensome to those suffering from psychopathology, studies have shown that patients find short-term EMA feasible and acceptable in research settings[19,32,33]. It thus comes as no surprise that EMA has been suggested to be a relevant clinical tool as well. Before we can turn to exploring the clinical utility of

EMA, however, we should examine how EMA data can be utilized to offer insight into psychopathology. Two theoretical approaches have such clinical potential: the network theory of psychopathology, and the understanding of psychopathology as a complex dynamical system.

Box 1.1. The relevance of EMA for people with mood disorders.

This thesis focuses on the utility of EMA for people with diverse types of mental disorders, such as mood disorders, anxiety disorders, and psychotic disorders, because it is expected that EMA can be relevant for any type of psychopathology. Most studies in this thesis, however, revolve around mood disorders: major depressive disorder (MDD) and bipolar disorder. In the Netherlands, approximately 20% of the population is diagnosed with MDD at one point in their lives2. Individuals diagnosed with MDD are characterized by prolonged and pervasive sad mood, reduced interest in activities, low self-esteem, and low energy for at least two weeks. A second, less prevalent type of mood disorder is bipolar disorder (lifetime prevalence 1.3%2). In addition to depressive episodes, people with this diagnosis tend to experience episodes where they feel abnormally happy, energetic, or irritable. Usually, this is paired with reduced need for sleep and increased activity patterns. Adequate assessment of mood problems is therefore highly important in the treatment of major depression as well as bipolar disorder. EMA can play an important role in this assessment, by providing estimates of the severity of mood problems, variability of mood, and the influence of context (activities, social interactions, events) on mood. Mood disorders therefore form an ideal starting point to examine the clinical utility of EMA.

Psychopathology as a network

The network theory was first proposed by Cramer and colleagues in 2010[34] and has been readily adopted by psychopathology researchers due to its intuitive appeal. Rather than assuming that symptoms arise from an underlying common cause (e.g., depression), the network theory proposes that symptoms can trigger each other up to the point that the individual presents with the symptoms of a full-blown mental disorder[14]. For example, sleep problems could trigger the loss of energy, creating concentration problems, finally resulting in depressed mood and loss of interest, thus coming together as the syndrome depression. Together, these associations (edges) between symptoms (nodes) form a network structure. Because clinicians also tend to conceptualize patients' problems as a causal structure, it has been

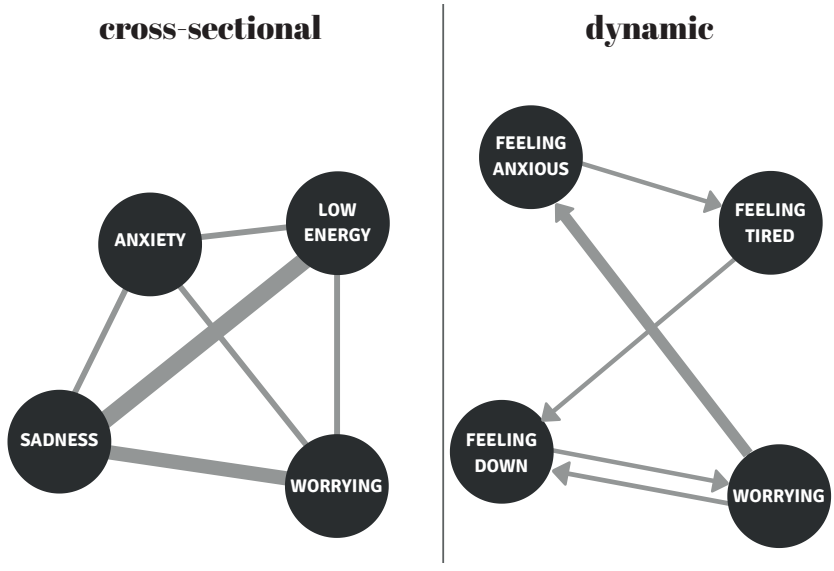
suggested that network methodology could test such implicit conceptualizations with actual data[35,36].

The network theory has important implications for the conceptualization of psychopathology. First, it assumes that mental disorders originate from a process of spreading activation in a symptom network. This activation can be triggered by an external event – for example, the loss of a loved one – after which the symptoms in the network keep reinforcing one another. This means that rather than searching for a common underlying cause, whether it be biological or environmental, research should focus on illuminating the causal interactions between symptoms or mental states. Second, it implies that network characteristics are as relevant as the symptoms themselves and could be used to inform treatment approaches. Indeed, many hypotheses derived from network theory have been brought forward in recent years. For example, it might be expected that comorbidity of psychiatric disorders can be explained through the existence of 'bridge' symptoms that activate symptoms of both disorders[37]. Likewise, symptoms with many strong connections in the network may be more influential or 'central', and could therefore constitute interesting treatment targets[38]. In this view, targeting a highly central symptom may simultaneously affect many other symptoms. Finally, research has examined whether a more connected or 'denser' network reflects increased vulnerability to psychopathology, also known as the connectivity-vulnerability hypothesis[39-41]. If symptoms are strongly connected to one another, (stressful) events may easily trigger many symptoms at once, making one vulnerable to psychopathology. These hypotheses are currently being tested empirically to determine their clinical implications, as well.

An important topic of discussion is the operationalization of the network theory. At what timescale do we expect to find associations relevant to the development of psychopathology? Should we focus on associations between symptoms when they are already present, or on associations between affective experiences? Research so far has adopted different methodologies to study this question (see Figure 1.2). The earliest empirical network studies examined network associations when people already experience symptoms, and constructed a network

of these symptoms (e.g., depressed mood, low energy, sleep). Usually, such studies correlate symptoms of retrospective questionnaires (e.g., how sad one has felt in the previous week) at one time-point (see e.g.[34,39,42]). This network methodology might therefore be termed a cross-sectional approach at the macro level. Associations between symptoms in macro-level networks are by definition between-person, which means that, for example, on average, depressed persons who are sadder are also likely to report lower energy. As such, macro-level cross-sectional studies may inform on the co-occurrence of symptoms in diverse psychiatric populations. However, the difficulty with this methodology is that associations in these networks cannot tell us *which* patients exhibit them, and whether they are reflective of how psychopathology develops over time in individual patients.

Figure 1.2. Exploration of the differences and similarities between the cross-sectional versus the dynamic network approach. The thickness of the lines represents the strength of the associations. Figure adapted from Wichers et al. (this thesis, chapter 5).



A second operationalization of the network theory has been proposed by Wichers (2014) and zooms in on smaller, more variable affective experience in daily life[9]. Therefore, this approach was termed the micro-level approach to network theory. This approach studies the dynamic interactions between momentary states

collected through EMA. Because EMA assesses a myriad of affective experiences at many time points, this approach enables the study of how these experiences follow one another over time. Associations in micro-level networks are usually within-person and mean, for instance, that if a depressed person feels sadder at one point in time, he or she will likely feel more tired the next moment in time. An advantage of dynamic micro-level EMA networks is that they fulfill an additional requirement to establish causality: temporal ordering of associations (experience A follows experience B). However, temporality cannot be taken to reflect causality; indeed, another unknown variable may be responsible for the associations (for example, the association between sadness and tired may be caused by a lack of sleep).

Another advantage of the micro-level network approach is that EMA data can be used to construct a personalized network model specific to one patient, which may have strong clinical utility. Consider the dynamic network of Figure 1.2. If this were the network of an individual patient, a clinician could discuss the relationships that emerge from this network. For example, for this person, worrying seems to increase feelings of anxiety and sadness several hours later. Treatment could therefore focus on decreasing worrying, using traditional treatment approaches such as cognitive behavioral therapy or mindfulness-based treatment. In networks with more nodes, one could identify central symptoms with the strongest outgoing connections and target those in treatment. In its most basic form, networks could thus facilitate discussion between patient and clinician on the relevant symptoms and affective experience to focus on. And if central symptoms indeed reflect effective treatment targets, this could help clinicians to determine their approach to treatment.

So far, although clinicians and patients are enthusiastic about the promise of networks for clinical practice, evidence of the clinical utility of networks has been mostly anecdotal[43-45]. Before networks can find large-scale implementation in clinical practice, there are several challenges that will need to be resolved. For instance, can we detect differences in the network structure before and after antidepressant treatment **(Chapter 2)**, or between individuals with versus those without anhedonia, a hallmark symptom of depression **(Chapter 3)**? And do the two

network methodologies (macro-level cross-sectional versus micro-level dynamic) result in similar, clinically relevant, conclusions[46,47] **(Chapter 4)**? And finally, what knowledge on psychopathology, in particular depression, has more than a decade of network research yielded, and is this suggestive of clinical usage of networks **(Chapter 5)**? The first part of this thesis revolves around these very questions, with the ultimate goal of determining the clinical utility of the network approach.

Psychopathology in the context of complex dynamical systems theory

Another promising way to utilize EMA data in personalized models can be derived from complex dynamical systems (CDS) theory. CDS theory is especially interesting because of its potential to use EMA data to alert patients and their clinicians to impending transitions in psychopathology: for example, a sudden increase in depressive symptoms, or the other way around, remission from depression. An important implication of applying CDS theory to psychopathology is that such upcoming shifts might be anticipated by examining the dynamics of time series data, the kind of data EMA provides.

Briefly, CDS theory proposes that many processes on earth can be characterized as dynamical systems with stable states[48,49]. Crucially, systems may undergo sudden switches, or transitions, between these states, that seem to happen out of the blue. For instance, historic data have shown that the climate exhibits sudden transitions between periods of relative warmth and ice ages. According to CDS theory, such transitions may be preceded by increasing instability, which is termed critical slowing down[48]. Although it may be counterintuitive that critical slowing down is reflected by increasing instability, it means that the system gets increasingly slower in recovering from minor perturbations. This can be detected in patterns in time series of the system, for example rising autocorrelations (e.g., the system's previous state increasingly lingers over time), variances (i.e., the system's previous states vary more widely around the mean over time), and (cross)-correlations between states[50]. Such indications that the system may be close to a

transition are termed early warning signals. Early warning signals have been shown to precede transitions in a wide variety of systems, ranging from changes in the climate[51], pollution in lakes[52], shifts in starlight[53], and financial crises[54].

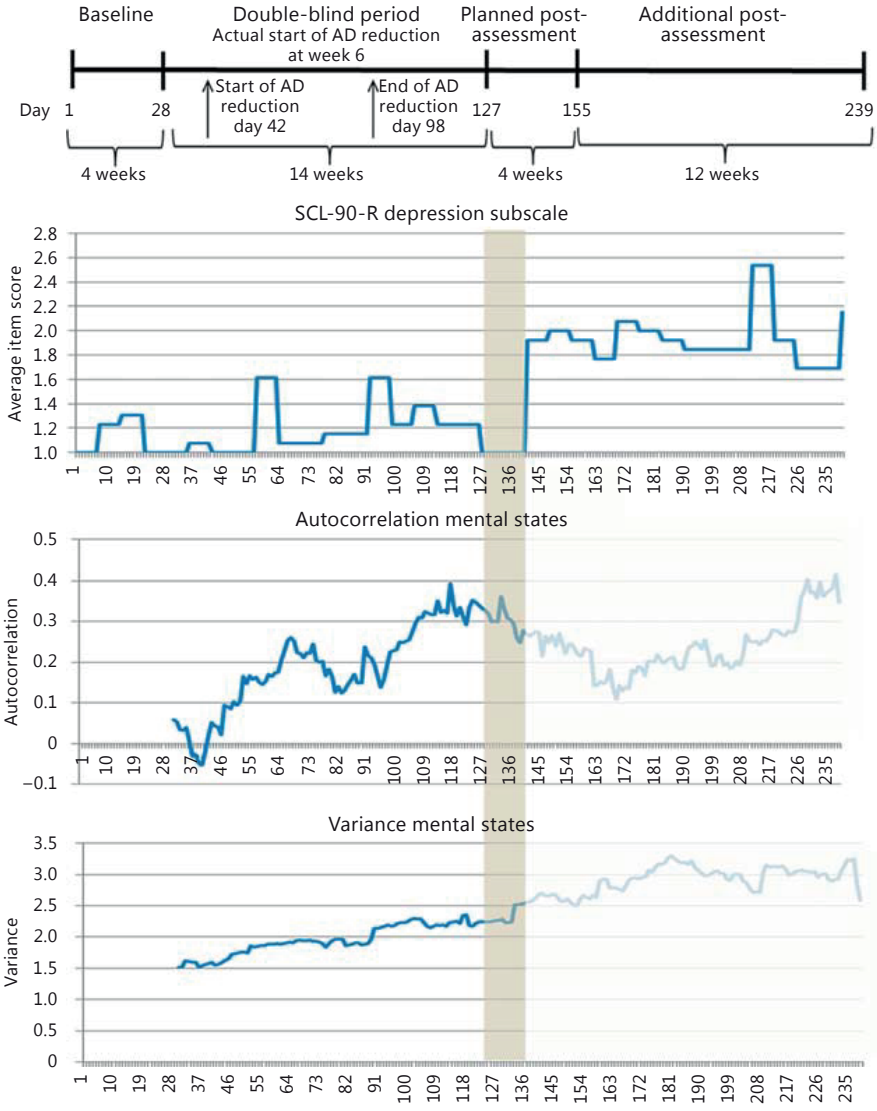
Recently, studies have proposed a complex dynamical systems approach to psychopathology[55]. If psychopathology can be characterized as an alternative stable state, we might be able to anticipate a transition from relative health to psychopathology (e.g., a recurrent depressive episode) by examining instability in EMA affective states[40,55]. In psychiatry, early warning signals indicate that external perturbations (e.g., stressful events) have an increasingly strong (variance) and lasting (autocorrelation) impact on one's wellbeing, thereby decreasing resilience and making the person vulnerable to the development of psychopathology[55-57]. The interesting part about CDS theory is its potential to anticipate transitions without a theoretical understanding of the system. We therefore do not have to fully understand the etiology of psychopathology and the complex interplay of genetic, psychosocial and biological factors resulting in mental disorders. Instead, CDS theory promises to be able to anticipate transitions in psychopathology using relatively simple measures like rising variances and autocorrelations.

So how would this work? Consider the first prospective empirical investigation into early warning signals by Wichers et al.[58] (see Figure 1.3). Here, a person in depressive remission completed a maximum of ten EMA diaries every day for 239 days. During this period, he gradually, and blindly, tapered his antidepressant medication. Around day 127, he reported a sudden increase in depressive symptoms on the weekly administered Symptom Checklist (SCL-90) depression subscale. Interestingly, we see that this sudden transition is preceded by early warning signals in his EMA affective experiences, namely a rising autocorrelation and variance. This suggests that we might be able to use EMA data to foresee a sudden transition towards psychopathology.

Recently, the findings of Wichers et al. were replicated in another patient with depression[58,59]. If these findings are found consistently in larger patient groups, the clinical impact could be monumental. Patients could monitor themselves longitudinally using EMA, enabling the detection of early warning signals to alert

patients and their clinicians to impending changes in symptoms. This would enable them to intervene early when symptoms emerge, thereby mitigating their impact and preventing their escalation. Such an approach could be especially relevant for

Figure 1.3. Demonstration of early warning signals in a person with depression. Source: Wichers et al., 2016[57].



persons with bipolar disorder. Persons with bipolar disorder frequently experience transitions from euthymic (neutral) mood to either depressed or elevated (manic) mood states. Most treatments of bipolar disorder already revolve around longitudinal mood monitoring and are focused on early recognition of mood episodes[60]. EMA compares favorably to those existing daily or weekly monitoring methods because it is more detailed and comprehensive as well as more frequent. Considering the enormous detrimental impact of depressive and manic episodes, it is vital that such shifts are recognized and acted upon as early as possible. Early warning signals based on EMA have the potential to be more sensitive to impending mood shifts, providing patients and their clinicians with an individual risk assessment of upcoming episodes, and enabling early intervention before a full-blown episode is developed.

So far, no longitudinal EMA data has been available to test whether EMA data can be used to detect early warning signals prior to upcoming mood episodes in people with bipolar disorder. Some studies, however, did already demonstrate that mood fluctuations of people with bipolar disorder follow principles of complex dynamical systems[61-63]. Furthermore, a study using simulated data suggested that early warning signals might be detected in actigraphy data[64]. A logical next step would be to gather longitudinal EMA data to determine whether early warning signals indeed precede transitions to depressive or manic episodes. Therefore, we have gathered such intensive longitudinal EMA data in a sample of patients with bipolar disorder, which will be discussed in **Chapter 9**. More generally, this will give insight into whether EMA data might be used as a tool to alert patients with diverse types of psychopathology and their clinicians to potential symptom changes.

Now that we have explored EMA as a self-monitoring method, and two promising approaches to construct personalized models based on EMA data, I will turn to the exploration of EMA as a clinical tool.

Ecological momentary assessment and personalized models as clinical tools

Early on, researchers and scientist-practitioners have started to highlight the potential of EMA as a clinical tool[13,65-67]. This potential lies in two main components of EMA. First, self-monitoring through EMA is suggested to improve patients' self-management by increasing insight into their well-being. By frequently reflecting on one's mood and symptoms, and learning what kind of activities or situations positively or negatively influence mood, patients might become more in control over their well-being[68,69]. Indeed, for this very reason, many existing treatment modalities such as cognitive behavioral therapy already employ some form of (paper-and-pencil) self-monitoring[70]. As such, EMA monitoring may already be an intervention in itself.

Secondly, EMA data can be visualized to enhance understanding of the patient's mechanisms of psychopathology. Such EMA feedback could be as simple as demonstrating variability in affect, or showing in which contexts patients experience complaints. Since EMA is suggested to be unaffected by memory biases[20], such feedback could provide a more reliable overview of how patients fared in between treatment sessions. Furthermore, when sufficient data is gathered with EMA, feedback could also consist of more complex statistical models such as a personalized network model or detection of early warning signals. EMA feedback could therefore form the basis of a more collaborative approach to diagnosis and intervention, where patients and therapists use EMA data to decide together what the next steps in treatment should be. It is expected that providing such personalized EMA feedback could balance the knowledge asymmetry that can characterize the patient-clinician relationship, and that EMA benefits shared decision making[13,68].

The intuitive appeal of EMA as a clinical tool has instigated the first empirical investigations of its effectiveness, resulting in a mixed picture. The first randomized controlled trial (RCT) by Kramer and colleagues in 2014 was promising: EMA monitoring and EMA feedback were more effective in reducing depressive complaints than antidepressant medication alone[71]. Furthermore, EMA monitoring and EMA feedback were found to improve patients' feelings of empowerment[72] and

instigated behavioral change[73]. Unfortunately, two more recent clinical trials were not able to replicate these beneficial effects in depressed individuals in routine care[74] and individuals with increased loss of interest and pleasure (anhedonia)[75]. Trials in other psychiatric populations are currently underway (see e.g.[76,77]), but their results so far suggest that the benefits of EMA might not lie in the reduction of symptoms, but more in increased self-management and feelings of empowerment.

This brings us to the question if, when, and how EMA has clinical value. When the promise of EMA as a clinical tool is discussed, many describe the universal potential of EMA, suggesting that it applicable to any clinical context (from diagnosis, to intervention, to relapse prevention), any kind of treatment (psychological or psychopharmalogical), and any psychiatric population (from anxiety disorder to bipolar disorder). The question is whether this is indeed true. Until recently, most information regarding the feasibility and applicability of EMA originated from (short-term) research contexts, which are arguably highly different from (long-term) clinical contexts. Such studies can be useful to inform on potential clinical applications of EMA, as is evaluated in the context of psychopharmacology in **Chapter 6**. However, it remains unclear to what extent such EMA diaries need to be adapted to be usable in treatment. For example, can EMA diaries remain standardized as is common practice in scientific studies, or is personalization of diary content, frequency, and duration necessary? Furthermore, although research has so far reported little evidence of potential negative side effects of EMA[78-80], we do not yet know whether we can expect any negative effects when EMA is used more intensively as a clinical tool. The field is therefore in need of a comprehensive investigation of the clinical utility of EMA: potential clinical applications in diverse settings and populations.

Qualitative research is optimally suited to answer such exploratory questions. As a starting point, the two most important stakeholders in the clinical use of EMA should be invited to the drawing board: patients and clinicians. They will be the ones directly using EMA. Their involvement is important for several reasons. First, their expectations and experiences can inform our understanding of how

EMA may benefit (or harm) treatment, how it may (or may not) be integrated in diverse phases of care, and what factors should be addressed when considering the implementation of EMA. Addressing such expectations early on can inform us on the optimal development of a clinically relevant EMA tool. Second, we know from implementation science and e-health research that it is notoriously difficult to implement e-health into clinical practice[81]. This is because of one of two important pitfalls: e-health tools either have limited scientific foundation, or they originated from a research methodology and are insufficiently translated to clinical contexts. Therefore, involving users (patients and clinicians) in the development process is crucial for the adoption of an e-health tool in practice[82]. Finally, patients and clinicians may help us determine relevant areas for future research, for example the identification of potential personalized models that could have clinical utility. For this reason, **Chapter 7** and **Chapter 8** form a qualitative investigation on patients' and clinicians' expectations and experiences with EMA as a clinical tool.

Outline of this thesis

The present thesis has adopted a multi-method approach to investigate the clinical utility of EMA. It has aimed to build bridges: between research and clinical practice, between different methodological approaches to personalized models, across diverse psychiatric populations, using both qualitative and quantitative research methods, with the ultimate aim of obtaining a comprehensive overview of the promise, pitfalls, and possibilities of EMA and EMA-based personalized models for psychiatric care.

Part I of this thesis investigates the network theory. The clinical utility of EMA is closely tied to the development of personalized models using EMA data, for example through network analysis. However, it is yet unclear how such personalized EMA networks relate to more easily obtained cross-sectional symptom networks. Therefore, **Chapter 2** starts by investigating whether the effects of antidepressant medication treatment can be inferred from cross-sectional symptom networks. We use data of two large RCTs to examine differences in network structure before

and after eight weeks of antidepressant medication. In **Chapter 3,** we utilize EMA data to investigate micro-level affective dynamics of individuals with subclinical depression with or without loss of interest. Can we understand the poorer outcomes associated with anhedonia by studying the impact of stress and physical activity on affect? **Chapter 4** delves further into the question whether cross-sectional and dynamic networks yield similar results and conclusions. Finally, **Chapter 5** reviews the literature on network studies in depression, summarizing what the network approach has so far taught us about the depressive syndrome, and reflecting on the utility of the network approach for clinical practice.

Part II of this thesis explores the clinical utility of EMA and EMA feedback. **Chapter 6** systematically reviews studies that have utilized EMA in the context of psychopharmacology, thereby summarizing several potential clinical applications of EMA. **Chapter 7** is a qualitative investigation of patient and clinician perspectives on the promise of EMA for clinical practice and what is required for its implementation. **Chapter 8** extends these findings by investigating patients and clinicians experiences with the addition of long-term (4-month) EMA to the treatment of bipolar disorder. Finally, in **Chapter 9** we investigate whether early warning signals in EMA data can inform patients with bipolar disorder of upcoming mood shifts approach.

In **Chapter 10**, I will summarize and discuss the main findings described in this thesis, and explore their clinical and academic implications. Furthermore, I will describe the development of a clinical EMA tool that incorporates the recommendations brought forward by the research in this thesis and recent other studies. Finally, this thesis will conclude by exploring directions for future research into the clinical utility of EMA.

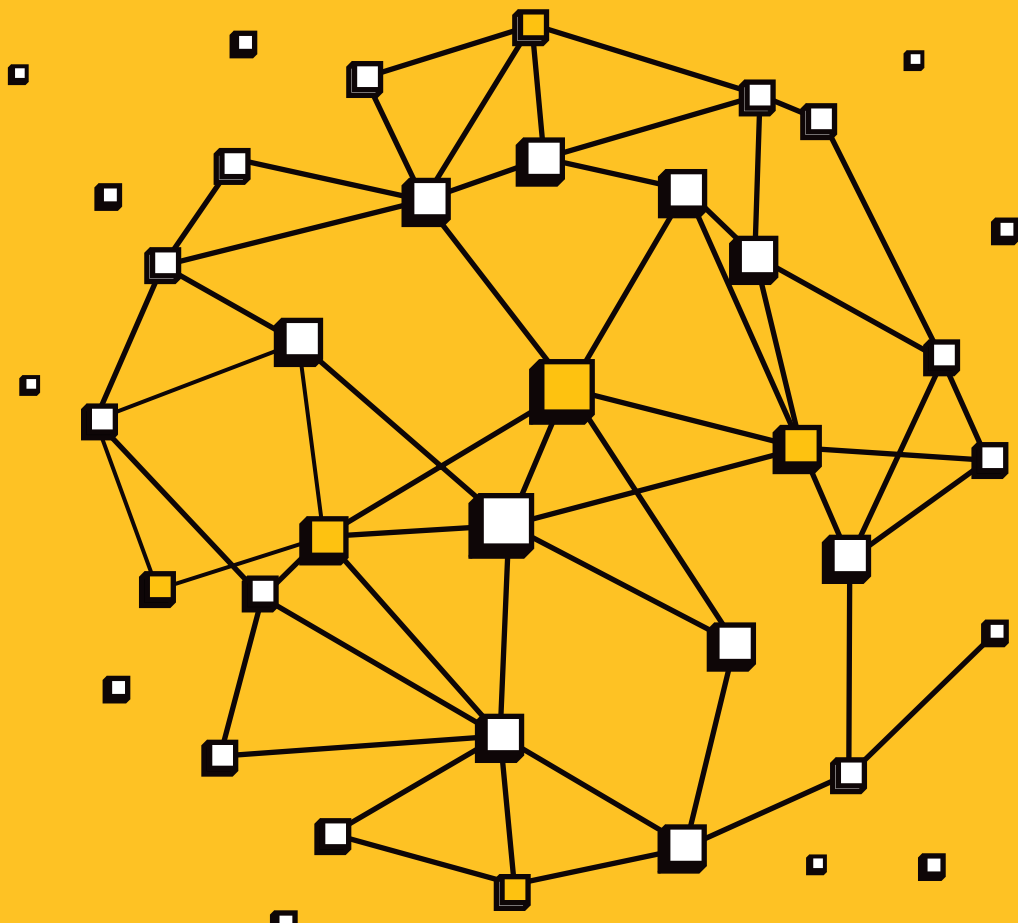
References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, Silove D. The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology.* 2014;43(2):476-493.
2. De Graaf R, Ten Have M, Van Gool C, Van Dorsselaer S. Prevalentie van psychische aandoeningen en trends van 1996 tot 2009; resultaten van NEMESIS-2. = Prevalence of mental disorders, and trends from 1996 to 2009. Results from NEMESIS-2. *Tijdschrift voor Psychiatrie.* 2012;54(1):27.
3. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet.* 2012;380(9859):2197-2223.
4. Pini S, de Queiroz V, Pagnin D, Pezawas L, Angst J, Cassano GB, Wittchen H-U. Prevalence and burden of bipolar disorders in European countries. *European Neuropsychopharmacology.* 2005;15(4):425-434.
5. Kendler KS, Zachar P, Craver C. What kinds of things are psychiatric disorders? *PsycholMed.* 2011;41(6):1143-1150.
6. Moskowitz DS, Young SN. Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience.* 2006;31(1):13.
7. 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.
8. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STARD study. *Journal of Affective disorders.* 2015;172:96.
9. Wichers M. The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *PsycholMed.* 2014;44(7):1349.
10. Insel TR. The NIMH research domain criteria (RDoC) project: Precision medicine for psychiatry. *AmJPsychiatry.* 2014;171(4):395-397.
11. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology.* 2008;4:1.
12. Csikszentmihalyi M, Larson R, Prescott S. Flow Experience in the Daily Lives of Older Adults: An Analysis of the Interaction between Flow, Individual Differences, Serious Leisure, Location, and Social Context. *Journal of Youth and Adolescence.* 1977;6:281-294.
13. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, Drukker M, Reininghaus U, Jacobs N, Lataster T, Simons C, Lousberg R, Gülöksüz S, Leue C, Groot PC, Viechtbauer W, Delespaul P. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *DepressAnxiety.* 2017;34(6):481.
14. Borsboom D. A network theory of mental disorders. *World Psychiatry.* 2017;16(1):5.
15. 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.
16. Delespaul P. Assessing schizophrenia in daily life: The experience sampling method. *Maastricht, The Netherlands: Maastricht Univ Press.* 1995.
17. Van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. *ACM Computing Surveys.* 2017;50(6):1-40.
18. Janssens KA, Bos EH, Rosmalen JG, Wichers MC, Riese H. A qualitative approach to guide choices for designing a diary study. *BMC Medical Research and Methodology.* 2018;18(1):1-12.
19. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *JMedInternet Res.* 2019;21(12):e14475.
20. Reis HT. Why researchers should think 'real-world': A conceptual rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:3.

21. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *PsycholAssess.* 2009;21(3):425.
22. Barge-Schaapveld D, Nicolson NA. Effects of antidepressant treatment on the quality of daily life: An experience sampling study. *JClinPsychiatry.* 2002;63(6):477.
23. 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. *PsycholMed.* 2009;39(9):1533-1547.
24. Zuidersma M, Riese H, Snippe E, Booij S, Wichers M, Bos E. Single-subject research in psychiatry: Facts and fictions. *Frontiers in Psychiatry.* 2020;11:1174.
25. Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives.* 2004;2(4):201.
26. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences.* 2018;115(27):E6106-E6115.
27. Wright AG, Woods WC. Personalized models of psychopathology. *Annual Review of Clinical Psychology.* 2020;16.
28. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, Reininghaus U. Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry.* 2018;17(2):123-132.
29. 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. *JAnxiety Disord.* 2014;28(8):925.
30. Oorschot M, Kwapil T, Delespaul P, Myin-Germeys I. Momentary assessment research in psychosis. *PsycholAssess.* 2009;21(4):498.
31. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. *PsycholAssess.* 2009;21(4):486.
32. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:636.
33. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul P, Lewis SW, Dunn G. Experience sampling research in individuals with mental illness: Reflections and guidance. *Acta PsychiatrScand.* 2011;123(1):12.
34. Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. Comorbidity: A network perspective. *BehavBrain Sci.* 2010;33(2-3):137.
35. Robinaugh D, Haslbeck J, Waldorp L, Kossakowski J, Fried EI, Millner A, McNally RJ, van Nes EH, Scheffer M, Kendler KS. Advancing the network theory of mental disorders: A computational model of panic disorder. *PsyArXiv.* 2019.
36. Burger J, van der Veen DC, Robinaugh DJ, Quax R, Riese H, Schoevers RA, Epskamp S. Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: A computational model of functional analysis. *BMC Med.* 2020;18:1-18.
37. Jones PJ, Mair P, Riemann BC, Mugno BL, McNally RJ. A network perspective on comorbid depression in adolescents with obsessive-compulsive disorder. *JAnxiety Disord.* 2018;53:1.
38. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks.* 2010;32(3):245.
39. van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, Schoevers RA. Association of symptom network structure with the course of longitudinal depression. *JAMA Psychiatry.* 2015;72(12):1219.
40. Cramer AOJ, van Borkulo CD, Giltay EJ, van der Maas HLJ, Kendler KS, Scheffer M, Borsboom D. Major depression as a complex dynamic system. *PLoS One.* 2016;11(12):e0167490.
41. Pe ML, Kircanski K, Thompson RJ, Bringmann LF, Tuerlinckx F, Mestdagh M, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Kuppens P, Gotlib IH. Emotion-network density in major depressive disorder. *Clinical Psychological Science.* 2015;3(2):292.
42. Fried EI, Bockting C, Arjadi R, Borsboom D, Amshoff M, Cramer AOJ, Epskamp S, Tuerlinckx F, Carr D, Stroebe M. From loss to loneliness: The relationship between bereavement and depressive symptoms. *JAbnormPsychol.* 2015;124(2):256.
43. Epskamp S, van Borkulo CD, van der Veen DC, Servaas MN, Isvoranu A-M, Riese H, Cramer AOJ.

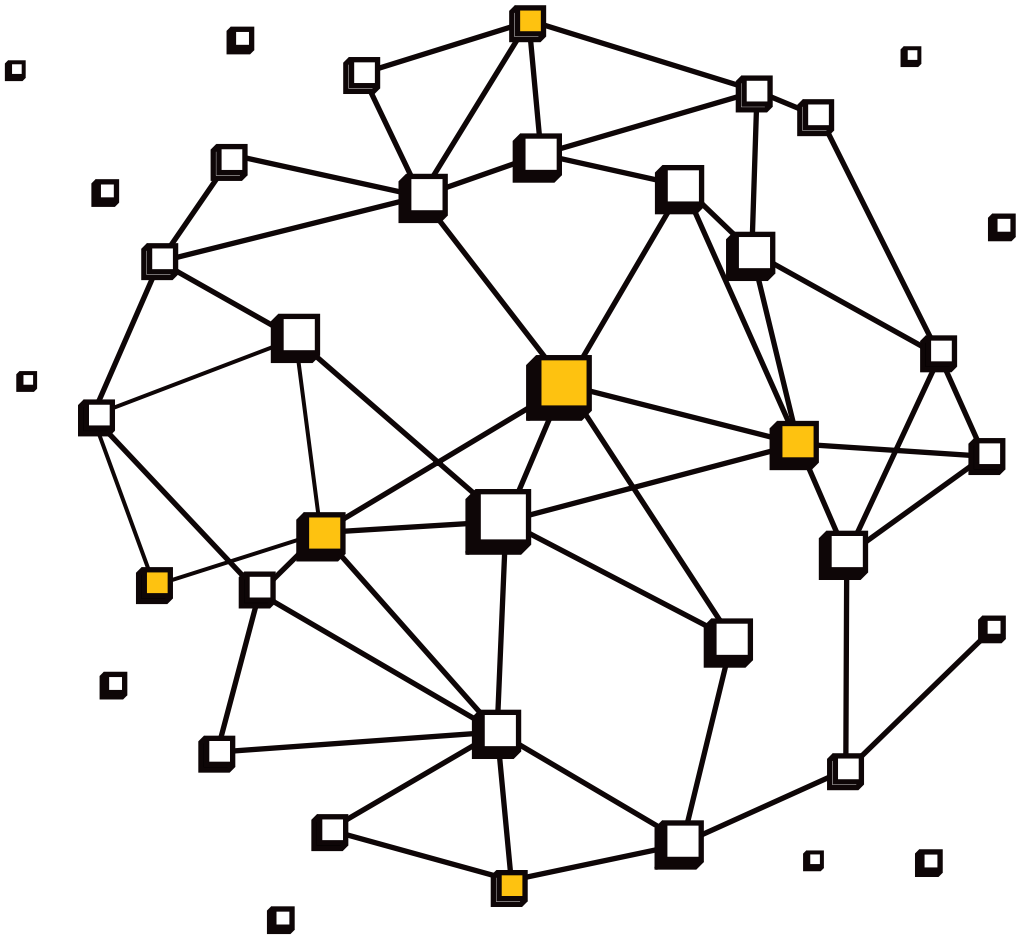
Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science.* 2018;6(3):416.
44. Kroeze R, van der Veen DC, Servaas MN, Bastiaansen JA, Oude Voshaar R, Borsboom D, Riese H. Personalized feedback on symptom dynamics of psychopathology: a proof-of-principle study. *Journal for Person-Oriented Research.* 2017;3(1):1.
45. 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.
46. Bringmann LF, Eronen MI. Don't blame the model: Reconsidering the network approach to psychopathology. *PsycholRev.* 2018;125(4):606.
47. Bos EH, Wanders RBK. Group-level symptom networks in depression. *JAMA Psychiatry.* 2016;73(4):411.
48. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, Van Nes EH, Rietkerk M, Sugihara G. Early-warning signals for critical transitions. *Nature.* 2009;461(7260):53.
49. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J. Anticipating critical transitions. *Science.* 2012;338(6105):344.
50. Dablander F, Pichler A, Cika A, Bacilieri A. Anticipating critical tansitions in psychological systems using early warning signals: Theoretical and practical considerations. *Preprint retrieved from psyarxiv.* 2020.
51. Dakos V, Scheffer M, van Nes EH, Brovkin V, Petoukhov V, Held H. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences.* 2008;105(38):14308-14312.
52. Donangelo R, Fort H, Dakos V, Scheffer M, Van Nes EH, Chaos. Early warnings for catastrophic shifts in ecosystems: comparison between spatial and temporal indicators. *International Journal of Bifurcation.* 2010;20(02):315-321.
53. George SV, Kachhara S, Misra R, Ambika G. Early warning signals indicate a critical transition in Betelgeuse. *Astronomy & Astrophysics.* 2020;640:L21.
54. Diks C, Hommes C, Wang J. Critical slowing down as an early warning signal for financial crises? *Empirical Economics.* 2019;57(4):1201-1228.
55. Hayes AM, Andrews LA. A complex systems approach to the study of change in psychotherapy. *BMC Med.* 2020;18(1):1-13.
56. 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.
57. Olthof M, Hasselman F, Strunk G, van Rooij M, Aas B, Helmich MA, Schiepek G, Lichtwarck-Aschoff A. Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science.* 2020;8(1):25-35.
58. Wichers M, Groot PC, Psychosystems, Group ESM, Group EWS. Critical slowing down as a personalized early warning signal for depression. *PsychotherPsychosom.* 2016;85(2):114.
59. Wichers M, Smit AC, Snippe E. Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: A confirmatory single-subject time-series study. *Journal for Person-Oriented Research.* 2020;6(1):1-15.
60. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, Matthews M. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: Implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association.* 2016;23(3):477-484.
61. Bonsall MB, Geddes JR, Goodwin GM, Holmes EA. Bipolar disorder dynamics: affective instabilities, relaxation oscillations and noise. *Journal of the Royal Society Interface.* 2015;12(112):20150670.
62. Ortiz A, Alda M. The perils of being too stable: mood regulation in bipolar disorder. *Journal of Psychiatry & Neuroscience.* 2018;43(6):363.
63. Curtiss J, Fulford D, Hofmann SG, Gershon A. Network dynamics of positive and negative affect in bipolar disorder. *Journal of Affective Disorders.* 2019;249:270-277.
64. Bayani A, Hadaeghi F, Jafari S, Murray G. Critical slowing down as an early warning of transitions in episodes of bipolar disorder: A simulation study based on a computational model of circadian activity rhythms. *Chronobiology International.* 2017;34(2):235-245.
65. Wichers M, Hartmann JA, Kramer IMA, Lothman C, Peeters F, van Bemmel L, Myin-Germeys I,

Delespaul P, van Os J, Simons CJP. Translating assessments of the film of daily life into person-tailored feedback interventions in depression. *Acta PsychiatrScand.* 2011;123(5):402.
66. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, van Bemmel AL, Peeters F, Delespaul P, van Os J. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta PsychiatrScand.* 2011;124(4):262.
67. 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.
68. Alpay L, van der Boog P, Dumaij A. An empowerment-based approach to developing innovative e-health tools for self-management. *Health Informatics Journal.* 2011;17(4):247-255.
69. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, Sanci L, Patton G. Self-monitoring using mobile phones in the early stages of adolescent depression: Randomized controlled trial. *JMedInternet Res.* 2012;14(3):e67.
70. Cohen JS, Edmunds JM, Brodman DM, Benjamin CL, Kendall PC. Using self-monitoring: Implementation of collaborative empiricism in cognitive-behavioral therapy. *Cognitive Behavioral Practice.* 2013;20(4):419-428.
71. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, van Bemmel AL, Myin-Germeys I, Delespaul P, van Os J, Wichers M. A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry.* 2014;13(1):68.
72. Simons CJP, Hartmann JA, Kramer I, Menne-Lothmann C, Höhn P, Van Bemmel AL, Myin-Germeys I, Delespaul P, Van Os J, Wichers M. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. *European Psychiatry.* 2015;30(8):900.
73. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, Viechtbauer W, Delespaul P, Myin-Germeys I, Wichers M. Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology.* 2016;35(5):433.
74. Bastiaansen JA, Ornee DA, Meurs M, Oldehinkel A. An evaluation of the efficacy of two add-on ecological momentary intervention modules for depression in a pragmatic randomized controlled trial (ZELF-i). *PsycholMed.* 2020:1-10.
75. 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. *Behavior Therapy.* 2017;48(1):76.
76. Bell IH, Rossell SL, Farhall J, Hayward M, Lim MH, Fielding-Smith SF, Thomas N. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): Feasibility, acceptability and preliminary clinical outcomes. *SchizophrRes.* 2020;216:479-487.
77. Gumley A, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Beattie L, Bell I, Birchwood M, Briggs A, Bucci S. Early signs monitoring to prevent relapse in psychosis and promote well-being, engagement, and recovery: Protocol for a feasibility cluster randomized controlled trial harnessing mobile phone technology blended with peer support. *JMIR Research Protocols.* 2020;9(1):e15058.
78. Barta WD, Tennen H, Litt MD. Measurement reactivity in diary research. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:108.
79. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors.* 2002;16(3):205.
80. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain.* 2003;104(1-2):343.
81. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: An explanatory systematic review. *Bulletin of the World Health Organization.* 2012;90:357-364.
82. Daniëls NE, Hochstenbach LM, van Bokhoven MA, Beurskens AJ, Delespaul PA. Implementing Experience Sampling technology for functional analysis in family medicine–a design thinking approach. *Front Psychol.* 2019;10:2782.



Part I

Empirical network studies and methodological challenges



Chapter 2

Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment

Published as:

Bos FM, Fried EI, Hollon SD, Bringmann LF, Dimidjian S, DeRubeis RJ, Bockting CLH. Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *SocPsychiatry PsychiatrEpidemiol.* 2018;53(6):617-627.

Abstract

Purpose. Recent reviews have questioned the efficacy of selective serotonin reuptake inhibitors (SSRIs) above placebo response, and their working mechanisms remain unclear. New approaches to understanding the effects of SSRIs are necessary in other to enhance their efficacy. The aim of this study was to explore the possibilities of using cross-sectional network analysis to increase our understanding of symptom connectivity before and after SSRI treatment.

Methods. In two randomized controlled trials (total N=178), we estimated Gaussian graphical models among 20 symptoms of the Beck Depression Inventory-II before and after 8 weeks of treatment with the SSRI paroxetine. Networks were compared on connectivity, community structure, predictability (proportion explained variance), and strength centrality (i.e. connectedness to oth ler symptoms in the network).

Results. Symptom severity for all individual BDI-II symptoms significantly decreased over 8 weeks of SSRI treatment, whereas interconnectivity and predictability of the symptoms significantly increased. At baseline, 3 communities were detected; 5 communities were detected at week 8.

Conclusions. Findings suggest the effects of SSRIs can be studied using the network approach. The increased connectivity, predictability, and communities at week 8 may be explained by the decrease in depressive symptoms rather than specific effects of SSRIs. Future studies with larger samples and placebo controls are needed to offer insight into the effects of SSRIs.

Introduction

Currently, selective serotonin reuptake inhibitors (SSRIs) are recommended as the first treatment of choice by the American Psychiatric Association (APA) and the National Institute of Clinical Excellence (NICE) for patients with major depressive disorder (MDD)[1,2]. However, recent reviews have questioned the efficacy of SSRIs[3-5]. Given that there is a substantial proportion of patients who do not seem to benefit from acute antidepressant treatment[6], it is important to increase our understanding of the way SSRIs work so as to enhance their efficacy.

A relatively novel conceptual and statistical approach may provide an interesting framework to study the working mechanisms of SSRIs. The network approach conceptualizes symptoms of depression as part of a network of co-occurring symptoms that interact to create the symptom profile that is termed depression[7]. Instead of assuming that symptoms passively originate from a common cause (e.g. depression), the network approach proposes a focus on how individual symptoms are related to each other[8]. The network approach is thus fundamentally different from the routine modus operandi of monitoring treatment response at the aggregate level based on a large number of disparate psychiatric symptoms that are added up to a sum-score, which may obfuscate important insights[9]. The network approach also can help to identify 'central' symptoms—symptoms that often co-occur with other symptoms and thus are assumed to easily trigger or be triggered by other symptoms[10].

Prior studies that have used network analysis to characterize symptom patterns in depressed patients have yielded interesting results. Loss of energy and anhedonia appear to be central symptoms in analyses of both cross-sectional data and dynamic data using experience sampling methodology (ESM), the latter of which allows for the study of predictive effects of the symptoms on each other[11-13]. Further, depressed patients appear to have increased network connectivity or denser networks in cross-sectional data when compared to remitted patients[14] and in dynamic data when compared to healthy controls[15], suggesting that once one symptom is present, other symptoms might be easily triggered. However, in studies

following patients at different time points, the opposite pattern has been found: increased network connectivity when depressive symptom severity decreases[16].

With regards to potential effects of antidepressant treatment, only two studies examined the effects of antidepressants on the network structure of depressive symptoms. A recent study using dynamic data could not find evidence that the tricyclic antidepressant imipramine changed the dynamic associations between mental states[17]. In the STAR*D sample, using cross-sectional data, sad mood was found to be the most central symptom after several weeks of treatment with the SSRI citalopram[18]. However, the absence of a pretreatment condition, where no SSRI treatment was administered yet, limits insight into the effects of SSRIs.

Until now, the network approach has not been applied to explore the relationships among depressive symptoms before and after SSRI treatment. Such an approach may offer insight into the interrelations among symptoms before and after SSRI treatment and thus yield interesting directions for further study of how SSRIs may exert their effects. Therefore, the present study aimed to explore the possibilities of using cross-sectional network analysis before and after SSRI treatment by looking at the most recent and insightful network metrics. Specifically, we will examine (1) the interrelations, community structure, and connectivity of depression symptoms and (2) their centrality and predictability estimates in acutely depressed patients treated with the SSRI paroxetine, a commonly prescribed antidepressant. We combined the data of two large randomized controlled trials (RCTs) comparing paroxetine to pill-placebo and psychological treatment to retain sufficient power for the analyses. Focusing on the patients treated with paroxetine (N=178), we constructed networks before the start of treatment and after 8 weeks of treatment. We compared the two resulting networks and their metrics to explore the possibilities of network analysis for increasing our understanding of symptom connectivity before and after SSRI treatment.

Method

Participants

Patients with moderate to severe major depressive disorder participated in one of two similar RCTs for treatment of MDD in the United States, which were combined to increase statistical power[19,20]. To be included, patients had to (1) fulfill the DSM-IV criteria for MDD, as assessed by the Structured Clinical Interview for the DSM-IV[21], (2) 18-70 years of age, (3) English speaking, (4) and willing and able to give informed consent. The first RCT, conducted at the University of Pennsylvania and Vanderbilt University (N=240) between 1996-2001, only enrolled moderate to severely depressed patients (17-item Hamilton Rating Scale for Depression (HRSD-17)[22] ≥20). The second RCT, conducted at the University of Washington (N=241), enrolled patients between 1998-2001 and included mildly, moderately, and severely depressed patients (Beck Depression Inventory (BDI-II)[23] ≥20 and HRSD-17 ≥14).

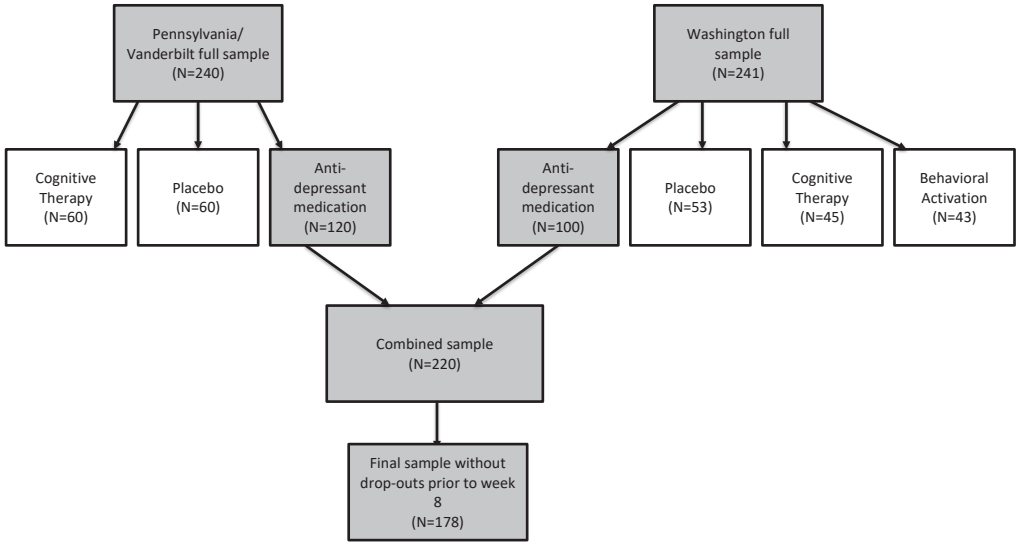
Patients were excluded in both trials if they had a lifetime diagnosis of psychosis, bipolar disorder, organic brain syndrome, or mental retardation. Further excluded were patients presenting with imminent suicide risk requiring immediate hospitalization, a current Axis I disorder judged to be primary, substance abuse, or antisocial, borderline, or schizotypal personality disorder. Lastly, participants who had not responded favorably to an adequate trial of either CT or paroxetine were excluded. The studies were conducted in accordance with the Declaration of Helsinki and institutional review boards of each of the three study sites approved the study protocols. All patients provided written informed consent.

Treatment

The present study focuses on patients randomized to antidepressant medication (see Figure 2.1), given our assumption that different treatments have different working mechanisms. Unfortunately, power was insufficient to construct separate networks for the other trial conditions. Patients randomized to one of the pill conditions were treated with the SSRI paroxetine or pill-placebo for 8 weeks in a double-blind manner. All patients receiving SSRI or pill-placebo received weekly sessions

with their prescribing clinician for the first 4 weeks and every other week thereafter. These sessions consisted of medication management, which involved education, adjustment of dosage and dosage schedules, and discussions of adverse effects; and of clinical management, which involved a review of the patient's functioning in major life domains, brief supportive counseling, and limited advice giving[24]. The usage of specific psychotherapy techniques was not allowed.

Figure 2.1. Overview of flow of participants throughout the two trials.



Measures

The BDI-II, a widely used self-report questionnaire with adequate psychometric properties[23,25], was administered at baseline (before the start of treatment) and week 8 (mid-treatment). Each item is rated on a 4-point Likert-scale, ranging from 0 to 3. Networks were constructed of 20 of the 21 individual items of the BDI-II at each timepoint. The BDI-II item on suicidal thoughts or wishes was removed from the network estimations given its low variance and positively skewed distribution[26] (see Appendix).

Statistical analyses

Missing data

The individual SSRI treatment arms of each trial were not sufficiently large to estimate networks due to the large number of individual symptoms. We therefore combined the data from the Pennsylvania/Vanderbilt and Washington studies, resulting in a total of 220 patients randomized to SSRI treatment. Of those, 42 participants dropped out prior to week 8. To ensure both networks consisted of the same people, those 42 participants were removed from the network analyses, resulting in a final sample of 178 participants at both time points. Seven individuals had partially missing data. The Gaussian graphical models (GGM) were estimated on the full data set (N=178) using pairwise complete observations (i.e., using all available information from all participants)[27]. Analyses of predictability and the network comparison test cannot deal with missing data, reducing the analytic data from 178 to 171 participants for those analyses.

Network estimation

First, we used the *R*-package *qgraph* (version 1.4.4) to estimate the network structures of the 20 BDI-II symptoms at baseline and at week 8[28]. Networks contain *nodes* (symptoms) and *edges* (cross-sectional associations among symptoms). We estimated GGMs in which the edges represent partial correlation coefficients. In order to reduce false positive edges, we applied the least absolute shrinkage and selection operator (lasso)[29]. This procedure penalizes very small edges by setting them to zero. The shrinkage parameter is chosen to minimize the extended Bayesian Information Criterion (EBIC) parameter[30] and has been shown to accurately recover underlying network structures[11]. We applied a conservative graphical tuning parameter of 0.5; false positive edges are very unlikely, while very small actual edges may not be captured. Position of the nodes in the networks was initially based on the Fruchterman-Reingold algorithm, which places the nodes with stronger and/or more connections closer together[31]. We averaged the layout of the networks to facilitate visual comparison of the timepoints. To examine robustness of our findings, we compared the standard deviations (SDs) of the BDI-II sum score and

all individual BDI-II items between the two time points by means of Levene's test (corrected for chance capitalization by the conservative Bonferroni-Holm method). If SDs change significantly, differences in the network structure might be a result of increased variation[32].

Communities

Within the GGMs as specified above, we explored the way nodes within the networks cluster together through exploratory graph analysis (EGA)[33]. Nodes that cluster together in communities may be part of the same latent variable or dimension. EGA estimates communities in networks via a random walk algorithm (walktrap).

Connectivity

Network connectivity (or density) for each network was calculated by summing all absolute edge weights. The difference in network connectivity for baseline and week 8 was tested for significance via the R-package network comparison test (NCT), which uses permutation testing to compare networks[34].

Centrality

Strength centrality was calculated for all symptoms at baseline and week 8 by summing the absolute values of the edges of a given node to other nodes[10]. The higher strength centrality, the more strongly is this symptom connected to other symptoms. The stability of strength centrality was calculated at baseline and week 8 through the correlation stability coefficient (CS-coefficient), which is a method based on bootstraps (N=1000) using the R-package *bootnet*[35]. The CS-coefficient represents the maximum proportion of cases that can be dropped, such that with 95% probability the correlation between the original strength centrality estimates and the bootstrapped estimates is 0.7 or higher[35]. The higher the CS-coefficient, the more reliable the interpretation of the order of centrality estimates. The R-package *bootnet* was further used to test strength centrality among nodes for significance at baseline and week 8.

Predictability

We further estimated the predictability of each of the nodes in the network, which is an estimate of how much of the variance of a node is explained by neighboring nodes[36]. The R-package *mgm*[37] was used to estimate the proportion of explained variance for each node, which could range from zero (a node cannot be predicted by other nodes in the network) to one (a node is perfectly predicted by other nodes).

Table 2.1. Demographic and clinical characteristics before randomization to SSRI treatment.

Variable	Combined sample (N = 170)		Pennsylvania/ Vanderbilt study (N = 95)		University of Washington study (N = 75)		P-value
Female, N (%)	93	(55%)	56	(59%)	54	(72%)	<.001
Caucasian, N (%)	140	(82%)	81	(85%)	59	(79%)	*n.s.*
Age	40.0	(10.9)	40.7	(11.3)	39.1	(10.4)	*n.s.*
Married/cohabitating, N (%)	66	(39%)	41	(43%)	25	(33%)	*n.s.*
Employed, N (%)	135	(79%)	84	(88%)	51	(68%)	.018
Chronic MDD, N (%)	85	(50%)	56	(59%)	29	(39%)	*n.s.*
Age of onset	23.8	(12.5)	21.0	(12.4)	27.3	(11.9)	.001
Number of previous episodes	1.9	(2.5)	2.7	(3.0)	1.0	(1.3)	<.001
Duration of current episode (months)	48.4	(74.5)	52.6	(77.1)	43.2	(71.2)	*n.s.*
History of psychiatric hospitalization, N (%)	15	(9%)	10	(11%)	5	(7%)	*n.s.*
Axis I comorbidity, N (%)	91	(54%)	71	(75%)	20	(27%)	<.001
Axis II comorbidity, N (%)	65	(38%)	50	(53%)	15	(20%)	<.001
BDI-II score baseline	31.6	(8.9)	31.2	(9.8)	32.1	(7.6)	*n.s.*
BDI-II score week 8	13.8	(10.2)	13.6	(10.6)	14.1	(9.8)	*n.s.*
HRSD-17 score baseline	22.3	(3.9)	23.8	(3.2)	20.4	(4.0)	<.001
HRSD-17 score week 8	12.3	(6.7)	12.4	(6.7)	12.2	(6.7)	*n.s.*
Daily dosage of paroxetine baseline	12.2	(2.7)	14.0	(4.9)	10.0	(0)	-
Daily dosage of paroxetine week 8	35.7	(11.2)	38.8	(11.0)	31.7	(11.5)	-

Note. Descriptive statistics represent mean (SD) unless otherwise stated. The table only includes individuals without missing data. Abbreviations: BDI-II = Beck Depression Inventory II; HDRS-17= 17-item Hamilton Depression Rating Scale. P-values represent tests between the two studies and were corrected for chance capitalization with the Bonferroni-Holm method. *n.s.* = non-significant.

Results

Participant characteristics

Table 2.1 depicts the demographic and clinical characteristics of the two study samples and the combined sample. The RCT samples were found to be quite similar with a few exceptions indicating that the Pennsylvania/Vanderbilt study included a more severely depressed sample as is appropriate given the inclusion criteria. BDI-II scores at baseline did not differ significantly between the two studies and percent change of BDI-II sum-scores from baseline to week 8 was similar (Pennsylvania/Vanderbilt study: 56.4%; Washington study: 56.1%); as was the percent of missing data (Pennsylvania/Vanderbilt: 20.8%; Washington: 25%).

Means and variation of all 21 BDI-II symptoms

In the combined sample, mean BDI-II sum score decreased significantly after 8 weeks by 44% (W=26,052, p<.001). Further, as depicted in Table 2.2, mean scores for all individual BDI-II items decreased significantly as well (p-values were corrected for chance capitalization via the conservative Bonferroni-Holm method). Loss of interest in sex (30% change) and appetite (38%) changed the least, whereas the largest changes were evident in suicidal ideation (74%), crying (72%), and feelings of punishment (70%). SD of the BDI-II sum score did not change significantly after 8 weeks of SSRI treatment (Levene's test = 1.34, p = .25). Five SDs of the individual BDI-II items decreased significantly (punishment feelings, suicidality, crying, sleep, and irritability); two increased (self-dislike and concentration difficulty).

Network connectivity and community detection

Figure 2.2 displays the GGM networks of the 20 BDI-II symptoms at baseline and week 8 (N=178). The network model for week 8 showed significantly higher connectivity: the sum of all edges was 9.1 compared to 7.3 for the baseline network (p = .02).

At baseline, three symptom communities emerged: (1) a cluster of

cognitive symptoms such as worthlessness, but also sad mood (orange nodes), (2) an affective behavioral cluster of symptoms such as loss of pleasure, energy, and agitation (blue), and (3) a cluster of sleep and appetite symptoms (green).

Table 2.2. Overview of means and standard deviations of the individual BDI-II items for baseline and week 8 (N = 178).

		Baseline		Week 8	
Item	**Abbreviation**	**Mean**	**SD**	**Mean**	**SD**
Sadness	SAD	1.4	0.8	0.5	0.6
Pessimism	PES	1.4	0.8	0.7	0.7
Past failure	FAI	1.8	0.7	0.8	0.8
Loss of pleasure	LPLE	1.8	0.7	0.9	0.8
Guilty feelings	GUI	1.3	0.8	0.5	0.6
Punishment feelings	PUN	1.0	1.1	0.3	0.7
Self dislike	DIS	1.9	0.8	0.9	0.9
Self criticalness	CRI	1.7	0.9	0.7	0.7
Suicidal thoughts or wishes*		0.7	0.6	0.2	0.4
Crying	CRY	1.6	1.1	0.5	0.9
Agitation	AGI	1.0	0.9	0.5	0.7
Loss of interest	LINT	1.9	0.9	0.8	0.8
Indecisiveness	IND	1.8	1.0	0.7	0.8
Worthlessness	WOR	1.6	0.8	0.5	0.7
Loss of Energy	LENE	1.7	0.8	0.9	0.8
Change in sleep	SLE	1.4	1.1	0.8	0.9
Irritability	IRR	1.5	0.9	0.5	0.7
Change in appetite	EAT	0.9	0.9	0.6	0.7
Concentration difficulty	CON	1.8	0.6	0.8	0.8
Fatigue	FAT	1.9	0.9	0.9	0.8
Loss of interest in sex	LSEX	1.5	1.0	1.0	1.1

*Because of its problematic distribution and small standard deviation, this BDI-II item was not included in the network analyses.

After 8 weeks of SSRI treatment, five communities emerged. The sleep and appetite cluster remained (orange nodes). The cognitive cluster also remained but lost symptoms of sad mood and crying (green). Three other clusters were identified: a cluster of irritation, agitation, and crying (yellow), a cluster of fatigue and

loss of energy (pink), and an affective behavioral cluster of symptoms such as sad mood, loss of interest, and agitation (blue).

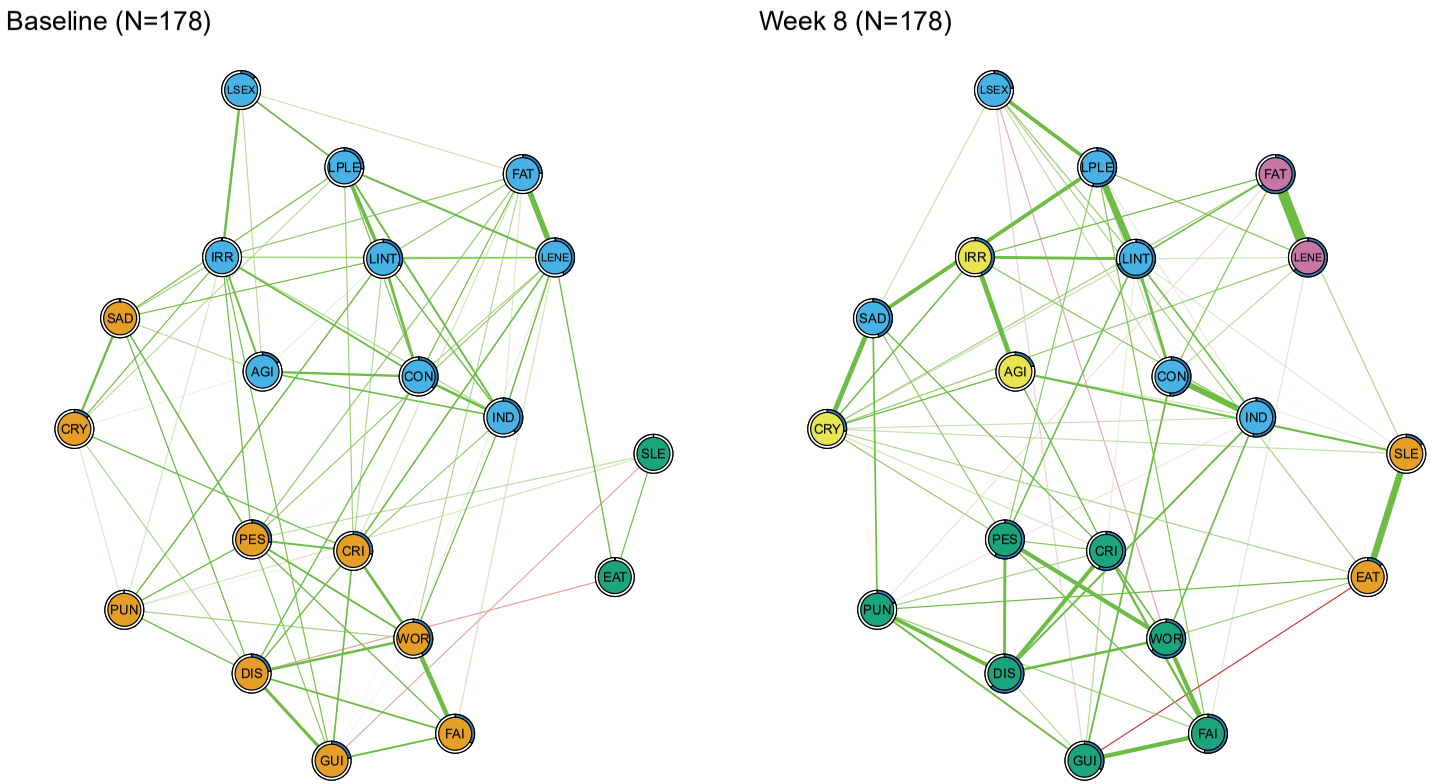
Symptom centrality and predictability

Figure 2.3 shows bootstrapped difference tests of strength centrality between nodes within the baseline network and the week 8 network (N=178). At baseline, the most central symptom was loss of energy. Loss of energy showed a significantly higher strength centrality than 9 other symptoms. In contrast, sleep had the lowest strength centrality and was significantly more weakly connected than 12 other symptoms. Mean severity of the symptoms was strongly associated with strength centrality (r=.69, p <.001), indicating that symptoms with higher severity were also more strongly interconnected.

At week 8, sleep remained the most weakly connected symptom; the most central symptom was loss of interest. However, no symptoms showed significantly stronger connections than other nodes in the network. Strength centrality (r=.33, p=.23) was no longer associated with mean symptom severity. The CS-coefficient was 0.13 for both baseline and week 8. This indicates the centrality estimates need to be interpreted with caution.

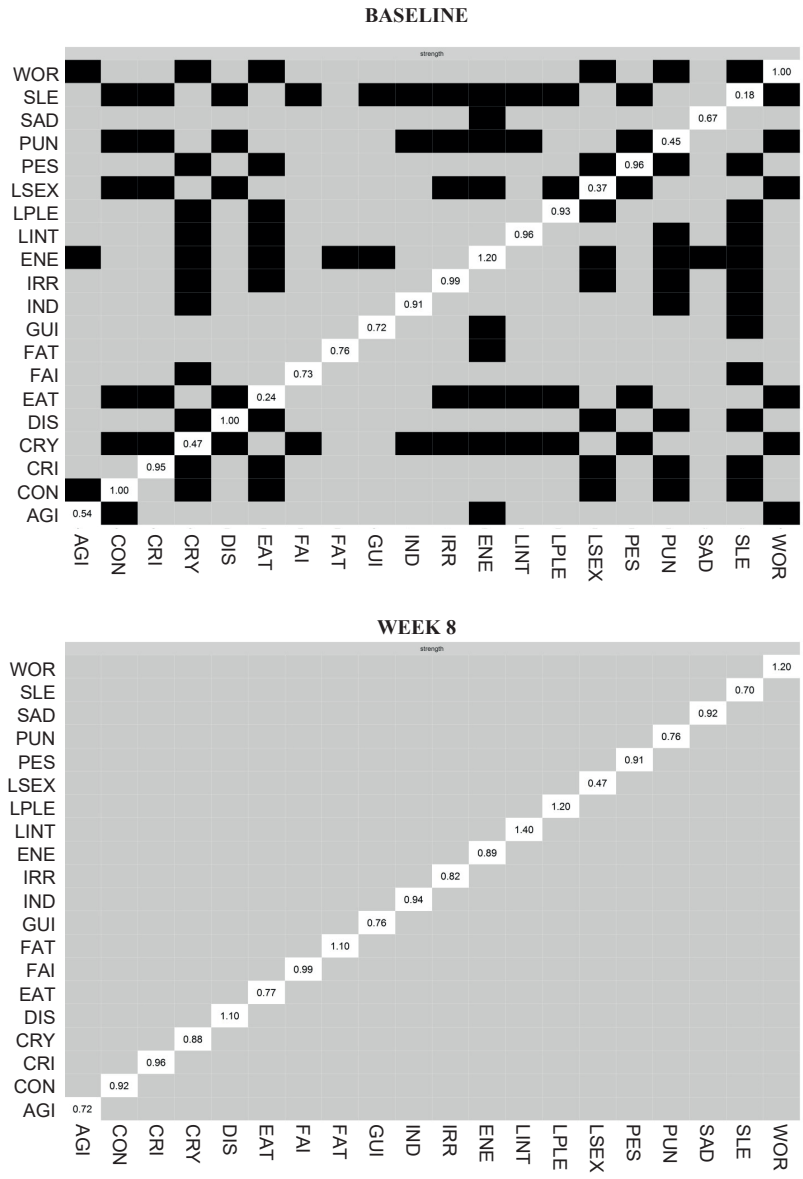
Finally, the overall proportion of explained variance (predictability) of nodes increased significantly from 0.21 at baseline to 0.45 at week 8 (t = -8.6, p <.001). At baseline, worthlessness and loss of energy showed the largest predictability relative to the other nodes; whereas for week 8, these were loss of interest and fatigue. The correlation between the predictability of the two time points was .76 (p < .001), indicating that if a node has high predictability at baseline, it tends to also have a high predictability at week 8. As expected, strength centrality was strongly related to the predictability of nodes at baseline (r = .73, p <.001) and week 8 (r = .80, p<.001), suggesting that if a node is strongly connected to other variables, it tends to be predicted by other nodes as well. Finally, although predictability was related to mean symptom severity at baseline (r = .68, p<.001), they were not significantly related at week 8 (r = .44, p = .053).

Figure 2.2. EBIC gLasso network of BDI-II symptoms before the start of treatment (left) and after 8 weeks of paroxetine treatment (N=178).



Note. Abbreviations can be found in Table 2.2. Green lines represent a positive association between two symptoms. The thicker the edge (line), the stronger the relationship between two symptoms. The color of the nodes represent the community the symptom belongs to; nodes with similar color belong to the same community. The proportion of explained variance (predictability) can be derived from the blue ring surrounding the node.

Figure 2.3. Bootstrapped difference tests of strength centrality between nodes within the baseline network and the week 8 network (N=178).



Note. Gray boxes indicate nodes that did not significantly from one-another and black boxes indicate nodes that do differ significantly from one-another. White boxes show the value of strength centrality for a given node.

Discussion

This study examined changes in the relationships among depressive symptoms after 8 weeks of SSRI treatment in a sample of outpatients treated for moderate to severe depression. To our knowledge, this is the first study to compare network structures before and after SSRI treatment.

The severity of all individual depressive symptoms significantly decreased after 8 weeks of paroxetine treatment and they became more strongly associated. The latter finding contrasts with previous network studies reporting that increased network connectivity is associated with acute depression[14,15], but is consistent with findings of increased network connectivity paralleling decreases in symptom severity[16,38,39]. Corresponding to a large body of literature on the factor solution of the BDI-II[13,23,40,41], we found a cluster consisting of cognitive symptoms both at baseline and week 8. However, the affective behavioral cluster usually also reported was split into several separate clusters in our sample, which also differed somewhat across the two time points.

Based on network theory, one would expect network connectivity to decrease as symptoms are alleviated; when symptoms become more weakly connected, they may be less easily triggered by other symptoms[7]. Our findings that network connectivity increased after depressive symptoms decreased during SSRI treatment therefore seems counterintuitive. However, previous studies with contrasting findings used between-subject rather than within-subject designs: van Borkulo et al. compared baseline networks of participants with persistent depression to participants who later remitted[14], and Pe et al. compared depressed patients versus healthy controls using dynamic data[15]. The present study examined connectivity over time within the same participants, which may explain the difference in results.

Indeed, our pattern of results, increased network connectivity as symptom severity decreased, is highly consistent with previous studies reporting a different network structure paralleling a decrease in symptom severity[16,38,39]. Fried et al. have suggested this pattern of results may have several possible explanations[16]. First, interpretation of items (of the BDI-II) may change due to psychological treatment, which is termed response bias. This seems less likely in our sample, given that our

sample did not receive psychological treatment. Second, variability of symptoms may change over time, explaining differences in network connectivity[32]. Although the standard deviations of the BDI-II individual items did not change significantly for 14 of the 20 items used in the network analyses, the proportion of increased variance (predictability) did increase at week 8. Finally, the hypothesis that increased network connectivity may be the result of SSRI treatment seems less likely, given parallel findings in other samples in which not all participants were on antidepressant medication[16,38,39]. Unfortunately, the lack of a control group in the present study precludes insight in potential unique effects of SSRIs on network connectivity.

Another aim of this study was to examine changes in the order of strength centrality after 8 weeks of SSRI treatment. This would enable us to investigate if and how SSRIs change the way symptoms are related to one another. However, due to the small sample size, power was too low to detect significant differences in the ordering of strength centrality at the two time points, especially at week 8. Indeed, correlation stability coefficients (0.13) were lower than the recommended lower bound of 0.25[35], suggesting that our estimation of the ordering of the centrality of symptoms may be less reliable. Therefore, interpretation of the order of symptoms remains speculative.

However, at baseline, some symptoms appear to be significantly more central than others, indicating they are more strongly connected. Indeed, our finding that loss of energy had the highest strength centrality at baseline, significantly higher than 9 other symptoms, is not completely surprising. The high centrality of loss of energy in depressed patients has been found in other network studies[11,13,42] and is in line with evolutionary theory, which proposes that depressive symptoms force an individual to conserve energy during adverse situations and reallocate it to solving the problem[43,44]. In a prospective study, strength centrality of fatigue was a strong predictor of depression onset[42]. It is speculated that by targeting and reducing central symptoms, because of its strong connections, other symptoms are reduced in parallel, which may prove to be a valuable treatment strategy[45]. However, other explanations may be possible as well, such as that highly central symptoms are part of the same latent variable (for example, loss of energy and fatigue may tap into the

same construct), and because they are so strongly interconnected, all estimates of strength centrality are biased[46].

The centrality of loss of energy decreased relative to the other symptoms at week 8; instead, loss of interest became the most central symptom. However, since strength centrality was not significantly higher for loss of interest than any other symptom, we cannot conclude with certainty that it is indeed more central. If our findings are replicated, a speculative thought may be that SSRIs weaken the connections of energy loss to other depressive symptoms, thereby alleviating depression. The finding that loss of interest became the most central symptom may suggest that if anhedonia is still present after treatment, other symptoms are still present too. Indeed, studies have shown that symptoms of anhedonia are resistant to SSRI treatment and predictive of poorer outcome[47-49]. SSRIs do not appear to target dopamine systems; thus they might not target symptoms related to the experience of positive affect and reward[50,51]. Speculatively, SSRIs may sever the links between symptoms related to negative affect and energy, whereas the relationships between anhedonia and other symptoms remain. However, power issues and lack of a comparison group prevented us from elucidating specific effects of SSRI treatment on the strength centrality of symptoms.

Important strengths of our study include the use of state-of-the-art network analysis techniques, our specific investigation of one type of SSRI, and the study design, which enabled us to compare symptom networks before and after treatment. However, our findings should also be considered in light of several limitations. First, because of insufficient power, we were not able to take advantage of the placebo controls. Thus, any observed change may not be due to specific effects of the SSRI. Second, our sample size was suboptimal for a network analysis with such a large number of nodes, limiting insight into potential significant differences in symptom centrality. Third, the positive skewness of the individual BDI-II items at week 8 may have affected our results[32]. Future investigations into how skewness should be addressed in Gaussian network models are warranted. Fourth, the specific characteristics of our sample may have limited generalizability. Also, BDI-II sum score was used as an inclusion criterion. Using the same instrument

for inclusion and network analyses may have influenced the results[52], although we expect this effect to be small given that participants were also selected based on other instruments (HRSD, SCID-I). Given these limitations, future studies with larger sample sizes and placebo controls are necessary before conclusions on the effects of SSRI on symptom interrelations can be drawn.

A final note on the interpretation of cross-sectional networks should be made considering recent concerns on the replicability of networks[53]. Indeed, the network literature will benefit from insight into whether reported network structures replicate across data sets before clinical inferences can be made. First attempts to study replicability have thus far yielded conflicting results (see[53-55]). Furthermore, cross-sectional networks face the limitation that they may only reveal the co-occurrence of symptoms, not how they follow each other over time[56], and it has been questioned whether cross-sectional group-level associations can be generalized to the level of the individual[57]. Our findings therefore need to be replicated, preferably by designs that offer more insight into the temporal dynamics of relationships between symptoms at the individual level[58]. Specifically, such temporal data enables investigation into changes in which symptoms precede changes in other symptoms. Consequently, one can be more certain about the direction of change, allowing the identification of symptoms that precede change in other symptoms[13]. We therefore consider this study an exploration of the possibilities of applying the network approach to uncover novel insights into the way antidepressant treatment works. Follow-up studies are required to examine whether central symptoms and changes over time generalize to the dynamics of symptoms taking place within individual patients.

To our knowledge, this was the first study to examine changes in symptom interrelations and their centrality after 8 weeks of SSRI treatment. Our findings demonstrated an increase in the interrelations between depressive symptoms after SSRI treatment, which adds to previous research reporting increased connectivity after symptom severity decreased. The current findings highlight the potential of the network approach in providing insight into the working mechanisms of SSRIs on the interrelations between depressive symptoms, which should be explored in future research.

References

1. NICE. *Depression: The treatment and management of depression in adults.* Updated ed. www.nice.org.uk/CG90: NICE clinical guideline 90; 2009.
2. American Psychiatric Association. *Practice guideline for the treatment of patients with major depressive disorder.* 3 ed: American Psychiatric Association; 2010.
3. Khan A, Brown WA. Antidepressants versus placebo in major depression: An overview. *World Psychiatry.* 2015;14(3):294.
4. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine.* 2008;5(2):e45.
5. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *NEnglJMed.* 2008;358(3):252.
6. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: Current status of research. *PsychotherPsychosom.* 2010;79(5):267.
7. Borsboom D. A network theory of mental disorders. *World Psychiatry.* 2017;16(1):5.
8. Borsboom D, Cramer AOJ, Schmittmann VD, Epskamp S, Waldorp LJ. The small world of psychopathology. *PLoS One.* 2011;6(11).
9. Fried EI, Nesse RM. Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Med.* 2015;13(1):72.
10. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks.* 2010;32(3):245.
11. van Borkulo C, Borsboom D, Epskamp S, Blanken TF, Boschloo L, Schoevers RA, Waldorp LJ. A new method for constructing networks from binary data. *Sci Rep.* 2014;4:5918.
12. Fried EI, Epskamp S, Nesse RM, Tuerlinckx F, Borsboom D. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders.* 2016;189:314.
13. Bringmann LF, Lemmens LHJM, Huibers MJH, Borsboom D, Tuerlinckx F. Revealing the dynamic network structure of the Beck Depression Inventory-II. *PsycholMed.* 2015;45(4):747.
14. van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, Schoevers RA. Association of symptom network structure with the course of longitudinal depression. *JAMA Psychiatry.* 2015;72(12):1219.
15. Pe ML, Kircanski K, Thompson RJ, Bringmann LF, Tuerlinckx F, Mestdagh M, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Kuppens P, Gotlib IH. Emotion-network density in major depressive disorder. *Clinical Psychological Science.* 2015;3(2):292.
16. Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, Borsboom D. Measuring depression over time... Or not? *PsycholAssess.* 2016;28.
17. Snippe E, Viechtbauer W, Geschwind N, Klippel A, de Jonge P, Wichers M. The impact of treatments for depression on the dynamic network structure of mental states: Two randomized controlled trials. *Sci Rep.* 2017;7:46523.
18. Madhoo M, Levine SZ. Network analysis of the Quick Inventory of Depressive Symptomatology: Reanalysis of the STAR* D clinical trial. *European Neuropsychopharmacology.* 2016;26(11):1768.
19. Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL, Jacobson NS. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *JConsultClinPsychol.* 2006;74(4):658.
20. DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, O'Reardon JP, Lovett ML, Gladis MM, Brown LL, Gallop R. Cognitive therapy vs medications in the treatment of moderate to severe depression. *ArchGenPsychiatry.* 2005;62(4):409.
21. First MB, Gibbon M, Spitzer RL, Williams JBW. User's guide for the structured clinical interview for DSM-IV axis I Disorders—Research version. *New York: Biometrics Research Department, New York State Psychiatric Institute.* 1996.
22. Hamilton M. A rating scale for depression. *JNeurolNeurosurgPsychiatry.* 1960;23:56.

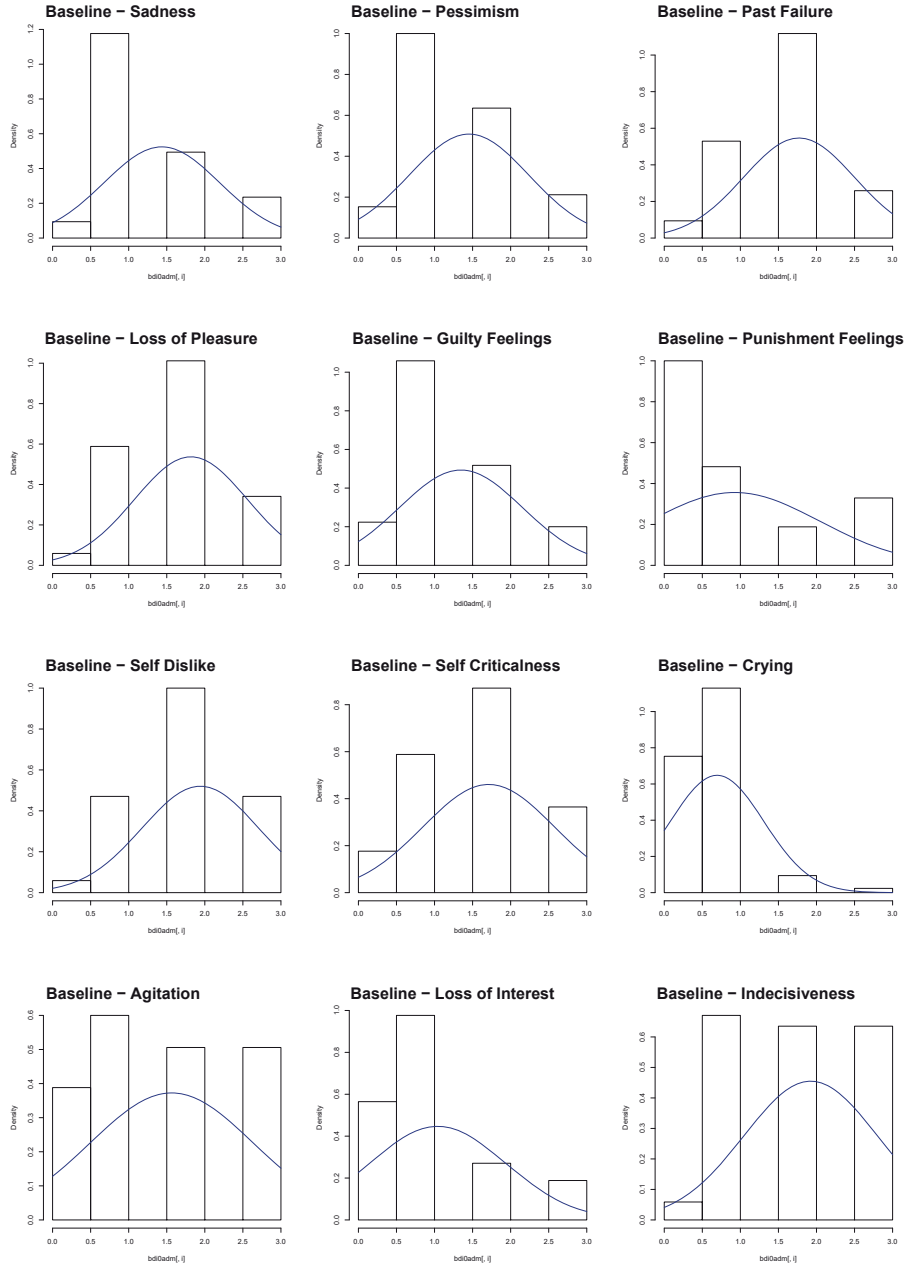
23. Beck AT, Steer RA, Brown GK. *Manual for the BDI-II.* San Antonio, TX: Psychological Corporation; 1996.
24. Fawcett J, Epstein P, Fiester SJ, Elkin I, Autry JH. Clinical management--imipramine/placebo administration manual. NIMH Treatment of Depression Collaborative Research Program. *PsychopharmacolBull.* 1987;23(2):309.
25. Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review.* 1988;8(1):77.
26. Epskamp S, Fried EI. A tutorial on regularized partial correlation networks. *PsycholMethods.* 2018.
27. Santos H, Fried EI, Asafu-Adjei J, Ruiz RJ. Network structure of perinatal depressive symptoms in Latinas: Relationship to stress and reproductive biomarkers. *ResNursHealth.* 2017;40(3):218.
28. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software.* 2012;48:1.
29. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society.* 1996;B58:267.
30. Foygel R, Drton M. Extended Bayesian Information Criteria for Gaussian Graphical Models. *Advances in Neural Information Processing Systems.* 2010;23:2020.
31. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Software: Practice and Experience.* 1991;21(11):1129.
32. Terluin B, de Boer MR, de Vet HCW. Differences in connection strength between mental symptoms might be explained by differences in variance: Reanalysis of network data did not confirm staging. *PLoS One.* 2016;11(11):e0155205.
33. Golino HF, Epskamp S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One.* 2017;12(6):e0174035.
34. Van Borkulo CD, Boschloo L, Kossakowski J, Tio P, Schoevers R, Borsboom D, Waldorp LJ. Comparing network structures on three aspects: A permutation test. *Manuscript submitted for publication.* 2016.
35. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods.* 2017:1.
36. Haslbeck JMB, Fried EI. How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. *PsycholMed.* 2017:1.
37. Haslbeck J, Waldorp LJ. Structure estimation for mixed graphical models in high-dimensional data. *arXiv preprint arXiv:151005677.* 2015.
38. Fokkema M, Smits N, Kelderman H, Cuijpers P. Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *PsycholAssess.* 2013;25(2):520.
39. Quilty LC, Robinson JJ, Rolland JP, Fruyt FD, Rouillon F, Bagby RM. The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. *International Journal of Methods in Psychiatric Research.* 2013;22(3):175.
40. Steer RA, Ball R, Ranieri WF, Beck AT. Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *JClinPsychol.* 1999;55(1):117.
41. Huang C, Chen J-H. Meta-Analysis of the Factor Structures of the Beck Depression Inventory–II. *Assessment.* 2015;22(4):459.
42. Boschloo L, van Borkulo CD, Borsboom D, Schoevers RA. A Prospective Study on how Symptoms in a Network Predict the Onset of Depression. *PsychotherPsychosom.* 2016;85(3):183.
43. Andrews PW, Bharwani A, Lee KR, Fox M, Thomson JA. Is serotonin an upper or a downer? The evolution of the serotonergic system and its role in depression and the antidepressant response. *Neuroscience & Biobehavioral Reviews.* 2015;51:164.
44. Durisko Z, Mulsant BH, Andrews PW. An adaptationist perspective on the etiology of depression. *Journal of Affective Disorders.* 2015;172:315.
45. Kroeze R, van der Veen DC, Servaas MN, Bastiaansen JA, Oude Voshaar R, Borsboom D, Riese H. Personalized feedback on symptom dynamics of psychopathology: a proof-of-principle study. *Journal for Person-Oriented Research.* 2017;3(1):1.
46. Fried EI, Cramer AOJ. Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science.* 2017:1.
47. McMakin DL, Olino TM, Porta G, Dietz LJ, Emslie G, Clarke G, Wagner KD, Asarnow JR, Ryan

ND, Birmaher B, Shamseddeen W, Mayes T, Kennard B, Spirito A, Keller M, Lynch FL, Dickerson JF, Brent DA. Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment-resistant depression. *JAmAcadChild AdolescPsychiatry.* 2012;51(4):404.

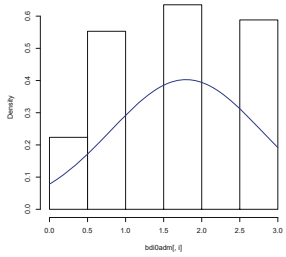
48. Yee A, Chin SC, Hashim AHB, Harbajan Singh MKA, Loh HS, Sulaiman AH, Ng CG. Anhedonia in depressed patients on treatment with selective serotonin reuptake inhibitor anti-depressant - A two-centered study in Malaysia. *IntJPsychiatry ClinPract.* 2015;19(3):1.
49. Vrieze E, Demyttenaere K, Bruffaerts R, Hermans D, Pizzagalli DA, Sienaert P, Hompes T, de Boer P, Schmidt M, Claes S. Dimensions in major depressive disorder and their relevance for treatment outcome. *Journal of Affective Disorders.* 2014;155:35.
50. Hyman SE, Nestler EJ. Initiation and adaptation: A paradigm for understanding psychotropic drug action. *AmJPsychiatry.* 1996;153(2):151.
51. Korte SM, Prins J, Krajnc AM, Hendriksen H, Oosting RS, Westphal KG, Korte-Bouws GAH, Olivier B. The many different faces of major depression: It is time for personalized medicine. *EurJPharmacol.* 2015;753:88.
52. Muthén BO. Factor structure in groups selected on observed scores. *BrJMathStatPsychol.* 1989;42(1):81.
53. Forbes MK, Wright AGC, Markon KE, Krueger RF. Evidence that psychopathology symptom networks have limited replicability. *JAbnormPsychol.* 2017;126(7):969.
54. Borsboom D, Fried EI, Epskamp S, Waldorp LJ, van Borkulo CD, van der Maas HLJ, Cramer AOJ. False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *JAbnormPsychol.* 2017;126(7):989-999.
55. Fried EI, Eidhof MB, Palic S, Constantini G, Huisman-van Dijk HM, Bockting CLH, Engelhard I, Armour C, Nielsen ABS, Karstoft KI. Replicability and generalizability of PTSD networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science.* 2018;6(3):335-351.
56. Bos FM, Snippe E, de Vos S, Hartmann JA, Simons CJP, van der Krieke L, de Jonge P, Wichers M. Can we jump from cross-sectional to dynamic interpretations of networks? Implications for the network perspective in psychiatry. *PsychotherPsychosom.* 2017;86(3):175-177.
57. Bos EH, Wanders RBK. Group-level symptom networks in depression. *JAMA Psychiatry.* 2016;73(4):411.
58. Wichers M. The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *PsycholMed.* 2014;44(7):1349.

Appendix

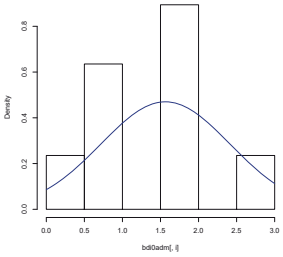
Figure 2.4. The distribution of all individual items of the Beck Depression Inventory II at baseline.



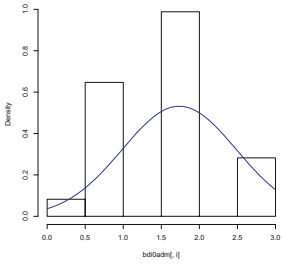
Baseline − Worthlessness



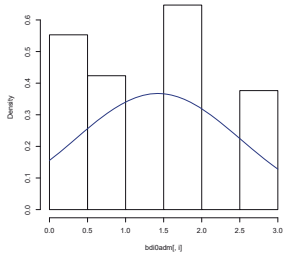
Baseline − Loss of Energy



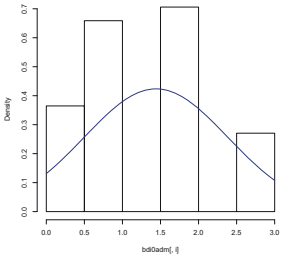
Baseline − Change in Sleeping Pattern



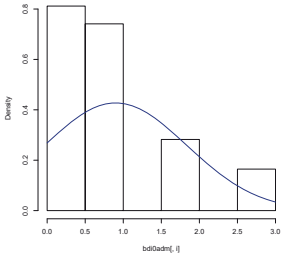
Baseline − Irritability



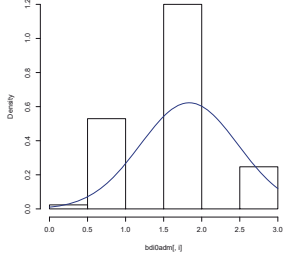
Baseline − Change in Appetite



Baseline − Concentration Difficulty



Baseline − Tiredness/Fatigue



Baseline − Loss of Interest in Sex

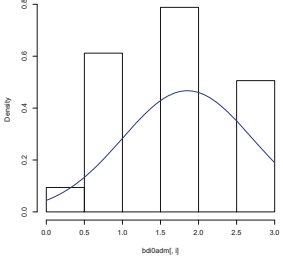
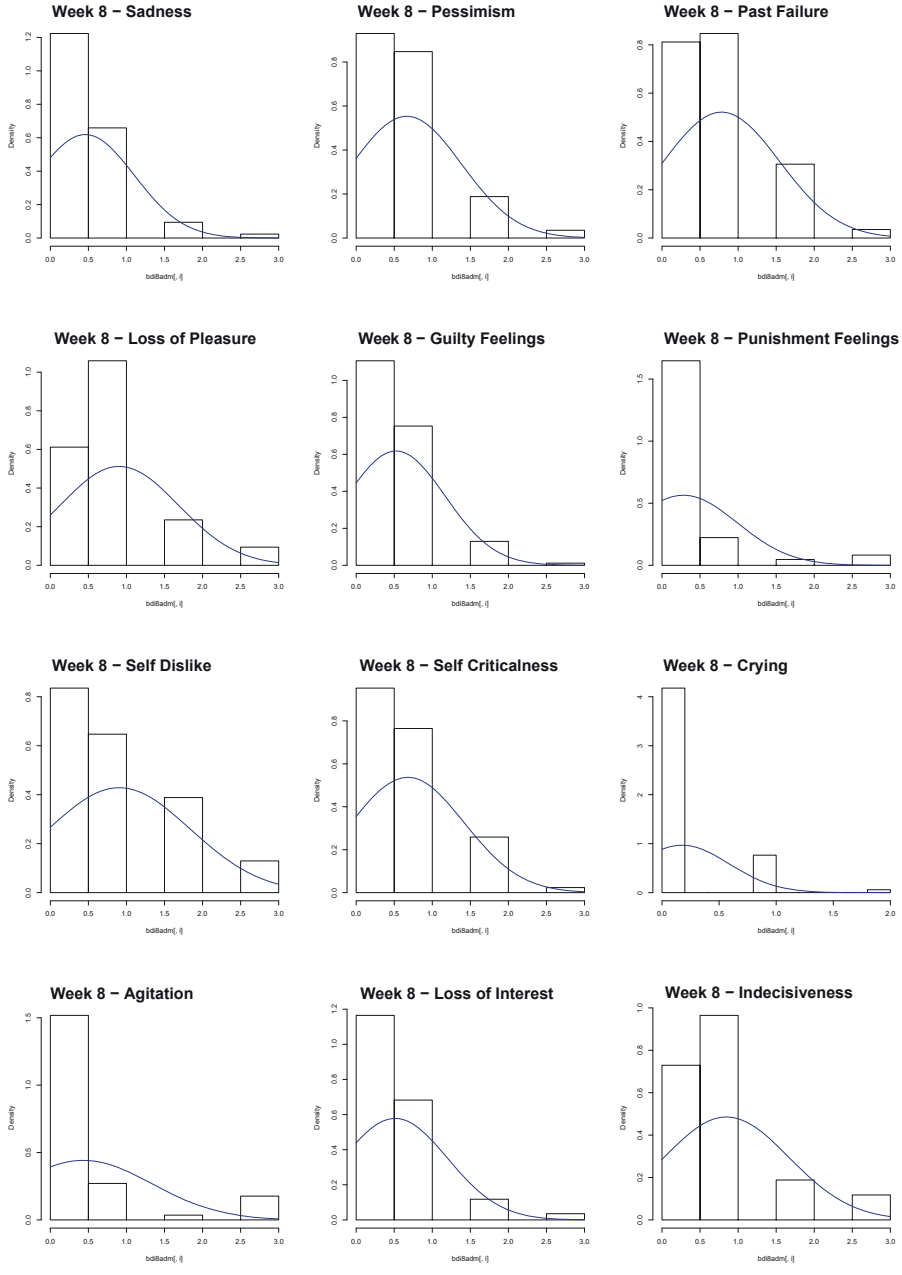
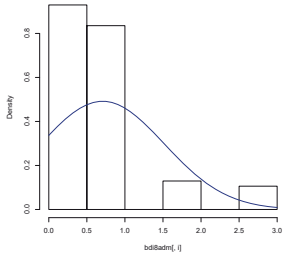


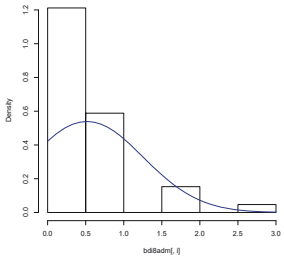
Figure 2.5. The distribution of all individual items of the Beck Depression Inventory II at week 8.



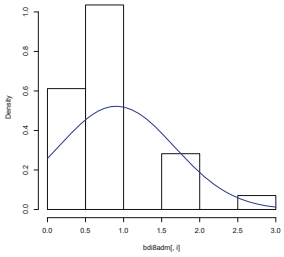
Week 8 − Worthlessness



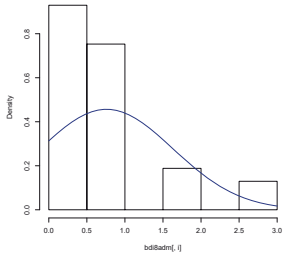
Week 8 − Loss of Energy



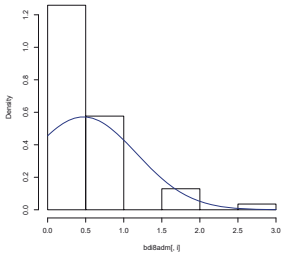
Week 8 − Change in Sleeping Pattern



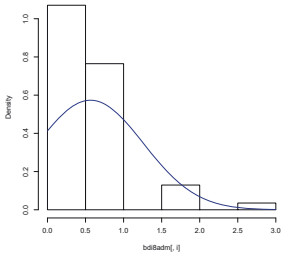
Week 8 − Irritability



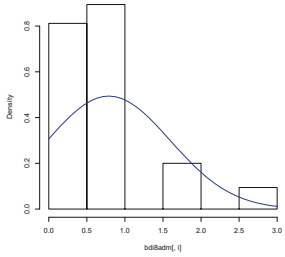
Week 8 − Change in Appetite



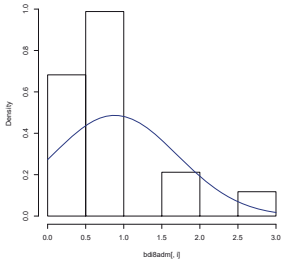
Week 8 − Concentration Difficulty

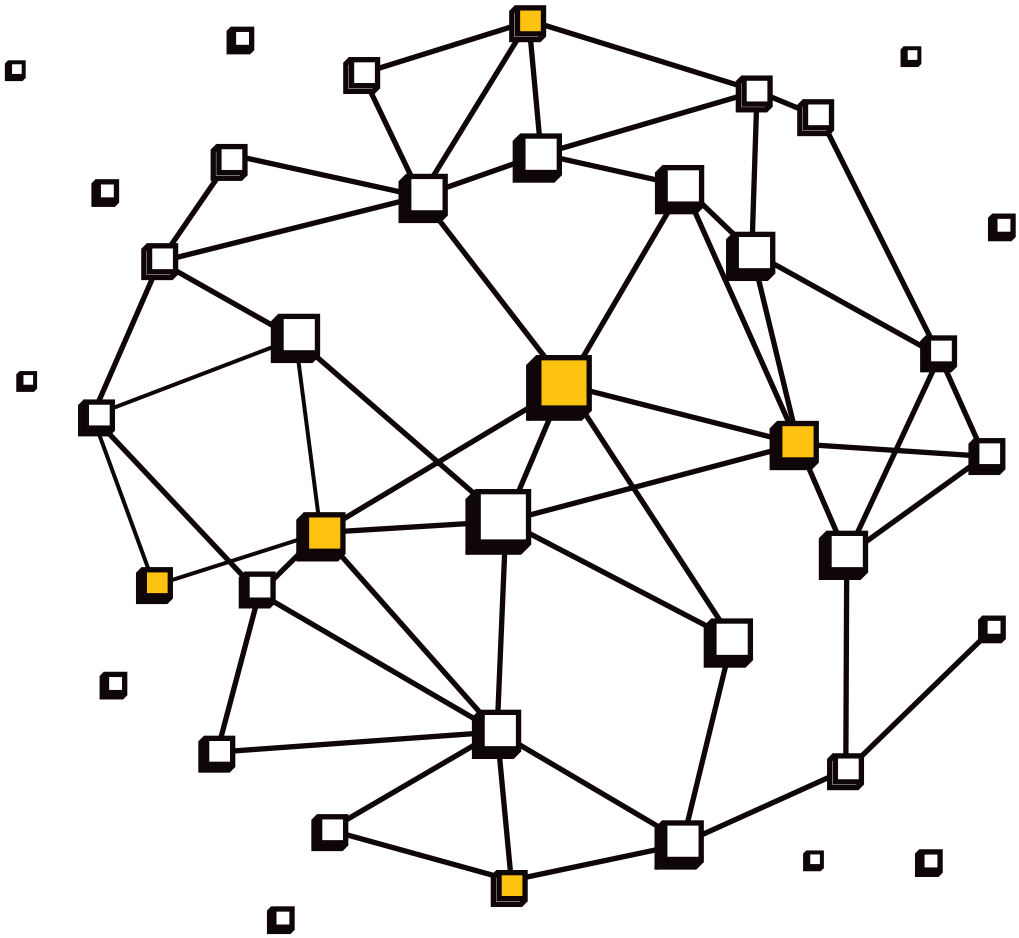


Week 8 − Tiredness/Fatigue



Week 8 − Loss of Interest in Sex





Chapter 3

Exploring the emotional dynamics of subclinically depressed individuals with and without anhedonia: An experience sampling study

Published as:

Bos FM*, Blaauw FJ*, Snippe E, van der Krieke L, de Jonge P, Wichers M. Exploring the emotional dynamics of subclinically depressed individuals with and without anhedonia: An experience sampling study. *Journal of Affective Disorders.* 2018;228:186-193.

*authors contributed equally

Abstract

Background. Anhedonia has been linked to worse prognosis of depression. The present study aimed to construct personalized models to elucidate the emotional dynamics of subclinically depressed individuals with versus without symptoms of anhedonia.

Methods. Matched subclinically depressed individuals with and without symptoms of anhedonia (N = 40) of the HowNutsAreTheDutch sample completed three experience sampling methodology assessments per day for 30 days. For each individual, the impact of physical activity, stress experience, and high/low arousal PA/NA on each other was estimated through automated impulse response function analysis (IRF). These individual IRF associations were combined to compare anhedonic versus non-anhedonic individuals.

Results. Physical activity had low impact on affect in both groups. In non-anhedonic individuals, stress experience increased NA and decreased PA and physical activity more strongly. In anhedonic individuals, PA high arousal showed a diminished favorable impact on affect (increasing NA/stress experience, decreasing PA/physical activity). Finally, large heterogeneity in the personalized models of emotional dynamics were found.

Limitations. Stress experience was measured indirectly by assessing level of distress; the timeframe in between measurements was relatively long with 6h; and only information on one of the two hallmarks of anhedonia, loss of interest, was gathered.

Conclusions. Our results suggest different pathways of emotional dynamics underlie depressive symptomatology. Subclinically depressed individuals with anhedonic complaints are more strongly characterized by diminished favorable impact of PA high arousal and heightened NA reactivity, whereas subclinically depressed individuals without these anhedonic complaints seem more characterized by heightened stress reactivity. The automatically generated personalized models may offer patient-specific insights in emotional dynamics, which may show clinical relevance.

Introduction

Major depressive disorder (MDD) is a highly disabling disorder characterized by considerable heterogeneity[1]. It has been suggested that anhedonia, one of the two core symptoms of MDD[2], constitutes a distinct endophenotype of MDD[3,4]. Anhedonia is the inability to experience interest in or pleasure from activities usually found enjoyable and is reported by roughly one third of MDD patients[5]. It has been linked to poorer prognosis of MDD[6,7], poorer treatment response[8-10], and increased risk of suicide[11].

Despite its debilitating influence, relatively little is known about underlying mechanisms of anhedonia. In order to bridge this gap in our knowledge, we need to find better and more direct ways to study the differences between subclinically depressed individuals with and without anhedonic symptoms. By studying individuals with subclinical levels of symptoms, mechanisms that underlie the future development of clinical symptoms and disorders may be uncovered. Indeed, the dimensional perspective on psychopathology assumes that the underlying mechanisms for subclinical and clinical levels of depression and anhedonia are at least partially shared[12]. Further, such an approach requires a translation from abstract measures of anhedonia (e.g. in the laboratory) to specific emotional responses to situations in daily life. Such knowledge potentially helps in targeting anhedonia more directly and effectively.

The hypothesis that anhedonia is a distinct MDD endophenotype[3] suggests that different daily life dynamics underlie depressive symptoms in individuals with anhedonic symptoms versus those without. Given that anhedonia is characterized by less enjoyment of activities, subclinically depressed individuals with anhedonic symptoms might benefit less from pleasurable behaviors, as indicated by smaller increases in positive affect (PA) and smaller reductions in negative affect (NA). Physical activity might be such a pleasurable behavior, since it is generally viewed as a behavior that increases PA and is often advised to depressed patients by clinicians[13]. In anhedonic individuals, we would expect that the favorable impact of physical activity on affect is diminished. Further, anhedonia has been related to

higher perceived stress[14] and the experience of stress has been found to worsen hedonic capacity and responsiveness to positive events[3]. We would therefore expect that the experience of stress exerts a stronger unfavorable impact on affect (i.e., in reducing PA and increasing NA) for individuals with anhedonia.

Previous research has primarily focused on group-level results, e.g. mean associations that do not necessarily represent associations of individuals[15,16]. Research so far may thereby have overlooked important heterogeneity in emotional dynamics. MDD is highly heterogeneous[1] and the effects of physical activity have been found to vary widely across individuals[17-19]. Thus, in contrast to previous research, we will examine mechanisms of anhedonia in daily life on a case-by-case basis so as to account for and gain insight into this heterogeneity. Based on individual models, we will discern more general patterns. Such a personalized approach may also have relevance for clinical practice in understanding emotional dynamics of individual patients.

Aims of the study

The present study aimed to examine emotional dynamics in the flow of daily life in subclinically depressed individuals with versus without anhedonic symptoms. Specifically, we will study the possibly differential impact of physical activity and stress experience on positive and negative affect in subclinically depressed individuals with versus without anhedonic symptoms. Such an investigation in a general population sample can be the starting point to investigate micro-level dynamics that may underlie the future development of clinical symptoms. These dynamics can be optimally measured through the ecologically valid experience sampling method (ESM)[20]. With ESM, individuals can record their affect, stress level, and level of physical activity multiple times a day in their own environments[21,22], to prospectively examine emotional responses to physical activity and the experience of stress. We will use an advanced extension of vector autoregressive (VAR) modelling called impulse-response function (IRF) analysis[23,24] to compare the impact of a hypothetical increase in physical activity or stress experience on affect for both subgroups. To this end, we used automated impulse-response analysis (AIRA), a

novel and sophisticated *R*-package that automates IRF analyses[25]. AIRA estimates network models for each individual, after which these models can be combined into aggregated models to compare the two groups. This approach accounts for and offers insight into individual differences in daily dynamics and depressogenic mechanisms.

Method

Participants

Participants are 629 individuals from the general Dutch population who participated in an ESM protocol of the study "HowNutsAreTheDutch?" (Dutch: HoeGekIsNL?) between May 22nd, 2014 and December 13th, 2014 (end of the first-year wave of the website[26,27]. In order to be included, participants had to indicate they (1) were at least 18 years of age, (2) could start with the study within five days (3) possessed a smartphone with a mobile internet connection, (4) were not engaged in shift work, (5) did not anticipate a major disruption of daily routines within the study period, (6) were aware that their results would be useless if too many assessments were missed, and (7) consented to having their anonymous data used for research purposes.

For the present paper, we selected individuals who (1) were at least mildly depressed, as indicated by a Quick Inventory for Depressive Symptomatology (QIDS-SR[28]) score of 6 or higher, and (2) completed at least 67 (75%) of the diary assessments (for a flow-chart, see Appendix). Given that anhedonia is defined as loss of interest or pleasure, we used the QIDS-SR item on loss of interest ("I notice that I am less interested in people or activities") as a proxy for anhedonia. Although this is a single item, this item seems to be a relatively valid measure of anhedonia given its high correlates to anhedonia items of Depression and Anxiety Stress Scales (DASS[29]). In the HowNutsAreTheDutch sample (N=8575), the QIDS-SR loss of interest item correlated 0.74 with the more general loss of interest item of the DASS[30] and 0.66-0.70 with the three DASS items on anhedonia (on enjoyment,

experience of positive affect, and enthusiasm). Participants who endorsed this item (scored at least '1') are henceforth referred to as 'anhedonic', participants who reported no loss of interest as 'non-anhedonic'. All anhedonic individuals were matched to non-anhedonic individuals based on their QIDS-SR score, sex, and education level, respectively. This resulted in 50 matched individuals, 25 in each group.

Measures

Depressive symptoms

Depressive symptoms at the time of study entry were assessed through the QIDS-SR, a 16-item self-report questionnaire. The QIDS-SR covers all depressive symptoms as described by the DSM and shows adequate validity and reliability[28].

Diary items

Participants completed 43 items on affect, behavior, cognitions, and activities through an electronic diary three times a day for 30 consecutive days, resulting in a maximum of 90 assessments. These assessments were completed online; links to the assessments were sent via text messages. Participants had one hour to complete an assessment after receiving the notification. In the present sample, on average 76 diary assessments (*SD* = 5.3) were completed per participant. Diary items were rated on visual analogue scales (VAS) ranging from 0 ('not at all') to 100 ('very much'). To accommodate the two dimensions of affect, valence and arousal [31], four affective variables were constructed. The mean score of the emotional items 'energetic', 'enthusiastic', and 'cheerful' was taken to reflect positive affect (PA) high-arousal. PA low-arousal was assessed by 'relaxed', 'content', and 'calm'. Likewise, negative affect (NA) high-arousal was assessed by 'anxious', 'nervous', and 'irritable', and NA low-arousal by 'gloomy', dull', and 'tired'. Participants further indicated their level of physical activity of the last six hours ('since the last measurement I was physically active', item no 41) and subjective experience of stress ('I am upset', item no 25).

Analyses

Personalized models of the dynamics between physical activity, stress experience, and affect in subclinically depressed individuals with versus without anhedonic complaints were estimated. Based on these models, we first examined our hypotheses on the potentially differential impact of activity and stress experience on the affective variables in subclinically depressed individuals with versus without anhedonic symptoms. Next, we explored other relevant differences in emotional dynamics between the two groups. Finally, we illustrated the individual differences in emotional dynamics.

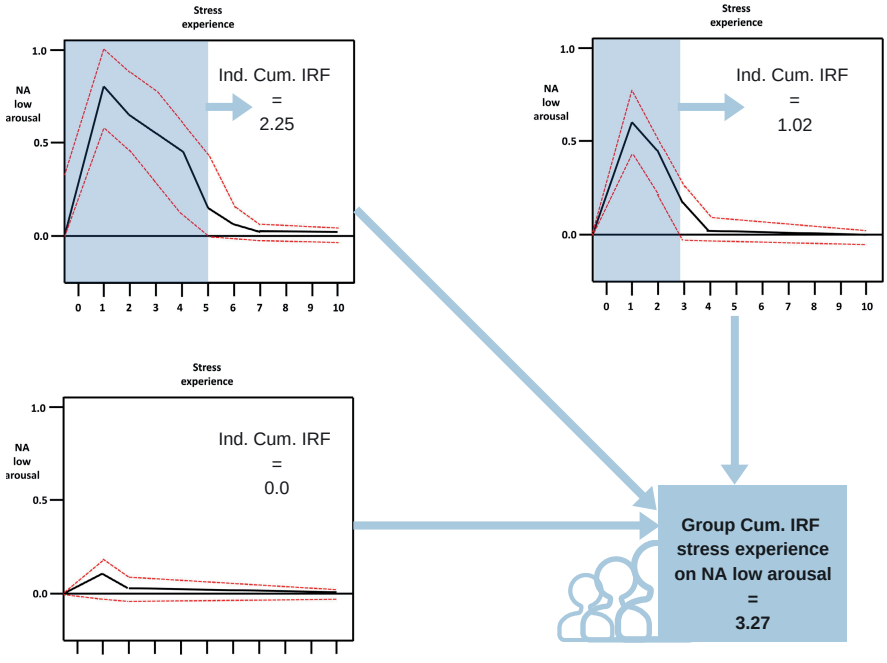
First, we fitted a vector autoregression (VAR) model for every participant. In a VAR model, each variable is regressed on its own lagged values (autocorrelation) as well as the lagged values of the other variables[23], resulting in a set of regression coefficients for each individual. As such, one can examine the dynamic effect of the variables on each other (e.g. the effect of physical activity at one moment in time (t) on high-arousal positive affect at the next moment in time (t+1)). Given that the dynamic effects of physical activity, stress experience, and affect on each other were expected to occur within the six hours between the measurement points, and to reduce risk of overparametrization of the VAR-models, a lag of 1 was chosen for all cross-correlations[23]. For all autocorrelations, a lag of 1 or 2 was chosen dependent on the most optimal model for the participant. The VAR models were fit using the R-package AutovarCore[32]. AutovarCore is an algorithm to automatically estimate vector autoregression (VAR) models for a participant. In our VAR models, we included six endogenous variables: PA high and low arousal, NA high and low arousal, physical activity, and stress experience. Measurement moment was included as an exogenous variable, weekday and study day were modeled if they improved the model for an individual, as well as linear and quadratic trends. Missing data was imputed using the R-package Amelia II, which is a well-validated approach to missing data handling[33]. AutovarCore automatically checks assumptions for a VAR model of stability, serial independence, homoscedasticity, and normality of the residuals[23,32]; which resulted in 42 valid models (no anhedonia: 22; anhedonia: 20). Two individuals could no longer be matched, resulting in a final

sample of 40 individuals; 20 in each group.

Second, our VAR models were analyzed automatically by means of impulse response function analysis (IRF) using the *R*-package AIRA (automated impulse response analysis)[25]. VAR models provide an overview of how the modeled time lagged variables are related to each other. However, it is the behavior of the combination of the coefficients (i.e., the model as a whole) that describes the dynamicity of the model[23]. One way to analyze the model as a whole is by simulating a sudden increase in one variable (or 'shock' in IRF parlance), and investigating how this sudden increase is propagated through the model, i.e., how it affects the other variables both in terms of duration and magnitude. This is known as IRF analysis. IRFs show the hypothetical change in a variable over a horizon of several time points in response to an isolated shock in one of the other variables (see Figure 3.1 for an example). AIRA performs IRF analysis on each of the variables in the VAR model in isolation to determine how much each variable affects the other variables.

For every person and every association between variables, we calculated cumulative IRFs[17], which were constructed by summing all impacts within the horizon of ten time points that are significant (i.e., the confidence interval does not include zero for that particular step, see Figure 3.1). These individual cumulative IRFs reflect the impact of all variables on each other over time, which was then visualized in 40 individual network models, one for each participant. Next, we constructed *group* cumulative IRFs by summing all individual cumulative IRFs for each association, to enable us to compare the non-anhedonic versus the anhedonic group. This was done separately for individual positive cumulative IRFs and individual negative cumulative IRFs, because combining both would cancel out present associations. Thus, the higher the positive or negative group cumulative IRF, the stronger the impact of one variable on another.

Figure 3.1. Example of how individual cumulative impulse response functions (IRFs) and group cumulative IRFs are constructed. This figure shows the impact of an impulse in stress experience on NA low arousal, over a horizon of 10 time points, for three hypothetical individuals. Dashed lines indicate the confidence intervals around the IRF. For the first individual, stress experience first increases NA low arousal at step 1-5 (grey transparent area), after which the impact of stress experience on PA high arousal is no longer significant (from step 6 onwards). To construct the individual cumulative IRF for the impact of stress experience on NA low arousal for this individual, the values of step 1-5 are summed. To construct the group cumulative IRF for the impact of stress experience on NA low arousal, the individual cumulative IRFs for all individuals are summed.



We used three approaches to compare emotional dynamics between the non-anhedonic group and the anhedonic group as described above. First, we compared the group cumulative IRFs for each association. Such a comparison would indicate whether the impact of physical activity and stress experience is *stronger* in one of the two groups. Second, we compared the *number* of individuals who showed a given IRF association by examining the individual models. Third, we compared the

importance of the variables as node in the network by comparing network centrality (node strength) indices between the two groups for each variable. Strength centrality is the sum of the connection strength values (based on the cumulative IRF scores) of all IRF associations that a given variable has within the network[34]. Thus, a high strength centrality of a variable indicates that this variable has a strong impact on other variables or is impacted by many variables. We focused on "outstrength" centrality, which is the total impact of a given variable on all other variables in the network (sum of outgoing cumulative IRF associations). We further examined whether each variable impacted other variables in a favorable manner (resulting in an increase of PA and activity or decrease of NA and stress) or unfavorable manner (resulting in a decrease in PA and activity or an increase in NA and stress).

Finally, we explored individual differences in emotional dynamics displayed in the individual network models. We will depict two of these individual models to illustrate existing individual emotional dynamics and how the use of such personalized networks may possibly inform on choice of intervention type.

Results

Mean levels of affect, stress and activity

Multilevel analyses indicated no significant differences in mean levels of affect, physical activity, and stress experience between the anhedonic group and the non-anhedonic group over the 30-day study period (for the means, standard deviations, and p-values, see Appendix). As the groups were matched, level of depression was the same in both groups (mean QIDS score = 9.1; range 6-17), as well as the distribution of gender (19 females and 1 male), and education level (non-anhedonic group: N = 17 with higher education; anhedonic group: N=18 with higher education). Groups were of similar age (non-anhedonic: M = 43.6, SD = 13.2; anhedonic: M = 39.5, SD = 11.7, p of difference =.302).

Impact of physical activity and stress experience

Table 3.1 and Figure 3.2 show the *strength* of the IRF associations through the group cumulative IRFs, which are composed of the individual cumulative IRFs, split into positive and negative associations for each possible association within the network. It also shows the range in individual cumulative IRFs. Further, it shows the *number* of individuals who showed a particular significant IRF association. Table 3.2 shows the *importance* of each of the variables in the network.

In both groups, the impact of physical activity on affect was *weak*, as shown by the small positive and negative group cumulative IRFs and the small *number* of individuals with significant IRFs (see Table 3.1). Further, the groups did not differ on the *importance* of physical activity in the network (non-anhedonic: outstrength = 0.98; anhedonic: outstrength =1 .04). In both groups, physical activity seemed to have a more unfavorable (non-anhedonic: unfavorable outstrength = 0.83; anhedonic: unfavorable outstrength = 0.61) than favorable impact (non-anhedonic: favorable outstrength = 0.15; anhedonic: unfavorable outstrength = 0.43) on affect and stress experience (see Table 3.2).

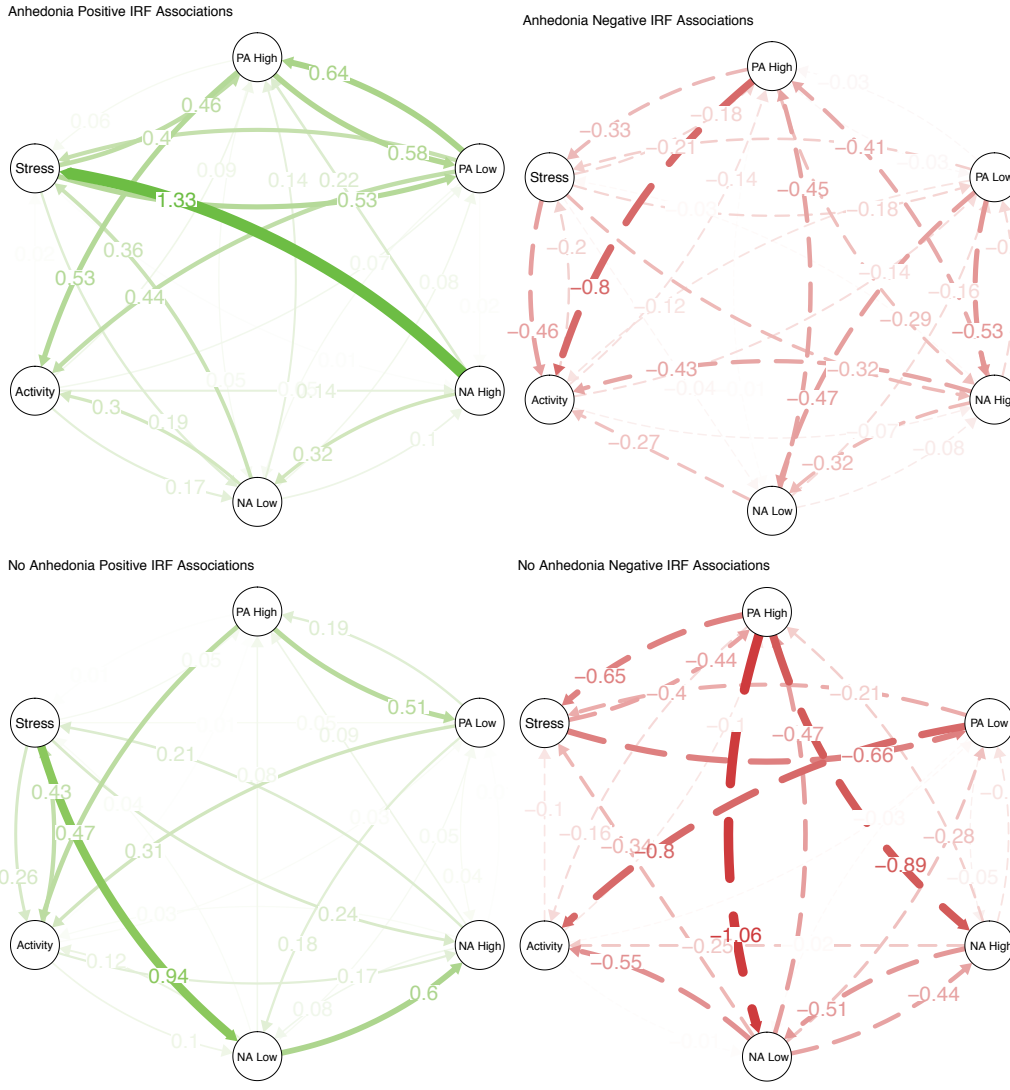
Table 3.1. Group cumulative IRF associations per group (*strength*), the *number* of individuals showing a given association significantly, and the range in individual cumulative IRFs.

		No anhedonia						Anhedonia					
		Positive IRF associations			Negative IRF associations			Positive IRF associations			Negative IRF associations		
Effect of	**On**	**GC IRF**	**N**	**Range**	**GC IRF**	**N**	**Range**	**GC IRF**	**N**	**Range**	**GC IRF**	**N**	**Range**
PA high arousal	PA low arousal	0.51	5	0.05 - 0.25	0.00	0	-	0.58	2	0.05 - 0.53	-0.03	1	-0.03
	NA high arousal	0.00	0	-	-0.89	4	-0.37 - -0.10	0.00	0	-	-0.29	4	-0.13 - -0.004
	NA low arousal	0.00	0	-	-1.06	4	-0.40 - -0.12	0.05	1	0.05	-0.01	1	-0.01
	Physical activity	0.47	2	0.21 - 0.26	-0.16	2	-0.13 - -0.03	0.53	3	0.01 - 0.43	-0.80	2	-0.74 - -0.06
	Stress experience	0.01	1	0.01	-0.65	7	-0.26 - -0.01	0.06	2	0.002 - 0.05	-0.33	2	-0.19 - -0.14
PA low arousal	PA high arousal	0.19	2	0.02 - 0.17	0.00	0	-	0.64	3	0.07 - 0.33	-0.03	1	-0.03
	NA high arousal	0.04	1	0.04	-0.05	1	-0.05	0.02	1	0.02	-0.53	3	-0.33 - -0.06
	NA low arousal	0.18	2	0.05 - 0.13	-0.02	1	-0.02	0.05	1	0.05	-0.47	4	-0.21 - -0.06
	Physical activity	0.31	1	0.31	-0.80	2	-0.78 - -0.02	0.44	2	0.13 - 0.31	-0.12	2	-0.08 - -0.03
	Stress experience	0.00	0	-	-0.40	2	-0.38 - -0.02	0.40	1	0.4	-0.21	3	-0.14 - -0.03
NA high arousal	PA high arousal	0.09	2	0.004 - 0.09	-0.21	2	-0.19 - -0.02	0.22	1	0.22	-0.41	3	-0.27 - -0.002
	PA low arousal	0.01	1	0.01	-0.12	2	-0.11 - -0.008	0.00	0	-	-0.17	1	-0.17
	NA low arousal	0.08	2	0.01 - 0.08	-0.51	4	-0.41 - -0.009	0.32	1	0.32	-0.32	3	-0.13 - -0.06
	Physical activity	0.03	1	0.03	-0.25	1	-0.25	0.00	0	-	-0.43	2	-0.22 - -0.21
	Stress experience	0.21	3	0.02 - 0.13	0.00	0	-	1.33	6	0.08 - 0.39	-0.03	1	-0.03

		No anhedonia						Anhedonia					
		Positive IRF associations			Negative IRF associations			Positive IRF associations			Negative IRF associations		
Effect of	**On**	**GC IRF**	**N**	**Range**	**GC IRF**	**N**	**Range**	**GC IRF**	**N**	**Range**	**GC IRF**	**N**	**Range**
NA low arousal	PA high arousal	0.08	2	0.001 - 0.08	-0.47	4	-0.30 - -0.02	0.14	2	0.007 - 0.14	-0.45	3	-0.33 - -0.05
	PA low arousal	0.05	2	0.008 - 0.04	-0.28	4	-0.11 - -0.03	0.08	1	0.08	-0.16	2	-0.11 - -0.05
	NA high arousal	0.60	5	0.07 - 0.17	-0.44	2	-0.39 - -0.05	0.10	1	0.1	-0.08	1	-0.08
	Physical activity	0.12	2	0.002 - 0.12	-0.55	3	-0.25 - -0.09	0.30	1	0.3	-0.27	1	-0.27
	Stress experience	0.04	1	0.04	-0.34	2	-0.33 - -0.01	0.36	2	0.04 - 0.32	0.00	0	-
Physical activity	PA high arousal	0.01	2	0.002 - 0.007	-0.10	2	-0.09 - -0.01	0.09	3	0.002 - 0.05	-0.14	1	-0.14
	PA low arousal	0.03	1	0.03	-0.03	4	-0.01 - -0.003	0.07	2	0.01 - 0.06	-0.14	2	-0.11 - -0.04
	NA high arousal	0.17	3	0.005 - 0.14	0.00	0	-	0.14	4	0.02 - 0.04	-0.07	2	-0.05 - -0.02
	NA low arousal	0.10	3	0.02 - 0.05	-0.01	1	-0.01	0.17	2	0.05 - 0.12	0.00	0	-
	Stress experience	0.43	5	0.002 - 0.30	-0.10	2	-0.10 - -0.00008	0.02	2	0.000007 - 0.02	-0.20	3	-0.09 - -0.04
Stress experience	PA high arousal	0.05	2	0.003 - 0.05	-0.44	4	-0.35 - -0.003	0.46	2	0.05 - 0.41	-0.18	4	-0.12 - -0.004
	PA low arousal	0.05	1	0.05	-0.66	3	-0.46 - -0.04	0.53	2	0.04 - 0.49	-0.18	3	-0.16 - -0.002
	NA high arousal	0.24	4	0.009 - 0.21	0.00	0	-	0.01	2	0.004 - 0.01	-0.32	1	-0.32
	NA low arousal	0.94	6	0.006 - 0.27	0.00	0	-	0.19	2	0.03 - 0.16	-0.04	2	-0.04 - -0.008
	Physical activity	0.26	3	0.03 - 0.15	0.00	0		0.00	0	0	-0.46	2	-0.39 - -0.07

Note. Abbreviations: PA = positive affect, NA = negative affect, GC IRF = group cumulative impulse response function.

Figure 3.2. Networks per group showing the *strength* of the IRF associations, by displaying the group cumulative IRFs, i.e., the sum of all positive and negative individual IRF associations of all participants of each group.



Note. Each association shown in the group networks reflects the total impact one variable has on another over time for the individuals in that group (group cumulative impulse response function). Green (solid) arrows indicate positive associations between variables, red (dashed) arrows negative ones. The stronger a particular association, the brighter the color of the arrow.

Table 3.2. Centrality estimates per group showing the *importance* of a variable in the network.

	No anhedonia			**Anhedonia**		
Variable	**Outstrength**			**Outstrength**		
	Total	Favorable	Unfavorable	Total	Favorable	Unfavorable
PA high arousal*	**3.75**	**3.58**	0.17	2.68	1.74	0.94
PA low arousal*	1.99	0.97	1.02	2.91	**2.29**	0.62
NA high arousal	1.51	0.64	0.87	**3.23**	0.57	**2.66**
NA low arousal	2.97	1.03	1.94	1.94	0.6	1.34
Physical activity*	0.98	0.15	0.83	1.04	0.43	0.61
Stress experience	2.64	0.36	**2.28**	2.19	1.35	1.02

Note: * indicates this is considered a positive variable. Bolded numbers reflect the highest estimate per group, indicating that this variable has the strongest impact on all other variables (outstrength). Outstrength was split into favorable and unfavorable impact of the variables. For example, the favorable outstrength of PA high arousal for the non-anhedonic group was constructed by summing all positive group cumulative IRFs for positive variables and all negative group cumulative IRFs for negative variables (0.51 + 0.47 + 0.89 + 1.06 + 0.65 = 3.58, see Table 3.1).

The unfavorable impact of stress experience on affect was more profound among non-anhedonic individuals compared to anhedonic individuals. For non-anhedonic individuals, an increase in stress experience resulted in more NA high arousal (non-anhedonic: group cumulative IRF = 0.24; anhedonic: group cumulative IRF = 0.01) and more NA low arousal (non-anhedonic: group cumulative IRF = 0.94; anhedonic: group cumulative IRF = 0.19) than for anhedonic individuals. Further, for non-anhedonic individuals, stress experience more strongly decreased PA high arousal (non-anhedonic: group cumulative IRF = -0.44; anhedonic: group cumulative IRF = -0.18) and PA low arousal (non-anhedonic: group cumulative IRF = -0.66; anhedonic: group cumulative IRF = -0.18) than for anhedonic individuals. However, the individual models (see Appendix) show that the *number* of individuals demonstrating an unfavorable impact of stress (i.e., these individuals showed at least one unfavorable IRF association of stress) was similar for both groups (non-anhedonic: N = 7; anhedonic: N = 5). The strong negative impact of

stress experience for non-anhedonic individuals is further reflected by their high unfavorable outstrength centrality (see Table 3.2), which was doubled for anhedonic individuals (non-anhedonic: unfavorable outstrength centrality = 2.28; anhedonic: unfavorable outstrength centrality = 1.02).

Network dynamics: role of other variables

As the other dynamic IRF associations may provide additional insight in the mechanisms underlying anhedonia, we also conducted exploratory analyses to examine the roles of other variables in the network.

For non-anhedonic individuals, PA high arousal showed a favorable impact on the other variables, which was evident in the strength as well as the number and the importance of the impact of PA high arousal. Regarding *strength*, for non-anhedonic individuals, PA high arousal resulted in less NA high arousal (non-anhedonic: group cumulative IRF = -0.89; anhedonic: group cumulative IRF = -0.29), less NA low arousal (non-anhedonic: group cumulative IRF = -1.06; anhedonic: group cumulative IRF = -0.01), and less stress (non-anhedonic: group cumulative IRF = -0.65; anhedonic: group cumulative IRF = -0.33). Further, the individual models show that the *number* of individuals with IRF associations originating from PA high arousal was larger in the non-anhedonic group (non-anhedonic: N = 13, anhedonic: N = 8). Finally, in terms of centrality measures, the favorable outstrength of PA high arousal was more than twice as high for non-anhedonic individuals (non-anhedonic: favorable outstrength = 3.58; anhedonic: favorable outstrength = 1.74) and was by far the most *important* variable in the network.

For anhedonic individuals, rather than PA low arousal, PA high arousal showed a favorable impact on the other variables, as indicated in the *strength*, the *number*, and the *importance* of PA low arousal in the network. This indicates that certain positive emotions have a very different role in the network of anhedonic compared to non-anhedonic individuals with depressive symptoms. Further, NA high arousal showed a stronger unfavorable impact on the other variables for anhedonic individuals relative to non-anhedonic individuals. This was reflected in the *strength,* the *number*, and the *importance* of NA high arousal in the network.

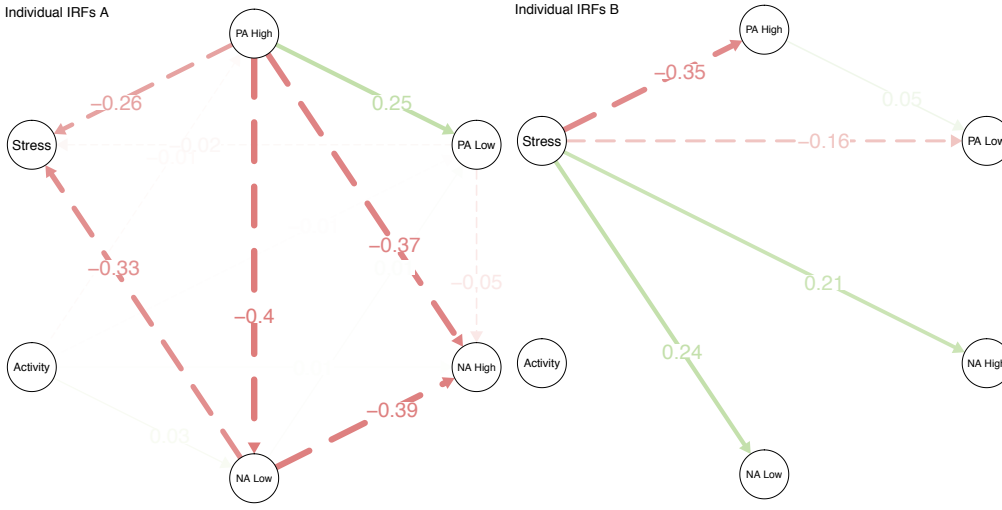
The strong unfavorable impact of NA high arousal mainly seemed to stem from six individuals showing a strong impact of NA high arousal on stress experience (see Table 3.1). No other important and consistent patterns emerged from the data.

Exploration of individual networks of emotional dynamics

All individual models per group can be found in the Appendix. The individual models reveal large individual differences in the dynamic associations between physical activity, stress experience, and affect within the groups of people with and without anhedonia. Three individuals (non-anhedonic: N = 1; anhedonic: N = 2) had no IRF associations, indicating that their physical activity, stress experience and affect did not have a dynamic impact on each other in these individuals. Nine individuals (non-anhedonic: N = 4; anhedonic: N = 5) only showed one or two IRF associations. Seven individuals (non-anhedonic: N = 3; anhedonic: N = 4) showed ten or more IRF associations.

Figure 3.3 illustrates an example of two participants who differ in their emotional dynamics. Both individual A and B were non-anhedonic and had equal levels of depression severity (QIDS = 6). However, for individual A, PA high arousal had a strong favorable impact on the other variables in the network (i.e., it decreased NA high and low arousal and stress, and increased PA low arousal). For individual B, stress experience had a strong unfavorable impact on the other variables (i.e., it increased NA high and low arousal, and decreased PA high and low arousal).

Figure 3.3. Individual IRF networks for two non-anhedonic individuals with equal levels of depression (QIDS = 6), female, who both received higher education. This Figure illustrates that although clinical characteristics are highly similar, emotional dynamics can show very different patterns, warranting a personalized approach to treatment.



Note. Each association shown in the individual networks reflects the total impact one variable has on another over time (individual cumulative impulse response function). Green (solid) arrows indicate positive associations between variables, red (dashed) arrows negative ones. The stronger a particular association, the brighter the color of the arrow.

Discussion

This study investigated the impact of physical activity and stress experience on affect in daily life, and explored other relevant differences in emotional dynamics, in subclinically depressed individuals with anhedonia versus without anhedonia. We used personalized IRFs analyses to study the dynamic impact of the variables on the network as a whole for each individual separately. To our knowledge, this is the first study that maps individual models of the dynamic associations between physical activity, stress, and affect to understand the mechanisms of anhedonia.

Contrary to our hypotheses, the impact of physical activity on affect was low for both anhedonic and non-anhedonic individuals. Thus, when a sudden increase in physical activity was simulated, the other variables only marginally changed in response. Furthermore, also against our expectations, stress experience demonstrated a stronger unfavorable impact on affect in non-anhedonic individuals compared to anhedonic individuals.

In addition, the exploratory analyses revealed that positive affect states played a very different role in the network dynamics of subclinically depressed people with versus without anhedonic complaints: PA high arousal showed a much stronger favorable impact on affect, physical activity and stress experience for non-anhedonic individuals. The finding that positive affect, although present to the same extent in both groups, had a *different dynamic impact* in daily life in the context of anhedonia shines a new light on what anhedonia may represent. Finally, this study reveals the presence of large heterogeneity in emotional dynamics within the anhedonic and non-anhedonic group.

We know of no other studies that examined the effects of physical activity in subclinically depressed individuals with versus without anhedonic symptoms. In depressed individuals, ESM studies have generally shown a favorable effect of physical activity on PA[18,35,36]. In the present study, the impact of physical activity was surprisingly small for all participants and did not differ between the two groups. However, in line with a previous ESM study, we detected large individual differences in whether this impact was favorable or unfavorable[37]. The small impact of physical

activity might partially be due to the relatively large time window of six hours between measurements; studies reporting larger effects had less time in between measurements[35,36].

Contrary to our expectations, stress showed a more profound unfavorable effect for non-anhedonic individuals: stress more strongly decreased PA and increased NA in this group than in the anhedonic group. In the anhedonic group, this was the other way around: NA high arousal demonstrated a more profound unfavorable impact on stress experience. Thus, in non-anhedonic individuals, stress experience seems to generate NA; whereas in anhedonic individuals, NA seems to generate stress experience. Previous ESM studies have consistently shown that MDD is associated with increased reactivity to stress[38,39]. The current study builds on these findings by showing that increased stress reactivity is especially profound in subclinically depressed individuals without anhedonic symptoms.

Further, our findings show that even though PA high arousal was experienced to similar extent in the two groups, the impact of PA high arousal on subsequent emotional and behavioral states was considerably lower for individuals with anhedonic symptoms. Research suggests that specifically the high arousal component of PA is associated with readiness for action, motivation, and goal-directed behavior[40,41]. The finding that PA high arousal does not have a favorable impact on NA and stress experience may help explain why anhedonic individuals in general tend to show poorer prognosis[6,7]. By reducing the impact of daily stressors and NA, PA high arousal may constitute a resilience factor that buffers against depressive symptoms. In line with this proposition, previous research has shown that PA may buffer against stress sensitivity[42].

Together with a close inspection of the individual models, these results may give rise to the hypothesis that different pathways underlie depressive symptoms. The individual models demonstrated that these pathways may be present to different extent in subclinically depressed individuals with and without anhedonia. For some individuals, this pathway may be heightened reactivity to stress or NA, whereas for others, this may be diminished favorable impact of PA. Interestingly, the extent to which these pathways were present differed for the anhedonic group versus

the non-anhedonic group. Where more individuals in the anhedonic group showed diminished favorable impact of PA and heightened reactivity to NA, individuals in the non-anhedonic group showed heightened reactivity to stress.

The large heterogeneity in the extent to which these pathways of emotional dynamics were present in individuals suggest that interventions need to be personalized in order to adequately target the relevant pathway for each patient. If specific pathways of emotional dynamics can be linked to different courses of MDD, and if intervening on central nodes is found to be effective, these individual models might guide the clinician towards a more informed choice for effective interventions. For example, for individuals demonstrating deficient PA high arousal dynamics, interventions may focus on enhancing the favorable effects of PA high arousal to render the individual more resilient (Figure 3.3). For individuals exhibiting strong unfavorable effects of stress experience (or NA high arousal), the clinician may concentrate on strategies to prevent or reduce stress experience, such as through mindfulness techniques. This call for personalized medicine is underscored by studies demonstrating large heterogeneity of MDD[1] and strong indications that group-level findings may not generalize to individual patients[16]. Future studies should reveal whether targeting the most central element of a personalized dynamic network indeed optimizes treatment outcomes.

In order for clinicians to be able to implement this personalized approach to treatment, it is paramount that these complex statistical analyses are automated, so the clinician can easily produce personalized models of emotional dynamics. The *R*-package AIRA automatically generates personalized IRF models, and thus facilitates implementation of these analyses in clinical practice[25]. Although the implementation of personalized networks in clinical practice is yet to receive empirical support, this approach shows promise in making more informed decisions on the focus of treatment.

This study had several notable strengths. First, our ESM design ensured that emotional dynamics were studied ecologically valid, in participants' daily lives and their natural environments. Second, we used a sophisticated and personalized statistical approach, automated IRF analyses (AIRA). Uniquely, AIRA examines the

impact of a variable on the network as a whole rather than on distinct variables and offers insight into individual differences in daily dynamics. Third, we distinguished between high and low arousal PA and NA, thereby shedding light on relevant differences in emotional dynamics that have been overlooked in studies excluding the arousal dimension.

However, our findings should also be considered in light of several limitations. First, the presence of anhedonia was indicated by endorsement of the QIDS-item on loss of interest, but the QIDS does not contain an item on the other hallmark of anhedonia, loss of pleasure. Second, our sample is drawn from the general population. Patients with clinical depression or more severe anhedonia may show a different pattern of results than the subclinically depressed individuals under study here. Third, our timeframe of six hours was relatively long, which may explain why the associations under study were only present in a small part of the sample. Fourth, given that our sample consisted mostly of highly educated women, results may not generalize to other populations. Fifth, stress experience was measured indirectly by assessing level of distress, rather than the direct impact of stressors. Thus, where the different role of PA in the anhedonic versus non-anhedonic group stands out more clearly and reliably, it remains difficult to unravel the difference in associations between NA and stress experience between the two groups. Finally, other factors than anhedonia may also explain the differences found between the anhedonic and non-anhedonic group, such as the presence of sad mood. Future studies may use a 2 by 2 design focusing on the two core symptoms of depression to fully disentangle their influence on emotional dynamics.

Our results suggest different emotional dynamics may underlie depressive symptomatology. Subclinically depressed individuals with anhedonic complaints may be characterized by lowered favorable impact of PA high arousal on affect and behavior, and heightened reactivity to NA. On the other hand, subclinically depressed individuals without anhedonic complaints may be characterized by heightened stress reactivity. The large heterogeneity in the extent to which these pathways were present in individuals advocates a personalized approach to gain insight in how depressive symptomatology is maintained in daily life. Future studies may relate different pathways of emotional dynamics to future course of depression.

References

1. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STARD study. *Journal of Affective disorders.* 2015;172:96.
2. American Psychiatric A. *Diagnostic and Statistical Manual of Mental Dsorders.* 5 ed. Arlington, VA: American Psychiatric Publishing; 2013.
3. Pizzagalli DA. Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology.* 2014;10:393.
4. Vrieze E, Claes SJ. Anhedonia and increased stress sensitivity: Two promising endophenotypes for major depression. *Current Psychiatry Reviews.* 2009;5(3):143.
5. Pelizza L, Ferrari A. Anhedonia in schizophrenia and major depression: State or trait? *Annals of General Psychiatry.* 2009;8.
6. Moos RH, Cronkite RC. Symptom-based predictors of a 10-year chronic course of treated depression. *JNervMentDis.* 1999;187(6):360.
7. Wardenaar KJ, Giltay EJ, van Veen T, Zitman FG, Penninx BWJH. Symptom dimensions as predictors of the two-year course of depressive and anxiety disorders. *Journal of affective disorders.* 2012;136(3):1198.
8. Vrieze E, Demyttenaere K, Bruffaerts R, Hermans D, Pizzagalli DA, Sienaert P, Hompes T, de Boer P, Schmidt M, Claes S. Dimensions in major depressive disorder and their relevance for treatment outcome. *Journal of Affective Disorders.* 2014;155:35.
9. Wichers M, Barge-Schaapveld D, Nicolson NA, Peeters F, de Vries M, Mengelers R, van Os J. Reduced stress-sensitivity or increased reward experience: The psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology.* 2009;34(4):923.
10. Yee A, Chin SC, Hashim AHB, Harbajan Singh MKA, Loh HS, Sulaiman AH, Ng CG. Anhedonia in depressed patients on treatment with selective serotonin reuptake inhibitor anti-depressant - A two-centered study in Malaysia. *IntJPsychiatry ClinPract.* 2015;19(3):1.
11. Damen NL, Pelle AJ, Boersma E, Serruys PW, van Domburg RT, Pedersen SS. Reduced positive affect (anhedonia) is independently associated with 7-year mortality in patients treated with percutaneous coronary intervention: Results from the RESEARCH registry. *Eur J Prev Cardiol.* 2013;20(1):127.
12. Krueger RF, Piasecki TM. Toward a dimensional and psychometrically-informed approach to conceptualizing psychopathology. *Behaviour research and therapy.* 2002;40(5):485.
13. Backhouse SH, Ekkekakis P, Biddle SJH, Foskett A, Williams C. Exercise makes people feel better but people are inactive: Paradox or artifact? *J Sport Exerc Psychol.* 2007;29(4):498.
14. Horan WP, Brown SA, Blanchard JJ. Social anhedonia and schizotypy: The contribution of individual differences in affective traits, stress, and coping. *Psychiatry Res.* 2007;149(1-3):147.
15. Hamaker EL. Why researchers should think 'within-person': A paradigmatic rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:43.
16. Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives.* 2004;2(4):201.
17. Rosmalen JGM, Wenting AMG, Roest AM, de Jonge P, Bos EH. Revealing causal heterogeneity using time series analysis of ambulatory assessments: Application to the association between depression and physical activity after myocardial infarction. *PsychosomMed.* 2012;74(4):377.
18. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, Viechtbauer W, Delespaul P, Myin-Germeys I, Wichers M. Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology.* 2016;35(5):433.
19. Stavrakakis N, Booij SH, Roest AM, de Jonge P, Oldehinkel AJ, Bos EH. Temporal dynamics of physical activity and affect in depressed and nondepressed individuals. *Health Psychology.* 2015;34:1268.
20. Reis HT. Why researchers should think 'real-world': A conceptual rationale. In: Mehl MR, Conner

TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:3.
21. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:636.
22. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology.* 2008;4:1.
23. Brandt P, Williams JT. *Multiple time series models.* Thousand Oaks, CA, US: Sage Publications Inc; 2007.
24. Lütkepohl H. *New introduction to multiple time series analysis.* Berlin: Springer; 2005.
25. Blaauw FJ, van der Krieke L, Emerencia AC, Aiello M, de Jonge P. Personalized advice for enhancing well-being using automated impulse response analysis — AIRA. *Submitted for publication.* 2017.
26. van der Krieke L, Blaauw FJ, Emerencia AC, Schenk HM, Slaets JPJ, Bos EH, de Jonge P, Jeronimus BF. Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *PsychosomMed.* 2016;79(2):213-223.
27. van der Krieke L, Jeronimus BF, Blaauw FJ, Wanders RBK, Emerencia AC, Schenk HM, De Vos S, Snippe E, Wichers M, Wigman JTW, Bos EH, Wardenaar KJ, De Jonge P. HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research.* 2016;25(2):123.
28. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *BiolPsychiatry.* 2003;54(5):573.
29. Lovibond PF, Lovibond SH. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy.* 1995;33(3):335.
30. Wardenaar KJ, Wanders RBK, Jeronimus BF, de Jonge P. The psychometric properties of an internet-administered version of the Depression Anxiety and Stress Scales (DASS) in a sample of dutch adults. *Journal of Psychopathology and Behavioral Assessment.* 2017:1.
31. Watson D, Tellegen A. Toward a consensual structure of mood. *PsycholBull.* 1985;98(2):219.
32. Emerencia AC, van der Krieke L, Bos EH, de Jonge P, Petkov N, Aiello M. Automating vector autoregression on electronic patient diary data. *IEEE J Biomed Health Inform.* 2016;20(2):631.
33. Honaker J, King G. What to do about missing values in time-series cross-section data. *American Journal of Political Science.* 2010;54(2):561.
34. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks.* 2010;32(3):245.
35. Wichers M, Peeters F, Rutten BPF, Jacobs N, Derom C, Thiery E, Delespaul P, van Os J. A time-lagged momentary assessment study on daily life physical activity and affect. *Health Psychology.* 2012;31(2):135.
36. Mata J, Thompson RJ, Jaeggi SM, Buschkuehl M, Jonides J, Gotlib IH. Walk on the bright side: Physical activity and affect in major depressive disorder. *JAbnormPsychol.* 2012;121(2):297.
37. Stavrakakis N, Roest AM, Verhulst F, Ormel J, de Jonge P, Oldehinkel AJ. Physical activity and onset of depression in adolescents: A prospective study in the general population cohort TRAILS. *JPsychiatrRes.* 2013;47(10):1304.
38. Myin-Germeys I, Peeters F, Havermans R, Nicolson NA, deVries MW, Delespaul P, van Os J. Emotional reactivity to daily life stress in psychosis and affective disorder: An experience sampling study. *Acta PsychiatrScand.* 2003;107(2):124.
39. Wichers M, Geschwind N, Jacobs N, Kenis G, Peeters F, Derom K, Thiery E, Delespaul P, van Os J. Transition from stress sensitivity to a depressive state: Longitudinal twin study. *The British Journal of Psychiatry.* 2009;195(6):498.
40. Bradley MM, Lang PJ. Motivation and emotion. *Handbook of psychophysiology.* 2007:581.
41. Harmon-Jones E, Gable PA, Price TF. Does negative affect always narrow and positive affect always broaden the mind? Considering the influence of motivational intensity on cognitive scope. *Current Directions in Psychological Science.* 2013;22(4):301.

42. van Winkel M, Peeters F, van Winkel R, Kenis G, Collip D, Geschwind N, Jacobs N, Derom C, Thiery E, van Os J, Myin-Germeys I, Wichers M. Impact of variation in the BDNF gene on social stress sensitivity and the buffering impact of positive emotions: Replication and extension of a gene–environment interaction. *European Neuropsychopharmacology.* 2014;24(6):930.

Appendix

Table 3.3. Mean levels of affect, physical activity, and stress experience of the two groups.

	No anhedonia (N=20)		Anhedonia (N=20)		
	Mean	SD	Mean	SD	P-value
PA high arousal	46.6	(15.7)	52.3	(11.0)	.155
PA low arousal	52.0	(16.6)	57.5	(8.9)	.166
NA high arousal	29.5	(18.6)	27.2	(13.1)	.607
NA low arousal	40.8	(18.0)	37.8	(11.3)	.479
Physical activity	33.4	(11.1)	39.1	(9.0)	.057
Stress experience	32.4	(20.7)	26.0	(13.2)	.217

Note. Scores could range between 0-100. Multilevel analyses were conducted to test for significant differences in mean levels of the two groups.

Figure 3.4. Overview of participant selection.

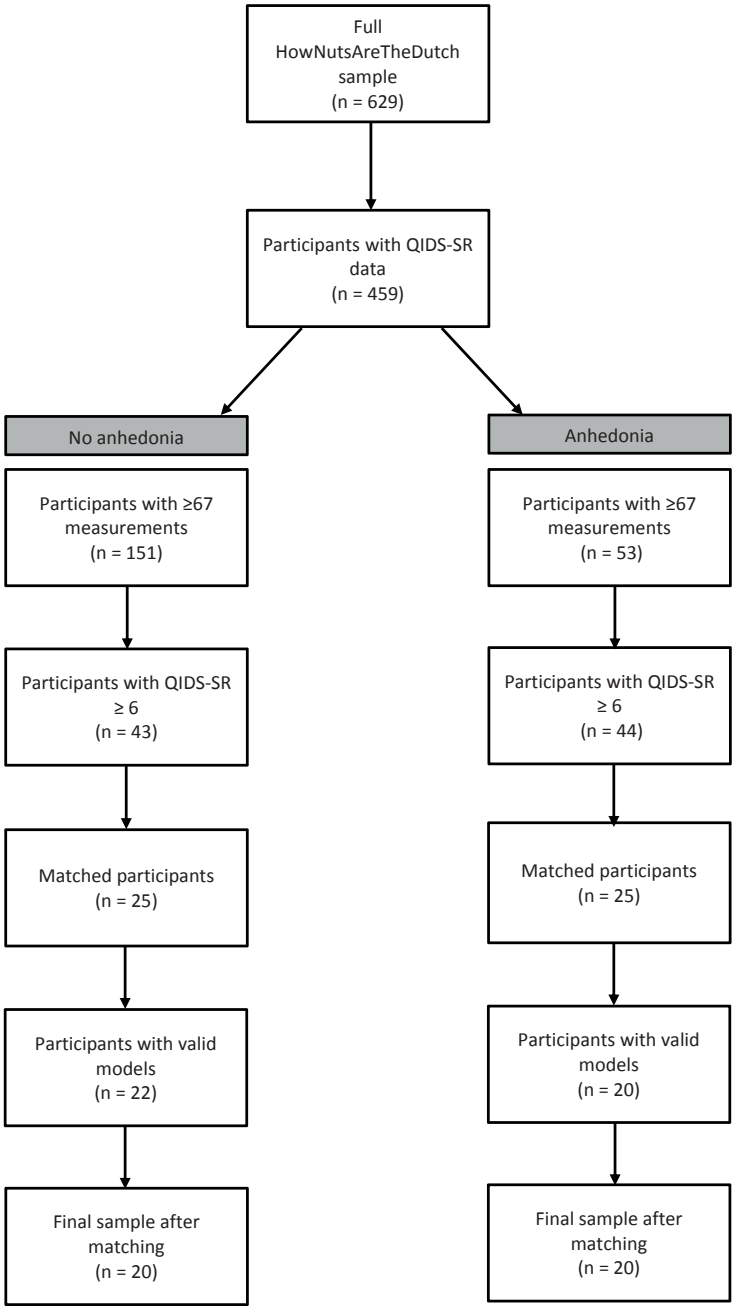
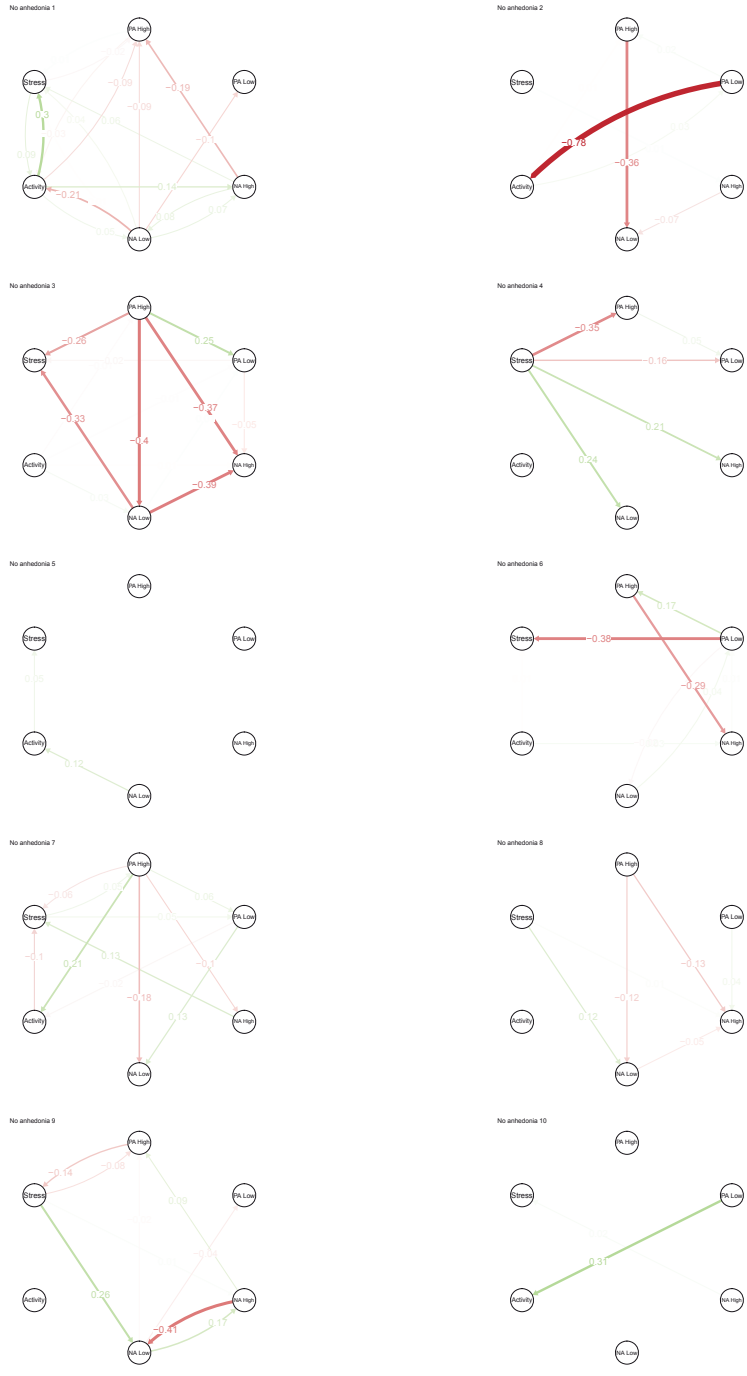
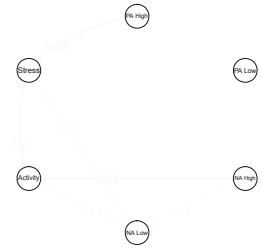
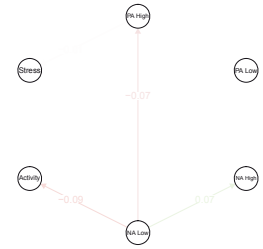


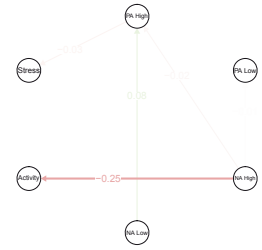
Figure 3.5. IRF models for all 20 individuals without anhedonia separately.



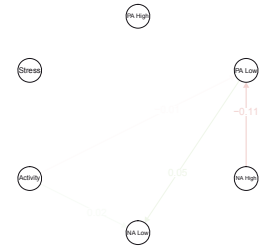




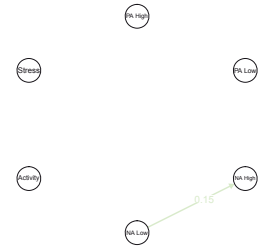
No anhedonia 13



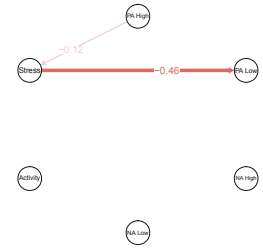
No anhedonia 14



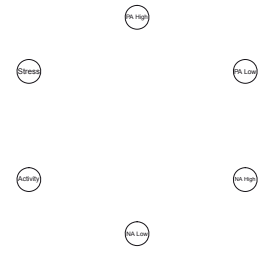
No anhedonia 15



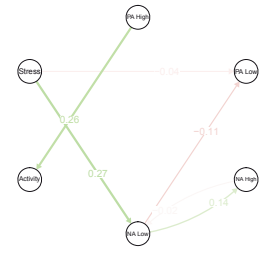
No anhedonia 16



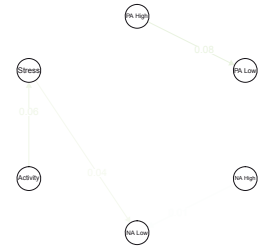
No anhedonia 17



No anhedonia 18



No anhedonia 19



No anhedonia 20

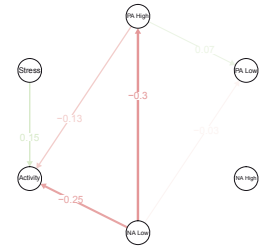
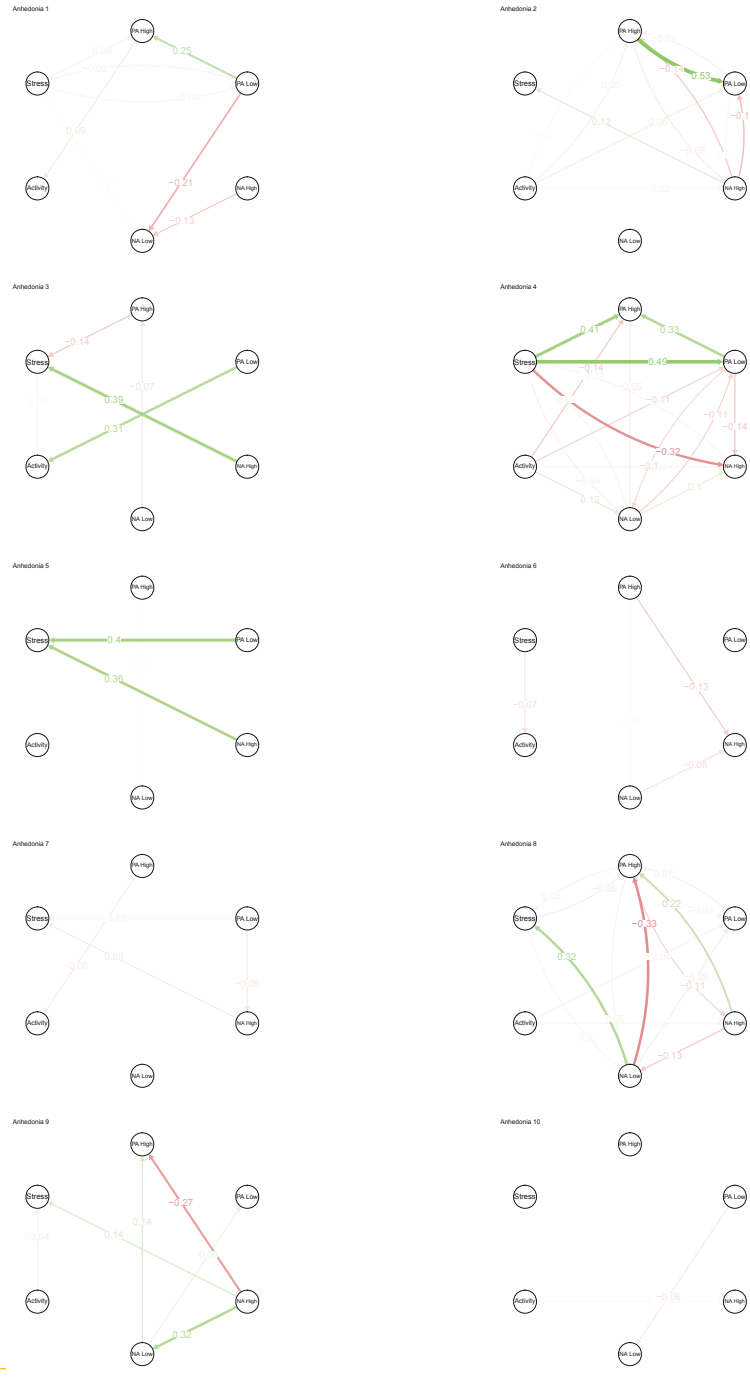
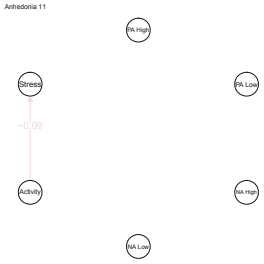
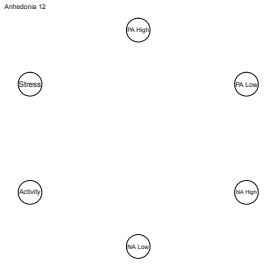
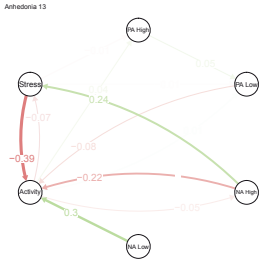


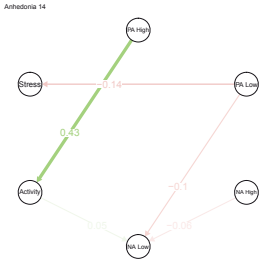
Figure 3.6 IRF models for all 20 individuals with anhedonia separately.

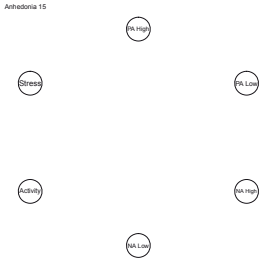


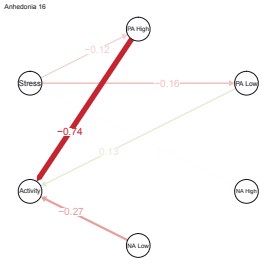

Anhedonia 11
PA High
Stress
PA Low
−0.09
Activity
NA High
NA Low

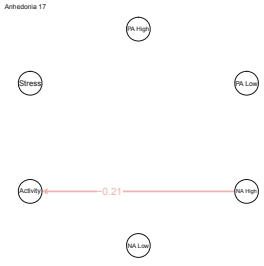

Anhedonia 12
PA High
Stress
PA Low
Activity
NA High
NA Low

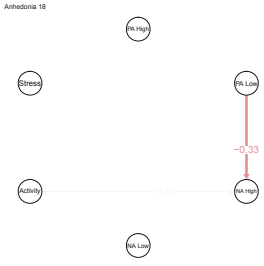

Anhedonia 13
PA High
Stress
PA Low
0.24
−0.39
−0.22
Activity
NA High
0.3
NA Low

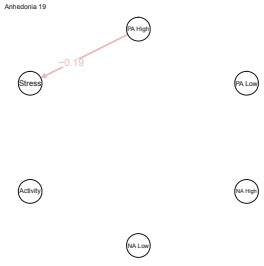

Anhedonia 14
PA High
Stress
PA Low
−0.14
0.43
Activity
NA High
−0.1
NA Low

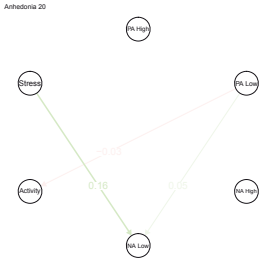

Anhedonia 15
PA High
Stress
PA Low
Activity
NA High
NA Low

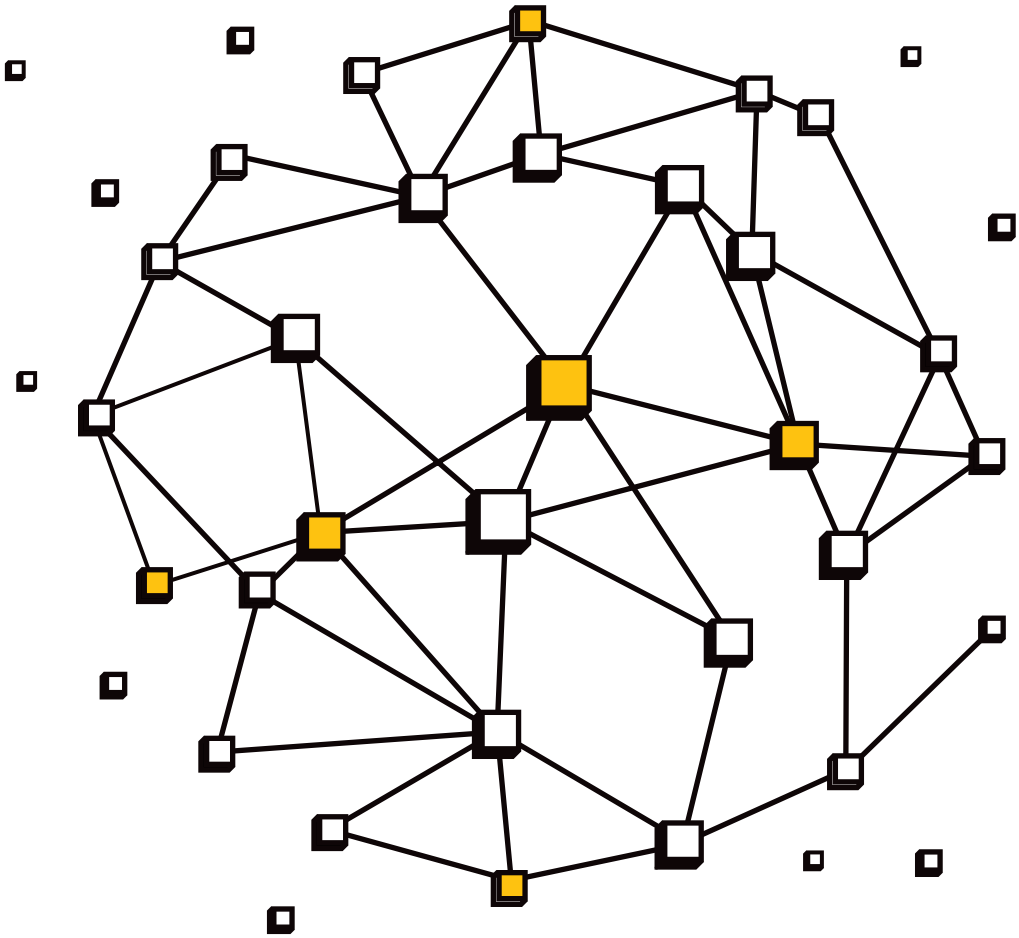

Anhedonia 16
PA High
−0.12
Stress
−0.16
PA Low
−0.74
0.13
Activity
NA High
−0.27
NA Low


Anhedonia 17
PA High
Stress
PA Low
Activity
−0.21
NA High
NA Low


Anhedonia 18
PA High
Stress
PA Low
−0.33
Activity
NA High
NA Low


Anhedonia 19
PA High
−0.19
Stress
PA Low
Activity
NA High
NA Low


Anhedonia 20
PA High
Stress
PA Low
0.16
Activity
NA High
NA Low



Chapter 4

Can we jump from cross-sectional to dynamic interpretations of networks? Implications for the network perspective in psychiatry

Published as:

Bos FM, Snippe E, de Vos S, Hartmann JA, Simons CJP, van der Krieke L, de Jonge P, Wichers M. Can we jump from cross-sectional to dynamic interpretations of networks? Implications for the network perspective in psychiatry. *PsychotherPsychosom.* 2017;86(3):175-177.

Letter to the editor

Interest in the network perspective of psychopathology is rapidly growing. This theory conceptualizes mental disorders as networks in which symptoms can trigger the presence of other symptoms[1]. It thus theorizes that symptoms actively generate other symptoms, and that this process can eventually lead to a full-blown mental disorder. If this is true, then symptom networks may be informative for clinical practice; symptoms that are more central in the network, and are thus assumed to influence many other symptoms, seem to be a logical starting point for intervention. However, in order to know how to optimally investigate the network perspective empirically, it is crucial to first establish whether different network approaches (cross-sectional vs. dynamic) result in similar conclusions.

Although the network theory described above proposes that symptoms are causally related to one other, most of the published studies use cross-sectional data on symptoms to find empirical support for this[2,3]. These studies thereby assume, implicitly or explicitly, that group-level contemporaneous associations between symptoms reflect causal influences between symptoms over time[1,3]. To acquire support for causality, it is important to establish a timeline, i.e., that some symptoms temporally precede other symptoms[4]. It has been questioned, however, whether concurrent, group-level associations among symptoms inform us on how symptoms follow each other over time within individuals[5,6].

We therefore aimed to investigate to what extent a cross-sectional network yields the same conclusion as a network on the same data that includes dynamic (temporal) associations between symptoms (i.e., dynamic networks)[7]. If both approaches result in similar conclusions, this would greatly facilitate future research and clinical applications because cross-sectional data can be more easily obtained. If not, then cross-sectional symptom networks are unlikely to reflect causal symptom dynamics, as postulated by this network theory.

To this end, we used the same experience sampling methodology (ESM) data on mood states to compare the 2 network approaches. Data came from the baseline ESM measurements of an interventional trial including 104 patients with a

DSM-IV-TR diagnosis of a major depressive episode[8]. Patients rated their momentary mental states on 7-point Likert scales 10 times a day for 5 days, resulting in a maximum of 50 measurement points per individual. Here, we focus on the 7 mood items that reflect symptoms of depression and that showed sufficient within-person variance over time, namely, sadness, irritation, loneliness, restlessness, worry, self-doubt, and anhedonia (i.e., cheerfulness reverse coded).

Given that traditional cross-sectional networks are based on (1) a single measurement point per individual, and (2) perceived symptoms that are not momentary but estimated over a somewhat longer period, we estimated 2 cross-sectional networks, each reflecting 1 of the above situations. First, we estimated partial correlations between the first observation of the first day of each participant for all 7 symptoms (Figure 4.1a). Second, we estimated partial correlations between the person-means of the symptoms (Figure 4.1b), i.e., the mean of all 50 measurement points of a symptom provided by 1 individual. Third, a dynamic network was estimated by examining the within-person time-lagged associations among the symptoms in 7 multilevel vector autoregressive models (Figure 4.1c), including 1 of the symptoms as the dependent variable and all symptoms at a previous moment in time (t - 1) as fixed and random effects[7]. Variables were detrended to remove time trends (not detrending the variables yielded similar results) and person-means were centered to disaggregate within-person effects from between-person effects[9]. The analyses were conducted in STATA (v14.1) and the network graphs were made using the qgraph package in R. For clarity, only significant associations are shown in the figures of each network (see Figure 4.1).

For each network, we computed node strength centrality, which is the sum of the absolute value of the strength of all associations of a given symptom with all other symptoms (also nonsignificant ones)[10]. In the dynamic network, strength can be split into “instrength” (the total weight of incoming arrows, not including self-loops) and “outstrength” (the total weight of outgoing arrows). As betweenness and closeness centrality correlated strongly with strength centrality, we decided to not additionally report these measures here.

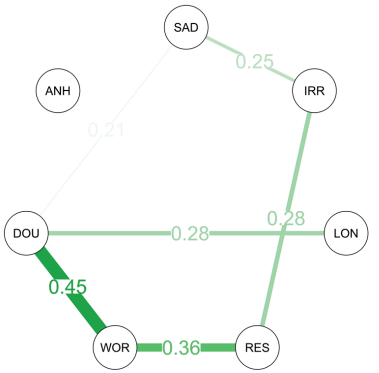
The results showed that the network approaches not only identified

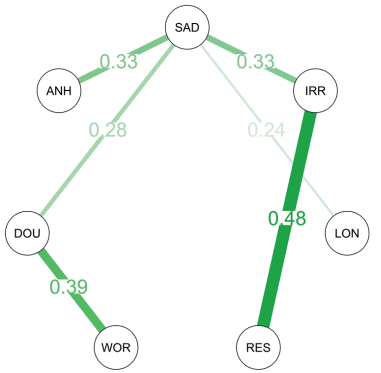
different associations between symptoms, they also yielded different conclusions with regards to which symptom is the most central (Figure 4.1). This would have clinical implications if centrality is used to guide targets for treatment. Whereas the first observations network finds self-doubt to be the most central symptom, the person-means network indicates that sadness is the most central symptom. The dynamic network also points towards sadness, but sadness only has the highest instrength, indicating that it is mostly influenced by other symptoms. It does not have the highest outstrength, suggesting that it does not have the strongest impact on other symptoms. Anhedonia has the highest outstrength in the dynamic network, suggesting that any change here would strongly influence the occurrence of other symptoms. Thus, the cross-sectional and dynamic networks would all indicate different targets for intervention.

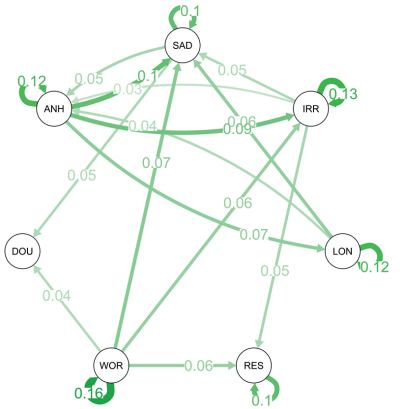
In terms of order of rank, the node strengths of the different networks only correlated modestly or even negatively with each other. The node strengths of the first observations network correlated negatively with the node instrengths (r = -0.24) and negatively with the node outstrengths (r = -0.33) of the dynamic network. The person-mean node strengths correlated negatively with the node outstrengths (r = -0.43) and positively with the node instrenghts (r = 0.62) of the dynamic network.

Our results strongly suggest that cross-sectional networks do not reflect how symptoms trigger each other over time, and therefore may not be interpreted as such. Cross-sectional networks may, however, be useful to examine the co-occurrence of symptoms, for example, to offer an insight into patterns of current symptom comorbidity across individuals[1]. Thus, if we would like to empirically test whether causal symptom dynamics are responsible for the development of mental disorders, the use of dynamic network analysis is advised. If future studies yield support for the network theory, then dynamic network techniques may also have relevance for clinical practice in pointing towards promising targets for intervention.

Figure 4.1. The first observation network (n = 104) (a), the person-mean network (n = 104) (b), and the dynamic network (n = 104 × 50) (c), including measures of node strength centrality for each approach. SAD, sadness; IRR, irritation; LON, loneliness; RES, restlessness; WOR, worry; DOU, self-doubt; ANH, anhedonia. Values in bold type denote the centrality estimates that are the most central symptoms within each network approach.







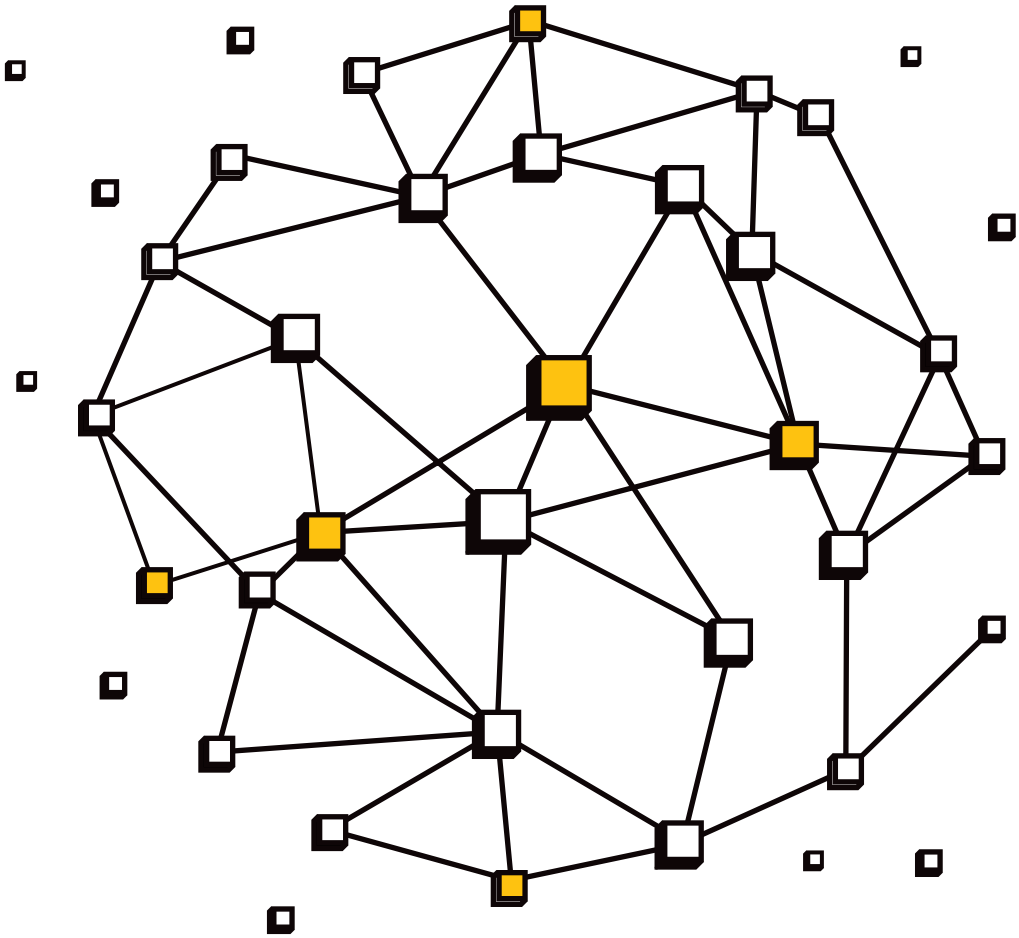
Symptom	Node strength centrality
SAD	0.69
IRR	0.89
LON	0.79
RES	0.89
WOR	1.13
DOU	**1.15**
ANH	0.61

Symptom	Node strength centrality
SAD	**1.44**
IRR	1.08
LON	0.74
RES	1.16
WOR	0.97
DOU	0.91
ANH	0.74

Symptom	Node instrength centrality	Node outstrength centrality
SAD	**0.33**	0.18
IRR	0.21	0.16
LON	0.14	0.21
RES	0.21	0.10
WOR	0.18	0.29
DOU	0.14	0.13
ANH	0.20	**0.34**

References

1. Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. Comorbidity: A network perspective. *Behav Brain Sci*. 2010;33(2-3):137-150.
2. van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, Schoevers RA. Association of symptom network structure with the course of longitudinal depression. *JAMA Psychiatry*. 2015:1219-1226.
3. McNally RJ, Robinaugh DJ, Wu GWY, Wang L, Deserno MK, Borsboom D. Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*. 2015;3(6):836-849.
4. Bolger N, Laurenceau J. *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY, US: Guilford Press; 2013.
5. Bos EH, Wanders RBK. Group-level symptom networks in depression. *JAMA Psychiatry*. 2016;73(4):411-411.
6. Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*. 2004;2(4):201-218.
7. Bringmann LF, Vissers N, Wichers M, et al. A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One*. 2013;8(4):e60188-e60188.
8. 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.
9. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu Rev Psychol*. 2011;62:583-619.
10. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*. 2010;32(3):245-251.



Chapter 5

A narrative review of network studies in depression: What different methodological approaches tell us about depression

Submitted as:

Wichers M, Riese H, Snippe E, Bos FM. A narrative review of network studies in depression: What different network methodologies tell us about depression. Submitted.

Abstract

The network theory of psychopathology proposes that mental disorders arise from direct interactions between symptoms. This theory provides a promising framework to understand the development and maintenance of mental disorders such as depression. In this narrative review, we summarize the literature on network studies in the field of depression. Four methodological network approaches are distinguished: i) studies focusing on symptoms at the macro-level versus ii) on momentary states at the micro-level, and iii) studies based on cross-sectional versus iv) time-series (dynamic) data. 56 studies were identified. We found that different methodological approaches to network theory yielded largely inconsistent findings on depression. Centrality is a notable exception: the majority of studies identified either positive affect or anhedonia as central nodes. To aid future research in this field, we outline a novel complementary network theory, the momentary affect dynamics (MAD) network theory, to understand the development of depression. Furthermore, we provide directions for future research and discuss if and how networks might be used in clinical practice. We conclude that more empirical network studies are needed to determine whether the network theory of psychopathology can indeed enhance our understanding of the underlying structure of depression and advance clinical treatment.

Introduction

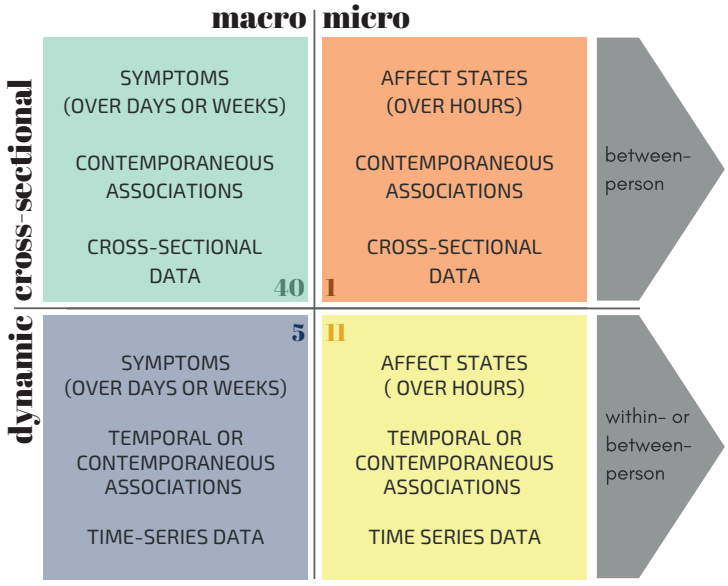
The network theory of psychopathology has gained increasing popularity in recent years[1,2]. This theory postulates that a psychiatric syndrome, such as depression, arises not because of the presence of a latent cause, but rather due to a process in which psychological states or symptoms trigger each other. The theory assumes that this process will eventually result in a cluster of co-occurring symptoms, which we call mental disorders. Empirical research regarding the application of the network theory and its techniques is exponentially growing[3]. The popularity of the network theory is also expressed in the eagerness of healthcare professionals to apply these ideas within clinical practice[4-7]. Thus, the field is in urgent need of a comprehensive evaluation of the findings of empirical network studies.

Such an evaluation should take into account that different methodological approaches have been used within each network study, each yielding their own interpretations and implications. These methodological approaches are different, but complementary, operationalizations of how network dynamics may explain the development and maintenance of psychopathology. Networks consists of nodes and edges (the connections between nodes). In networks on psychopathology, nodes have so far signified either clinical symptoms that supposedly operate at the macro-level[1], or momentary affective states operating at a much smaller timescale, termed the micro-level[8]. Typically, the macro-level network approach is used to investigate between-person relationships among symptoms cross-sectionally at a given time point in a large group of individuals. The micro-level network approach, on the other hand, is often used to examine the processes underlying the development of clinical symptoms by studying the dynamics between everyday life momentary affective states.

In addition to differences in macro- versus micro-level approaches, network studies also differ in the type of data on which they base their network analyses, namely: cross-sectional (data with one assessment per participant) versus time-series data (data with multiple assessments per participant). Often, cross-sectional network studies examine symptoms on the macro-level, whereas

dynamic network studies use time-series data to examine momentary affect at the micro-level. Although it might be expected that the distinction between micro- and macro-level experiences is rather a continuum than distinct categories, for the purposes of this narrative review, we have placed each network study in one of four quadrants: cross-sectional[1] versus dynamic associations and micro- versus macro-level approaches (see Figure 5.1).

Figure 5.1. Overview of characteristics of the four methodological approaches to network theory: micro-level affective states versus macro-level symptoms, based on cross-sectional versus dynamic (time-series) data. The number of studies are indicated in the corner of each quadrant (total N = 56). Note that one study constructed both a cross-sectional and a dynamic network, and is therefore referenced twice in this Figure.



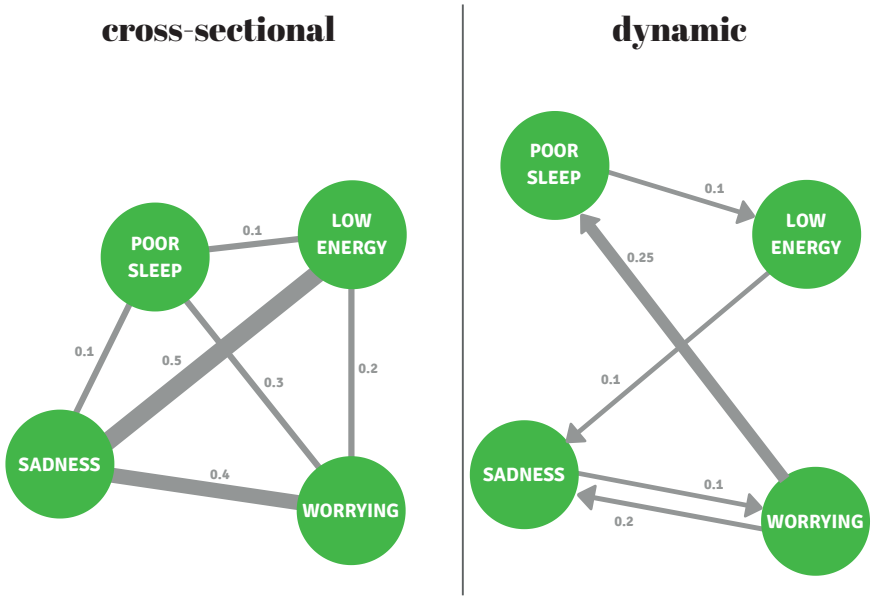
[1] The term 'cross-sectional' here refers to data observations collected at the same time point in the research design.

Distinguishing between these four methodological approaches is relevant as results of these varying network approaches should be interpreted differently[9-11]. Most cross-sectional network studies utilize data from symptom questionnaires to estimate between-person correlations of symptoms at one point in time for a given group of individuals. The associations between variables in such a network demonstrate, at the group level, the probability that these variables tend to occur together (controlling for all other associations). If, for example, those who worry more than others also suffer from sad moods more than others, then the connection between these symptoms will be stronger in such a network. In contrast, networks utilizing time-series data are typically based on the within-person dynamic associations among momentary affective states. Whereas most dynamic networks are based on temporal associations, some studies have also examined the contemporaneous (concurrent) associations. Connections in dynamic networks based on temporal associations show how changes relative to a person's average in one variable follow changes in the other variables within that person. If, for example, individuals start to worry more than usual every time they feel sadder than they usually do, these symptoms will be more strongly connected in the network (see Figure 5.2). It is important to note here, however, that temporality does not imply causality; temporal associations could also be the result of an unknown third variable. In order to construct dynamic networks, time-series analyses techniques, such as (multilevel) vector auto regression (VAR) analyses[12,13], are often used. Thus, whereas cross-sectional models mainly express something about the coexistence of different variables at one moment in time at the group-level, dynamic network models say something about how these variables relate to each other *over* time, within individuals.

Unfortunately, network methodologies are often not clearly separated in discussions on empirical support for networks. Furthermore, overviews of findings of empirical network studies in individuals suffering from depression are lacking. Two recent reviews have provided a first overview of the network literature on psychopathology in general[3,14]. Such systematic reviews are important because they advance our understanding of how the network theory has, thus far, been operationalized. However, both reviews have not explicitly summarized and compared

findings between the different methodological approaches. Furthermore, as these reviews did not solely focus on network studies in depression, a comprehensive evaluation and discussion thereof was beyond their scope. Therefore, we aimed to address this gap in the literature by providing a clear and in-depth overview of the findings of network studies in depression while distinguishing between the aforementioned methodological approaches. Subsequently, we discuss the current status of network research and challenges for both future research and application in clinical practice.

Figure 5.2. Examples of a cross-sectional and a temporal (dynamic) network. In the cross-sectional network, *feeling down* is the most central node (0.1 + 0.4 + 0.5 = 1.0); it has the strongest connections with the other nodes. In the temporal network, *worry* has the highest outgoing centrality; this node strongly predicts other nodes (0.25 + 0.2 = 0.45). The node with the greatest incoming centrality is *feeling down* (0.1 + 0.2 = 0.3), because this node is most strongly predicted by other nodes.



Methods

We conducted a systematic search in PUBMED and PsycINFO to identify empirical studies that have applied network analysis to investigate (risk of) depression. We searched for papers published before October 2020, in which abstracts included the terms; i) 'depression or depressive or MDD or major depression' or 'psychopathology', combined with one of the following terms 'network' or 'impulse response' or 'vector autoregression' or 'VAR' or 'qgraph', or 'network approach', or 'network intervention' and ii) 'affect' or 'mental states' or 'emotion' or 'experience sampling' or 'momentary assessment' or 'depressive symptoms' or 'network analysis'. Papers with the terms 'fMRI' or 'connectomics' or 'functional connectivity' or 'mouse' or 'rat' or 'social network' were excluded.

So far, network studies have primarily focused on investigating the comorbidity of depression with other mental disorders, node centrality, and network connectivity. Therefore, papers had to meet the following inclusion criteria: i) the paper reported empirical results on a network of symptoms or affect states, with at least three nodes pertaining to depressive constructs; ii) the sample population was selected based on their experience of depressive symptoms in the past, present or future, and iii) the paper included results on comorbidity, node centrality and/or network connectivity. As a consequence of this focus on depression, some well-cited network papers fell out of the final selection[2,15,16]. Furthermore, given that our narrative review focuses on the relationships between symptoms and/or affect states, and the inclusion of other nodes might confound estimates of comorbidity, centrality or connectivity, we excluded papers that included contextual factors as nodes such as treatment, genes, coping strategies, or activities[17-19].

Results

Our search resulted in 56 network papers spanning the years 2014-2020. Table 5.1 provides an overview of studies and the methodological approach used to construct networks. Most of these studies used network nodes representing macro-level depressive symptoms[1], assessed with retrospective measures (assessing symptoms of the past week or weeks) via diagnostic interviews or questionnaires (N=45 see Figure 5.1). Of these studies at the macro-level, most (N=40) constructed cross-sectional networks, although there were also five dynamic studies at the macro-level. Eleven studies used momentary affective states as nodes in the network. These states were assessed with frequent questions, often multiple times a day, using the experience sampling methodology (ESM). With ESM, participants indicate their affect states in that particular moment, such as sadness, irritation, or cheerfulness. We refer to this as network research at the 'micro-level' because the examined processes occur on a much smaller time scale[8]. Of the eleven micro-level studies, the majority examined dynamic temporal associations (N=8). One study compared three networks: one based on contemporaneous associations, one based on temporal connections, and one based on cross-sectional micro-level data. Another study also constructed both a temporal as well as a contemporaneous network. The final micro-level study solely examined contemporaneous associations in two separately estimated networks. In the following paragraphs, we will synthesize the results of the network studies regarding comorbidity, centrality, and connectivity, while distinguishing between the different methodological appraches.

Table 5.1. Overview of network studies on depression (N = 56). Overviews are given for the studies on comorbidity (upper panel), centrality (middle panel), and connectivity (lower panel), respectively. Findings of studies that have constructed multiple networks are given separate rows to distinguish their findings.

Comorbidity (N=23)*					
Authors (Year)	**N**	**Time-scale**	**Cross/ dyn**	**Population**	**Bridge symptoms**
Afzali et al. (2017)[20]	909	Macro	Cross	PTSD and depression screening	Sense of foreshortened future, guilt, sadness
An et al. (2019)[21]	776	Macro	Cross	MDD diagnosis & high/low anxiety (BAI≥16 Or BAI<16)	No bridge symptoms identified
Beard et al. (2016)[22]	742	Macro	Cross	MDD, bipolar disorder, anxiety disorder, personality disorder and/or psychotic disorder diagnosis	No bridge symptoms identified
Bekhuis et al. (2016)[23]	2,704	Macro	Cross	MDD or GAD diagnosis now or in remission and healthy controls	Anxiety, fatigue, psychomotor agitation
Curtiss and Klemanski (2016)[24]	111	Macro	Cross	MDD or GAD diagnosis	No bridge symptoms identified
de Haan et al. (2020)[25]	2,313	Macro	Cross	PTSD screening and depressive symptoms	Bridge estimates too unstable
de la Torre-Luque and Essau (2019)[26]	1,494	Macro	Cross	MDD and social phobia diagnosis	No bridge symptoms identified
Djelantik et al. (2020)[27]	458	Macro	Cross	PGD, PTSD or MDD symptoms	No bridge symptoms identified
Garabiles et al. (2019)[28]	355	Macro	Cross	PHQ-9≥6 and GAD-7≥7	Fatigue, sadness, anhedonia
Heeren et al. (2018)[29]	174	Macro	Cross	Social anxiety disorder diagnosis and (some) MDD diagnosis	Avoidance, fear of working while observed, suicidal ideation, anhedonia
Jones et al. (2018)[30]	87	Macro	Cross	OCD diagnosis and (some) MDD diagnosis	Obsessional problems, concentration, guilt, and sadness
Lazarov et al. (2020)[31]	1,795	Macro	Cross	PTSD diagnosis and depression screening	Sleep, sadness, tension, avoidance, upset due to trauma reminders
Levinson et al. (2017)[32]	196	Macro	Cross	Bulimia Nervosa and (some) MDD or anxiety diagnosis	No bridge symptoms identified

Comorbidity (N=23)*					
Authors (Year)	**N**	**Time-scale**	**Cross/ dyn**	**Population**	**Bridge symptoms**
Lorimer et al. (2020)[33]	867	Macro	Cross	Relapse: PHQ-9≥10 and GAD-7≥8 Remission: PHQ-9≤10 and GAD-7≤8	No bridge symptoms identified
McNally et al. (2017)[34]	408	Macro	Cross	OCD diagnosis and (some) MDD diagnosis	No bridge symptoms identified
Park and Kim (2020)[35]	223	Macro	Cross	MDD diagnosis and anxiety symptoms	No bridge symptoms identified
Rogers et al. (2019)[36]	167	Macro	Cross	MDD, substance use, PTSD or bipolar disorder diagnosis	No bridge symptoms identified
Shim et al. (2020)[37]	907	Macro	Cross	MDD and AUD screening	Men: sadness, suicidal ideation, attempt Women: worthlessness, suicidal ideation
Tundo et al. (2020)[38]	241	Macro	Cross	MDD or bipolar disorder diagnosis and anxiety symptoms	No bridge symptoms identified
van Heijst et al. (2020)[39]	618	Macro	Cross	MDD, dysthymia or depressive disorder not otherwise specified diagnosis and autism symptoms	No bridge symptoms identified
Wang et al. (2020)[40]	2,542	Macro	Cross	PHQ-9≥5 and GAD-7≥5	During outbreak**: psychomotor symptoms After peak phase: irritability, energy loss
Kaiser and Laireiter (2018)[41]	10	Macro	Dyn-C	MDD, social anxiety disorder, or GAD or a combination	Large individual differences in bridge symptoms
Groen et al. (2020)[42]	220	Micro	Dyn-T	MDD and anxiety disorder	No evidence for bridge symptoms
Centrality (N=30)					
Authors (year)	**N**	**Time-scale**	**Cross/ dyn**	**Population**	**Most central in network**
Berlim et al. (2020)[43]	151	Macro	Cross	MDD diagnosis	Before AD: fatigue, cognitive disturbance, suicidality After AD: sadness, psychomotor disturbance, cognitive disturbance

Centrality (N=30)					
Authors (Year)	**N**	**Time-scale**	**Cross/ dyn**	**Population**	**Bridge symptoms**
Bos et al. (2018)[44]	178	Macro	Cross	MDD diagnosis	Before AD: energy loss, concentration, worthlessness After AD: anhedonia, worthlessness
Boschloo et al. (2016)[45]	501	Macro	Cross	Future MDD diagnosis	Energy loss, concentration, anhedonia
Carney et al. (2018)[46]	125	Macro	Cross	MDD diagnosis	In remission: energy loss, hypersomnia, concentration Not in remission: fatigue, energy loss, hypersomnia
de la Torre-Luque et al (2020)[47]	427	Macro	Cross	MDD diagnosis	Concentration, energy loss, slow thinking
Fried et al. (2016)[48]	3,463	Macro	Cross	MDD diagnosis	Energy loss, anhedonia, appetite
Hakulinen et al. (2020)[49]	595	Macro	Cross	MDD diagnosis	Anhedonia, sadness, energy loss
Kendler et al. (2018)[50]	5,952	Macro	Cross	MDD diagnosis	Psychomotor skills, hopelessness, reduced self-confidence
Madhoo et al. (2016)[51]	2,876	Macro	Cross	MDD diagnosis	Before AD: hypersomnia After AD: sadness
Murri et al. (2018)[52]	8,557	Macro	Cross	EURO-D≥1	Suicidality, anhedonia, sadness
Park et al. (2020)[53]	1,174	Macro	Cross	MDD diagnosis	Sadness, fatigue, anhedonia
Santos et al. (2017)[54]	515	Macro	Cross	BDI≥10 and healthy controls	Sadness, joy
Santos et al. (2018)[55]	306	Macro	Cross	CES-D≥16	Feeling unwanted, concentration
Semino et al. (2017)[56]	264	Macro	Cross	GDS>5	Full of energy, hopeless, happy
van Borkulo et al. (2014)[57]	1,108	Macro	Cross	MDD diagnosis now or in remission and healthy controls	Sadness, energy loss, self-criticism
van Borkulo et al. (2015)[58]	515	Macro	Cross	MDD diagnosis	Anhedonia, energy loss, concentration
van Loo et al. (2018)[18]	5,784	Macro	Cross	MDD diagnosis	Psychomotor skills, hopelessness, decreased self-confidence
Bringmann et al. (2015)[59]	182	Macro	Dyn-T	MDD diagnosis	Out: suicidality and anhedonia In: indecision and anhedonia

Centrality (N=30)					
Authors (Year)	**N**	**Time-scale**	**Cross/ dyn**	**Population**	**Bridge symptoms**
Groen et al. (2019)[60]	60	Macro	Dyn-T	MDD diagnosis	Out: worrying and energy loss In: feeling trapped and energy loss
Komulainen et al. (2020)[61]	3,559	Macro	Dyn-T	MDD diagnosis	AD Out: Suicidality, work/activity difficulties, weight loss In: Insight, suicidality, work/activity difficulties Placebo Out: Insight, suicidality, retardation In: Insight, retardation, genital symptoms
Savelieva et al. (2020)[62]	72,971	Macro	Dyn-T	EURO-D≥1	Out: suicidality, fatigue, sadness In: suicidality, anhedonia, sadness
Bos et al. (2017)[9]	104	Micro	Cross	MDD diagnosis	Worry and self-doubt
Bos et al. (2017)[9]	104	Micro	Dyn-C	MDD diagnosis	Sadness and restlessness
Bos et al. (2017)[9]	104	Micro	Dyn-T	MDD diagnosis	Out: cheerful | In: sadness
David et al. (2018)[63]	1	Micro	Dyn-T	MDD diagnosis	Out: tension | In: concentration
de Vos et al. (2017)[64]	54	Micro	Dyn-T	MDD diagnosis	Out: cheerful | In: tense
Fisher et al. (2017)[65]	1	Micro	Dyn-C	MDD diagnosis	Hopeless and guilty
Fisher et al. (2017)[65]	1	Micro	Dyn-T	MDD diagnosis	Out: positivity | In: content
Wichers et al. (2016)[66]	1	Micro	Dyn-T	MDD diagnosis in remission	Positive affect
Wigman et al. (2015)[67]	129	Micro	Dyn-T	MDD diagnosis	Out: cheerful | In: content

Connectivity (N=17)					
Authors (year)		**Time-scale**	**Cross/ dyn**	**Population**	**Hypothesis connectivity**
Baez and Heller (2020)[68]	3,184	Macro	Cross	MDD diagnosis	Higher connectivity for younger ages of onset
Berlim et al. (2020)[43]	151	Macro	Cross	MDD diagnosis	No support
Bos et al. (2018)[44]	178	Macro	Cross	MDD diagnosis	No support
Hakulinen et al. (2020)[49]	6,593	Macro	Cross	MDD diagnosis and healthy controls	No support
Madhoo et al. (2016)[51]	2,876	Macro	Cross	MDD diagnosis	Supports hypothesis
Santos et al. (2017)[54]	515	Macro	Cross	≥10 BDI and healthy controls	Supports hypothesis
Schweren et al. (2018)[69]	465	Macro	Cross	MDD diagnosis	No support
van Borkulo et al. (2015)[58]	515	Macro	Cross	MDD diagnosis	Supports hypothesis
Groen et al. (2019)[60]	60	Macro	Dyn-T	MDD diagnosis	No support
de Vos et al. (2017)[64]	54	Micro	Dyn-T	MDD diagnosis	Supports hypothesis depending on method
Pe et al. (2015)[70]	106	Micro	Dyn-T	MDD diagnosis	Supports hypothesis
Snippe et al. (2017)[71]	169	Micro	Dyn-T	MDD diagnosis	No support
van de Leemput et al. (2014)[72]	93	Micro	Dyn-C	MDD diagnosis	No support
van de Leemput et al. (2014)[72]	621	Micro	Dyn-C	Future increase in SCL-90 depressive symptoms	Supports hypothesis
Wichers et al. (2016)[66]	1	Micro	Dyn-T	MDD diagnosis in remission	Supports hypothesis
Wichers et al. (2020)[73]	6	Micro	Dyn-T	MDD diagnosis in remission	Supports hypothesis
Wigman et al. (2015)[67]	129	Micro	Dyn-T	MDD diagnosis	Supports hypothesis

Note. 'And' denotes populations consisting of individuals with both diagnoses or symptom types; 'Or' indicates populations consisting of individuals with either one of the diagnoses or symptom types.

*For the category comorbidity, inclusion criteria were broadened to also include studies with additional participants who experience other sorts of psychiatric symptoms than depressive symptoms (only).

**Refers to the COVID-19 outbreak.

Abbreviations: AD = use of antidepressants; AUD = alcohol use disorder; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; CES-D = Center for Epidemiological Studies Depression; Cross = cross sectional associations; Dyn-C = dynamic contemporaneous associations; Dyn-T = dynamic temporal associations; GAD = generalized anxiety disorder; GDS = Geriatric Depression Scale; In = incoming centrality; MADRS = Montgomery-Åsberg Depression Rating Scale; OCD = obsessive compulsive disorder; PGD = prolonged grief disorder, PHQ-9 = Patient Health Questionnaire; PTSD = posttraumatic stress disorder; Out = outgoing centrality.

Comorbidity

Cramer and colleagues[2] were the first to construct a cross-sectional network with clinical symptoms as nodes (macro-level) to map comorbidity of psychiatric symptoms of depression and generalized anxiety disorder (GAD). Although conducted in a general population sample, the study convincingly demonstrated that symptoms attributed to one diagnostic label (e.g., depression) often co-occur with symptoms associated with differing diagnoses (e.g., GAD). Subsequent studies have since then investigated the comorbidity of depression with other forms of psychopathology. With two exceptions, all studies (N=19) were cross-sectional network studies at the macro-level. Depressive symptoms were most often investigated alongside symptoms of posttraumatic stress disorder (PTSD) and anxiety disorders such as GAD. Other forms of psychopathology that have also been explored include obsessive-compulsive disorder (OCD), prolonged grief disorder, autism, alcohol use disorder (AUD), and somatic symptomatology (see the heading 'Comorbidity' in Table 5.1 for an overview of these studies). Together, these studies confirm that symptoms of depression often co-occur with symptoms of other disorders, which strengthens the conceptualization of psychopathology as transcending boundaries of diagnostic categories[74].

Relatively recently, studies have started to identify so-called 'bridge symptoms'. Bridge symptoms are symptoms in the network that strongly relate to symptoms of two distinct disorders, for example depression and PTSD or depression and GAD. Symptoms with high bridge centrality have the strongest connections to nodes of symptoms belonging to another diagnosis[75,76]. Based on the network theory of psychopathology, it can be hypothesized that comorbidity arises because a bridge symptom of one disorder (e.g., PTSD) also activates symptoms of the other disorder (e.g., depression). Thereafter, activation of symptoms can further expand within the cluster of depression symptoms. Although difficult to directly test this hypothesis, several studies have attempted to identify such bridge symptoms. Notably, the comparison of these studies is complicated because they have operationalized bridge symptoms in different ways: some followed the original operationalization as proposed by Jones and colleagues (2019), whereas others

relied on visual inspection or other more general centrality measures. For clarity, we will only discuss studies here that have followed the original operationalization, because this stays closest to the original bridge symptoms hypothesis based on network theory[75,76].

In macro-level studies (N=8), depressive symptoms that were most often identified as a bridge between depression and other disorders were *sadness*, followed by *loss of interest* and/or *pleasure* (anhedonia), *energy loss* or *fatigue*, and *guilt*. It is notable that only two dynamic network studies have been conducted on identifying bridge symptoms between symptom clusters. Such dynamic network studies could further illuminate whether within-person connections between symptoms of different diagnoses might help to explain comorbidity. A first dynamic network study at the macro-level examined the contemporaneous associations among daily reported symptoms for ten patients separately and reported large individual differences in identified bridge symptoms[41]. The second dynamic network study examined temporal associations at the micro-level in patients diagnosed with depression and anxiety disorder. No evidence for bridge symptoms in overlapping momentary states (*irritated* and *worrying*) was found[42]. Thus, methodological approaches regarding bridge symptoms have so far yielded inconsistent findings. Further studies investigating bridge symptoms are needed to further validate the bridge symptom hypothesis within persons.

Centrality

Research has further focused on the centrality of nodes in networks. Centrality measures have been suggested to indicate how influential nodes are in transmitting information to other nodes within the network[77]. Most studies have focused on three centrality measures: strength, betweenness, closeness. However, the latter two have been argued to be unsuitable as measures of node importance in psychological networks[78]. Strength centrality, on the other hand, is easy to interpret and is therefore used most often. Strength centrality reflects the weight of connections between a certain node and all other nodes in the network. In cross-sectional networks, this is calculated by summing the weights of all connections of

that node. In temporal networks, a distinction can also be made between outgoing and incoming connections of a node (see Figure 5.2 for further explanation).

Again, most network studies investigated centrality in cross-sectional networks at the macro-level (N=17, see the heading 'Centrality' in Table 5.1), with depressive symptoms as nodes in the network. Four symptoms were most often identified as highly central symptoms in the network, namely: *energy loss* (or *fatigue*), *anhedonia* (loss of pleasure or interest), *depressed mood (*or sadness), and *concentration problems*. One of three studies at the micro-level studying contemporaneous associations also indicated *sadness* as one of the most central symptoms in the networks. As such, it seems that *sadness* is often associated with other symptoms of depression, both at weekly and momentary levels. There are multiple interpretations of this result. *Sadness* and *anhedonia* may be central because they trigger other symptoms of depression. However, another explanation is that *sadness* and *anhedonia* are considered the core symptoms of depression and therefore necessary to be able to receive the diagnosis. They may therefore have the greatest chance of frequently coexisting with other symptoms. This speculation is plausible given that many cross-sectional studies included individuals who met the clinical diagnosis of depression, for whom by definition *sadness* or *anhedonia* should be present. A similar explanation may be applicable for the high centrality of *energy loss* and *concentration problems* in depressive symptom networks.

In dynamic networks based on temporal associations we can distinguish between incoming and outgoing centrality. That is, whether a symptom predicts other symptoms at a later time (outgoing) or is itself predicted by other symptoms at an earlier moment in time (incoming). In particular, symptoms with high outgoing centrality might be interesting in the context of possible interventions within a network structure. This would be the case when one would assume that a node with many outgoing connections strongly influences the rest of the network. Comparing dynamic network studies to cross-sectional ones, we see both similarities and differences. Four studies (see Table 5.1) looked at the centrality in temporal networks of macro-level depressive symptoms assessed both daily and weekly. In these studies, *suicidality* often ranked among the symptoms with the highest

outgoing centrality, followed by *energy loss*, similar to cross-sectional networks. In terms of incoming connections, these studies found that *suicidality* and *anhedonia* were the most central. Here, again, two of the four symptoms correspond to the most central symptoms seen in the cross-sectional studies. The idea that *sadness* would be central, in the sense that this symptom may have a major role in triggering other symptoms, is not supported in these dynamic macro-level network studies based on temporal associations.

The dynamic studies examining temporal associations at the micro-level consisted of three group-based studies and three single-subject network analyses. The findings of the three group-based studies were all consistent concerning outgoing centrality of momentary affect: all indicated *cheerfulness*, or *positive affect* in general, as the most central node. This means that short-term changes in *cheerfulness* or *positive affect* most often preceded short-term changes in other affective states within persons. One of the single-subject studies even showed that as relapse into depressive symptoms drew closer, *positive affect* became increasingly central in this individual's network[66]. The finding that *cheerfulness* and *positive affect* were consistently identified as the most influential nodes is interesting. This could indicate that, for many patients, intervening on this node by increasing positive affect could have the greatest impact on the rest of their affective states. Whether this will also be the case in clinical practice still needs to be tested. This finding is in line with psychological theories, such as the "Broaden-and-Build" hypothesis[79,80], which postulates that positive emotions play an important role in regulating negative emotions, and protect against the negative effects of stress. This finding also partly corresponds to findings in macro-level studies, which often identified *anhedonia* as a central symptom. *Anhedonia*, as defined by a lack of interest or pleasure, could be viewed as a macro-level expression of a lack of momentary positive affect in daily life[81]. Interestingly, in the dynamic macro-level networks, *anhedonia* was mostly identified as being the symptom most often influenced by other symptoms. For now, we can conclude that *anhedonia* and *positive affect* play an important role in the syndrome depression.

As such, converging evidence from different symptom network approaches

is found on the relative importance of *anhedonia* and *positive affect* in networks of depressed individuals. Even still, a critical note is in order. Most studies have not tested the ordering of centrality estimates for statistical significance. This limits our ability to assess whether nodes that are identified as central symptoms actually are significantly more central than others. The findings of our review should be interpreted in light of this limitation. We therefore urge researchers to test for ordering which can be done in cross-sectional statistical designs by bootstrapping[82].

Connectivity

The third topic that has been investigated in network studies is connectivity (also known as density). Connectivity is calculated by summing the weights of all edges within the network. Network theory postulates that depression develops because symptoms or momentary affective states trigger each other over and over again. Therefore, this theory predicts that greater vulnerability to depression is directly related to stronger connectivity within the network[83].

However, studies have reported mixed evidence for this hypothesis. At the macro-level, eight cross-sectional and one dynamic study were conducted. Of these, three cross-sectional studies reported an association between larger connectivity and more current or future depressive symptoms[51,54,58]. Another study provided indirect support for the hypothesis by relating higher connectivity to an earlier age of onset of depression[68]. However, the results of the other five studies were inconsistent with the connectivity hypothesis. Two studies reported increased connectivity after antidepressant treatment[43,44] and one study did not find stronger connectivity for depressed patients when compared to healthy controls[49]. The final cross-sectional macro study attempting to replicate the results of van Borkulo et al.[58] found no indications of higher connectivity predicting worse future development of depression[69]. And, finally, the only dynamic (temporal) study at the macro-level also could not confirm that increased connectivity distinguished a worse course from a more favorable course of depression[60].

Therefore, at the macro-level, support for the hypothesis that stronger network connectivity is associated with increased vulnerability to depression is

inconsistent. However, these inconsistent findings may simply be due to study design. Studies that failed to find support for this hypothesis compared networks of depressed individuals to networks of remitted patients or healthy controls. It may be expected that samples of healthy or remitted individuals have larger variability in their extent of depressive symptoms than depressed individuals, who may demonstrate mostly high levels of depressive symptoms. Such so-called floor effects are presumed to reduce the strength of associations in a network, and may therefore result in an opposite pattern than would be expected based on network theory[84]. In line with this, studies supporting the connectivity hypothesis compared individuals who could be expected to show similar variability in symptoms, by comparing baseline network characteristics between individuals who would later develop depression or stay healthy. More studies are needed to confirm this notion.

Eight studies at the micro-level looked at the association between network connectivity and depression, based on the temporal or contemporaneous associations between momentary affective states. Six of them supported the assumption that higher network connectivity indicates a higher vulnerability for depression. Three studies found that depressed patients had a higher network connectivity than healthy controls[64,67,70]. Two single-subject studies showed that the network connectivity between momentary affective states increased precisely in the weeks preceding a sharp rise in depressive symptoms[66,73]. This finding strongly supports the idea that connectivity between momentary affective states causally affects the development of depressive symptoms. This is especially due to the close temporal association between the increase in connectivity and the moment of sudden change in symptoms. Lastly, one dynamic network study examining contemporaneous associations at the micro-level showed that higher connectivity predicted future depressive symptoms[72].

Two micro-level studies did not support the connectivity hypothesis of the network theory. The first study examined temporal associations and found that treatment with mindfulness or antidepressants did not decrease network connectivity in depressed patients[71]. An alternative explanation, also for some of the negative findings regarding treatment effects at the macro-level, could be that treatment

does not address underlying vulnerability reflected by the network structure, but rather the symptoms themselves. Moreover, treatment outcomes may differ for each individual. Such heterogeneity may explain why effects are not visible at the group level[71,85]. The second micro-level study that did not support the connectivity assumption examined contemporaneous associations between momentary states. Higher network connectivity at baseline was found in depressed patients who showed higher declines in symptoms in the following year[72]. Although inconsistent with the connectivity hypothesis of the network theory, this finding corresponds to another hypothesis based on complex systems theory. From this complex systems theory it is derived that higher connectivity indicates higher instability of the system. An instable system has a higher likelihood to suddenly shift to an alternative stable state, which can be either better or a worse in nature[72,86,87]. This means that higher levels of connectivity between symptoms or affective states are hypothesized to occur before sudden transitions to alternative states *in general*. Such a transition to an alternative state could entail a sudden increase in symptom levels, similar as in the network theory, but could also entail a decrease in symptom levels. Thus, it may be that system *stability* is more relevant to network connectivity levels, rather than vulnerability per se. This might provide an alternative explanation for the mixed findings regarding the connectivity hypotheses both at the macro- and micro-level.

Altogether, although most macro-level studies did not find support for the connectivity hypothesis, the results from most micro-level studies did support this hypothesis. Results largely support the idea that the continuous dynamics between momentary affective states may play an important role in the development of clinical depressive symptoms. Whether this conclusion holds up awaits findings from future research.

Discussion

This review focused on three hypotheses based on network theory: comorbidity, centrality, and connectivity. First, regarding comorbidity, macro-level cross-sectional studies supported the hypothesis that depressive symptoms tend to co-occur with symptoms of other disorders, and that bridge symptoms that connect depression with other psychiatric disorders can be identified. Such bridge symptoms were often found in *sadness*, *anhedonia*, *energy loss* or *fatigue*, and *guilt*. However, given the lack of micro-level temporal studies on this topic, we do not yet know whether such bridges can also be detected within individuals; and whether momentary states associated with different syndromes indeed follow upon each other over time.

Second, in research on the centrality of network nodes, we see different outcomes for macro- versus micro-level research on the most influential symptoms or affective states. In macro-level studies, most often identified central symptoms are *energy loss*, *anhedonia*, *sadness*, and *concentration problems*. At least partly corresponding to these findings, micro-level networks have consistently identified *positive affect* as having high outgoing centrality, suggesting that changes in positive affect strongly influence the rest of the network. However, the number of micro-level network studies is still small. Moreover, it is unknown whether highly central momentary states at the micro-level actually trigger the future development of symptoms assessable at the macro-level.

Third, regarding network connectivity, we found that symptom networks at the macro-level do not consistently support the network theory's connectivity hypothesis. This means that stronger network connectivity was not associated with higher symptom levels or future depressive symptoms in more than half of the studies. In contrast, most results from micro-level studies, examining associations between momentary affective states, did support the connectivity hypothesis. This mixed support for the connectivity hypothesis for depression and depressive symptoms is conform findings for psychopathology in general[3].

The above review of the existing literature shows that the network theory of psychopathology has yielded several interesting areas for further research. As

our review has demonstrated, network studies have used different methodological approaches to network theory and, although findings at least partly overlap, each has yielded different conclusions regarding comorbidity, centrality, and connectivity. In future evaluations of the network literature, findings derived from different methodological approaches should be clearly distinguished from one another. We propose that researchers should distinguish between the network theory of psychopathology and the complementary momentary affective dynamics (MAD) network theory, which we will introduce in the next section. After elaborating on the MAD theory, we will discuss four important points to be considered regarding methodological approaches to both network theories. Finally, we will elaborate on the application of network theory in clinical practice.

Proposing the momentary affect dynamics (MAD) network theory

An important question is at which level network dynamics operate. Or more specifically, whether network dynamics operate at the level of clinical symptoms, momentary affective states, or both, to result in the depressive syndrome. In this review, we have reviewed empirical research conducted at both levels. However, most systemic evaluations of the network literature have not taken this distinction into account[3,14]. One reason for this may be that both methodological approaches are described under the theoretical concept of 'the network theory of psychopathology', even though their focus and assumptions are slightly different. For the purpose of transparency, we will therefore name the theoretical approach on the relevance of network dynamics at the micro-level, as described by Wichers[8], the '*momentary affect dynamics (MAD)*' *network theory* of psychopathology. We will shortly summarize the main similarities and differences of this approach to the traditional macro-level network approach as first described by Cramer and colleagues[2] and expanded upon by Borsboom[1].

Both network theories share the assumption that psychological states causally influence one another and that these dynamics play an important role in the further development of psychopathology. An important difference, however, is the level at which network dynamics are assumed to exert their influence, as referenced

before. The two theories focus on a different part of the developmental process of psychopathology. The macro-level network theory focuses on the relationships between depressive symptoms. The MAD theory proposes that dynamics between micro-level momentary affective states are actually the building blocks for the development or maintenance of these macro-level symptoms.

Furthermore, the traditional macro-level network theory describes the process of causal influence between symptoms as a serial process with feedback loops. In other words, symptoms develop, which cause other symptoms to develop, and this process continues until it eventually leads to a mental disorder[1]. The MAD network theory does not assume such a serial process per se. Rather, it assumes that macro-level network connections result from continuously repeating minor impacts of one momentary affective state onto another[8]. Since affect states, such as feeling down, cheerful, or irritated, and their fluctuations are frequent everyday experiences, this assumption makes sense at this level of investigation. For example, within a person, affect state A (feeling down) may often impact affect state B (worrying), which often impacts affect state C (feeling less energetic). The higher the connectivity between these negative momentary affective states, the more these dynamics reinforce one another and draw individuals into cycles of persistent and negative psychological states. It is hypothesized that these persistent psychological states are then eventually experienced as symptoms that can be rated on a traditional psychopathology questionnaire. Note that here, dynamic effects are also likely to occur in parallel and that the whole cycle of dynamic effects does not necessarily need to be precisely timed one after another in order to develop psychopathology. Instead, the MAD network theory assumes that when multiple affect states negatively repeatedly impact many other affect states, the risk of getting stuck in a persistent negative psychological state increases.

In this review, we have seen that findings differ between studies based on the macro-level and micro-level network theory of psychopathology. Now that we have outlined the underlying assumptions of both network theories, at both the macro- and micro-level, it may become clearer that differences in findings are simply due to inherent differences between these approaches. By separating

the methodological approaches underlying both theories, and naming them the macro-level network theory and the MAD network theory, we aim to facilitate future systematic evaluations of empirical network research.

The importance of individual models

So far only a few studies have modelled networks per individual[41,64,66,73]. These studies show a large heterogeneity in network structure between individuals. Interestingly, within-person network structures also appeared to change over a period in which vulnerability for depression changed[66,73]. At the group-level, connections in time-series networks reflect the average outcome of within-person effects across the entire group. However, the question is then what a group-level network would tell us about the network structure of each individual person within that group[88,89]. Vicious cycles, for example, arising from the dynamics between certain symptoms or momentary affective states, are assumed to be an important risk factor for psychopathology. For instance, an individual could be in a cycle where poor sleep leads to lower energy, which lowers cheerfulness, which triggers worrying, resulting even poorer sleep (see Figure 5.2). Models at the group-level, however, cannot offer insight into whether these network connections co-occur within every individual within that group. Indeed, some connections may occur only in certain individuals in the group while other connections occur only in other individuals. More idiographic research is thus needed[90], to verify whether presumed vicious cycles found at a group-level actually exist at an individual level. To conclude, there is still an important gap in the testing of assumptions of the network theory using individual models. This is not only a scientific issue, but also a clinical one, as individual models have the potential to bring novel personalized scientific insights into clinical practice.

The importance of studying processes of change using networks

All but two publications[66,73] discussed here estimated network structures during a period in which the predicted network parameters were expected to remain constant. These networks were either modelled before symptoms developed or when the symptoms were already present. The process of change over time, thus

during a period in which depressive symptoms increase or decrease, has hardly been mapped. If we want to know how symptoms develop and remit we will have to focus on this dynamic process. This is achieved by creating movies of networks that allow parameters to change in order to visualize their developments[66]. If we can demonstrate that network structures already change prior to a change in symptoms, this would strongly support the hypothesis that network structures expose processes that are important in the development of psychopathology. Essential questions remain as to whether or not network connectivity, indeed, increases shortly before the start of a depressive episode and whether the network structure changes as expected (e.g., that vicious cycles decrease in strength or disappear prior remission). These kinds of studies will be essential to enhance our understanding of the developmental process of psychopathology. Recently, it has also been shown to be feasible to collect necessary longitudinal time-series data to examine such questions[66,73].

The importance of the selection of network nodes

Another point of discussion concerns the choice of nodes that are included in a network. First, when aiming to draw conclusions on whether or not symptoms trigger each other, we assume that network connections are not the result of reasons other than causality. Many studies have used the complete list of items from depressive symptom questionnaires or ESM diary questionnaires to construct networks. For a number of these items, however, it seems likely that a third variable is responsible for the co-occurrence of these symptoms[11,91]. For example, reward dysfunction is a likely latent cause for both loss of appetite and loss of pleasure or reduced interest[11]. Therefore, to test hypotheses derived from network theory, both the macro-level theory and the MAD theory, it is desirable to prevent the inclusion of nodes with such a common latent cause in the network. Preferably, future studies devise new questionnaires that are designed for network modeling purposes[3] and only include symptom or affect states that really represent distinct facets of depression. The construction of such new questionnaires could be informed by important depressive constructs identified by patients and caregivers[92]. Furthermore, networks should focus on the inclusion of contextual variables as

nodes in the network, for example treatment[17,93,94], life events[95] or social activities[8]. Although not the focus of the present review, depressive symptoms do not develop within a vacuum and are likely to be strongly influenced by context. Unfortunately, in micro-level studies, such variables are often measured with categorical response scales, making it difficult to include them in the network. Future research should develop ways to assess context alongside momentary states, to be able to assess their interplay.

The importance of statistical choices and pre-processing steps

The operationalization of network theory faces several statistical challenges that will need to be resolved, some specific to cross-sectional macro-level networks and others specific to micro-level dynamic networks. First, pertaining to both methodological approaches, several empirical studies compared network connectivity between groups of individuals with and without depressive symptoms. However, this is problematic as floor effects in symptoms or affect states may bias estimates of associations between network nodes[84]. This problem can be overcome by comparing network connectivity between individuals with equal levels of depression, but different future outcomes of symptom development[58,69].

Second, estimates of network associations depend on the statistical method used[9,64,96]. In recent years, a large variety of statistical methods have been developed to model networks. Cross-sectional macro-level networks are mostly based on partial correlations between symptoms, using regularization techniques to identify only significant associations[97]. Other methods do exist however; for example, relative importance networks or directed acyclic graphs. It is yet unclear how results from these methods relate to one another[98]. Since cross-sectional networks have received strong criticisms regarding their stability and generalizability[99,100], many studies have since then increased their sample size and started to report measures of stability, refraining from reporting network estimates when they can be considered unstable[82,101]. Temporal networks, however, can be estimated on the basis of several techniques, such as multilevel VAR[13], the sparse VAR technique[12], impulse response functions[102], or the graphical VAR model[103]. The choice of statistical

method can greatly influence the results and thus the conclusions of a study. This is demonstrated by the study of de Vos and colleagues[64], who showed that multilevel VAR and sparse VAR techniques resulted in conflicting results regarding network connectivity in the same data. This inconsistency in statistical models hampers our understanding of the results of networks.

A third challenge is specific to micro-level MAD networks and pertains to the preprocessing steps to be carried out on the data. This occurs even before one has selected the statistical method of choice. These steps may seem trivial, and are often unfortunately left out in publications, but can greatly affect the conclusions drawn from network results[64,96,104]. Examples of such preprocessing steps are decisions on person-mean centering and the removal of time trends. Others concern decisions regarding whether or not to impute missing values and apply regularization techniques for model optimization[105]. To date, there is no gold standard for preprocessing decisions because optimal choices can differ per study design. Further, we simply do not yet know which choices result in network coefficients that adequately reflect psychopathological vulnerability. We therefore urge researchers to be transparent in their publications regarding preprocessing steps and analytical models, and their rationale. This will enable us to better understand conflicting or replicated results. Furthermore, systematic empirical research is needed as this may reveal what type of models and statistical choices yield networks that expose information with true clinical value.

Where do we stand in terms of the application of networks in clinical practice?

There is rightly a lot of enthusiasm about the network approach in psychopathology among mental health professionals[5,6]. Network theory could be an interesting clinical application in several ways. First, in clinical practice the intuitive idea prevails that the network theory is in line with how psychopathology is expressed, and that these networks can important processes contributing to the development of psychopathology. It is thus a natural fit to clinical frameworks[5]. A second advantage of network models is that, hypothetically, they can identify how relevant contextual

factors, such as physical activity or social behavior, influence a patient's well-being. Individual networks could also expose the presence of certain vicious cycles of psychological states and behaviors, as has been demonstrated in a patient with psychosis[4], and in a patient with a panic disorder[7]. A final advantage lies in the potential of increased network connectivity to alert patients and clinicians of symptom relapse in the near future[66]. Applications that use real-time monitoring and detection can thus be envisioned to apply these novel insights into clinical practice[73]. In this way, the network approach could help patients and their caregivers to intervene more quickly and in a more focused manner. This is the promise of the network approach for clinical practice.

However, it is apparent from our review that network research is still in its infancy. The empirical testing of network theory has barely begun and still has many challenges. As stated above, it is still unclear which statistical network models can best inform clinical issues and how. Given that the choice of statistical models and preprocessing steps in data analyses can determine the conclusions drawn from networks, this is not a trivial question. Thus, at present, it is not yet possible to guarantee that individual network models based on the data of patients are valid and trustworthy. An illustrative example in this regard is the suggestion that highly central symptoms or momentary states represent good initial treatment targets. However, the use of centrality estimates in psychological networks has been criticized, since the flow process in psychopathology may be radically different from flow process in other types of networks from which it stems (e.g., social networks or virus infections[78]). It is therefore questionable whether centrality can be equated with influence, and whether it is indeed informative for intervention targets. Likewise, it is also highly difficult to empirically test whether intervening on a central symptom does directly improve the rest of the system[106]. Such hypotheses will need to be empirically tested before they can be implemented in treatment.

Nonetheless, several studies have conducted proof-of-principle experiments to examine if networks could improve diagnostics and treatment[7], monitor the development of symptoms[4], and inform on the risk for depressive relapses[66]. These applications might be considered successful in the sense that they facilitated the dialogue between patient and therapist, increased motivation

for trying out different treatment techniques, and improved the self-management of the patient[5,6,107]. Therefore, despite the fact that the validity of networks is still under discussion, the network approach could be valuable in improving the patient-therapist alliance and encouraging more active involvement of patients in their treatment processes. However, the danger of current applications of network feedback in clinical practice is that patients may get the false impression that these personalized networks provide scientifically validated information, which cannot yet be guaranteed[108]. It is, therefore, important that research clarifies, as quickly as possible, under what conditions networks make proper predictions about vulnerability to psychopathology and result in useful patient-specific knowledge on the best intervention targets.

Conclusion

The potential of the network theory is large and cannot be denied. It has both scientific and clinical face validity. This justifies intensive scientific explorations into operationalizations of the network theory. In this narrative review, we have outlined the current state of empirical network studies within the field of depression. We made explicit that at least two conceptually different, but complementary, network theories have been investigated: the traditional macro-level network theory of psychopathology focusing on clinical symptoms, and the MAD network theory focusing on affective states at the micro-level. By systematically differentiating findings of these methodological approaches, we have structured the current empirical support for assumptions of the network theory. Importantly, we argue that we need more empirical studies, and careful systematic evaluation of their findings, to conclude whether network studies can be considered to illuminate the development of psychopathology of individual patients, and whether they provide novel and clinically useful information. Future research may focus on constructing individual rather than group models, illuminating processes of change, defining relevant network nodes, and systematic testing of the impact of various statistical specifications on network models. These steps will ensure that the network theory is further consolidated as both a research methodology as well as a clinical instrument.

References

1. Borsboom D. A network theory of mental disorders. *World Psychiatry.* 2017;16(1):5.
2. Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. Comorbidity: A network perspective. *BehavBrain Sci.* 2010;33(2-3):137.
3. Robinaugh DJ, Hoekstra RH, Toner ER, Borsboom D. The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *PsycholMed.* 2020;50(3):353-366.
4. 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.
5. von Klipstein L, Riese H, van der Veen DC, Servaas MN, Schoevers RA. Using person-specific networks in psychotherapy: challenges, limitations, and how we could use them anyway. *BMC Med.* 2020;18.
6. 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;70(11):983-991.
7. Kroeze R, van der Veen DC, Servaas MN, Bastiaansen JA, Oude Voshaar R, Borsboom D, Riese H. Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research.* 2017;3(1):1.
8. Wichers M. The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *PsycholMed.* 2014;44(7):1349.
9. Bos FM, Snippe E, de Vos S, Hartmann JA, Simons CJP, van der Krieke L, de Jonge P, Wichers M. Can we jump from cross-sectional to dynamic interpretations of networks? Implications for the network perspective in psychiatry. *PsychotherPsychosom.* 2017;86(3):175-177.
10. Bringmann LF, Eronen MI. Don't blame the model: Reconsidering the network approach to psychopathology. *PsycholRev.* 2018;125(4):606.
11. Wichers M, Wigman JTW, Bringmann LF, de Jonge P. Mental disorders as networks: some cautionary reflections on a promising approach. *SocPsychiatry PsychiatrEpidemiol.* 2017;52(2):143.
12. Abegaz F, Wit E. Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics.* 2013;14(3):586.
13. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, Borsboom D, Tuerlinckx F. A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One.* 2013;8(4):e60188.
14. Contreras A, Nieto I, Valiente C, Espinosa R, Vazquez C. The study of psychopathology from the network analysis perspective: A systematic review. *PsychotherPsychosom.* 2019;88(2):71.
15. Boschloo L, van Borkulo CD, Rhemtulla M, Keyes KM, Borsboom D, Schoevers RA. The network structure of symptoms of the Diagnostic and Statistical Manual of Mental Disorders. *PLoS One.* 2015;10(9):e0137621.
16. Goekoop R, Goekoop JG. A network view on psychiatric disorders: network clusters of symptoms as elementary syndromes of psychopathology. *PLoS One.* 2014;9(11):e112734.
17. Blanken TF, van der Zweerde T, van Straten A, van Someren EJW, Borsboom D, Lancee J. Introducing Network Intervention Analysis to Investigate Sequential, Symptom-Specific Treatment Effects: A Demonstration in Co-Occurring Insomnia and Depression. *PsychotherPsychosom.* 2019;88(1):52.
18. van Loo HM, van Borkulo CD, Peterson RE, Fried EI, Aggen SH, Borsboom D, Kendler KS. Robust symptom networks in recurrent major depression across different levels of genetic and environmental risk. *Journal of Affective Disorders.* 2018;227:313-322.
19. Bakker JM, Goossens L, Lange I, Michielse S, Schruers K, Lieverse R, Marcelis M, van Amelsvoort T, van Os J, Myin-Germeys I, Wichers M. Real-life validation of reduced reward processing in emerging adults with depressive symptoms. *JAbnormPsychol.* 2017;126(6):713.
20. Afzali MH, Sunderland M, Teesson M, Carragher N, Mills K, Slade T. A network approach to the comorbidity between posttraumatic stress disorder and major depressive disorder: The role of

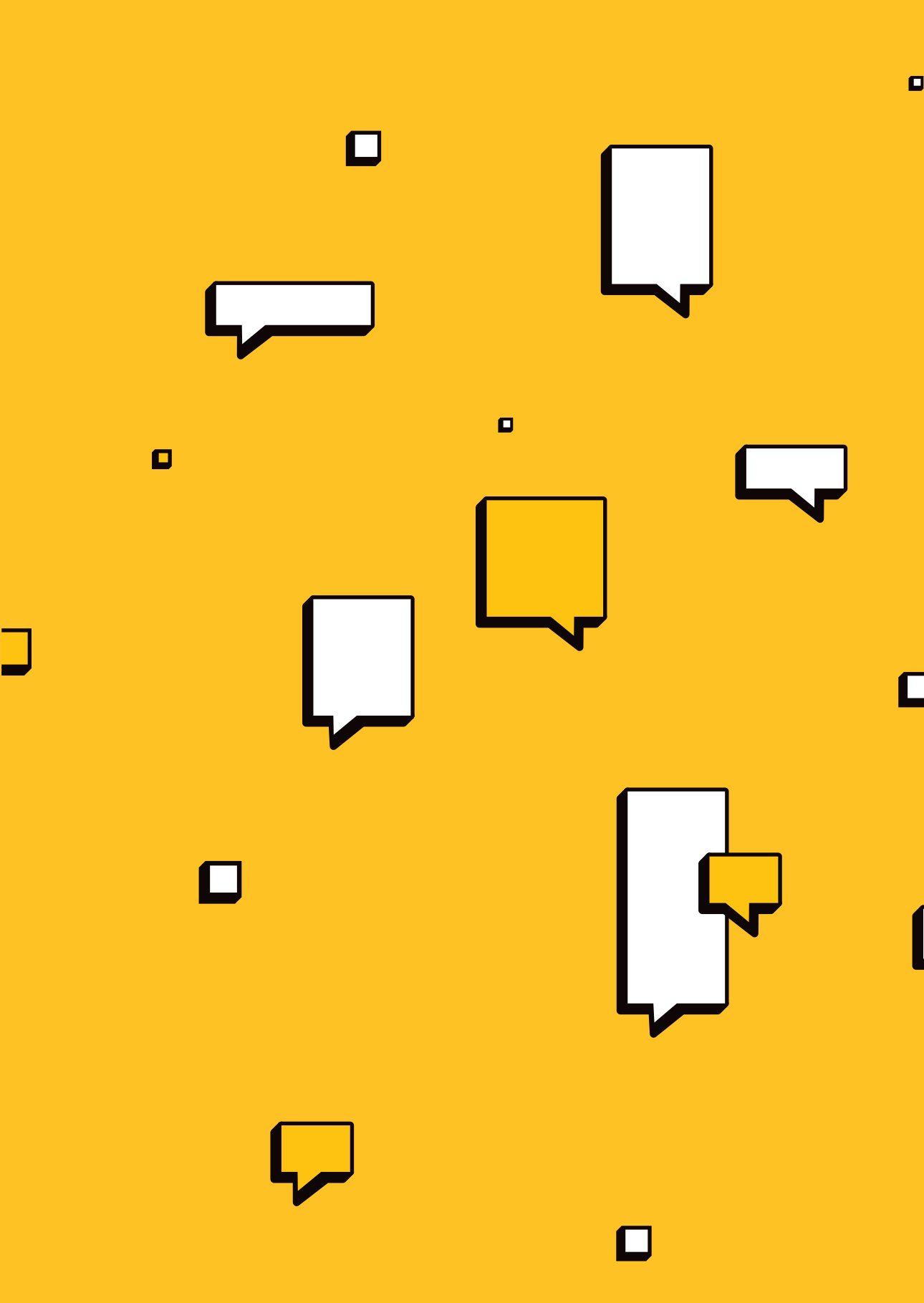
overlapping symptoms. *Journal of Affective Disorders.* 2017;208:490.
21. An MH, Park SS, You SC, Park RW, Kim HK, Son SJ. Depressive symptom network associated with comorbid anxiety in late-life depression. *Frontiers in Psychiatry.* 2019;10:856.
22. Beard C, Millner AJ, Forgeard MJC, Fried EI, Hsu KJ, Treadway MT, Leonard CV, Kertz SJ, Björgvinsson T. Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *PsycholMed.* 2016;46(16):3359.
23. Bekhuis E, Schoevers RA, van Borkulo CD, Rosmalen JGM, Boschloo L. The network structure of major depressive disorder, generalized anxiety disorder and somatic symptomatology. *PsycholMed.* 2016;46(14):2989.
24. Curtiss J, Klemanski DH. Taxonicity and network structure of generalized anxiety disorder and major depressive disorder: An admixture analysis and complex network analysis. *Journal of Affective Disorders.* 2016;199:99.
25. De Haan A, Landolt MA, Fried EI, Kleinke K, Alisic E, Bryant R, Salmon K, Chen SH, Liu ST, Dalgleish T. Dysfunctional posttraumatic cognitions, posttraumatic stress and depression in children and adolescents exposed to trauma: A network analysis. *Journal of Child Psychology.* 2020;61(1):77-87.
26. de la Torre-Luque A, Essau CA. Symptom network connectivity in adolescents with comorbid major depressive disorder and social phobia. *Journal of Affective Disorders.* 2019;255:60-68.
27. Djelantik AAAMJ, Robinaugh DJ, Kleber RJ, Smid GE, Boelen PA. Symptomatology following loss and trauma: Latent class and network analyses of prolonged grief disorder, posttraumatic stress disorder, and depression in a treatment-seeking trauma-exposed sample. *DepressAnxiety.* 2020;37(1):26.
28. Garabiles MR, Lao CK, Xiong Y, Hall BJ. Exploring comorbidity between anxiety and depression among migrant Filipino domestic workers: A network approach. *Journal of Affective Disorders.* 2019;250:85-93.
29. Heeren A, Jones PJ, McNally RJ. Mapping network connectivity among symptoms of social anxiety and comorbid depression in people with social anxiety disorder. *Journal of Affective Disorders.* 2018;228:75.
30. Jones PJ, Mair P, Riemann BC, Mugno BL, McNally RJ. A network perspective on comorbid depression in adolescents with obsessive-compulsive disorder. *JAnxiety Disord.* 2018;53:1.
31. Lazarov A, Suarez-Jimenez B, Levi O, Coppersmith DD, Lubin G, Pine DS, Bar-Haim Y, Abend R, Neria Y. Symptom structure of PTSD and co-morbid depressive symptoms–a network analysis of combat veteran patients. *PsycholMed.* 2020;50(13):2154-2170.
32. Levinson CA, Zerwas S, Calebs B, Forbush K, Kordy H, Watson H, Hofmeier S, Levine M, Crosby RD, Peat C. The core symptoms of bulimia nervosa, anxiety, and depression: A network analysis. *JAbnormPsychol.* 2017;126(3):340.
33. Lorimer B, Delgadillo J, Kellett S, Brown G. Exploring relapse through a network analysis of residual depression and anxiety symptoms after cognitive behavioural therapy: A proof-of-concept study. *Psychotherapy Research.* 2020;30(5):650-661.
34. McNally RJ, Mair P, Mugno BL, Riemann BC. Co-morbid obsessive–compulsive disorder and depression: a Bayesian network approach. *PsycholMed.* 2017;47(7):1204.
35. Park S-C, Kim D. The centrality of depression and anxiety symptoms in major depressive disorder determined using a network analysis. *Journal of Affective Disorders.* 2020.
36. Rogers ML, Hom MA, Joiner TE. Differentiating Acute Suicidal Affective Disturbance (ASAD) from anxiety and depression symptoms: A network analysis. *Journal of Affective Disorders.* 2019;250:333-340.
37. Shim E-J, Ha H, Park JE, Kim B-S, Chang SM, Hong JP, Cho MJ, Hahm B-J. Gender-based examination of the association between individual symptoms of alcohol use disorder, major depressive disorder, and suicidal behaviors: A network analysis of data from the Korean Epidemiologic Catchment Area Study. *Journal of Affective Disorders.* 2020;272:432-439.
38. Tundo A, Musetti L, Del Grande C, de Filippis R, Proietti L, Marazziti D, Gibertoni D, Dell'Osso L. The relationship between depression with anxious distress DSM-5 specifier and mixed depression: A network analysis. *CNS Spectrums.* 2020:1-7.
39. van Heijst BF, Deserno MK, Rhebergen D, Geurts HM. Autism and depression are connected: A

report of two complimentary network studies. *Autism.* 2020;24(3):680-692.
40. Wang Y, Hu Z, Feng Y, Wilson A, Chen R. Changes in network centrality of psychopathology symptoms between the COVID-19 outbreak and after peak. *MolPsychiatry.* 2020:1-10.
41. Kaiser T, Laireiter A-R. Process-symptom-bridges in psychotherapy: An idiographic network approach. *Journal for Person-Oriented Research.* 2018;4(2):49-62.
42. Groen RN, Ryan O, Wigman JT, Riese H, Penninx BW, Giltay EJ, Wichers M, Hartman CA. Comorbidity between depression and anxiety: Assessing the role of bridge mental states in dynamic psychological networks. *BMC Med.* 2020;18(1):1-17.
43. Berlim MT, Richard-Devantoy S, dos Santos NR, Turecki G. The network structure of core depressive symptom-domains in major depressive disorder following antidepressant treatment: A randomized clinical trial. *PsycholMed.* 2020:1-15.
44. Bos FM, Fried EI, Hollon SD, Bringmann LF, Dimidjian S, DeRubeis RJ, Bockting CLH. Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *SocPsychiatry PsychiatrEpidemiol.* 2018;53(6):617-627.
45. Boschloo L, van Borkulo CD, Borsboom D, Schoevers RA. A prospective study on how symptoms in a network predict the onset of depression. *PsychotherPsychosom.* 2016;85(3):183.
46. Carney RM, Freedland KE, Steinmeyer BC, Rubin EH, Rich MW. Residual symptoms following treatment for depression in patients with coronary heart disease. *PsychosomMed.* 2018;80(4):385.
47. de la Torre-Luque A, Ojagbemi A, Caballero FF, Lara E, Moreno-Agostino D, Bello T, Olaya B, Haro JM, Gureje O, Ayuso-Mateos JL. Cross-cultural comparison of symptom networks in late-life major depressive disorder: Yoruba Africans and the Spanish population. *IntJGeriatrPsychiatry.* 2020.
48. Fried EI, Epskamp S, Nesse RM, Tuerlinckx F, Borsboom D. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders.* 2016;189:314.
49. Hakulinen C, Fried EI, Pulkki-Råback L, Virtanen M, Suvisaari J, Elovainio M. Network structure of depression symptomology in participants with and without depressive disorder: The population-based Health 2000-2011 study. *SocPsychiatry PsychiatrEpidemiol.* 2020:1-10.
50. Kendler KS, Aggen SH, Flint J, Borsboom D, Fried EI. The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. *Journal of Affective Disorders.* 2018;227:739.
51. Madhoo M, Levine SZ. Network analysis of the Quick Inventory of Depressive Symptomatology: Reanalysis of the STAR* D clinical trial. *European Neuropsychopharmacology.* 2016;26(11):1768.
52. Murri MB, Amore M, Respino M, Alexopoulos GS. The symptom network structure of depressive symptoms in late-life: Results from a European population study. *MolPsychiatry.* 2018.
53. Park SC, Jang EY, Xiang YT, Kanba S, Kato TA, Chong MY, Lin SK, Yang SY, Avasthi A, Grover S. Network analysis of the depressive symptom profiles in Asian patients with depressive disorders: Findings from the Research on Asian Psychotropic Prescription Patterns for Antidepressants (REAP-AD). *Psychiatry ClinNeurosci.* 2020.
54. Santos H, Fried EI, Asafu-Adjei J, Ruiz RJ. Network structure of perinatal depressive symptoms in Latinas: relationship to stress and reproductive biomarkers. *ResNursHealth.* 2017;40(3):218.
55. Santos H, Kossakowski JJ, Schwartz TA, Beeber L, Fried EI. Longitudinal network structure of depression symptoms and self-efficacy in low-income mothers. *PLoS One.* 2018;13(1):e0191675.
56. Semino L, Marksteiner J, Brauchle G, Danay E. Networks of Depression and Cognition in Elderly Psychiatric Patients. *GeroPsych.* 2017;30:89-96.
57. van Borkulo C, Borsboom D, Epskamp S, Blanken TF, Boschloo L, Schoevers RA, Waldorp LJ. A new method for constructing networks from binary data. *Sci Rep.* 2014;4:5918.
58. van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, Schoevers RA. Association of symptom network structure with the course of longitudinal depression. *JAMA Psychiatry.* 2015;72(12):1219.
59. Bringmann LF, Lemmens LHJM, Huibers MJH, Borsboom D, Tuerlinckx F. Revealing the dynamic network structure of the Beck Depression Inventory-II. *PsycholMed.* 2015;45(4):747.
60. Groen RN, Snippe E, Bringmann LF, Simons CJP, Hartmann JA, Bos EH, Wichers M. Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences. *Psychiatry Res.* 2019;271:640.

61. Komulainen K, Airaksinen J, Savelieva K, Gluschkoff K, Velázquez RG, Elovainio M, Jokela M. Network dynamics of depressive symptoms in antidepressant medication treatment: secondary analysis of eight clinical trials. *MolPsychiatry.* 2020:1-8.
62. Savelieva K, Komulainen K, Elovainio M, Jokela M. Longitudinal associations between specific symptoms of depression: Network analysis in a prospective cohort study. *Journal of Affective Disorders.* 2020;278:99-106.
63. David SJ, Marshall AJ, Evanovich EK, Mumma GH. Intraindividual dynamic network analysis–implications for clinical assessment. *Journal of Psychopathology and Behavioral Assessment.* 2018;40(2):235.
64. de Vos S, Wardenaar KJ, Bos EH, Wit EC, Bouwmans MEJ, de Jonge P. An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *PLoS One.* 2017;12(6):e0178586.
65. Fisher AJ, Reeves JW, Lawyer G, Medaglia JD, Rubel JA. Exploring the idiographic dynamics of mood and anxiety via network analysis. *JAbnormPsychol.* 2017;126(8):1044.
66. Wichers M, Groot PC, Psychosystems, Group ESM, Group EWS. Critical slowing down as a personalized early warning signal for depression. *PsychotherPsychosom.* 2016;85(2):114.
67. Wigman JTW, van Os J, Borsboom D, Wardenaar KJ, Epskamp S, Klippel A, Viechtbauer W, Myin-Germeys I, Wichers M. Exploring the underlying structure of mental disorders: cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. *PsycholMed.* 2015;45(11):2375.
68. Baez LM, Heller AS. Impact of age at onset on the phenomenology of depression in treatment-seeking adults in the STAR* D trial. *Journal of Affective Disorders.* 2020;262:381-388.
69. Schweren L, van Borkulo CD, Fried E, Goodyer IM. Assessment of symptom network density as a prognostic marker of treatment response in adolescent depression. *JAMA Psychiatry.* 2018;75(1):98.
70. Pe ML, Kircanski K, Thompson RJ, Bringmann LF, Tuerlinckx F, Mestdagh M, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Kuppens P, Gotlib IH. Emotion-network density in major depressive disorder. *Clinical Psychological Science.* 2015;3(2):292.
71. Snippe E, Viechtbauer W, Geschwind N, Klippel A, de Jonge P, Wichers M. The impact of treatments for depression on the dynamic network structure of mental states: Two randomized controlled trials. *Sci Rep.* 2017;7:46523.
72. van de Leemput IA, Wichers M, Cramer AOJ, Borsboom D, Tuerlinckx F, Kuppens P, van Nes EH, Viechtbauer W, Giltay EJ, Aggen SH, Derom C, Jacobs N, Kendler KS, van der Maas HL, Neale MC, Peeters F, Thiery E, Zachar P, Scheffer M. Critical slowing down as early warning for the onset and termination of depression. *ProcNatlAcadSciUSA.* 2014;111(1):87.
73. Wichers M, Smit AC, Snippe E. Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: A confirmatory single-subject time-series study. *Journal for Person-Oriented Research.* 2020;6(1):1-15.
74. McGorry PD, Hartmann JA, Spooner R, Nelson B. Beyond the "at risk mental state" concept: transitioning to transdiagnostic psychiatry. *World Psychiatry.* 2018;17(2):133.
75. Jones PJ, Ma R, McNally RJ. Bridge centrality: A network approach to understanding comorbidity. *J Multivariate Behavioral Research.* 2019:1-15.
76. Jones P. networktools: Tools for identifying important nodes in networks. *R package version 12.* 2017;1.
77. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks.* 2010;32(3):245.
78. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, Wigman J, Snippe E. What do centrality measures measure in psychological networks. *JAbnormPsychol.* 2019;128(8):892.
79. Fredrickson BL. The broaden-and-build theory of positive emotions. *Philosophical transactions of the Royal Society of London Series B, Biological sciences.* 2004;359(1449):1367.
80. 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. *Clinical Psychology Review.* 2010;30(7):849.

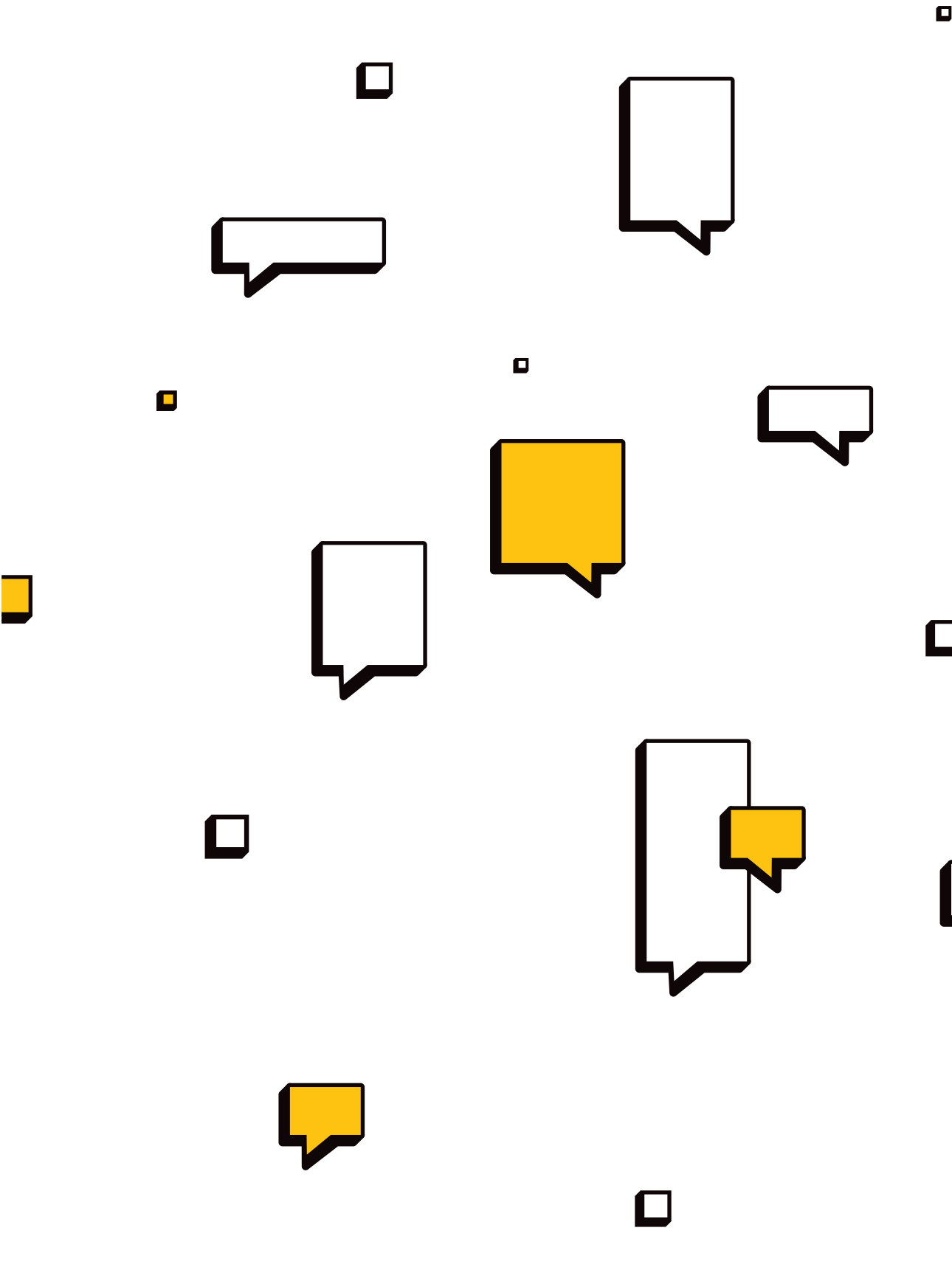
81. van Roekel E, Heininga VE, Vrijen C, Snippe E, Oldehinkel AJ. Reciprocal associations between positive emotions and motivation in daily life: Network analyses in anhedonic individuals and healthy controls. *Emotion.* 2019;19(2):292.
82. Epskamp S. Bootnet: Bootstrap methods for various network estimation routines. *R package.* 2015.
83. Cramer AOJ, van Borkulo CD, Giltay EJ, van der Maas HLJ, Kendler KS, Scheffer M, Borsboom D. Major depression as a complex dynamic system. *PLoS One.* 2016;11(12):e0167490.
84. Terluin B, de Boer MR, de Vet HCW. Differences in connection strength between mental symptoms might be explained by differences in variance: Reanalysis of network data did not confirm staging. *PLoS One.* 2016;11(11):e0155205.
85. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, Viechtbauer W, Delespaul P, Myin-Germeys I, Wichers M. Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology.* 2016;35(5):433.
86. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, Van Nes EH, Rietkerk M, Sugihara G. Early-warning signals for critical transitions. *Nature.* 2009;461(7260):53.
87. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J. Anticipating critical transitions. *Science.* 2012;338(6105):344.
88. Wright AG, Woods WC. Personalized models of psychopathology. *Annual Review of Clinical Psychology.* 2020;16.
89. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences.* 2018;115(27):E6106-E6115.
90. Zuidersma M, Riese H, Snippe E, Booij SH, Wichers M, Bos EH. Single-subject research in psychiatry: Facts and fictions. *Frontiers in Psychiatry.* 2020;11:1174.
91. Fried EI, Cramer AOJ. Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science.* 2017:1.
92. Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, Fried EI, Gartlehner G, Cipriani A, Tran VT. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *The Lancet Psychiatry.* 2020;7(8):692-702.
93. Bekhuis E, Schoevers R, de Boer M, Peen J, Dekker J, Van H, Boschloo L. Symptom-specific effects of psychotherapy versus combined therapy in the treatment of mild to moderate depression: A network approach. *PsychotherPsychosom.* 2018;87(2):121.
94. Boschloo L, Bekhuis E, Weitz ES, Reijnders M, DeRubeis RJ, Dimidjian S, Dunner DL, Dunlop BW, Hegerl U, Hollon SD. The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: results from an individual patient data meta-analysis. *World Psychiatry.* 2019;18(2):183-191.
95. Cramer AOJ, Borsboom D, Aggen SH, Kendler KS. The pathoplasticity of dysphoric episodes: Differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *PsycholMed.* 2012;42(5):957.
96. Bulteel K, Tuerlinckx F, Brose A, Ceulemans E. Using Raw VAR Regression Coefficients to Build Networks can be Misleading. *Multivariate Behav Res.* 2016;51(2-3):330.
97. Epskamp S, Fried EI. A tutorial on regularized partial correlation networks. *PsycholMethods.* 2018.
98. McNally RJ. Network analysis of psychopathology: controversies and challenges. *Annual Review of Clinical Psychology.* 2020;17.
99. Forbes MK, Wright AGC, Markon KE, Krueger RF. The network approach to psychopathology: promise versus reality. *World Psychiatry.* 2019;18(3):272.
100. Forbes MK, Wright AGC, Markon KE, Krueger RF. Evidence that psychopathology symptom networks have limited replicability. *JAbnormPsychol.* 2017;126(7):969.
101. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Research Methods.* 2017:1.
102. Brandt P, Williams JT. *Multiple time series models.* Thousand Oaks, CA, US: Sage Publications Inc; 2007.

103. Wild B, Eichler M, Friederich H-C, Hartmann M, Zipfel S, Herzog W. A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology.* 2010;10(1):28.
104. Bastiaansen JA, Kunkels YK, Blaauw FJ, Boker SM, Ceulemans E, Chen M, Chow S-M, de Jonge P, Emerencia AC, Epskamp S. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *JPsychosomRes.* 2020:110211.
105. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society.* 1996;B58:267.
106. Eronen MI. Causal discovery and the problem of psychological interventions. *New Ideas in Psychology.* 2020;59:100785.
107. Frumkin MR, Piccirillo ML, Beck ED, Grossman JT, Rodebaugh TL. Feasibility and utility of idiographic models in the clinic: a pilot study. *Psychotherapy Research.* 2020:1-15.
108. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: A qualitative study of patient and clinician experiences. *International Journal of Bipolar Disorders.* 2020;8(1):1-14.



Part II

Insights on the utility of ecological momentary assessment for clinical practice



Chapter 6

Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review

Published as:

Bos FM, Schoevers RA, aan het Rot M. Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review. *European Neuropsychopharmacology.* 2015;25(11):1853-1864.

Abstract

Experience sampling methods (ESM) and ecological momentary assessment (EMA) offer insight into daily life experiences, including symptoms of mental disorders. The application of ESM/EMA in psychopharmacology can be a valuable addition to more traditional measures such as retrospective self-report questionnaires because they may help reveal the impact of psychotropic medication on patients' actual experiences. In this paper we systematically review the existing literature on the use of ESM/EMA in psychopharmacology research. To this end, we searched the PsycInfo and Medline databases for all available ESM/EMA studies on the use of psychotropic medication in patients with DSM-III-R and DSM-IV disorders. Dissertations were excluded. We included 18 studies that applied ESM/EMA to study the effects of medication on patients with major depressive disorder, substance use disorder, attention-deficit hyperactivity disorder, psychotic disorder, and anxiety disorder. We found that ESM/EMA may allow researchers and clinicians to track patients during different phases of treatment: before treatment to predict outcome, during treatment to examine the effects of treatment on symptoms and different aspects of daily life experience, and after treatment to detect vulnerability for relapse. Moreover, ESM/EMA can potentially help determine how long and in what contexts medications are effective. Thus, ESM/EMA may benefit both researchers and clinicians and might prove to be an effective tool for improving the treatment of psychiatric patients.

Introduction

Retrospective clinician-administered and self-report questionnaires are the golden standard in human psychopharmacology. Usually, to evaluate treatment progression and outcome in clinical practice and randomized controlled trials, past symptoms are assessed over a period of several days or weeks. However, this golden standard is not undisputed[1]. Retrospective measures are subject to memory distortions. They reveal how patients have reconstructed the past, not how they were experiencing it *in situ*[2]. In contrast, experience sampling methods (ESM) and ecological momentary assessment (EMA) enable clinicians and researchers to tap into daily life processes in real time. The importance of ESM/EMA to capture the daily experience of symptoms is increasingly being recognized in psychiatric research[3]. Consequently, we believe the time has come to systematically review ESM/EMA studies on the pharmacological treatment of patients diagnosed with a mental disorder.

Advantages of ESM/EMA for the field of human psychopharmacology

As mentioned above, one advantage of ESM/EMA for the field of human psychopharmacology is the reduction in memory bias. Retrospective questionnaires are well-suited when clinicians evaluate how patients experienced treatment after the treatment has been completed. However, patients' reflections on past experiences may not correspond with actual experiences during treatment. For example, current mood influences the type of information that is recalled[4]. Additionally, Barge-Schaapveld and Nicolson[5] studied depressed individuals and reported low agreement between side effects recorded in real time using ESM/EMA and side effects reported to a general practitioner. This is highly relevant to psychopharmacology, as it implies that side effects may lead to treatment discontinuation even when they are not recognized as such by the prescribing clinician. Further, retrospective recall of average levels of mood or symptoms might be more difficult than considering the present moment, particularly for individuals with psychiatric diagnoses[3]. Furthermore, retrospective recall of *fluctuations* in

mood or symptoms is highly unreliable. For example, Solhan et al.[6] found that retrospective reports of extreme mood changes, over the previous month and even over the preceding week, were largely unrelated to reports obtained *in situ*. As ESM/EMA questions pertain to the present moment or to a recent interval, memory biases are expected to be minimal.

A second advantage of ESM/EMA is that it taps into processes and experiences as they occur in real life. As a result, the conclusions drawn from psychopharmacology studies that used ESM/EMA are highly ecologically valid[2]. In contrast, measurements obtained in laboratory settings may not always generalize to daily life. The relevance for the field of human psychopharmacology is exemplified by a recent study which found highway driving performance, assessed *in situ* using a specially outfitted vehicle, to remain impaired 3-4 hours after intake of zolpidem[7], even though laboratory tests conducted at this time no longer indicate any deficiencies[8].

Thirdly, when ESM/EMA is used to test the effects of a pharmacological agent, these effects are assessed repeatedly, at different times and in different situations. Thus, ESM/EMA can provide psychopharmacologists with valuable information regarding patients' treatment response in multiple relevant contexts. This information can even be offered as feedback to individual patients[9]. Providing patients with insight into how their feelings, thoughts, and behaviors are influenced by contextual factors enhances control over their treatment process and prognosis.

Finally, ESM/EMA offers methodological flexibility such that it can be tailored to the research or clinical question at hand. By using signaling devices, participants can be asked to provide recordings or assessments at unpredictable intervals. This signal-contingent approach aims to increase insight in ongoing daily-life processes, such as fluctuations in mood or physiology[10]. Other approaches sample at specific times of the day (i.e., time-contingent recording[11]) or in proximity of events of interest such as cigarette smoking (i.e., event-contingent recording[12]). Early ESM/EMA studies used paper questionnaires, but computerized versions on personal digital assistants (PDAs) and smartphones have also become available. Thus, it is increasingly interesting for experimental psychopharmacologists to

incorporate ESM/EMA into their research, and for clinical psychopharmacologists to incorporate ESM/EMA into their treatment.

The present review

ESM/EMA has been used extensively in psychiatry. This indicates that the intensive sampling procedure of ESM/EMA can provide reliable and valid data in patients with severe mental illness. For example, there are reviews on ESM/EMA studies on major depressive disorder (MDD[13], psychotic disorders[14], substance use disorders[15], anxiety disorders[16], and eating disorders[17]. However, no previous review has focused on the use of ESM/EMA in psychopharmacology across disorders. Here we review how ESM/EMA has been used in studies that involved psychopharmacological treatment of patients with Axis I diagnoses according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)[18]. We describe how ESM/EMA has been applied before, during, and after treatment to enhance insight into predictors of treatment outcomes, processes of change, and residual symptoms. We subsequently discuss how these various uses of ESM/EMA can be applied across disorders in future studies and how psychopharmacology might benefit from using ESM/EMA along more traditional measurement approaches.

Experimental procedures

We searched the online PsycInfo and Medline databases, which yielded 1919 potentially relevant studies in our final search (August 2015). Our search terms can be found in the Appendix. Subsequently, FB and MahR independently reviewed the titles, abstracts, and full articles sequentially, while applying the following selection criteria: (1) examination of a pharmacological intervention; (2) repeated measures took place outside the laboratory more than once a day for at least 24h; and (3) DSM-III or DSM-IV diagnoses were established using validated clinical instruments. The PRISMA guidelines were followed during the search for and selection of studies[19]. Initial disagreements regarding the final selection were resolved by reaching consensus. Reference lists of the selected articles were inspected to determine whether eligible studies had been missed. We also examined the digest

of the website of the Society of Ambulatory Assessment. We excluded dissertations and studies not published in English. The Appendix includes details on the reasons for exclusion on the basis of the abstracts and the full articles.

Results

Table 6.1 provides an overview of the 18 selected studies. They are concerned with the treatment of MDD, substance use disorders, attention-deficit hyperactivity disorder (ADHD), psychotic disorders, and anxiety disorders. While we set out to include ESM/EMA studies on all DSM-IV Axis I disorders, there were no eligible studies for most disorders.

Sample sizes of the selected studies ranged from 10-173 patients. Compliance with the ESM/EMA procedures was unknown for five studies but otherwise considered adequate. Most studies included affect variables as a proxy of mood state, with items such as happy, pleased, and excited reflecting positive affect and items such as sad, anxious, and angry reflecting negative affect.

Table 6.1. Overview of included studies.

	Patient Group		Design	Comparison Group(s)		Measurement method						Population
	N	*Medication (n)*		*N*	*Description*	*Format*	*Contingency*	*Observations per day*	*Days*	*Observations*	*Compliance*	
Major depressive disorder												
Barge-Schaapveld and Nicolson, 2002[5]	63	Imipramine (32) or placebo (31)	B	22	Healthy	Paper	Signal	10	21	210	At least 50%, for 80% of included patients	Adult
Barge-Schaapveld et al., 1995[20]	21	Fluvoxamine (11) or amitriptyline (10)	B & W	-	-	Paper	Signal	10	12	120	On average 81%	Adult
Geschwind et al., 2011[21]	63	Imipramine (17) or placebo (22)	B	-	-	Paper	Signal	10	9	90	At least 50%, for 80% of included patients**	Adult
Höhn et al., 2013[22] Study 2[b]	43	Serotonergic antidepressants*	W	39	Healthy	Paper	Signal	10	6	60	On average 82%**	Adult
Study 3[a]	50	Imipramine (23) or placebo (27)	B	21	Healthy	Paper	Signal	10	6	60	At least 50%, for 80% of included patients**	Adult
Peeters et al., 2010[23]	43	Serotonergic antidepressants*	W	39	Healthy	Paper	Signal	10	6	60	On average 82%	Adult
Wichers et al., 2009[24]	63	Imipramine (32) or placebo (31)	B	22	Healthy	Paper	Signal	10	9	90	At least 50%, for 80% of included patients**	Adult
Wichers et al., 2012[25]	47	Serotonergic antidepressants*	W	39	Healthy	Paper	Signal	10	6	60	On average 82%**	Adult

	Patient Group		Design	Comparison Group(s)		Measurement method						Population
	N	*Medication (n)*		*N*	*Description*	*Format*	*Contingency*	*Observations per day*	*Days*	*Observations*	*Compliance*	
Substance use disorders												
Cooney et al., 2007[26]	118	Nicotine patch + intensive smoking cessation treatment (55) or brief smoking cessation advice intervention only (63)	B	-	-	PDA	Signal Event	4 Variable?	14 14	52 Variable	On average 73%	Adult
Holt et al., 2012[27]	48	Nicotine patch + nicotine gum or nicotine patch + placebo gum	B	-	-	Phone	Signal	5	28	140	On average 65%	Adult
Muhonen et al., 2008[28]	80	Memantine (40) or escitolapram (40)	B	-	-	Paper	Event	Variable	182	Variable	-	Adult
Tidey et al., 2008[29]	173	Naltrexone (88) and placebo (85)	B	-	-	PDA	Signal Event Time	5 1 4	33 33 33	165 33 132	On average 79%	Adult
Attention-deficit hyperactivity disorder												
Gehricke et al., 2006[30]	10	1. Nicotine patch 2. Nicotine patch + stimulant medication [Dexedrine (2); Ritalin (2); Adderall (3); Concerta (3)]; 3. Placebo patch + stimulant medication; 4. Placebo	W	-	-	PDA	Signal	25-30	8	200-240	Unknown	Adult

Gehricke et al., 2011[31]	15	Individualized ADHD medication [Dextroamphetamine (1); Amphetamine (7); Atomoxetine (4); OROS-Methylphenidate (1); Lisdexamphetamine (2)] and placebo	W	-	-	PDA	Signal Event	~9 ~13	4 4	Variable Variable	Unknown	Adult

	Patient Group		**Design**	**Comparison Group(s)**		**Measurement method**						**Population**
	N	*Medication (n)*		*N*	*Description*	*Format*	*Contingency*	*Observations per day*	*Days*	*Observations*	*Compliance*	
Whalen et al., 2010[32]	51	Stimulant medication (26) or atomoxetine (25)	B	58	Healthy	PDA	Time	2	7	14	Unknown	Child and adolescent
Psychotic disorders												
Lataster et al., 2011[33]	13	Aripiprazole while discontinuing dopamine antagonists	W	-	-	Paper	Signal	10	12	120	On average 71.5%	Adult
Lataster et al., 2011[34]	109	Haloperidol (39), risperidone (35), or olanzapine (35)	B	-	-	Paper	Signal	10	6	60	Unknown	Adult
Anxiety disorders												
Forbes et al., 2012[35]	66	SSRI (21), CBT + SSRI (21), or CBT (24)	B	-	-	Phone	Signal	2-4	20	60-80	On average 88.5%	Child and adolescent
Sheehan et al., 2007[36]	30	Alprazolam-CT and Alprazolam-XR	W	-	-	Paper	Time	24	63	1512	Unknown	Adult

Note: Studies consisting at least partly of the same individuals are marked with identical superscript letters. B = between subjects design; W = within subjects design; PDA = personal digital assistant, handheld computer or smartphone; Phone = phone call; D-AMPH = Dextroamphetamine; Alprazolam-CT = compressed tablet; Alprazolam-XR = extended release. * = supplemented with supportive psychotherapy. ** = compliance data for this study was not mentioned but considered similar to other studies reporting on the same sample.

Major depressive disorder (MDD)

Most of the selected studies were conducted in patients with MDD. One study examined youth with MDD and/or an anxiety disorder. This study is reviewed under anxiety disorders.

There are seven MDD studies in adult populations. These studies were all conducted in one of three samples. Sample 1 included 49 first-episode and recurrently depressed individuals who were treated with monoaminergic antidepressants in a flexible dose and received supportive psychotherapy. ESM/EMA was conducted for six days prior to the start of treatment. Follow-up data (no ESM/EMA) were collected 1, 2, 3, 6, 12, and 18 months afterwards. Sample 2 included 63 currently depressed individuals treated with the tricyclic antidepressant imipramine or placebo. ESM/EMA was conducted for six days prior to the start of treatment, for the last three days of the first week of treatment, and for six days after six weeks of treatment. A subgroup of patients also provided follow-up ESM/EMA data at 18 weeks after the start of treatment. Finally, Sample 3 consisted of 21 MDD patients randomized to fluvoxamine or amitriptyline. Patients completed ESM/EMA procedures for six days before the start of treatment and for another six days after six weeks of antidepressant treatment. The researchers who studied Samples 1 and 2 operationalized several concepts related to mood, which are summarized in Table 6.2.

Samples 1 and 2 have both been used to study associations between baseline affect dynamics and treatment outcomes. Using Sample 1, Peeters et al.[23] examined emotional reactivity to daily events (cf. Table 6.2) before treatment as a predictor of depression severity after treatment. Importantly, the predictive value of emotional reactivity was tested after first including traditional clinical predictors of MDD course in the statistical model. Peeters et al.[23] also examined baseline stress sensitivity (Table 6.2) as a predictor of remission. Similarly, Wichers et al.[25] examined in Sample 1 the “dynamic interplay between negative and positive emotions” at baseline as a predictor of MDD course, treatment response, and remission. Additionally, baseline retention of positive affect (cf. Table 6.2) was examined as a predictor of treatment response in both Sample 1 and Sample 2[22].

Sample 2 has additionally been used to examine early responses after several days of treatment as predictors of treatment response and outcome. Barge-Schaapveld and Nicolson[5] examined side effects in the first treatment week as a predictor of dropout after six weeks. Geschwind et al.[21] examined changes in positive and negative affect levels during the first week of treatment as a predictor of depressive symptoms on the Hamilton Depression Rating Scale (HDRS), treatment response, and remission. As with the Peeters et al. study[23], the predictive value of early changes in positive affect was tested after including traditional clinical predictors of MDD course in the statistical model.

Finally, Samples 2 and 3 have been used to study treatment outcomes. Barge-Schaapveld et al.[20] compared in Sample 3 the effects of successful and unsuccessful antidepressant treatment on mood and on passive versus active time expenditure (i.e., reading or doing nothing versus engaging in sports or hobbies, respectively). In Sample 2, Wichers et al.[24] studied treatment-induced changes in stress sensitivity and reward experience (see Table 6.2). In the same sample, Barge-Schaapveld and Nicolson[5] examined the effects of antidepressant treatment on quality of life, mood, and enjoyment of activities. ESM/EMA was able to distinguish between remitted patients and healthy controls after 18 weeks of treatment, whereas the HDRS no longer indicated any differences.

In sum, past studies on the pharmacological treatment of MDD have used ESM/EMA in different phases of treatment: (1) before treatment to predict which patients will respond favorably; (2) early in treatment to detect the first changes in affect regulation, which may help predict later outcome; and (3) after treatment to study the effect of treatment on symptomatology and to differentiate between responders and nonresponders and between remitted patients and healthy individuals, which may help predict which patients are likely to relapse. Moreover, ESM/EMA may be a predictor of depressive symptoms above and beyond traditionally used depression measures.

Table 6.2. Terms used in ESM/EMA studies on major depressive disorder.

Term	Definition	Operationalization
Dynamic interplay between negative and positive emotions (Wichers et al., 2012[25])	Decreased negative affect after increases in positive affect	Course of negative affect one beep before and five beeps after the maximum within-person increase in positive affect per day
Emotional reactivity (Peeters et al., 2010[23])	Reacting with increased positive affect or negative affect to positive or negative events, respectively	Predictive value of affective response after a positive or negative event at the same beep
Retention of positive affect (Höhn et al, 2012[22])	Ability to savor positive affect for longer periods of time	Value of positive affect at the previous beep in predicting positive affect at the next beep
Reward experience (Wichers et al., 2009[24])	Increased positive affect after positive events	Effect of positively appraised events (events that are enjoyable, do not require effort and that the participant is skillful at) on positive affect
Stress sensitivity (Wichers et al., 2009[24])	Increased negative affect after stressful events	Effect of stressful events (events that are not enjoyable, require effort, and that the participant is not skillful at) on negative affect

Note. ESM/EMA = Experience sampling methods/Ecological momentary assessment.

Substance use disorders

Pharmacological treatment for substance use disorders has been investigated in four ESM/EMA studies. Two studies focused on alcohol dependence. Muhonen et al.[28] compared memantine to escitalopram in alcohol-dependent patients with comorbid MDD and used event-contingent recording to collect real-time data on alcohol consumption. No other ESM/EMA data were collected.

Tidey et al.[29] studied the effects of naltrexone in heavy drinkers, of whom more than one-third were diagnosed with alcohol dependence. ESM/EMA was conducted for one week before and for four weeks after the start of treatment. Patients recorded drinking urges, mood, situational variables, and the effects of alcohol. The data were used to examine the effects of naltrexone on the percentage of drinking days, the amount of time between drinks, the intensity of drinking urges, and the stimulating effects of alcohol. Further, the authors conducted moderator analyses to examine the effects of naltrexone on specific subgroups (i.e., gender,

genetic subtypes, early onset versus late onset drinkers, patients with versus without a family history of alcohol problems).

The other two studies focused on concurrent alcohol and nicotine dependence. Holt et al.[27] studied patients with alcohol or dependency and nicotine dependency. Patients were given a nicotine patch and additionally randomized to nicotine gum or placebo gum. ESM/EMA was conducted for one month during treatment. Cooney et al.[26] also examined concurrent alcohol and nicotine treatment in patients enrolled in an intensive substance abuse treatment program. Patients were randomized to a smoking cessation intervention, consisting of behavioral counseling and transdermal nicotine replacement, or to brief smoking cessation advice. ESM/EMA was conducted for two weeks post-treatment. In both studies, ESM/EMA was used to assess smoking and drinking urges and behavior, mood, and abstinence self-efficacy (i.e., the confidence to resist drinking and smoking urges). Both studies used ESM/EMA data to identify both risk and protective factors for smoking and alcohol relapse in patients' natural environments, during treatment[27], and after treatment[26].

In sum, ESM/EMA studies of pharmacological treatment in substance use disorder have shown ESM/EMA can be used (1) to estimate the number of drinks consumed during the day, (2) to reveal what types of patients benefit the most from a particular medication by studying moderation effects, and (3) to prospectively study antecedents of smoking and drinking lapse in the preceding hours, in the presence of patients' own smoking and drinking cues and situations, in different phases of treatment.

Attention-Deficit Hyperactivity Disorder (ADHD)

There have been two ESM/EMA studies on the treatment of ADHD in adults[30,31] and one in youth[32].

Gehricke et al.[30] examined the effects of four pharmacological interventions in adult smokers with ADHD: a transdermal nicotine patch or a placebo patch with or without stimulant medication (details in Table 6.1). ESM/EMA was used during each intervention to monitor ADHD symptoms and a self-defined core symptom,

which was daydreaming for 50% of patients and zoning out for the rest. The authors examined the effects of the four two-day interventions on core symptoms, mood, arousal, self-control, and smoking urges.

In a second study, Gehricke et al.[31] compared the effects of ADHD medication (see Table 6.1) and placebo on smoking and withdrawal symptoms. Patients recorded ADHD symptoms, smoking urges, and stress levels for two days during both interventions. In addition to the effects of medication on symptoms and smoking, the authors studied in which contexts the medication was most effective.

Whalen et al.[32] compared atomoxetine to stimulant medication in children with ADHD. The ESM/EMA questionnaires were completed in the mornings and evenings by the children's mothers and assessed the mothers' moods as well as their perceptions of the moods and behaviors (e.g. inattention, hyperactivity, impulsivity, opposition) of their children. Ratings of mothers of children with ADHD were compared to mothers of children without ADHD. Moreover, as medication was administered once per day and the assessments were conducted twice per day, duration of the medication effect could be established.

So far, studies have shown that ESM/EMA can be applied (1) to monitor the immediate effects of ADHD medications in daily life, (2) to reveal situation-specificity in the effects of ADHD medication, (3) to distinguish between medicated children with ADHD and healthy controls, and (4) to provide insight into duration of effect of medication.

Schizophrenia and other psychotic disorders

Two studies have examined the effects of medication on emotional experience in patients with a psychotic disorder[33,34]. Lataster et al.[34] compared haloperidol and risperidone, considered tight-binding agents to the dopamine D_2 receptor, to olanzapine, a loose-binding agent. ESM/EMA was used to assess psychotic symptoms, mood, context, and appraisals of the current situation. The authors examined whether and how the relationship between dose and emotional experience and symptoms differed for tight- and loose-binding agents.

In another study, Lataster et al.[33] examined emotional responses

to switching from a traditional dopamine antagonist (olanzapine, pimozide, haloperiodol, or quetiapine) to a partial dopamine agonist (aripiprazole). ESM/EMA was conducted at baseline when patients were still taking the traditional medication and again after five weeks of aripiprazole. The authors examined how mood and psychotic symptoms changed as a result of switching the medication.

In sum, ESM/EMA has been used in patients with a psychotic disorder to increase insight into the working mechanisms of medication with respect to emotional experience and symptoms in daily life.

Anxiety disorders

To date, there has been one ESM/EMA study on the pharmacological treatment of anxiety disorders in adults[36] and one in youth[35].

In adults with panic disorder, Sheehan et al.[36] compared the extended release (XR) formulation of alprazolam to the compressed tablet (CT) formulation. Patients were prescribed alprazolam-CT for three weeks and were then switched to six weeks of alprazolam-XR. For the entire treatment period, patients indicated every waking hour how much anxiety relief they experienced from the medication. The study examined the rate at which peak benefit was obtained as well as the duration of effectiveness.

Forbes et al.[35] studied the effects of pharmacological treatment on the affective and social dynamics of youth with an anxiety disorder and/or MDD. Participants could choose among a selective serotonin reuptake inhibitor, cognitive-behavioral therapy, or a combination. Before, during, and immediately after treatment, ESM/EMA was used to monitor mood, current activities, and the presence of others, during five weekends (Friday-Monday). Baseline ESM/EMA variables such as negative affect, positive affect, and social time expenditure (e.g., amount of time spent with parents versus peers) were used as predictors of changes during treatment and outcomes after treatment. Importantly, baseline momentary affect offered information on treatment response and course above and beyond the traditional baseline questionnaires.

In sum, studies on the psychopharmacology of anxiety disorders have

shown that ESM/EMA can be applied (1) to study the effect duration of medication, (2) to study the role of social time expenditure on treatment course and response, and (3) to predict the response to and course of treatment by assessing the dynamics of mood before and during treatment. This yields additional information over and beyond traditional measures.

Discussion

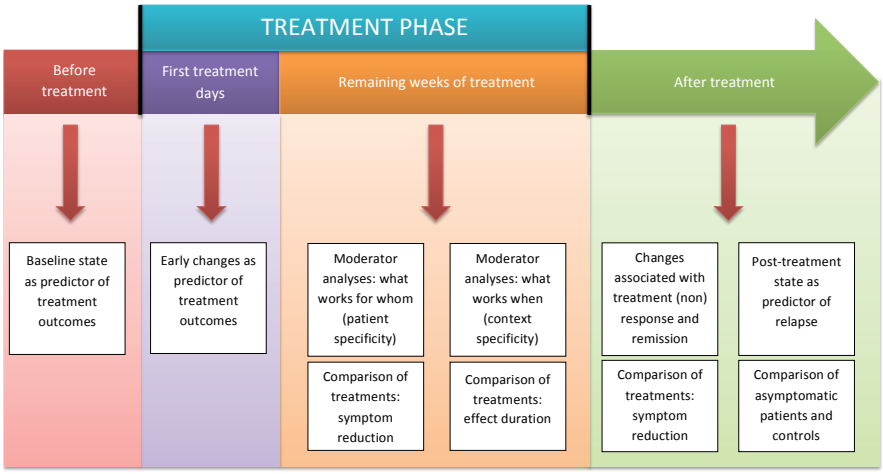
ESM/EMA has been applied to study pharmacological interventions for MDD, substance use disorders, ADHD, psychotic disorders, and anxiety disorders. Here we consider those applications that are particularly relevant to psychopharmacology. We also mention directions worth pursuing in future research and we touch upon the potential of ESM/EMA to be integrated into clinical practice.

Applications of ESM/EMA for psychopharmacology

Our review shows that ESM/EMA might help characterize patients before, during, and after pharmacological treatment (Figure 6.1). Before treatment, baseline affective characteristics assessed using ESM/EMA were found to predict treatment outcomes[22,23,25,35]. This has mainly been studied in patients with MDD. For example, Höhn et al. found that the ability to savor positive affect for longer periods of time was associated with better treatment response[22].

During treatment, ESM/EMA has been used to examine subtle changes in affect and side effects in the first week of treatment as a predictor of later response and remission[5,21]. Further, ESM/EMA might be used to identify predictors of and protectors against relapse in the patients' natural environments[27]. This opens the possibility for clinicians to offer patients additional guidance in high-risk situations during treatment. Furthermore, ESM/EMA during treatment might offer psychopharmacologists better insight into when doses wear off. For example, using ESM/EMA, Sheehan et al.[36] were able to show that extended-release alprazolam is longer-lasting in anxiety disorder patients than the compressed-tablet formulation. Similarly, ESM/EMA data have revealed that atomoxetine may have a more enduring effect in children with ADHD than stimulants[32].

Figure 6.1. The value of ESM/EMA in different phases of treatment.



Another potential advantage of applying ESM/EMA during treatment may be its sensitivity to context-specific treatment effects. For example, using ESM/EMA, Gehricke et al.[31] were able to reveal that ADHD medication may specifically reduce concentration problems during stressful situations and smoking abstinence. Similarly, using ESM/EMA, Whalen et al.[32] showed that atomoxetine was more effective than stimulants in reducing ADHD symptoms in the morning (but not in the evening). This suggests that ESM/EMA can be used to illuminate the effects of medication in different situations and at different time points, which might ultimately help inform clinicians in which contexts additional help is needed.

Finally, after treatment, ESM/EMA might illuminate the effects of medication on several aspects of daily life, for example on symptoms[33], positive and negative affect[24], quality of life[5], the amount of time spent on daily activities[20], and the number of drinks consumed[29]. Moreover, ESM/EMA has been shown to help detect differences between healthy individuals and patients in remission even when traditional clinical predictors indicate normalization[5,32] and identify predictors of relapse after treatment termination in the patients' natural environments[26]. For patients receiving long-term treatment, ESM/EMA may help identify vulnerabilities that might need additional attention. For example, medicated children with ADHD

continue to show more negative affect and behavioral symptoms than healthy controls, even after having been successfully medicated for at least two months[32].

Across treatment phases, ESM/EMA measures have shown predictive value above and beyond traditional measures[21,23,35]. Peeters et al.[23] found that emotional reactivity predicted depression severity after one month of treatment over and above traditional clinical predictors, such as baseline depression severity, episode duration, and mean levels of reported mood. Also, adding early changes in positive affect to a model already containing early changes in HDRS scores significantly improved the accuracy of the prediction of depressive symptoms, response to treatment, and remission after six weeks of treatment[21]. Lastly, Forbes et al.[35] found that ESM/EMA-derived affect was predictive of the rate of change in symptoms over the course of treatment, whereas traditional self-report assessments of depressive and anxiety symptoms were not.

Limitations of this review

With this review, we aimed to give an overview of the use of ESM/EMA in examining the pharmacological treatment for all DSM-IV disorders. However, there are some limitations.

First, we did not include daily diary studies. In these studies, assessments are conducted only once per day, which is fundamentally different from ESM/EMA in that daily diaries cannot be used to examine within-day fluctuations in symptoms and mood. Further, daily diaries require participants to reflect on a day at the end of that day, rather than report about their current state, which might yield biased results[3]. Furthermore, unlike ESM/EMA studies, daily diary studies cannot offer insight into the context-specificity of medication effects and the immediate precipitants of drinking and smoking. In spite of these limitations, however, since daily diaries are less time-consuming and burdening for participants and researchers, they may still be used instead of ESM/EMA. For example, Lenderking et al.[37] found that diaries completed at the end of the day detected antidepressant onset and response more quickly than standard weekly assessments.

Second, our findings are based on a limited number of ESM/EMA studies,

some of which were conducted in the same data sets. As pointed out by an anonymous reviewer, due to the variation in the terminology that has been used to describe ESM/EMA, we might have overlooked some eligible studies. Nevertheless, these numbers in the field of psychopharmacology are likely very low, considering the growing popularity of ESM/EMA in the fields of psychiatry, clinical psychology, and biopsychology. We encourage replication of the present findings so conclusions about the usefulness of ESM/EMA can be more powerfully drawn. Moreover, as most ESM/EMA pharmacology studies were conducted in patients with MDD, they should be extended to studies conducted in other clinical populations.

Limitations of ESM/EMA

Before discussing how ESM/EMA may be used in future human psychopharmacological studies, we wish to point out that we believe these studies are very feasible. It has been argued that ESM/EMA may be too demanding for patients with severe mental illness, who are particularly likely to receive pharmacological treatment, resulting in a selection bias. Nevertheless, compliance levels were acceptable in patients with psychiatric disorders who did participate in ESM/EMA studies[38]. Some studies did not report any data on compliance, which limits our understanding of the reliability and validity of the conclusions that may be drawn. However, most other studies reported adequate compliance (Table 6.1). One study did report a high dropout rate, but this was unrelated to the inability to comply with the ESM/EMA protocol[33]. Overall, we believe it is feasible to include ESM/EMA in psychopharmacology research and practice. The study length and data sampling rate can be adjusted to the specific questions and the patients under study.

While we do not have any reservations with respect to feasibility, we do have some reservations about the validity of the ESM/EMA questionnaire items used to date. Past ESM/EMA studies have often adapted items from retrospective questionnaires. For example, questions about mood tend to be derived from the Positive and Negative Affect Scale (PANAS)[39]. One problem could be that the meaning of a certain item changes over the course of treatment for a given patient, a phenomenon known as response shift[40]. For example, at the end of the treatment

patients might feel as sad as at the beginning of the treatment, but more able to cope with it. If they would score the same ESM/EMA questionnaire item differently while they are feeling the same, then this would limit insight into the processes of change that occurred during treatment. Future research should determine the degree of response shift that might occur in ESM/EMA questionnaires applied to clinical psychopharmacology research.

Previous ESM/EMA research often took a paper-and-pencil approach to data collection, which precludes insight into whether patients actually complete the questionnaires as instructed. It has been suggested that compliance with paper-and-pencil ESM/EMA approaches is poor: patients may complete questionnaires with a significant delay or complete all delayed questionnaires at once, which retains memory biases[41]. These issues can be solved with electronic devices such as PDAs and mobile phones, which time-stamp entries and limit access to the questions to certain time intervals. Electronic data collection is thought to yield data that are as valid as those obtained using paper-and-pencil approaches[42]. Thus, collecting data electronically seems to have definite benefits. ESM/EMA software and technology that lessen the burden of data collection to the researcher have become increasingly available[43].

Interestingly, in spite of the growing interest to use ESM/EMA in pharmacological research, the Food and Drug Administration (FDA) guidelines on Patient-Reported Outcome Measures offer little guidance on how to properly conduct ESM/EMA in pharmacological research[44]. We believe clear guidelines are necessary to help researchers determine the appropriate methodological details and derive valid conclusions from the data.

Future directions

Studies on MDD indicate that ESM/EMA may be used to monitor patients at different phases of treatment, that this monitoring can help predict clinical outcomes, and that ESM/EMA may show superiority over traditional measures in identifying individuals at risk for relapse[5,21-25]. ESM/EMA may be used for similar purposes in patients with diagnoses other than the ones studied to date.

Two studies found that ESM/EMA was able to reveal the effectiveness of medications in different situations[31,45]. This context-sensitivity of ESM/EMA is highly relevant to psychopharmacology. For example, in future research, ESM/EMA might also be used to investigate the extent to which ADHD medication has different effects at home and at school.

In two other studies, ESM/EMA helped reveal the duration of the effects of psychotropic medication[32,36]. This opens up exciting possibilities of adjusting treatment to accommodate individual differences in treatment response due to varying metabolic rates, for example. Patients may then be able to benefit from treatment more optimally. Relatedly, as ESM/EMA involves the sampling of data at short intervals, it could help chart the effects of rapid-acting compounds with short-lasting effects such as ketamine for MDD[46]. Adding ESM/EMA to studies on experimental therapeutics might generate valuable novel insights.

ESM/EMA studies often yield many observations per participant and thus allow for the study of within-person processes, which is difficult to impossible with traditional retrospective measures[47]. Studies comparing groups assume that findings at the population level are generalizable to the individual, but this assumption is almost never appropriate when studying dynamic processes that change over time[48]. Consequently, the within-person approach is increasingly recognized as an important addition to the traditional between-groups approach[47]. This is highly relevant for psychopharmacology as it might lead to more person-tailored interventions. Moreover, adding ESM/EMA to psychopharmacology studies might generate new insights into the causal mechanisms underlying the effects of treatment on specific symptoms of a mental disorder. For example, Young et al.[49] argued that antidepressants might partially work by altering how depressed patients interact with others and that this could be studied using ESM/EMA.

Lastly, we encourage psychopharmacologists to consider how ESM/EMA could be applied to study the effects of medications on aspects of daily life other than affect and symptoms, which have been the focus to date. Social interactions constitute one relevant example, since they form a large part of daily experience and are often altered in individuals with psychiatric diagnoses (e.g.[50,51]). ESM/EMA

has already been used to examine the effect of tryptophan on social interactions in a non-clinical population reporting interpersonal problems[52]. Yet other aspects of life can be studied using ambulatory devices and even personal smartphone applications that measure e.g. physical activity, light exposure, sleep, and aspects of cardiovascular function[53], variables that may also be relevant to the field of psychopharmacology. For example, sleep disturbances occur in many mental disorders.

Clinical implications

ESM/EMA may become an effective tool in clinical practice[54]: ecological momentary interventions (EMI) can be integrated with pharmacological treatment to provide both psychopharmacologists and patients with person-tailored feedback on progress (EMI[55]). As EMI teaches patients how their responses to their usual environments change in the context of pharmacological treatment, it can bring about long-lasting psychological changes. Interestingly, the detailed feedback that ESM/EMA provides may stimulate shared decision making by patients and their clinicians[9]. Kramer et al.[56] recently investigated the effectiveness of adding EMI to pharmacological treatment. Results indicated that this was clinically more effective in reducing depressive symptoms than pharmacological treatment alone, and than pharmacological treatment plus ESM/EMA without person-tailored feedback.

Conclusion

We have shown how ESM/EMA has been used in psychopharmacology research to date. We have described how ESM/EMA can be applied in multiple phases of treatment to examine what happens at the micro-level and to predict future outcomes. It is becoming increasingly clear that ESM/EMA can provide unique insights in how daily life experiences might change as a result of treatment. With this in mind, researchers and clinicians may have obtained a powerful tool that can help optimize the care of individuals diagnosed with mental illness.

References

1. Moskowitz DS, Young SN. Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience.* 2006;31(1):13.
2. Reis HT. Why researchers should think 'real-world': A conceptual rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:3.
3. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:636.
4. Koster EHW, De Raedt R, Leyman L, De Lissnyder E. Mood-congruent attention and memory bias in dysphoria: Exploring the coherence among information-processing biases. *Behaviour Research and Therapy.* 2010;48(3):219.
5. Barge-Schaapveld D, Nicolson NA. Effects of antidepressant treatment on the quality of daily life: An experience sampling study. *JClinPsychiatry.* 2002;63(6):477.
6. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *PsycholAssess.* 2009;21(3):425.
7. Vermeeren A, Vuurman EFPM, Leufkens TRM, Van Leeuwen CJ, Van Oers ACM, Laska E, Rico S, Steinberg F, Roth T. Residual effects of low-dose sublingual zolpidem on highway driving performance the morning after middle-of-the-night use. *Sleep.* 2014;37(3):489.
8. Roth T, Hull SG, Lankford DA, Rosenberg R, Scharf MB. Low-dose sublingual zolpidem tartrate is associated with dose-related improvement in sleep onset and duration in insomnia characterized by middle-of-the-night (MOTN) awakenings. *Sleep.* 2008;31(9):1277.
9. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, van Bemmel AL, Peeters F, Delespaul P, van Os J. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta PsychiatrScand.* 2011;124(4):262.
10. Wilhelm P, Perrez M, Pawlik K. Conducting research in daily life: A historical review. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:62.
11. Conner TS, Lehman BJ. Getting started: Launching a study in daily life. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:89.
12. Moskowitz DS, Sadikaj G. Event-contingent recording. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:160.
13. 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.
14. Oorschot M, Kwapil T, Delespaul P, Myin-Germeys I. Momentary assessment research in psychosis. *PsycholAssess.* 2009;21(4):498.
15. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. *PsycholAssess.* 2009;21(4):486.
16. 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. *JAnxiety Disord.* 2014;28(8):925.
17. Haedt-Matt A, Keel PK. Hunger and binge eating: A meta-analysis of studies using ecological momentary assessment. *IntJEatDisord.* 2011;44(7):573.
18. American Psychiatric A. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* American Psychiatric Publishing; 2000.
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj.* 2009;339:b2535.
20. Barge-Schaapveld D, Nicolson NA, van dH, DeVries MW. Changes in daily life experience associated with clinical improvement in depression. *Journal of affective disorders.* 1995;34(2):139.

21. 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. *EurNeuropsychopharmacol.* 2011;21(3):241.
22. Höhn P, Menne-Lothmann C, Peeters F, Nicolson NA, Jacobs N, Derom C, Thiery E, van Os J, Wichers M. 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.
23. Peeters F, Berkhof J, Rottenberg J, Nicolson NA. Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. *Behaviour Research and Therapy.* 2010;48(8):754.
24. Wichers M, Barge-Schaapveld D, Nicolson NA, Peeters F, de Vries M, Mengelers R, van Os J. Reduced stress-sensitivity or increased reward experience: The psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology.* 2009;34(4):923.
25. Wichers M, Lothmann C, Simons CJP, Nicolson NA, Peeters F. The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: a momentary assessment study. *BrJClinPsychol.* 2012;51(2):206.
26. Cooney NL, Litt MD, Cooney JL, Pilkey DT, Steinburg HR, Oncken CA. Alcohol and tobacco cessation in alcohol-dependent smokers: Analysis of real-time reports. *Psychology of Addictive Behaviors.* 2007;21(3):277.
27. Holt LJ, Litt MD, Cooney NL. Prospective analysis of early lapse to drinking and smoking among individuals in concurrent alcohol and tobacco treatment. *Psychology of Addictive Behaviors.* 2012;26(3):561.
28. Muhonen LH, Lahti J, Sinclair D, Lönnqvist J, Alho H. Treatment of alcohol dependence in patients with co-morbid major depressive disorder—Predictors for the outcomes with memantine and escitalopram medication. *Substance Abuse Treatment, Prevention, and Policy.* 2008;3(20).
29. Tidey JW, Monti PM, Rohsenow DJ, Gwaltney CJ, Miranda R, Jr., McGeary JE, MacKillop J, Swift RM, Abrams DB, Shiffman S, Paty JA. Moderators of naltrexone's effects on drinking, urge, and alcohol effects in non-treatment-seeking heavy drinkers in the natural environment. *Alcoholism: Clinical and Experimental Research.* 2008;32(1):58.
30. Gehricke J-G, Whalen CK, Jamner LD, Wigal TL, Steinhoff K. The reinforcing effects of nicotine and stimulant medication in the everyday lives of adult smokers with ADHD: A preliminary examination. *Nicotine Tobacco Res.* 2006;8(1):37.
31. Gehricke J-G, Hong N, Wigal TL, Chan V, Doan A. ADHD medication reduces cotinine levels and withdrawal in smokers with ADHD. *Pharmacology, Biochemistry and Behavior.* 2011;98(3):485.
32. Whalen CK, Henker B, Ishikawa SS, Emmerson NA, Swindle R, Johnston JA. Atomoxetine versus stimulants in the community treatment of children with ADHD: An electronic diary study. *J Atten Disord.* 2010;13(4):391.
33. Lataster J, Myin-Germeys I, Wichers M, Delespaul PAEG, 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. *Ther Adv Psychopharmacol.* 2011;1(5):145.
34. Lataster J, van Os J, de Haan L, Thewissen V, Bak M, Lataster T, Lardinois M, Delespaul PAEG, Myin-Germeys I. Emotional experience and estimates of D2 receptor occupancy in psychotic patients treated with haloperidol, risperidone, or olanzapine: An experience sampling study. *JClinPsychiatry.* 2011;72(10):1397.
35. Forbes EE, Stepp SD, Dahl RE, Ryan ND, Whalen D, Axelson DA, Birmaher B, Silk JS. Real-world affect and social context as predictors of treatment response in child and adolescent depression and anxiety: An ecological momentary assessment study. *JChild AdolescPsychopharmacol.* 2012;22(1):37.
36. Sheehan DV, Sheehan KH, Raj BA. The speed of onset of action of alprazolam-XR compared to alprazolam-CT in panic disorder. *PsychopharmacolBull.* 2007;40(2):63.
37. Lenderking WR, Hu M, Tennen H, Cappelleri JC, Petrie CD, Rush AJ. Daily process methodology for measuring earlier antidepressant response. *Contemp Clin Trials.* 2008;29(6):867.
38. Hufford MR, Shields AL. Electronic diaries: Applications and what works in the field. *Applied Clinical Trials.* 2002;11:46.
39. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and

negative affect: The PANAS scales. *JPersSocPsychol.* 1988;54(6):1063.
40. Barclay-Goddard R, Epstein JD, Mayo NE. Response shift: A brief overview and proposed research priorities. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation.* 2009;18(3):335.
41. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *Bmj.* 2002;324(7347):1193.
42. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health.* 2008;11(2):322.
43. Intille SS. Emerging technology for studying daily life. In: Mehl MR, Conner TS, eds. New York, NY, US: Guilford Press; 2012:267.
44. Food and Drug Administration. Guidance for industry on patient-reported outcome measures: Use in medical product development to support labeling claims. *Federal Register 74.* 2009:65132-65133.
45. Whalen CK, Henker B, Jamner LD, Ishikawa SS, Floro JN, Swindle R, Perwien AR, Johnston JA. Toward mapping daily challenges of living with ADHD: Maternal and child perspectives using electronic diaries. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology.* 2006;34(1):115.
46. aan het Rot M, Zarate CA, Charney DS, Mathew SJ. Ketamine for depression: Where do we go from here? *BiolPsychiatry.* 2012;72(7):537.
47. Hamaker EL. Why researchers should think 'within-person': A paradigmatic rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:43.
48. Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the terson back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives.* 2004;2(4):201.
49. Young SN, Moskowitz DS, aan het Rot M. Possible role of more positive social behaviour in the clinical effect of antidepressant drugs. *Journal of Psychiatry & Neuroscience.* 2014;39(1):60.
50. Brown LH, Strauman T, Barrantes-Vidal N, Silvia PJ, Kwapil TR. An experience-sampling study of depressive symptoms and their social context. *JNervMentDis.* 2011;199(6):403.
51. Kashdan TB, Farmer AS. Differentiating emotions across contexts: Comparing adults with and without social anxiety disorder using random, social interaction, and daily experience sampling. *Emotion.* 2014;14(3):629.
52. aan het Rot M, Moskowitz DS, Pinard G, Young SN. Social behaviour and mood in everyday life: The effects of tryptophan in quarrelsome individuals. *Journal of Psychiatry & Neuroscience.* 2006;31(4):253.
53. Bruining N, Caiani E, Chronaki C, Guzik P, van der Velde E. Acquisition and analysis of cardiovascular signals on smartphones: Potential, pitfalls and perspectives: by the Task Force of the e-Cardiology Working Group of the European Society of Cardiology. *Eur J Prev Cardiol.* 2014;21(25):4.
54. Wichers M, Hartmann JA, Kramer IMA, Lothman C, Peeters F, van Bemmel L, Myin-Germeys I, Delespaul P, van Os J, Simons CJP. Translating assessments of the film of daily life into person-tailored feedback interventions in depression. *Acta PsychiatrScand.* 2011;123(5):402.
55. Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology.* 2010;15(1):1.
56. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, van Bemmel AL, Myin-Germeys I, Delespaul P, van Os J, Wichers M. A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry.* 2014;13(1):68.

Appendix

Table 6.3. Search terms.

Inclusion criterion	Search terms
ESM/EMA methods	(“diary” OR “momentary assessment” OR “experience sampling” OR “ecological momentary assessment” OR “event-contingent recording”)
DSM-IV axis I disorders	AND (“depressi*” OR “MDD” OR “major depressive disorder” OR “bipolar disorder” OR “mood disorder” OR “affective disorder” OR “anxiety disorder” OR “phobi*” OR “panic disorder” OR “PTSD” OR “post-traumatic stress disorder” OR “obsessive compulsive disorder” OR “acute stress disorder” OR “agoraphobi*” OR “OCD” OR “GAD” OR “psychosis” OR “schizoph*” OR “psychoti*” OR “eating disorder” OR “anorexia nervosa” OR “bulimia nervosa” OR “binge eating” OR “purging” OR “substance use” OR “substance misuse” OR “substance abuse” OR “drug use” OR “drug misuse” OR “drug abuse” OR “inhalants” OR “solvents” OR “impulse control disorder” OR ”conduct disorder” OR “oppositional defiant disorder” OR “ODD” OR “intermittent explosive disorder” OR “ADHD” OR “attention-deficit hyperactivity disorder” OR “autism” OR “Asperger” “pervasive developmental disorder” OR “PDD-NOS” OR “dissociation” OR “dissociative disorder” OR “sex*” OR “ericti*” OR “dyspareunia” OR “vaginism*” OR “orgasm*” OR “ejaculat*” OR “impoten*” OR “somatoform disorder” OR “somatization disorder” OR “conversion disorder” OR “pain disorder” OR “hypochondriasis” OR “body dysmorphic disorder” OR “sleep disorder” OR “insomnia” OR “hypersomnia”)
Psychopharmacological treatment	AND (“intervention” OR “treatment” OR “pharmacotherapy” OR “psychopharmacol*” OR “medication” OR “antidepressant” OR “antipsychotic” OR “anxiolytic” OR “benzodiazepine”)

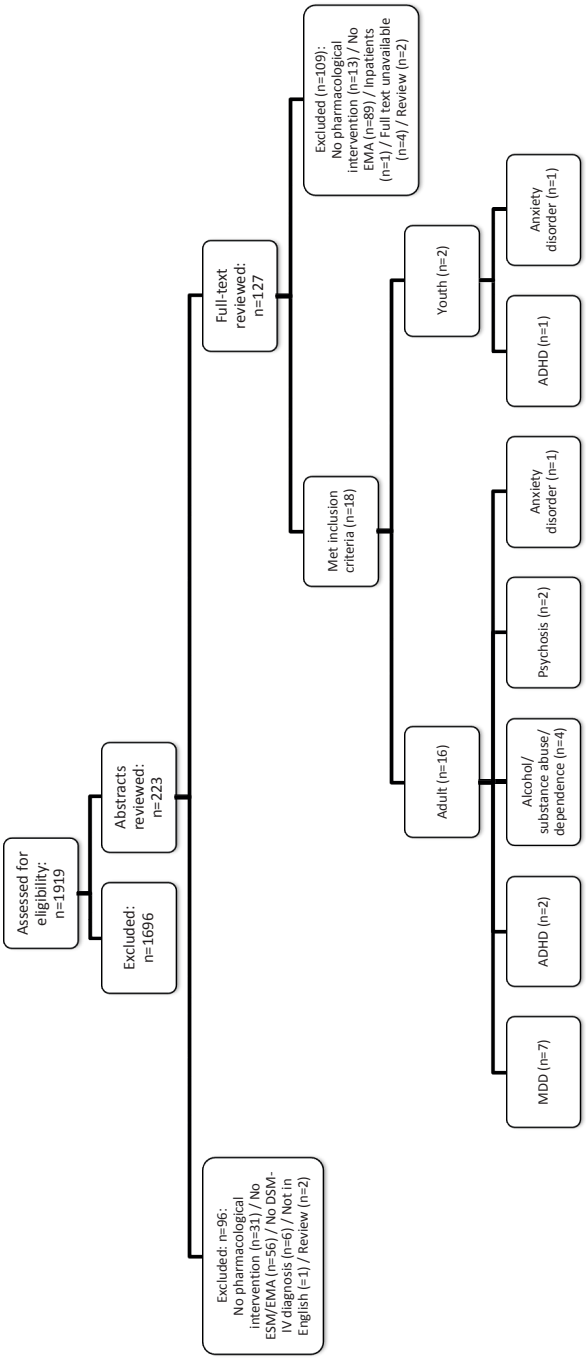
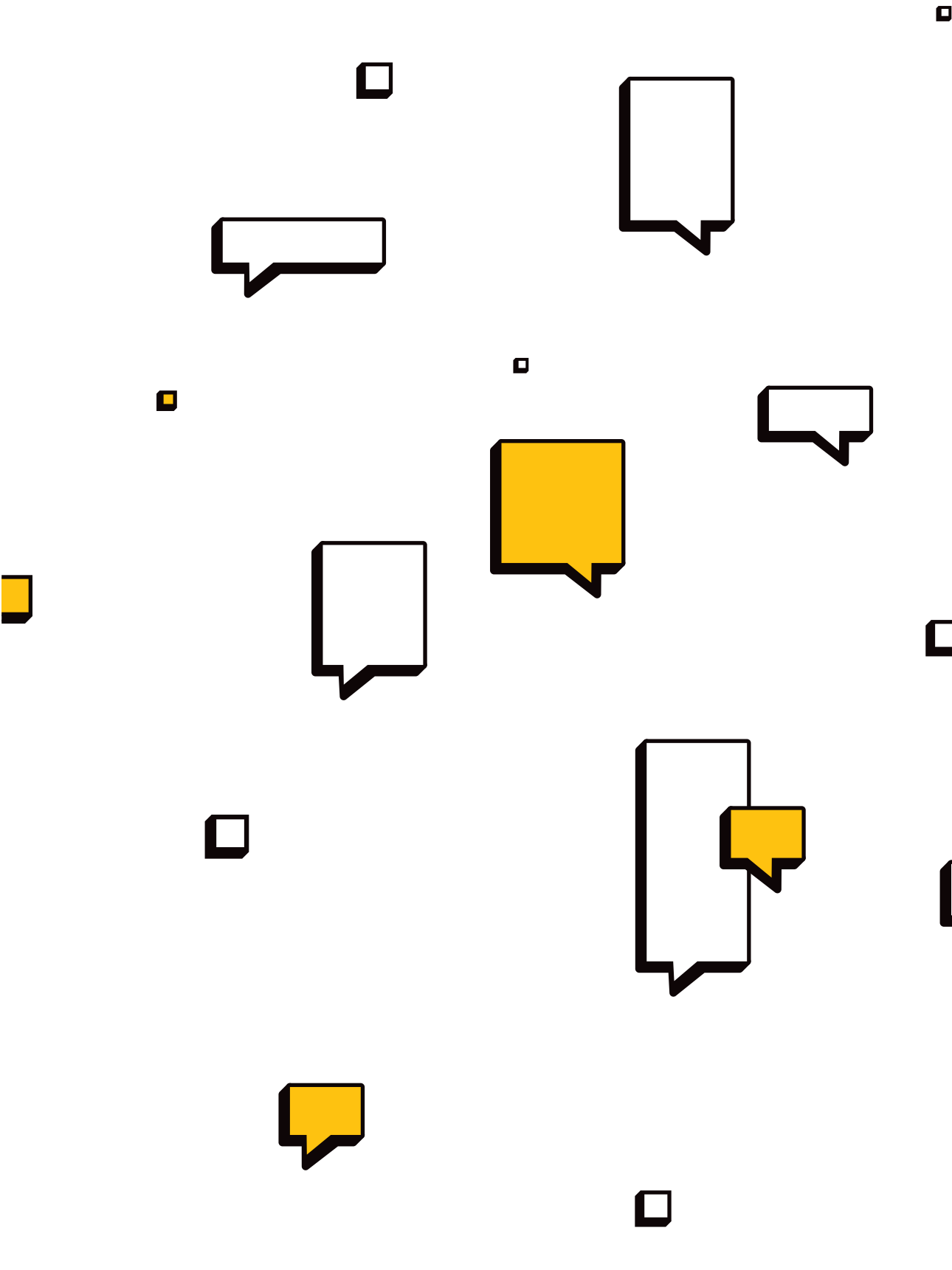


Figure 6.2. Overview of the selection procedure.



Chapter 7

Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care

Published as:

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;70(11):983-991.

Abstract

Objective. The current qualitative study aimed to map the relevance of the experience sampling method (ESM) for psychiatric practice and identify barriers and facilitators for implementation, as perceived by patients and clinicians.

Methods. Participants were 22 patients with various diagnoses and 21 clinicians (e.g., psychiatrists, psychologists) who participated in interviews or focus groups. Using Atlas.TI, qualitative thematic analysis was conducted to analyze the transcripts, resulting in four themes: 1) applications, 2) advantages, 3) undesirable effects, and 4) requirements for implementation of ESM in care.

Results. Clinicians and patients believed ESM could be relevant in every phase of care to increase patients' awareness, insight and self-management, personalize interventions, and alert patients to rising symptoms. Further, ESM was expected to improve the patient-clinician relationship, lead to objective, personalized, reliable and visual data, and increase efficiency of care. However, participants warned against high assessment burden and potential symptom worsening.

Conclusions. This study provides first evidence that the potential of ESM is recognized by both patients and clinicians. Key recommendations for optimal implementation of ESM in psychiatric care include flexible application of ESM, collaboration between patient and clinician, regular evaluation, awareness of negative reactivity, availability to patients with different psychiatric syndromes, and implementation by an interdisciplinary team of patients, clinicians, researchers, and information technology specialists.

Introduction

The experience sampling method (ESM) receives increasing attention in psychiatry and holds the promise to greatly advance personalized health care[1]. ESM involves the repeated sampling of people's moods, symptoms, experiences, behaviors, and contexts[2]. Research has thus far applied ESM to elucidate the daily life dynamics of a myriad of psychiatric disorders[3]. Because ESM entails intensive self-monitoring and the resulting data can reveal individual models of associations among daily life experiences, ESM is assumed to also have relevance for psychiatric practice[4].

Although monitoring in itself may already benefit emotional self-awareness[5,6], supplementing ESM monitoring with personalized feedback might improve feelings of empowerment[7] and even symptoms[8], suggesting potential usefulness for both patients and clinicians. Indeed, researchers believe that ESM may provide micro-level information, difficult to catch by clinical impression, that can add to processes of diagnostics, treatment choice, and relapse prevention[9,10]. However, the general assumption that ESM can be of value to psychiatric care lacks a solid evidence base, and it remains unclear how and when ESM should be applied. Thus far, ESM in research was often of short duration (e.g., 5-14 days) with intensive sampling (3-10 times a day) and without personalized feedback[11]. In clinical practice, ESM can be expected to require a different form[8].

For health care innovations to be effectively introduced in clinical practice, premier stakeholders need to be included and barriers to implementation addressed beforehand[12]. This requires a currently unavailable in-depth qualitative study into the views of patients and clinicians on the opportunities of ESM for psychiatry. Only one qualitative study reported that patients with psychosis recognized the advantages of ESM, but did not include clinician views[13]. Although patients will be the primary users of ESM, clinicians might be important stakeholders in introducing ESM, and might use ESM themselves to inform treatment decisions[9,14]. Therefore, the present study is the first to 1) map the relevance of ESM for psychiatric practice and 2) inquire into barriers and facilitators for implementation through focus groups and interviews with patients and clinicians.

Methods

Participants

Reporting of this study is done according to the Standards for Reporting Qualitative Research[15]. Participants were psychiatric patients and clinicians. Patients received mental health care during the study or in the recent past. Clinicians were psychiatrists, psychologists, psychiatric nurses, or job coaches. Participants were selected with the aim of achieving maximum variation on age, gender, diagnosis (if patient), experience with ESM and/or mobile technology, and affinity with research. Participants were recruited through posters and contacts at mental health institutions until no new information was heard (data saturation). After they signed up, participants were approached through e-mail to provide more information on the study. They were also invited (but not required) to participate in an open-source ESM study[16], to try out ESM before participation.

Of the 31 patients who signed up for participation, 22 showed up and provided informed consent. The remaining 9 did not provide reasons for the no-show. Of the 23 approached clinicians, 21 clinicians showed up and provided informed consent. The remaining 2 clinicians were unable to participate because of time constraints. The institutional review board approved of the study.

Interviewers

Interviews were conducted by authors FB (M.Sc., female, Ph.D. candidate) and LK (Ph.D., female, postdoctoral researcher and psychologist). Both were trained in qualitative interviewing and analysis. Focus groups were conducted by FB as moderator, with assistance of ES (Ph.D., female, postdoctoral researcher) or LK. There was no contact between researchers and patients before the study. LK and FB knew some of the clinicians.

Interviews and focus groups

Interviews (on average 57 minutes) and focus groups (92 minutes) were conducted in several mental health care institutions and private practices in the northern Netherlands between June 2016 and February 2017. One focus group participant was later individually interviewed to elaborate on a potential downside of ESM she was reluctant to share in the focus group. Interviewers explained the study rationale and what ESM entails (see Appendix). ESM was explained as a method by which individuals can record their moods, experiences, behaviors, contexts, and thoughts multiple times per day on their smart phones[2]. Example items (e.g., "I feel relaxed") and possible ESM-derived feedback were shown, such as mood variation, mood during activities, and associations between mood and behavior[16].

A semi-structured interview guide was used to ask open questions (see the Appendix), covering 1) the usefulness of ESM in general and specific phases of care, 2) possible consequences of using ESM, 3) implementation in care, and 4) design of the ESM protocol. Example questions include: "What do you think of ESM?", "How do you view the implementation of ESM in clinical care?" and "Do you see possible risks or downsides to ESM?" All interviews and focus groups were audio-recorded and field notes were made.

Data analysis

The digital audio recording of each interview and focus group was transcribed verbatim. Thematic analysis was applied by FB and LK according to the Qualitative Analysis Guide of Leuven[17]. This approach involves the identification of central themes in the transcripts, which are iteratively verified against the data.

First, all transcripts were summarized in conceptual interview schemes and narrative reports to gain a holistic understanding of the participants' experiences. Next, a concept code list was constructed based on subthemes identified in the data (e.g., time-investment). FB and LK used this code list to independently code the transcripts in Atlas.TI (version 8). Throughout this first round of coding, new codes were created when previously unidentified themes were encountered, and existing codes were more clearly defined through consensus. Hereafter, the code

list was finalized and used in a second round of coding.

The codes were grouped in four overarching coding categories or central themes. These central themes were verified against all transcripts and discussed with ES and MW. Participants were invited to provide feedback on a summary of the central themes.

Results

Four themes were identified (see Figure 7.1). Participant characteristics are described in Table 7.1. For illustrative quotes related to the themes, see Table 7.2 and 7.3.

Figure 7.1. Schematic overview of all themes, theme 1 (applications), theme 2 (advantages), theme 3 (undesirable effects and limitations), and theme 4 (requirements for implementation).

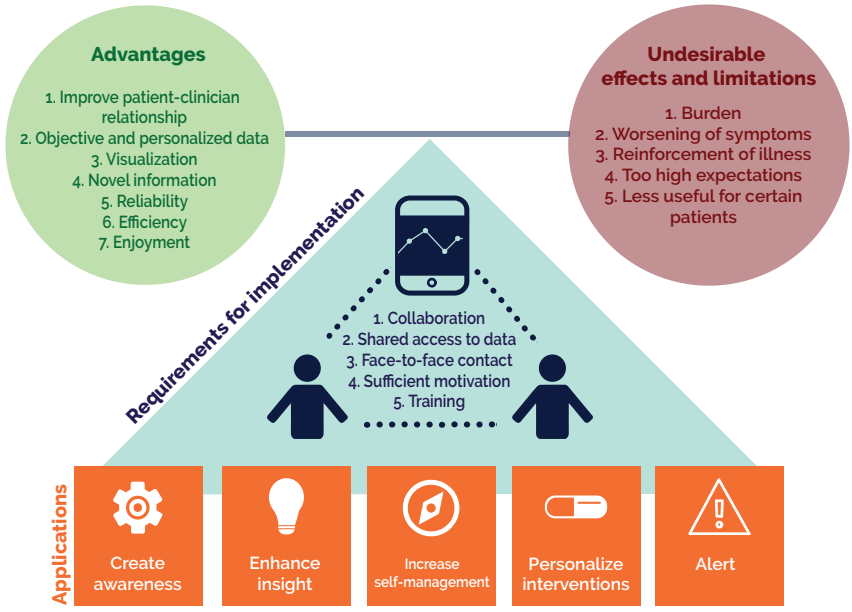


Table 7.1. Demographic and clinical characteristics of patients (N=22) and clinicians (N=21).

	Patients	Clinicians
Gender		
Male	8	13
Female	14	8
Age		
20-35	6	8
36-50	7	8
51-65	7	5
66 or older	2	0
Experience with ESM		
No previous experience with ESM	16	17
Started participation in ESM try-out study	6	4
Used ESM in clinical practice	0	0
Education level		
Higher education	12	
Secondary vocational education	9	
High school	1	
Profession		
Psychiatrist		4
Psychologist		13
Psychiatric nurse		3
Job coach		1
Self-reported diagnosis		
Depression	10	
Bipolar disorder	7	
Anxiety disorder	4	
Psychosis	3	
Eating disorder	1	
Autism spectrum disorder	1	
Unknown	1	
Years in treatment		
<1 year	4	
1-5 years	8	
>5 years	8	
Unknown	2	

Note. Most patients indicated multiple diagnoses. ESM = experience sampling method.

Theme 1: Applications

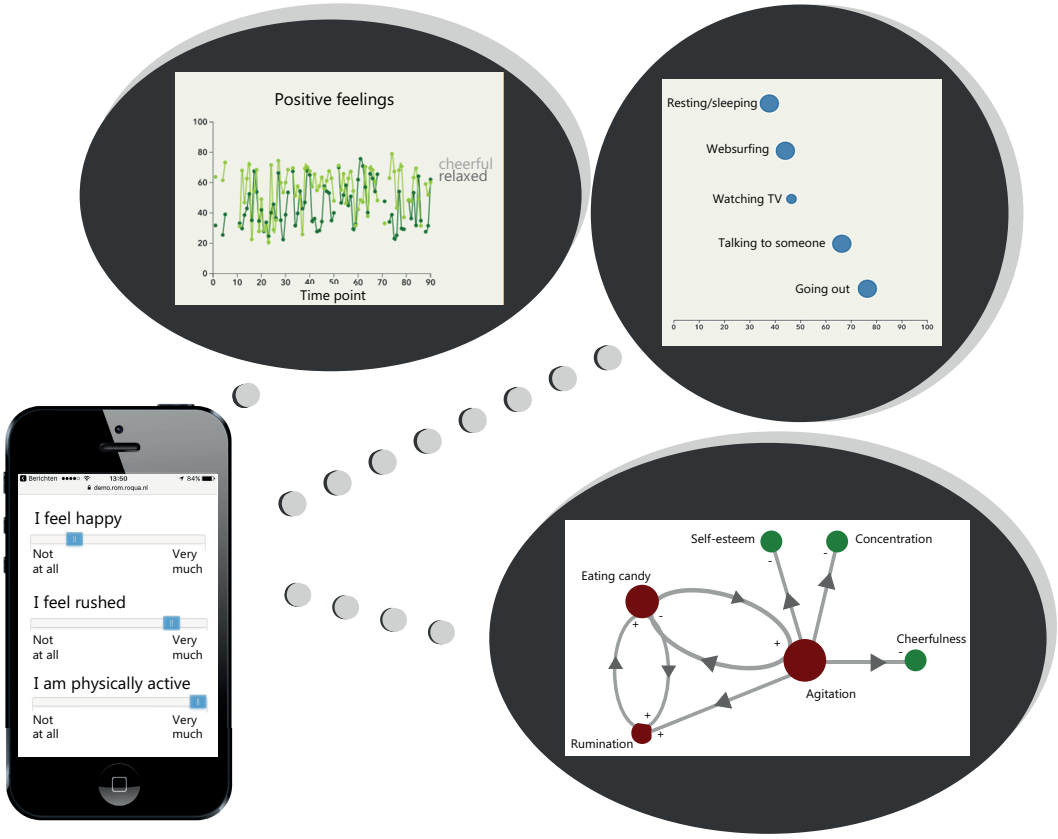
Most patients and clinicians believed ESM could be applied flexibly in every phase of care, from diagnosis to relapse prevention, depending on the patients' care needs. First, by monitoring symptoms, experiences, and contexts multiple times a day, many patients and clinicians suggested that ESM could be used to help the patient focus on the present and increase real-time awareness of what influences their symptoms.

Second, all patients and clinicians believed that ESM and ESM-derived feedback (see Figure 7.2) could offer relevant insights on 1) the severity of symptoms and variation therein, 2) short- and long-term associations between symptoms, experiences, behavior, context, medication, drugs, and life events, 3) symptom reduction, and 4) patterns building up to symptoms in smaller time-windows (e.g. panic attack) or larger ones (e.g. depressive episode). As such, most patients and clinicians believed that ESM could be applied to strengthen patients' self-management by providing them with concrete insights on how to cope with their symptoms.

Patients and clinicians also discussed employing ESM to determine intervention effects, thereby guiding decisions regarding future course of treatment. The majority of patients and clinicians suggested that the personalized nature of ESM has the potential to convince patients to start or continue interventions or behaviors if ESM-derived personalized feedback demonstrates its effectiveness.

Finally, multiple patients mentioned ESM might be used to alert patients and their clinicians of elevated ESM scores. Several patients argued that such alerts could help them notice the beginning of a downward spiral and could easily update clinicians on how they are doing. Possibly, (personalized) therapeutic advice could be attached to these alerts, to help patients directly alleviate symptoms and practice treatment strategies in daily life. However, several clinicians were hesitant of the possibility of receiving alerts, worrying about patient safety, responsibility, and time constraints.

Figure 7.2. Examples of ESM items and ESM-derived feedback that were shown to participants.



Note. Examples adapted from the HowNutsAreTheDutch study[16].

Theme 2: Advantages

Patients and clinicians identified several advantages of ESM for clinical practice. First, ESM may benefit the clinician-patient relationship by providing a framework for shared decision making. Multiple patients indicated that ESM may help articulate their experiences, consequently making them feel more heard and understood. As such, ESM was believed to lead to better mutual understanding between patient and clinician and provide a larger role for the patient perspective.

Second, ESM was generally seen as resulting in data that is 'personalized', 'neutral', 'objective' and 'nonjudgmental'. These characteristics of ESM were contrasted to receiving explicit advice or insights from clinicians, which patients do not always accept. Personalized and objective ESM data was perceived as convincing and seen as the key to gaining insight and changing behavior, especially if the interpretation of ESM-derived feedback is not imposed on patients by clinicians.

Third, the majority of patients and clinicians believed that ESM provides a more reliable overview of a given period than asking the patient or administering a retrospective questionnaire. These patients indicated a difficulty in stating how they have felt since the previous session, which is often influenced by current mood. Some clinicians and patients with bipolar disorder mentioned that ESM also maps mood fluctuations more accurately than once-a-day mood questionnaires such as the LifeChart[18].

Fourth, many patients and clinicians expected ESM to result in novel information because ESM 1) has more items than traditional registration strategies and focuses more on mood, experiences, behavior, and context rather than symptoms alone, 2) illuminates the time between treatment sessions, otherwise difficult to capture, 3) may lower the threshold to disclose sensitive information, and 4) offers the possibility of automatically generated models of symptoms and contexts (e.g., network analysis) otherwise unavailable to patients and clinicians. This may also enhance efficiency according to some patients, because problem areas can be found faster with ESM than with current, mostly retrospective, methods.

Fifth, some clinicians mentioned that the visual nature of ESM-derived feedback may help explicating associations normally verbally discussed in therapy.

Table 7.2. Quotes related to theme 1 (applications) and theme 2 (advantages) of interviews and focus groups with patients and clinicians on the use of the experience sampling method (ESM).

Participant	Quote
Theme 1: Applications	
Male patient in his fifties (ID15)	Well, actually I just wanted to react, because when you get a sms like that, that says, 'what have you done the past part of the day?', then you can really make a connection between your mood and what you are doing. For example if you, when you've been outdoors, or have met people, are energetic and happy because of that. Or, if that morning it happens to be the case that: 'I have not seen anyone, I am on my own at the computer, I am completely run down and irritable'. So a connection could very well be made between loneliness, or being alone, and a bad mood. That is of course very interesting. And also, the time of the day. It could well be that you are generally just a lot more energetic and cheerful in the afternoon than in the morning. And then you could also be able to make connections between... Well, that is very interesting.
Male psychologist in his sixties (ID37)	Because of that he also becomes more active in his own process and maybe also in his own mental state. And I think he is explicitly challenged to start making connections. That almost doesn't happen now. Now he is in a kind of, almost in a kind of depressed vacuum you know? Where really nearly everything is hidden under the mist [...] Get some nuance in the day. If you don't have an eye for that, then at the end of the day (…) it can indeed just seem very bleak, seem like a flat line, while, in terms of measurements, you can observe nuances in there.
Female patient in her twenties (ID12)	Yes, if for example you've filled in the whole week: 'I think life isn't worth living', that it then sends a signal and that an action like that might be taken. Because, well, what [ID9] already said, eventually you have reached a point that you don't ... that you can't fill in a list like that anymore. But I think that for a lot of people you can already somewhat notice that things are really going completely wrong.
Theme 2: Advantages	
Female patient in her fifties (ID17) and female patient in her sixties (ID14)	ID14: But what you said about, if you ... Look if you know that this app [ESM] is available, that does not mean that you will always use it. But when you think to yourself, hey, I think that I am doing a bit worse, you can start using the app again at that moment. To get some clarity on, well, how am I actually doing? Then you can ... ID17: Then you do not have to be dependent on your clinician. ID14: Yes, then you can indeed really put it to good use as a tool for yourself. (female patient in her sixties)
Female patient in her forties (ID7)	I think that for me it might result in me thinking: well, maybe I should try more to do something creative during the rest of the week, because apparently that helps me. Apparently it calms me down, I can relax more. So it can give you some insight into activities that you can undertake.
Male psychologist in his thirties (ID40)	Furthermore, what does get me enthusiastic, is the fact that a kind of network analysis is possible. How precisely that would go, I don't know. But I do think that you would often come up short with two people. In your knowledge, or in seeing connections. And if a bit of statistics can assist with that, then that is really good. It was almost a holistic theory, the way it was set out on paper. Those networks and so on. So diagnostically that could be very interesting. That you discover things, for example, which you at first you did not see at all. Like, if that, and that, that then it leads to that. And that then leads to something else. Well, that's fascinating.

Sixth, clinicians and patients expected smartphone-based ESM assessments to be less burdensome than paper-and-pencil registration, and less easily forgotten because patients are reminded through prompts. Some clinicians speculated that ESM may bring psychiatric care more 'up-to-date', thereby increasing resonance with patients' everyday environments.

Finally, some patients expected to enjoy the very process of monitoring, learning about themselves through ESM-derived feedback, and checking whether certain expectations are reflected in the data.

Theme 3: Undesirable effects and limitations

Patients and clinicians identified several potential undesirable effects and limitations of ESM monitoring or feedback. First, several patients and clinicians indicated that ESM could be burdensome when 1) assessments are too frequent or too long in duration, 2) assessments interfere with patients' activities, 3) patients already complete other questionnaires, 4) patients have to type in entries, and 5) ESM items are irrelevant to the patient. Burden was suggested to be reduced by clear delineation of the assessment period and letting the patient choose the timing and focus of the assessments.

Many patients and some clinicians feared that ESM monitoring will negatively influence patients' wellbeing or worsen symptoms. Some patients mentioned they might start dreading the assessments or feel guilty and incompetent if they miss assessments. Further, some participants mentioned that ESM may keep reminding patients of their symptoms rather than what goes well, which may worsen symptomatology, but could also help them acknowledge and handle their situation. Other plausible negative influences that were mentioned by one of the clinicians were 1) ESM monitoring becoming a ritual, 2) a constant focus on themselves rather than getting help, and 3) too much emphasis on symptom scores instead of the meaning of symptoms. Negative reactivity was suggested to be partially resolved by asking more neutral or positive questions.

Most patients and clinicians did not believe ESM-derived feedback will have negative consequences, but mentioned that these may arise when 1)

patients do not recognize themselves in the results, 2) ESM data does not reveal clear patterns, confirming patients' ideas that 'it does not matter what I do', or 3) important associations are uncovered, but impossible or difficult to change. Generally, clinicians believed it to be their task as a professional to help patients cope with these consequences, and indicated that this could also be a helpful learning process. Some clinicians warned for too high expectations of the relevance of ESM for clinical practice, emphasizing that it is only a tool and will not drastically change psychiatric care.

ESM was perceived to be applicable to all types of psychiatric syndromes, but some clinicians speculated it to be less suitable for 1) patients with limited insight in their symptoms (e.g., young children, patients with autism), 2) patients who prefer pills over psychological treatment, 3) patients with lower intelligence, 4) patients less comfortable with technology, 5) patients with insufficient mastery of the assessment language, 6) patients with neurocognitive deficits, 7) patients who keep asking for reassurance, and 8) psychotic patients for whom phone use may increase paranoia. Clinicians disagreed on the risks of ESM for patients with personality disorders, suicidal ideation, alcohol or substance use disorders, somatic symptom disorder and obsessive-compulsive disorder, wondering whether a constant focus on their symptoms worsens them.

Theme 4: Requirements for implementation

Several requirements for smooth implementation of ESM in clinical practice were described. First, all patients and clinicians agreed that ESM should be a collaborative process, where patients and clinicians decide together on 1) the relevance and feasibility of ESM, 2) clinician access to the data, 3) desirability of patient and clinician alerts 4) relevant items, 5) the frequency and duration of assessments, and 6) the interpretation of ESM-derived feedback. If not regularly evaluated, ESM might lose its advantages. Patients preferred ESM-derived feedback to be discussed by mental health professionals with whom they have a long-standing relationship, such as psychiatric nurses or experts by experience.

Further, it was generally viewed that both patients and clinicians should

have access to the patient's ESM data, and both should have a role in deciding when it is examined. Several patients assumed they will be the owner of their data, and that they can decide whether or not to share those data with others. Ideally, patients wanted to be able to initiate ESM monitoring themselves, but also recognized that without clinician involvement, ESM will be less effective in gaining insights and changing behavior. Some clinicians imagined that direct access to the data (not via the patient) is necessary to integrate ESM in treatment. However, some clinicians were concerned that continuous access to the patient's ESM data may enhance the power imbalance between the two. They further underscored that they cannot be expected to constantly monitor the data and act on elevated scores.

Third, a number of patients and clinicians stressed that ESM should never replace face-to-face contact. Contact with clinicians should not solely depend on ESM scores, and patients should be encouraged to ask for help directly rather than through ESM.

Fourth, a number of clinicians discussed how patients could be kept motivated. This starts with a proper rationale and patient input on relevant constructs. Some clinicians believed that certain patients will be sufficiently curious or in such distress that this in itself motivates them for ESM. Others argued that patients will need appropriate reward for their efforts, e.g. by continuous ESM-derived feedback, focusing on positive experiences, and giving advice and compliments. Motivation was believed to disappear if clinicians do not discuss feedback or when the patient has gained sufficient insights from ESM.

Fifth, several clinicians wanted to receive training on potential threats to the validity of ESM-derived feedback and the selection of the proper ESM protocol. This includes research-guided information on item formulation, assessment frequency and duration, minimum number of assessments, and feedback interpretation.

Finally, many patients and clinicians highlighted the limited time of clinicians, and indicated that user-friendly software and reimbursement from insurance companies might help clinicians to incorporate ESM in care.

Table 7.3. Quotes related to theme 3 (undesirable effects and limitations) and theme 4 (requirements for implementation) of interviews and focus groups with patients and clinicians on the use of the experience sampling method (ESM).

Participant	Quote
Theme 3: Negative consequences	
Male psychologist in his thirties (ID40)	I think the risk is that we will start hoping, or expect that the therapies will become more effective or something like that. But I am afraid that it isn't going to be like that....Something new hits the market and then all the attention is focused on it and all of a sudden everyone will have to do it. And then the insurers will back it. And then we all have to apply it. And that is a bit of a recurring wave in the whole health care system; that we then expect that this is going to do it. But I remain convinced that those kind of basic factors like motivation, discipline, mental distress and so on, that those will remain decisive for the success of therapy and not this kind of thing.
Female patient in her forties (ID22)	What I have noticed, and that's a bit of a drawback, is that each time there is a questions like 'I am tired', I discovered that I am actually always tired and I hadn't really expected that. I wasn't really aware of that. So since those questions I am much more aware, but now it also bothers me more. If I hadn't been made aware of it, I think it would not have bothered me so much. It is the other way around with other questions. It is also a bit more positive, 'oh how nice that I do still have that'. So there is that, but as far as tiredness goes, I really do think: 'yes, since I have been filling that in I actually noticed it'.
Theme 4: Requirements for implementation	
Female patient in the sixties (ID2)	Do you also include that it is a real issue? That for us it is not always really medical but it can be a very important contribution to our own sense of being in control of things. And that not everyone is used to that, so you have to be taught that, you have to be guided along in that, be guided along positively. That it is important that the therapist realises that. They do not have to do that all themselves, because some things you can delegate to other members of staff. But that even when you think, 'I can't take it anymore' that then a therapist just says: 'look, this is what you did it for.'
Female psychiatric nurse in her forties (ID43)	ID43: And I myself would not readily check it, irrespective of the patient. Because what would I do with it? As the therapist I can't interpret it. Because if this profile is the outcome for you, it means something different when it is the outcome for me. Interviewer: So, you should also do that interpretation with that person? ID43: Yes, I think so, yes. Really it belongs to the patient, but it can help me as a therapist to have the conversation with the patient.
Male psychiatrist in his forties (ID24)	So, how nice would it be if you could show a fantastic graph of the past months? That you can say to someone, just look at how you have filled it all in. So it should also be user-friendly for the therapist who can easily magic it up on his screen. That sort of thing is also a reward.
Male psychiatrist in his fifties (ID25) and male psychiatrist in his forties (ID24)	ID25: Yes, but it is very strongly a case of garbage in, garbage out, so when you put rubbish in..... ID24: You get rubbish out. ID25: Then you get rubbish out, and then you either see nothing, or you see things that are not right. So you have to carefully define what you are putting in before you put someone to work with it. And potentially it might not have any effect or even adverse effects. But I don't think anybody knows that.

Discussion

Main findings

The present qualitative study aimed to gain an in-depth understanding of 1) the relevance of ESM for psychiatric practice and 2) barriers and facilitators for implementation. Importantly, clinicians and patients recognized many of the applications and advantages of ESM also highlighted in research, such as the monitoring of treatment effects[19], the beneficial effects on awareness[5] and empowerment[7], the potential for shared-decision making[20], the increased reliability of the data compared to traditional assessment methods[21], and the possibility of real-time alerts on elevated scores[22]. The present study provides first evidence that these applications and advantages of ESM are indeed desired in practice. Our findings contrast to those of a previous qualitative study, which reported that although patients recognized the benefits of ESM, they were unsure of its relevance for their own situation[13]. However, the aforementioned study was limited to one specific 6-day application of ESM (without feedback) for a specific patient group (psychosis) and did not include the perspective of clinicians, which may explain the differing results.

Patients and clinicians stressed that successful use of ESM will depend on the active involvement of patients in the selection of the ESM protocol, interpretation of ESM-derived feedback, and subsequent action taken based on ESM. They further emphasized that the specific application of ESM should vary across treatment phases according to the patient's care needs. The need for clear agreements on data access became especially apparent when discussing real-time alerts. Although desired by patients, both patients and clinicians feared potentially adverse situations caused by not knowing whether the data were viewed and acted upon. Our findings are in line with research showing that tailored care and shared-decision making may improve patient satisfaction, treatment adherence, and health status[23]. They further highlight that patient involvement and flexible application are crucial factors for implementation of ESM.

Both patients and clinicians mentioned symptom worsening as a potential undesirable effect of ESM, because ESM may continuously make patients aware of their symptoms. However, studies among patients with substance abuse or pain disorder found little evidence of such negative reactivity in short-term ESM[24,25]; in fact, studies in psychiatric patients so far only reported favorable effects of self-monitoring[5,8]. Reactivity might vary according to specific patient characteristics, such as symptom severity, neuroticism, or readiness for change[26,27]. When implementing ESM in practice, reactivity will need to be controlled, as is also common practice in research settings, through careful construction and ordering of the items[3]. Nonetheless, some patients and clinicians worried that monitoring in itself might fixate patients on their illness, thereby hampering their autonomy. By providing constant reminders of their patient status, ESM walks a fine line between improving self-management and undermining it[28]. This potential downside was suggested to occur regardless of item content, and although ESM is suggested to benefit patient empowerment[7], future research will have to show for whom and under what circumstances this holds true.

General consensus was that most patients could benefit from ESM. However, clinicians expected that ESM might be less useful for patients with autism, paranoia, or substance abuse. Interestingly, patients themselves believed ESM could be relevant for all psychiatric syndromes, as is supported by research[3,29]. This suggests that the potential of ESM is not so much dependent on psychiatric syndrome, but rather on the willingness of the patient.

Finally, patients and clinicians highlighted that clinician training and research-guided advice are essential to guarantee the validity of ESM and minimize potential undesirable effects. These recommendations and our experiences with using ESM in practice have led us to believe that actual implementation of ESM can only be realized when researchers provide a framework that 1) translates clinical hypotheses to ESM protocols, 2) ensures that these protocols meet the strict rules also applied in research[27], and 3) provides valid interpretation of ESM-derived feedback.

Strengths of the current study include the in-depth nature of the interviews and focus groups, and the large and diverse participant sample, varying on age, gender, occupation, diagnosis, discipline, and experience with mobile technology. Further, by exploring the views of two premier stakeholders (patients and clinicians), our qualitative approach allowed us to formulate key recommendations on the utility and implementation of ESM.

In contrast to quantitative research, the goal of qualitative research is not to generalize but to describe and understand phenomena that may be time- and context-specific. As such, generalizing the results to other settings than The Netherlands should be done with caution. Furthermore, (most) participants in our study were asked to envision on the role of ESM in clinical care without having used the method; experiencing ESM might offer different results. Finally, most patients had mood disorders. Envisioned advantages and applications may differ for patients with other types of disorders.

Conclusion

This study provides first evidence that the relevance of ESM for psychiatric care is recognized by both patients and clinicians. Based on the study's findings, we suggest the following key recommendations for the optimal implementation of ESM. First, Patients and clinicians should apply ESM flexibly (across care phases) and collaboratively. Second, clinicians should make clear agreements with patients on data access. Third, patients and clinicians should be aware that patient moods or symptoms may worsen because ESM assessments remind patients of their symptoms. Fourth, patients and clinicians should regularly evaluate whether ESM helps or hinders patient self-management. Fifth, ESM should be applied to all psychiatric syndromes, and no patient group should be excluded a priori. Finally, ESM needs to be implemented by an interdisciplinary team of patients, clinicians, researchers, and information technology specialists. If these recommendations are followed, ESM might very well deliver on its promise for psychiatric care.

References

1. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, Drukker M, Reininghaus U, Jacobs N, Lataster T, Simons C, Lousberg R, Gülöksüz S, Leue C, Groot PC, Viechtbauer W, Delespaul P. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *DepressAnxiety.* 2017;34(6):481.
2. Larson R, Csikszentmihalyi M. The experience sampling method. *New Directions for Methodology of Social & Behavioral Science.* 1983.
3. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:636.
4. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, van Bemmel AL, Peeters F, Delespaul P, van Os J. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta PsychiatrScand.* 2011;124(4):262.
5. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, Viechtbauer W, Delespaul P, Myin-Germeys I, Wichers M. Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology.* 2016;35(5):433.
6. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, Sanci L, Patton G. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *JMedInternet Res.* 2012;14(3):e67.
7. Simons CJP, Hartmann JA, Kramer I, Menne-Lothmann C, Höhn P, Van Bemmel AL, Myin-Germeys I, Delespaul P, Van Os J, Wichers M. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. *European Psychiatry.* 2015;30(8):900.
8. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, van Bemmel AL, Myin-Germeys I, Delespaul P, van Os J, Wichers M. A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry.* 2014;13(1):68.
9. Kroeze R, van der Veen DC, Servaas MN, Bastiaansen JA, Oude Voshaar R, Borsboom D, Riese H. Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research.* 2017;3(1):1.
10. Fisher AJ, Boswell JF. Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment.* 2016;23(4):496.
11. 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.
12. Grol R, Grimshaw J. From best evidence to best practice: Effective implementation of change in patients' care. *The Lancet.* 2003;362(9391):1225.
13. Palmier-Claus J, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, Barkus E, Kapur S, Wykes T, Lewis SW. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: A qualitative study. *BMC Psychiatry.* 2013;13.
14. Kaiser T, Laireiter A-R. Process-symptom-bridges in psychotherapy: An idiographic network approach. *Journal for Person-Oriented Research.* 2018;4(2):49-62.
15. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine.* 2014;89(9):1245.
16. van der Krieke L, Jeronimus BF, Blaauw FJ, Wanders RBK, Emerencia AC, Schenk HM, De Vos S, Snippe E, Wichers M, Wigman JTW, Bos EH, Wardenaar KJ, De Jonge P. HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research.* 2016;25(2):123.
17. de Casterle BD, Gastmans C, Bryon E, Denier Y. QUAGOL: A guide for qualitative data analysis. *IntJNursStud.* 2012;49(3):360.
18. Leverich GS, Post RM. The NIMH life chart manual for recurrent affective illness: The LCM. *NIMH Monograph.* 1993.
19. Bos FM, Schoevers RA, aan het Rot M. Experience sampling and ecological momentary assessment

studies in psychopharmacology: A systematic review. *European Neuropsychopharmacology.* 2015;25(11):1853.
20. 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.
21. Schwarz N. Why researchers should think "real-time": A cognitive rationale. *Handbook of research methods for studying daily life.* 2012:22.
22. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: A systematic review and meta-analysis. *JMedInternet Res.* 2016;18(6):e152.
23. Joosten EA, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CP, de Jong CA. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. *PsychotherPsychosom.* 2008;77(4):219.
24. Sobell MB, Bogardis J, Schuller R, Leo GI. Is self-monitoring of alcohol consumption reactive? *Behav Assess.* 1989;11(4):447.
25. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain.* 2003;104(1-2):343.
26. Conner TS, Reid KA. Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science.* 2012;3(3):315.
27. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology.* 2008;4:1.
28. Pols J. *Care at a distance: on the closeness of technology.* Amsterdam University Press; 2012.
29. Chen Y-W, Cordier R, Brown N. A preliminary study on the reliability and validity of using experience sampling method in children with autism spectrum disorders. *Developmental Neurorehabilitation.* 2015;18(6):383.
30. Evers J. *Kwalitatief interviewen: kunst én kunde.* Boom Lemma uitgevers; 2015.
31. Weiss RS. *Learning from strangers: The art and method of qualitative interview studies.* Simon and Schuster; 1995.

Appendix

Description of the experience sampling method (ESM) to participants

The description of ESM was the same for all participants (patients and clinicians), whether they participated in focus groups or interviews. A PowerPoint presentation or handouts were used to show examples of items, delivery format, and ESM-derived feedback (summarized in Figure 7.2 in the main article).

ESM was explained as a method by which patients can record their moods, experiences, psychological/physical complaints, behaviors, experiences, activities, contexts, thoughts, or anything else of importance several times a day. Many participants were familiar with retrospective questionnaires in the context of treatment, such as the quick inventory of depressive symptomatology (QIDS). To contrast ESM to such questionnaires, we mentioned that most ESM studies so far had given prompts 3-10 times a day, but that the questionnaires took less time to answer (e.g., 1-2 minutes). Further, it was emphasized that ESM questions pertain to the present moment and not to longer periods (e.g., days/weeks/months). We also mentioned that patients could participate in ESM for as long as they wanted, ranging from days to months.

Regarding the content of ESM, a couple of example items were shown for clarity (e.g., I feel relaxed, I am upset, I experience physical complaints, I worry), but it was stressed that everything was possible here as long as items pertained to the present moment or the last couple of hours. If participants offered suggestions of things they wanted to measure, interviewers discussed if and how this was an appropriate ESM item.

Practically, we explained that participants would receive a text message on their smart phones with a link to the questionnaire and that they could use a slider to indicate the level of agreement to the items. Items were answered using a visual analogue scale, ranging from 0 (‘not at all’) to 100 (‘very much’). We explained that, in the example study (HowNutsAreTheDutch), participants had one hour to

complete the questionnaire, but that this could be a shorter or longer period.

After it was clear that participants understood the concept of ESM, and that all parameters (item content, schedule, measurement period, use in practice etc.) were subject to discussion, we showed several examples of graphical feedback (see Figure 7.2 in the main article). To briefly summarize, these feedback examples showed fluctuations in mood, mood patterns, frequency of activities, mood during activities, and associations between experiences/complaints in a network. The interviewers stressed that both the content and the graphical display of the feedback were subject to discussion.

Example ESM study (HowNutsAreTheDutch)

All patients and clinicians were invited (but not required) to participate in an open-source ESM study called HowNutsAreTheDutch (www.hoegekis.nl[16]). This invitation was intended to give participants an idea of what ESM could look like. In the HowNutsAreTheDutch study, participants complete assessments three times a day for thirty days on their own smart phones, after which they receive automated personalized feedback (see Figure 7.2 in the main article). Focus groups consisted of a mix of individuals that started the HowNutsAreTheDutch study (27% of patients and 19% of clinicians) and individuals with no previous knowledge about ESM. Given the short try-out period (most individuals started the HowNutsAreTheDutch study only a few days prior to the interview or focus group), responses to the interview and focus group questions were largely similar for patients and clinicians that did versus did not try out ESM. Further, this study's main focus of interest was the use of ESM in clinical practice, which was hypothetical for all participants.

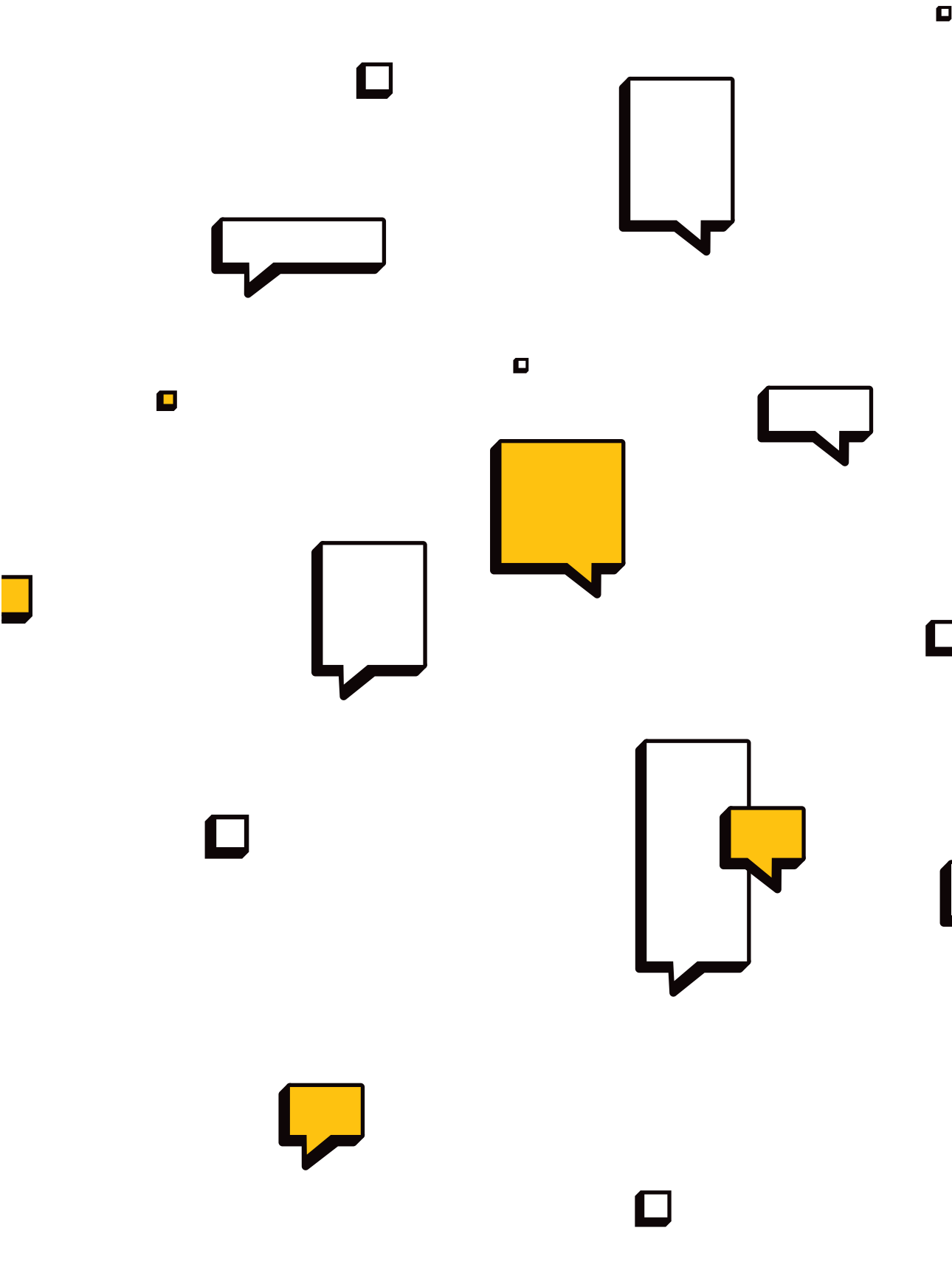
Interview questions

After the introduction, participants were asked the questions outlined below, not necessarily in this order. Participants could also raise the topic themselves and prompts were used to gain a detailed understanding of participants' thoughts[30,31].

Table 7.4. Interview questions and prompts.

Question		Specific prompts
1.	What do you think of ESM?	
2.	To what extent would you use ESM yourself?	How?
3.	What could be possible consequences of using ESM?	
4.	Do you see possible risks or downsides to ESM?	
5.	*Patients*: do you have an example of when you would use ESM yourself? *Clinicians*: do you have an example of a patient where you could use ESM?	
6.	*Clinicians*: are there patients where you would decide against using ESM?	What kind of patients? Why?
7.	*Patients*: would the way you get mental health care change through ESM? *Clinicians*: would the way you give mental health care change through ESM?	How?
8.	How do you view the implementation of ESM in mental health care?	Could you identify pitfalls?
9.	What would you want to do with ESM-derived feedback?	
10.	*Patients:* How would you want to receive ESM-derived feedback? *Clinicians:* How would you discuss the ESM-derived feedback?	*Patients:* Do you discuss it with your clinician or not? How?
10.	What kind of questions would you want to ask in the ESM-diaries?	
11.	What kind of clinical questions could you answer with ESM?	

Note: examples of generic prompts: "what does […] mean for you?", "can you elaborate?", "what do you mean by […]?", "can you give an example of […]?"



Chapter 8

Recommendations for the use of long-term experience sampling in bipolar disorder care: A qualitative study of patient and clinician experiences

Published as:

Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: a qualitative study of patient and clinician experiences. *International Journal of Bipolar Disorders.* 2020;8(1):1-14.

Abstract

Background. Self-monitoring has been shown to improve the self-management and treatment of patients with bipolar disorder. However, current self-monitoring methods are limited to once-daily retrospectively assessed mood, which may not suit the rapid mood fluctuations in bipolar disorder. The experience sampling method (ESM), which assesses mood in real-time several times a day, may overcome these limitations. This study set out to assess the experiences of patients and clinicians with the addition of ESM monitoring, real-time alerts, and personalized feedback to clinical care. Participants were twenty patients with bipolar disorder type I/II and their clinicians. For four months, patients completed five ESM assessments per day on mood, symptoms, and activities. Weekly symptom questionnaires alerted patients and clinicians to potential episodes. After the monitoring, a personalized feedback report based on the patient's data was discussed between patient and clinician. Three months later, patient and clinician were both interviewed.

Results. Thematic analysis of the transcripts resulted in four themes: perceived effects of the monitoring, alerts, and feedback, and recommendations for implementation of ESM. ESM was perceived as helping patients to cope better with their disorder by increasing awareness, offering new insights, and encouraging life style adjustments. ESM was further believed to facilitate communication between patient and clinician and to lead to new treatment directions. However, high assessment burden and pre-occupation with negative mood and having a disorder were also described. Patients and clinicians advocated for increased personalization and embedding of ESM in care.

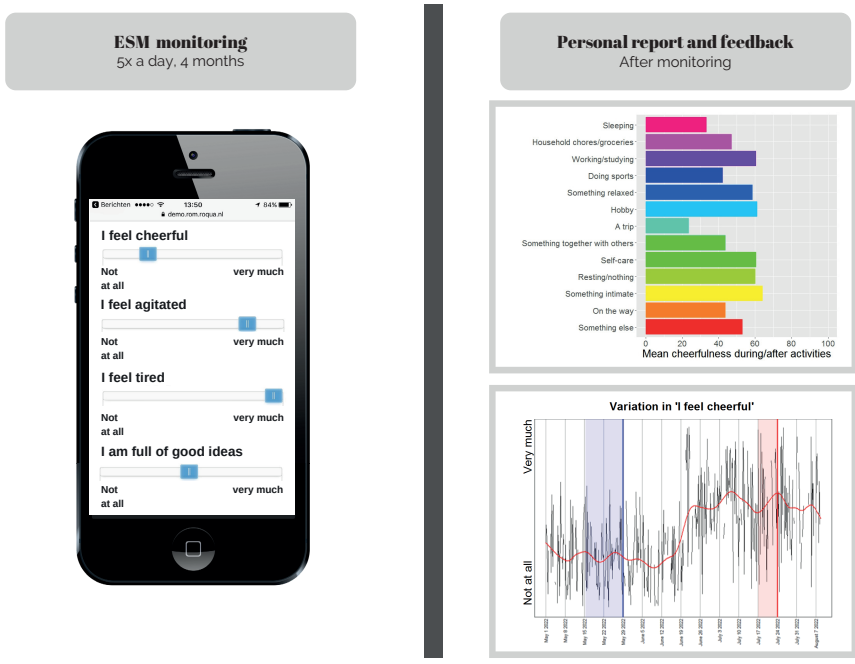
Conclusions. This study demonstrates that long-term ESM monitoring, alerts, and personalized feedback are perceived as beneficial to the treatment and self-management of patients with bipolar disorder. Future research should further test the clinical utility of ESM. Clinically relevant feedback and technology need to be developed to enable personalized integration of ESM in clinical care.

Background

Bipolar disorder is a severe and often lifelong affective disorder, involving depressive and (hypo)manic episodes, and is associated with tremendous burden for patients[1,2]. Many patients look for ways to successfully live and cope with the impact of the disorder[3,4]. An important strategy to bolster self-management is to self-monitor mood, for example with the NIMH prospective Life-Chart[5,6], ChronoRecord[7], smartphone applications, or rudimentary paper-based methods[8]. Indeed, research has suggested that self-monitoring may increase illness insight and self-management, by helping patients to make lifestyle adjustments and facilitate communication with clinicians[8,9]. However, existing methods such as the Life-Chart or ChronoRecord have two potential limitations. First, patients have to summarize their symptoms, experiences, and level of functioning into a *single* rating of their overall mood, which might not accurately capture the highly frequent and disparate mood swings patients experience throughout the day[8]. Second, the ratings are completed *retrospectively* over the previous 24h, which may lead to inaccurate information due to mood biases[10].

The experience sampling method (ESM) is an ecologically valid self-monitoring method that may overcome these shortcomings. Although originally developed for research[11], ESM has been advocated as a tool to promote self-management and resilience for patients in mental health care[12-14]. With ESM, patients monitor their daily experiences, moods, and symptoms on their smartphones several times a day (see Figure 8.1)[15]. ESM differs from more commonly used self-monitoring methods in that multiple distinct micro-level aspects of mood (e.g., feeling cheerful/down versus manic/depressed) are assessed prospectively, as they occur in daily life. Therefore, ESM may better match patient experiences, and may capture mood fluctuations more accurately[10]. The detailed ESM data that patients gather can be used to generate personalized feedback on mood variability or associations between mood and lifestyle, or alert patients to impending manic or depressive episodes.

Figure 8.1. Visual representation of the ESM monitoring and examples of personalized feedback given to participants.



Importantly, patients and clinicians recognize the potential of ESM for clinical care while also noting the possibility of high assessment burden and symptom worsening[13]. Still, little is known about how patients and clinicians actually experience the addition of ESM to clinical care. This is a crucial first step that will inform on future academic and practical innovations that are currently (technologically) unavailable but will need to be developed for clinical usage. Until now, qualitative studies have focused on feasibility and tolerability[16], evaluating only short-term (1-2 weeks) or low-intensity (1-2 assessments daily) ESM[17,18], whereas long-term ESM (e.g., several months) may better fit the prolonged nature of bipolar disorder treatments. Furthermore, although personalized feedback seems to be essential for the efficacy of ESM[19], previous qualitative work has been limited to the effects of ESM monitoring[20]. Finally, if ESM is to facilitate shared decision making

and patient-clinician communication[13], the combined experiences of patients and clinicians are needed.

The present qualitative study will therefore comprehensively assess patients' and clinicians' experiences with the addition of intensive long-term ESM monitoring, episode alerts, and personalized feedback to clinical care.

Methods

Participants

Reporting is done according to the Standards for Reporting Qualitative Research[21]. Participants are twenty patients with bipolar disorder type I or type II receiving treatment, and six clinicians who treated at least two patients participating in this study (see Table 8.1). Included patients met the following criteria: (1) ≥18 years of age, (2) diagnosed with and currently in treatment for bipolar disorder type I/II, and (3) demonstrate high occurrence of episodes (at least 2) in the previous year. Clinicians were selected based on their interest in using ESM in treatment.

Recruitment took place at two Dutch tertiary care institutions between May 2016 and July 2017. Six clinicians invited their patients to participate in the study until the intended cap of twenty participants was reached. After patients signed up, researchers telephoned them to provide more information. Interested patients were invited to the research facility, where the study was explained in detail.

Clinicians referred 28 patients to the study. Two patients were unreachable. Six patients declined participation during the first telephone call, expecting that study participation and the focus on mood would be too burdensome. This left a total of twenty patients that started and finished the study, of which eighteen were interviewed until data saturation was reached. All participants signed informed consent. The University Medical Center Groningen medical ethics committee approved of the study (201501161).

Table 8.1. Baseline demographic and clinical characteristics of patients (N=20) and clinicians (N=6).

Characteristic	Patients	Clinicians
Gender (N)		
Male	4	5
Female	16	1
Age (N)		
20-35 years	9	0
36-50 years	8	3
51-65 years	3	3
Education level (N)		
Higher education	9	
Secondary education	5	
Secondary vocational education	3	
Pre-vocational education	3	
Years in treatment or years of experience as clinician (*M, SD*)	10.6 (8.8)	16.4 (10.3)
Years since bipolar disorder diagnosis (*M, SD*)	6.4 (6.3)	
Bipolar disorder diagnosis (N)		
Bipolar disorder type I	9	
Bipolar disorder type II	11	
Comorbid diagnoses (N)		
No comorbid Axis I/II disorder	12	
Attention Deficit/Hyperactivity Disorder	1	
Autism Spectrum Disorder	1	
Sleep disorder	1	
Alcohol/drug dependence	1	
Personality disorder	6	
Medication use (N)		
None	2	
Amphetamine	1	
Anti-epileptic	10	
Atypical antipsychotic	10	
Benzodiazepine	9	
Thyreomimetica	2	
Lithium	5	
Monoamine oxidase inhibitor	3	
Selective serotonin reuptake inhibitor	4	
Tricyclic antidepressant	1	
Profession (N)		
Psychiatrist		3
Psychologist		1
Psychiatric nurse		2
Experience with technology in treatment (N)		
None		2
A little experience		4
A lot of experience		0

Note. ESM = experience sampling methodology; M= mean; N = number; SD = standard deviation.

Study design

The goal of this project was twofold: to examine whether ESM data can be used to detect early warning signals for mania and depression, and to qualitatively assess experiences with adding ESM to bipolar disorder treatment. The study design was developed to meet both goals.

Experience sampling methodology

Patients received five ESM prompts per day for at least four consecutive months. Every three hours (time-contingent schedule), patients received a text message containing a link to the ESM assessment on their smartphone. The assessments were securely administered and stored via RoQua (www.roqua.nl) in patients' personal health records (see the Appendix). Patients chose their own start and end time and had one hour to complete each assessment (which took approximately 1-2 minutes to complete). No reminder prompts were given. Researchers contacted patients after the first three days of monitoring and if compliance was low[22], if they preferred regular contact, or if the weekly questionnaires (see below) indicated a manic or depressive episode. No financial compensation was offered to participants.

Bipolar symptoms

Patients weekly completed the Dutch versions of the Altman Self-Rating Mania Scale (ASRM[23,24]) and the Quick Inventory for Depressive Symptomatology (QIDS[25,26]). Both patients and clinicians were e-mailed when scores exceeded cut-offs indicating potential manic (ASRM≥5[27]) or depressive (QIDS≥10[25]) episodes.

Experience sampling methodology items

The questionnaire consisted of 26 Dutch items pertaining to mood (e.g., cheerful, down), symptoms of bipolar disorder (e.g., racing thoughts, feeling inadequate), sleep, and activities (see the Appendix of Chapter 9). These items were based on previous ESM research[28-30] and interviews with three patients and a clinician on relevant constructs for people with bipolar disorder. Further, patients formulated one personal item they believed might be insightful to them. Most items were answered

on visual analogue scales ranging from 0 (“not at all”) to 100 (“very much”). Patients further had the option to write down anything in a comment field.

Compliance

Two patients were part of a pilot and therefore completed three instead of four months of the study. The eighteen other patients all completed at least four months of monitoring with an average of 18 weeks (*range*=16-32). Near study end, all patients were offered to continue the monitoring for their own benefit (without receiving feedback), which two did for an additional 4 and 14 weeks. Average compliance (number of completed assessments divided by the number of assessments participants received) was 76% (491 assessments, *SD*=137.8).

ESM feedback

Within one month after the end of the ESM monitoring, researcher FB constructed a personal feedback report (see the Appendix). This report contained information on (1) variation in mood and symptoms in general and related to time of day and activities, (2) frequency of activities, (3) occurrence of episodes in association with ESM items, (4) sleeping patterns, (5) personal questions, and (6) comment field entries. The feedback session took place on average 68 days (*SD*=33.0) after the monitoring period had ended. Clinicians explained the feedback, asked patients to interpret unexpected findings, and helped them formulate conclusions. All clinicians were briefed beforehand on the interpretation of the ESM feedback and researcher FB was present to provide explanations if necessary. If new questions arose (in five cases), the researcher ran additional analyses and e-mailed the results to both patient and clinician.

Interviews

Follow-up interviews were held with 18 patients and 6 clinicians and took place at the research facility between April 2017 and March 2018. The interviews were planned three months after the feedback session (*M*=107 days, *SD*=21.5) and had an average duration of 46.8 minutes. Participants were explained the goal of the

interview: to learn more about their experiences with adding ESM to clinical care. The interview started with the open question how participants reflected on their experiences with ESM. After fully exploring their first responses to this question, a semi-structured interview guide was used to ask open questions on (1) experiences with the ESM monitoring, the alerts, and the report and feedback session, (2) potential insights gained, (3) potential behavioral changes made, and (4) the utility of ESM for clinical care. Further, person-specific questions were asked using field notes made during the monitoring phase and the feedback session. All interviews were audio-recorded and transcribed verbatim. Interviews were conducted by FB (M.Sc., female, PhD-student) and ES (PhD, female, postdoctoral researcher), both trained in qualitative interviewing and analysis. At the time of the interview, all participants knew FB from the introduction and follow-up calls.

Qualitative data analysis

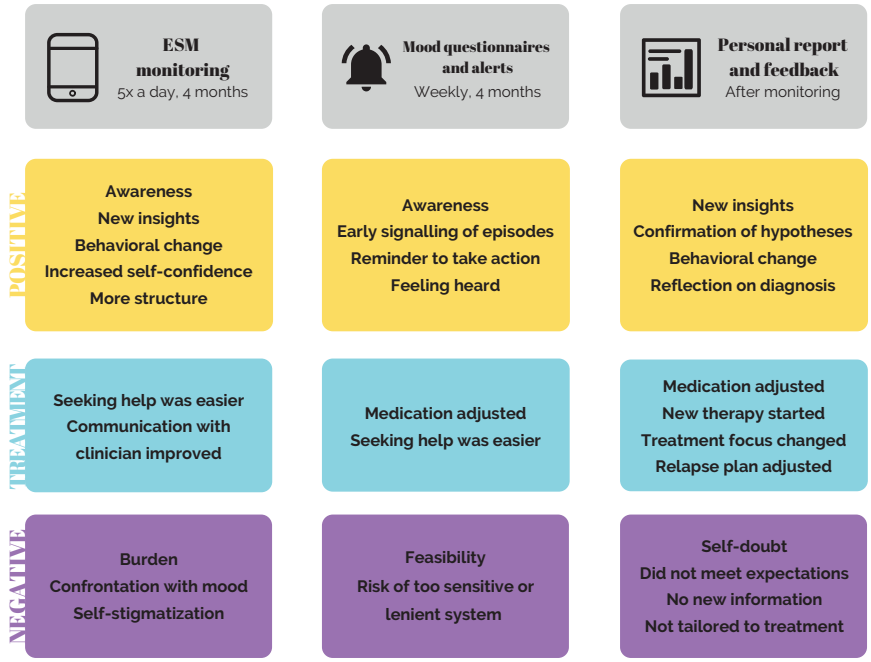
To understand patients' and clinicians' experiences with adding ESM to clinical care, thematic analysis was applied on the transcripts by FB, LK, and ES according to the Qualitative analysis Guide of Leuven[31]. Field notes and observations were analyzed to understand the effects of ESM monitoring while it took place.

First, transcripts were summarized in conceptual interview schemes and narrative reports to gain a holistic understanding of participants' experiences. Based on subthemes identified in the data (e.g., awareness), a concept code list was constructed. This code list was then used to code the transcripts in Atlas.TI (version 8), creating new codes when previously unidentified themes were encountered. The codes were then grouped in four central themes, which were verified against all transcripts and discussed with all authors. Participants were invited to provide feedback on a summary of the themes.

Results

Thematic analysis resulted in four themes (see Figure 8.2): 1) effects of ESM monitoring, 2) effects of the weekly symptom questionnaires and alerts, 3) effects of the personal report and feedback session, and 4) recommendations on the use of ESM in clinical practice. For each aspect of ESM, patients and clinicians described perceived positive and negative effects, and effects on treatment they attributed to ESM. See Table 8.2 and 8.3 for illustrative quotes related to the themes.

Figure 8.2. Schematic overview of the main findings of the first three themes: perceived effects of ESM monitoring, the weekly questionnaires (ASRM/QIDS) and alerts, and the personal report and feedback session.



1. Effects of ESM monitoring

Positive effects

First, almost all patients described increased awareness of their mood and symptoms, behavior, well-being, factors influencing mood, and positive aspects in their lives. Many patients contrasted the Life-Chart with ESM and highlighted the benefits of dividing their mood into multiple distinct components (e.g., agitated, easily distracted, racing thoughts) rather than focusing on mania or depression as a whole, making it easier to judge their mood and the likelihood of impending episodes. Several patients described heightened mood awareness even after the monitoring period was over. Clinicians confirmed seeing heightened awareness in their patients. Both patients and clinicians found this one of the most helpful aspects of ESM monitoring.

Second, many patients described gaining new insights on the impact of their behavior, activities, events, and time of day on their mood, which their clinicians confirmed. Other patients realized that their mood could also act independently of their behavior or context, which made mood swings easier to accept. Two patients learned that their mood fluctuated much more during the day than expected. Two other patients learned they had trouble recognizing their emotions, and felt that this improved during the monitoring period. One patient learned that her job was too heavy for her. One clinician gained new insights into his patients' coping strategies. Many patients and clinicians indicated that ESM monitoring helped patients discover these insights by themselves, thereby firmly consolidating them.

Third, more than half of the patients made concrete behavioral changes to influence mood and symptoms (e.g., becoming more active when feeling depressed, or slowing down when experiencing (hypo)mania). Clinicians confirmed this observation. Several patients continued monitoring their mood in a diary or the Life-Chart. One patient felt she could safely experiment with the strict behavioral rules she had learned during years of therapy and investigated the effects with the ESM monitoring. One patient described that the monitoring led to more meaningful conversations with her environment about what it means to have bipolar disorder.

Finally, a few patients described gaining more self-confidence in rating their mood and becoming more structured due to the steady rhythm of the assessments each day.

Effects on treatment

Some patients felt that the threshold to seek help was lowered, and that their increased awareness helped in the communication with their clinician.

Negative effects

Patients attributed several negative effects to the ESM monitoring. First, although many patients noted that the assessments were quick and easy to complete, most patients found five assessments per day burdensome because they felt disturbed in daily activities. Some felt they constantly kept (the possibility of) assessments in their minds and felt guilty or irritated when missing them. Nonetheless, most patients indicated that the benefits of the monitoring outweighed the burden, which aligned with the impression of their clinicians.

Secondly, approximately half of the patients felt negatively influenced by the frequent confrontation with their moods and symptoms, especially when already feeling depressed or (hypo)manic. When feeling depressed, several patients described that this confrontation made them feel restless, emotional, or irritated, and made them worry about their wellbeing. Alternatively, when experiencing (hypo)mania, several patients felt too busy and wanted to avoid reflection when feeling good. Of those describing a negative response, around half added that the confrontation actually helped them to become aware of their mood and either accept it, act upon it, or put it into perspective. The other half believed the assessments impaired their usual coping strategy, seeking distraction. Several clinicians agreed that confrontation might be difficult for patients, although increased awareness might mobilize patients into action.

Finally, a few patients felt that the assessments were a constant reminder of their diagnosis, the “unhealthy” part of themselves. They described that each assessment reminded them that they monitored themselves because of their bipolar disorder. In this way, ESM monitoring was perceived as self-stigmatizing, which two clinicians also warned against.

Table 8.2. Quotes related to theme 1 (effects of monitoring) and theme 2 (effects of weekly questionnaires and alerts).

Quote patients	Quote clinicians
Theme 1: effects of monitoring	
ID13: During the monitoring period I have mentioned this several times, but what really surprised me and helped me a lot was the compartmentalization in those five parts a day. That really was a revelation that I had never heard before in mental health care. Nobody had divided it in small pieces of three hours. Previously, I only had the Life Chart, once every 24 hours. A big yes, or a big no, or a big wow or a big 'bleh'. And now, something unpleasant could happen, and it would make sense that it makes me feel bad, or hyper, or sad. That will maybe last a part of three hours, but then my mood is… […] So this is what I learned from the monitoring, and I don't know if this was the intended effect of the study. But what I learned is to look at myself much more objectively, and much more relaxed. (female patient in her forties)	**ID22**: Yes, that they become much more aware of factors influencing their mood. And that really differs across persons. That's one thing. Or the fact that they become much more aware of their vulnerability in developing mood swings. An important part of treatment is about accepting that you are chronically instable. Some people keep wanting a sort of stable phase or that everything will be okay again, will return to how it was before. And that of course doesn't always work out like that. And this [ESM] holds up some sort of mirror for them, of course. (male psychiatrist in his thirties)
ID2: In the period that I filled in the assessments, I experienced several times that I answered that I hadn't been outside, for example, during a week I was at my mothers' place sitting around and being depressed. And then I had to answer three times that I hadn't yet been outside, and no, I didn't feel that well. You know, like that, and then I thought, well, okay I'll just go. That happened multiple times I think. (female patient in her twenties)	**ID25**: Yes, this really helps, you get to the point much more easily, and can give better targeted lifestyle advice. If they haven't already developed those insights themselves. That is what I believe to be the advantage of self-monitoring and assignments you can do at home, outside our conversations here in the clinic: that you can adapt your own behavior and make healthy choices, so that is a nice side effect of this study, I think. (male psychiatric nurse in his fifties)
ID17: Especially when the assessments come at an inconvenient moment. I find that very stressful. That is mostly the problem. The questions were completed in no time, but just, when I was in the car for example, a text comes, and I keep thinking, "I should not forget, I should not forget". That's it. No, the amount of work itself was not that much. (female patient in her forties)	**ID24**: On the Life Chart you can indicate that you score this or that, on average. A lot of people will then say that the actual situation is very different. So the micro-level is much more fine-grained. The danger is, though, that if people feel very bad, because their relationship has ended or I don't know, that they will immediately think that they have a depression. That the micro-level overshadows the macro-level. (male psychiatrist in his sixties)

Quote patients	Quote clinicians
ID8: By continually confronting you with it, you keep getting reminded of the fact you're doing badly. Or badly… Sad, that you're feeling sad. So then I found it hard to look at it another way. Because normally, I do that, I try to do things differently and find distraction and everything. But when I looked for distractions, I got a new assessment, making me think, "damn, I am indeed doing very badly". And that's what I found really annoying, or really annoying… I didn't like that. (female patient in her thirties)	**ID23**: Well, if there are people who keep getting hung up on it and keep feeling sad as a result, then I find that a negative consequence. But still, if that is the case, it suggests to me that we [patient and clinician] have to work on that. So in that sense, it can be helpful. (male psychiatrist in his forties)
ID15: Now I'm doing well, and I complete the Life Chart every morning. And then, for the rest of the day, I don't have to think about my having bipolar disorder. Because then I know that I'm okay, I don't have to pay attention to anything. But if you have to complete a questionnaire five times a day, then you really get confronted five times a day that you have that disorder. Throughout the day, you keep being confronted with 'you have a disorder'. (female patient in her fifties)	**Interviewer**: Was the burden too much? Weighing it against what patients gained from it? **ID23**: Maybe. Maybe it was too much, but you don't know that beforehand. That's why I think: you have to try. And self-management is a major step. So to invest a good amount of energy into that, because you have a severe disorder, you can invest a lot of energy into that, and then it is actually helpful to have something like this available to see if it gives you more insight. So in hindsight, yes it might have been burdensome, but I find that a bit too easy. Although, if you start using it now as a tool in clinical practice, then it might have to be toned down just a little. (male psychiatrist in his forties)
Theme 2: effects of weekly questionnaires and alerts	
ID14: I had expected that, when feeling more manic, or hypomanic, I would really find the questionnaire stupid. That's what I expected, but that happened actually right near the end of the ESM monitoring period, that I noticed "something is happening and I don't really trust it". And you also notified me of elevated things. And at that moment, that was actually really nice. Like, I really have to take step back. I feel fantastic, that's not it, but I hadn't realized yet that the scores were high until I saw it in the questionnaires. And then I could admit it more easily to myself, that maybe I had to take a step back. I will e-mail [clinician]. That was a really good experience that really helped me. Like: if I see it coming beforehand one way or another, because usually I notice it too late, then I experience everything less intense. (female patient in her twenties)	**ID22**: Well, I was really busy then, and then I also got those alerts and I thought: "do I have to do something with this as well?" That felt a bit as a responsibility, in a way. Whereas I always believe, you know, people really have to reach out themselves. That's what you teach them, that we don't take it all over and take care of them. So you really need to make clear agreements beforehand, like "what are we going to do when I see this?" And now, it just happened. I think it is something you can use in your treatment, but then you really have to discuss with patients, "what will we do, do you want me to reach out, or not? You get the alerts, do you appreciate that or not?" I think that is a good opportunity, but you have to think about this really well." (male psychiatrist in his thirties)

Quote patients	Quote clinicians
ID7: That I was not alone or let go in this. Because on the one hand, I am really inclined to go my own way and withdraw myself, really disregard everything and everyone. But back then I would consistently complete the questionnaires. And well, that by doing so I was not and could not be invisible. And actually, I like that. Because the withdrawing that I do, I actually don't want to do that. And then it helps if somewhere a graph shows: "this woman is not doing well. And I will e-mail her." (female patient in her fifties)	

2. Effects of the weekly questionnaires and alerts

Positive effects

First, several patients described that the alerts made them more aware of their mood and events of the past week, and possible ways to cope when mood worsened. For one patient, the alerts signaled a hypomanic state before realizing it herself, allowing her to get early treatment for the first time. Several others found that the alerts confirmed what they or their environment already suspected, which helped coax them into action. Second, many patients felt supported by the idea that someone (here, the researcher) would inform them of elevated symptoms. They described feeling heard and less alone.

Effects on treatment

For one patient, medication was occasionally adjusted upon receiving alerts. Several other patients found it easier to seek help and contacted their clinician. Clinicians differed in whether they contacted patients themselves or let patients seek contact and argued that this might differ across patients. The rationale given for not contacting patients was that recognizing and acting upon episodes are important treatment goals, and patients should be encouraged to seek help themselves. As such, several clinicians emphasized the need for making clear agreements with patients on who contacts whom in case of elevated scores. Clinicians also questioned the feasibility of acting upon alerts if sent for their entire caseload.

Negative effects

Patients did not describe negative effects. Two patients were curious of (but not bothered by) the timing of the alerts; they either received alerts when feeling well, or did not receive alerts when they signaled episodes themselves (e.g., QIDS score exceeded cut-off but for less than three weeks). Some clinicians suggested that a too sensitive alert system could lead to panic or demotivate patients.

3. Effects of the personal report and feedback session

Many patients and clinicians were positively surprised by the amounts of data collected by patients. The report was described as 'mapping yourself', and 'a personal history on a very small scale'. Many patients reread the report at home or showed it to close family or friends.

Positive effects

First, many patients described that the report helped them in gaining new insights on 1) the association between mood, behavior, sleep, and activities, 2) the presence or duration of emotions, 3) the benefits of healthy life style adjustments, 4) warning signals for manic or depressive episodes, and 5) the frequency and duration of behaviors and activities. Two patients used the personal report to specifically test ideas about themselves or their life style (e.g., effects of soft drugs on mood), which they saw confirmed. Many patients explicitly mentioned that the report was a more accurate, objective, and detailed reflection of their wellbeing during the monitoring period than their own account, which they described to be often discolored by present mood. Although clinicians did not describe new insights themselves, they believed that the report had been insightful to patients.

Second, several patients attributed lasting behavioral changes to the report. Two patients noticed the effects of sleep on their wellbeing, and started slowing down and cancelling appointments when they slept badly. Others learned how activities affected their mood, and started doing positive activities more often, even when feeling depressed. Finally, one patient concluded from the report that feeling tense preceded depressive episodes, and started taking medication when noticing tension. At the time of the interview, this had occurred several times, and

she felt that this strategy had made the episodes less intense. Many patients felt that the graphs objectively showed the benefits of certain behavioral changes in their own data, making it easier to accept and act on these insights. The visual nature of the report further helped some patients to 'get out of their head'.

Finally, the personal report made several patients reflect on their diagnosis: some described that it was helpful to once again realize and accept the impact of the disorder, which was also described by two clinicians. Although a few other patients had hoped that the report would give rise to a different diagnosis, they found it a helpful reminder that sometimes, they could feel bad for no apparent reason.

Effects on treatment

Five patients and their clinicians made changes in treatment that they attributed to the personal report. One patient started emotion regulation therapy based on her typed entries. For one patient, the relapse prevention plan was adjusted based on new insights in the personal report. For another patient, the focus of treatment shifted, because the report had shown that the previous focus was no longer relevant. Yet another patient was referred to a psychologist. Finally, medication was adjusted for one patient. Several other patients and clinicians mentioned that the report provided them with a helpful framework to discuss the course of treatment and the patient-clinician relationship. Several patients and clinicians had planned to adjust the relapse plan or refer the patient, but had not yet started the process.

Negative effects

A few negative effects were described. Some patients were apprehensive of the feedback session, thinking that the results might have implications for their diagnosis or treatment. One patient, whose data showed large mood fluctuations, started doubting her diagnosis and ability to judge her mood: contrary to her own experiences, she had inadvertently understood from the report that she had experienced no manic episodes.

Although many patients and clinicians mostly reflected positively on the personal report, they were also somewhat disappointed by it. Many had high

expectations, believing the report would suggest novel directions for treatment, indicate a different diagnosis, or provide the missing piece of the puzzle to help patients cope better with their complaints. They felt that the descriptive nature of the report did not meet their need for unambiguous conclusions. Some clinicians noted that so many factors appeared to influence mood and symptoms, that it was hard to find clear patterns in ESM data, especially because most patients showed erratic mood fluctuations. Some patients felt that the report only confirmed what they already knew or had learned during the monitoring period, or were unhappy that the monitoring took place during an unrepresentative period (e.g., during holidays or stable periods). A few patients further found the graphs difficult to understand.

Several clinicians added that the report was difficult to tie to treatment goals, and that it might have been more effective to discuss the feedback more frequently than just at the end of the monitoring period. Finally, two clinicians warned that ESM results could falsely suggest a hard truth, while ESM data supposedly only shows the patient's vision on their complaints. They stressed the need for clinicians to explain this to patients.

Table 8.3. Quotes related to theme 3 (effects of personal report and feedback session) and theme 4 (recommendations on the use of ESM in clinical practice).

Quote patients	Quote clinicians
Theme 3: effects of personal report	
ID2: Actually, last week a lot of people asked me to go here or go there, and I told them, "actually I'm not feeling so well, I'm not sure." And they started saying, "what does it matter, if you're already sick you should come anyway and be sick tomorrow". And then I thought, "no, I'm not doing this because I know how this will go, then it will happen that day and again and again the next days and before you know it, I'm really not doing well and that will have its effect on others as well." So in that way it worked maybe, a sort of small life lessons. Interviewer: Is that something that the report taught you, or? ID2: Well, to see that on paper, that really worked, those large mood swings. That you really have some sort of reflective moment. This is what happened then, and in that sense I think unconsciously shaped the way I think, I think. (female patient in her twenties).	ID25: For example, for one client, you later explicitly investigated sleeping, tiredness. After using cannabis, for example, the night before. Well, those very specific data coming out of the study are very helpful. Because I can have a very strong intuition that something is the case, but now we have it on paper, it is confirmed. Because she herself has supplied the data that shed a light on the situation. And there were more explicit outcomes: hours of sleep, energy, that's what we discussed, that's helpful to integrate in relapse prevention plans. (male psychiatric nurse in his fifties)

Quote patients	Quote clinicians
ID15: Now I know immediately that when I really start to worry and feel tense, it is the beginning of a depressive episode. If it lasts for about a week, I know that it is the start of a depressive episode, and I start taking medications, and I don't sink so deeply. Interviewer: Does this help? ID15: Yes, because I start taking my medications sooner. Because usually it takes a week or three before they start working, and then I notice a bit sooner that the tension starts to disappear. Normally I am really tense first and I no longer want anything at all, and because the medications take a while to start working, you reach the point that nothing works anymore, that you really have to fight to keep doing your daily activities. Now I get there on time, because I've already started taking my medications. So I don't sink so deeply anymore. (female patient in her fifties)	
ID14: Ah yes, I couldn't do anything with it [personal report]. But that maybe also has something to do with my expectations. I don't know what I'd expected. Probably I'd expected that I... that something about myself I didn't … That's what I'd hoped, maybe. I'd hoped that something would come out that would help. A piece of the puzzle. You know. You really want that it does something big. And it mostly was a confirmation of everything I already knew. And that is okay, but that is not what I'd hoped. Nice that I know myself better than I thought, I liked that, that there were no surprises. But I also thought, "and what now, now I have this, and what should I do?" So it didn't help me as much. (female patient in her twenties)	ID23: If you look at the results, it really is so hard to interpret them. It's still much more complicated than you had hoped beforehand. On the one hand, it's a lot of data and I like graphs and such, I think they're nice, you have a sort of overview, and well, about activities and such, it is solid. But what comes out as predictors disappoints me. Such that I think: it's not so unequivocal or it's not so easy to predict. Especially for people who are so instable in their mood, then the story gets even more unclear. (male psychiatrist in his forties)
Theme 4: recommendations for the use of ESM in clinical practice	
ID5: Like I'm saying, if you're young you're really inclined to go against everything. If somebody says something, you won't accept it, whereas if you've experienced it yourself, then you just know, you can't go around it. You see, without such a study it can take years before you've been through all that or have experienced a relapse or episode. And that's such a pity. Whereas if you can demonstrate such small changes with this study, they don't have to experience it all themselves. That they don't experience all the very heavy consequences, but see the small changes in themselves, which they have filled out themselves. (female patient in her forties)	ID25: I would really try to develop it tailored to the situation of the patient. And maybe link it to the relapse prevention plan. And it would be even better to also link it to the Life Chart method, for example. Or a sort of mood app, right? I mean, those exist, but are usually not so comprehensive. This way, you have all the information that you could use in treatment, and you have the aspect of self-management that can directly, in that moment, be adapted or stimulated even. (male psychiatric nurse in his fifties)

Quote patients	Quote clinicians
ID10: During therapy or something, it [ESM] might also be very easy. Then the system could directly inform your clinician, rather than bringing a copy yourself, so to speak. That they [the clinicians] could directly, if you give your consent, have insight in the data. And yes, the system doesn't lie. You can show that you have filled it out, at those moments. (male patient in his twenties)	ID26: The difficulty remains that this is a self-report measure, so people indicate their own visions on their problems. There are people, if you ask them a number between 1 and 10 to indicate their stress level, who will say a 10 with a very calm demeanor. Or the other way around, sitting there like thís [raising arms to indicate high stress level] saying, it's a 5. An app like this [ESM] will have it wrong too, people are not so good at judging themselves. So you always have to be aware that it is not a science, in fact it is their vision on their problems. (female psychologist in her forties)

4. Recommendations for the use of ESM in clinical care

Based on their experiences with ESM, patients and clinicians reflected on the use of ESM in clinical care.

For whom and when

Many patients and clinicians would especially recommend ESM for young, recently diagnosed people, for whom the insights yielded by ESM are still new. Clinicians also proposed patients in long-term care who keep having trouble and need a fresh perspective. Some patients and clinicians would refrain from ESM during 1) highly unstable or bad periods, because ESM could be too confrontational, 2) prolonged (depressive) periods without change, 3) good periods, when patients do not want to worry about their diagnosis, and 4) periods when there is insufficient suffering to motivate patients for ESM. Indeed, all but two patients opted not to continue the monitoring after four months, when they believed to have gained sufficient insights and additional data gathering was no longer necessary for the personal report. Several clinicians would not recommend ESM to patients already very pre-occupied with their disorder.

Goals

The majority of patients and clinicians advocated that ESM be tailored to the patient's situation and treatment goal. Many patients would consider doing ESM again when 1) symptoms increase, 2) specific hypotheses arise, 3) starting new

medications or treatments, or 4) they want to monitor for episodes. Clinicians highlighted the potential of providing alerts and advice in case of elevated ESM scores, to train patients in recognizing mood changes and getting help on time. Two clinicians deemed ESM more reliable and insightful than commonly used methods like paper-and-pencil registration or the Life-Chart [5].

ESM monitoring

Many patients and clinicians would have preferred the possibility to personalize the content, frequency, and duration of the ESM diary to better fit the patient's situation and treatment goal. Furthermore, many patients and clinicians liked the combination between micro-level ESM and macro-level symptom questionnaires.

Data access

Many patients and clinicians suggested discussing ESM feedback at every treatment session, to keep the feedback relevant and the patient motivated. Some patients would have liked to examine the feedback in between sessions. Others found an overview feedback report after several months more relevant. Some patients suggested that patients should be offered the choice to share their data with their clinicians, whereas most patients assumed that all data is automatically available to both themselves and their clinician.

Report

On top of the information included in the report, several patients and clinicians would like to have seen 1) effects of starting or stopping medication or treatment, 2) effects of life style adjustments, 3) testing of specific hypotheses, 4) decisions on diagnosis, or 5) simple explanations of the data analysis.

Clinician role

All patients and clinicians saw an important role for the clinician in the process of ESM, especially in discussing the personal feedback. Many patients preferred this clinician to be someone they know well and see frequently. Patients and clinicians further emphasized that clinicians should learn how to interpret ESM feedback and should believe in the potential of ESM. Patients and clinicians suggested the

feedback session is successful if the clinician 1) invests time to agree on the goal, content, and interpretation of ESM, 2) facilitates conversation on the meaning of the results, 3) teaches patients how to interpret ESM feedback, 4) is conscious of potential negative effects, and 5) makes clear agreements on what to do in case of elevated scores.

Discussion

The present qualitative study examined how patients with bipolar disorder and their clinicians experienced the addition of long-term ESM (i.e., 4 months, 5 assessments daily), episode alerts, and personalized feedback to clinical care. Confirming other qualitative work[20], most patients and clinicians reported that ESM monitoring and personalized feedback helped patients to cope better with their disorder by increasing awareness, offering new insights, and encouraging life style adjustments. In addition, as has been described as the promise of ESM[8,13,20], the monitoring and alerts were perceived as facilitating communication between patient and clinician. A relatively small group perceived the personalized feedback as helpful in informing treatment directions. As such, the present study is the first to demonstrate that long-term ESM is perceived to be a helpful tool for patients with bipolar disorder and their clinicians.

The present qualitative findings warrant further quantitative research to confirm both the positive and negative effects of adding ESM to bipolar disorder treatment. Potential beneficial effects of ESM should be weighed against the considerable burden and the risk of negative effects reported by several patients, namely the confrontation with (negative) mood and pre-occupation with having a disorder. Indeed, even though quantitative studies often report high compliance rates[32], qualitative work consistently demonstrates that patients and clinicians can be apprehensive of these effects[13,20]. For some patients, this may be sufficient reason not to participate in ESM. ESM seems to carry the risk of making a small yet significant group of patients hyperaware of their mood, especially during mood episodes[33,34], thereby cutting off adaptive coping strategies such as seeking distraction[35]. Therefore, when implementing ESM, clinicians have to be aware of

potential negative effects and discuss them with patients, whereas researchers have to ensure that ESM assessments minimize mood reactivity[36].

Notable was the finding that the descriptive nature of the personal feedback report disappointed patients and clinicians, who had hoped for more clear-cut advice on diagnosis and ways to improve well-being. Statistical analyses currently under development show promise in summarizing complex ESM data by predicting diagnosis[37], forecasting episodes[38], or testing associations[39]. These techniques could also inform on a personalized alert system that optimizes both sensitivity and specificity. However, these analyses are still relatively new and need empirical testing before they can be implemented in individual patients[40,41]. A recent study reported that, even when the same data was used, different analytical approaches led to highly disparate clinical advice across research groups[42]. Indeed, patients in our sample attributed high credibility to scientific data even when it contradicted their own beliefs. These results urge researchers to further develop valid models. Until they become available, researchers should carefully manage expectations of patients and clinicians on the possibilities of ESM feedback.

In the present study, the ESM diary, alerts, and personalized feedback were standardized and minimally embedded in clinical care. However, patients and clinicians indicated a desire for more personalization and integration in care, as these ESM components can be expected to differ across people, and even vary within patients, based on current treatment goals and the patient situation[13]. This is further exemplified by the finding that only 2 out of 20 patients continued monitoring when feedback and embedding in care were no longer provided. For the academic field, this means considerable effort needs to be invested into developing the necessary technology for the integration of personalized ESM in clinical care. This translation from research to practice is often overlooked[43,44] but imperative for its usability. Integration requires a user-friendly interface that helps patients and clinicians to construct a personalized and scientifically valid ESM diary, sets sensitive alerts if needed, and provides them with automatically available personalized feedback. Such an endeavor necessitates intensive collaboration between researchers, patients, clinicians, and software developers. Currently, our

research team is developing such a tool termed PETRA (PErsonalized Treatment by Real-Time Assessment, www.petrapsy.nl/en/).

Finally, if ESM should become available to clinical care, the question remains if, when, and how ESM can be useful and feasible for any given patient. For example, the low-intensity yearlong self-monitoring with the Life-Chart[5] or ChronoRecord[7] might suffice for some, whereas for others, the high-intensity and detailed ESM might better fit their needs. The sampling schedule will greatly depend on the intended goal of the self-monitoring: e.g., enhancing self-management, tracking treatment effects, or informing diagnostics. Therefore, many participants recommended involving both patient and clinician equally in all decisions regarding ESM: the goal and desirability of ESM, potential negative effects and burden, the ESM diary content and schedule, clinician involvement, and ESM feedback interpretation. Reaching agreement on these points beforehand is essential for enhancing patient self-management and empowerment through self-monitoring technology[35]. Based on their experiences, we have formulated crucial issues for patients and clinicians to discuss together when considering using ESM (see Table 8.4). These discussion points can help in deciding if and how to use ESM in treatment.

Strengths of our study include the prolonged ESM monitoring period, which so far has not been reported elsewhere, and our in-depth exploration of patients' and clinicians' experiences with all aspects of ESM (long-term monitoring, alerts, and personalized feedback). Their suggestions can inform future academic and clinical endeavors on the utility of ESM and patients and clinicians willing to experiment with ESM.

A limitation of this study is its convenience sample: 80% of included patients were women, relatively few clinicians were included, and participants were selected (in part) on their interest in self-monitoring. Given our sample of convenience, as well as the fast-moving field and the diversity in healthcare systems, generalizations to other settings should be done cautiously. Additionally, any positive or negative effects that patients and clinicians attributed to ESM could be due to other causes unknown to them. Finally, clinicians were less involved during the monitoring period, which limited integration of ESM in treatment.

Table 8.4. Practical discussion points for clinicians and patients to consider before starting ESM in treatment, based on the findings of the present study.

Topic	Discussion points
1: Determine rationale of ESM	
Desirability of ESM	Do both patient and clinician agree that ESM is helpful and doable?
Goal of ESM	What do patient and clinician hope to gain from ESM? How does it fit into the patients' current treatment goals?
2: Manage expectations	
Risk of negative effects	Are patient or clinician apprehensive of any negative effects (e.g., mood worsening, pre-occupation with disorder)? What can the patient do if these occur?
Burden	Is it okay if patients' occasionally miss assessments, and how often?
Feedback	What can patient and clinician expect to learn from the ESM feedback, and what not?
3: Determine the ESM protocol	
Feasibility of the monitoring	What frequency and duration of assessments is necessary to meet the goal and remain feasible for both patient and clinician?
Content of the assessments	What should the ESM diary include to meet the goal?
Need for weekly mood questionnaires	Is weekly monitoring for episodes necessary?
Desirability of alerts to patient and/or clinician	Do patient and clinician want to be informed of elevated scores?
4: Determine level of involvement of clinician	
Data access	What data is the clinician allowed to examine and how often?
Degree of contact through ESM	Does the patient contact the clinician in case of elevated scores or alerts, or vice versa? What happens if patients indicate elevated scores?
5: Facilitate interpretation of personalized feedback	
Frequency of feedback	How often is the personalized feedback discussed?
Content of feedback	Which data are discussed? Is it necessary to discuss all the feedback every session, or only parts of it?
Interpretation of feedback	Can the clinician help the patient to read the graphs? How do both interpret the feedback, and are there meaningful differences therein?
6: Evaluate regularly	
Usefulness and feasibility of ESM	Does the current ESM protocol still meet its intended goal or does it needs to be adapted? Is it still feasible for the patient?
Negative effects	Have any negative effects of ESM occurred and (how) can the patient cope with them?

Conclusions

This study demonstrates that long-term ESM monitoring, alerts, and personalized feedback are perceived as beneficial to the treatment and self-management of patients with bipolar disorder. To optimize clinical utility of ESM, our results suggest that future research should prioritize the development of clinically relevant data analyses and the necessary technology to enable personalized integration of ESM in clinical care. Furthermore, we recommend intensive collaboration between patient and clinician to ensure ESM fits the patient's situation and treatment goals. This way, ESM has the potential improve clinical care.

References

1. Pini S, de Queiroz V, Pagnin D, Pezawas L, Angst J, Cassano GB, Wittchen H-U. Prevalence and burden of bipolar disorders in European countries. *European Neuropsychopharmacology.* 2005;15(4):425-434.
2. Michalak EE, Yatham LN, Kolesar S, Lam RW. Bipolar disorder and quality of life: A patient-centered perspective. *Quality of Life Research.* 2006;15(1):25-37.
3. Lean M, Fornells-Ambrojo M, Milton A, Lloyd-Evans B, Harrison-Stewart B, Yesufu-Udechuku A, Kendall T, Johnson S. Self-management interventions for people with severe mental illness: Systematic review and meta-analysis. *The British Journal of Psychiatry.* 2019;214(5):260-268.
4. Murray G, Suto M, Hole R, Hale S, Amari E, Michalak EEJCP, Psychotherapy. Self-management strategies used by 'high functioning'individuals with bipolar disorder: From research to clinical practice. 2011;18(2):95-109.
5. Leverich GS, Post RM. The NIMH life chart manual for recurrent affective illness: The LCM. *NIMH Monograph.* 1993.
6. Schärer LO, Krienke UJ, Graf S-M, Meltzer K, Langosch JM. Validation of life-charts documented with the personal life-chart app–a self-monitoring tool for bipolar disorder. *BMC Psychiatry.* 2015;15(1):1-7.
7. Bauer M, Grof P, Gyulai L, Rasgon N, Glenn T, Whybrow PC. Using technology to improve longitudinal studies: Self-reporting with ChronoRecord in bipolar disorder. *Bipolar Disorders.* 2004;6(1):67-74.
8. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, Matthews M. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: Implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association.* 2016;23(3):477-484.
9. Bilderbeck AC, Saunders KE, Price J, Goodwin GM. Psychiatric assessment of mood instability: Qualitative study of patient experience. *The British Journal of Psychiatry.* 2014;204(3):234-239.
10. Reis HT. Why researchers should think 'real-world': A conceptual rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:3.
11. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, Reininghaus U. Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry.* 2018;17(2):123-132.
12. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, Drukker M, Reininghaus U, Jacobs N, Lataster T, Simons C, Lousberg R, Gülöksüz S, Leue C, Groot PC, Viechtbauer W, Delespaul P. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *DepressAnxiety.* 2017;34(6):481.
13. 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;70(11):983-991.
14. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, van Bemmel AL, Peeters F, Delespaul P, van Os J. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta PsychiatrScand.* 2011;124(4):262.
15. Larson R, Csikszentmihalyi M. The experience sampling method. *New Directions for Methodology of Social & Behavioral Science.* 1983.
16. Wenze SJ, Armey MF, Miller IW. Feasibility and acceptability of a mobile intervention to improve treatment adherence in bipolar disorder: A pilot study. *Behavior Modification.* 2014;38(4):497-515.
17. Wenze SJ, Armey MF, Weinstock LM, Gaudiano BA, Miller IW. An open trial of a smartphone-assisted, adjunctive intervention to improve treatment adherence in bipolar disorder. *Journal of Psychiatric Practice.* 2016;22(6):492.
18. Tsanas A, Saunders K, Bilderbeck A, Palmius N, Osipov M, Clifford G, Goodwin G, De Vos M.

Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders.* 2016;205:225-233.
19. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, van Bemmel AL, Myin-Germeys I, Delespaul P, van Os J, Wichers M. A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry.* 2014;13(1):68.
20. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes J, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: A qualitative study. *European Psychiatry.* 2017;41:115-121.
21. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine.* 2014;89(9):1245.
22. Baxter RJ, Hunton JE. Capturing affect via the experience sampling method: Prospects for accounting information systems researchers. *International Journal of Accounting Information Systems.* 2011;12(2):90-98.
23. Altman E, Hedeker D, Peterson JL, Davis JM. A comparative evaluation of three self-rating scales for acute mania. *BiolPsychiatry.* 2001;50(6):468.
24. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman self-rating mania scale. *BiolPsychiatry.* 1997;42(10):948.
25. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *BiolPsychiatry.* 2003;54(5):573.
26. Bernstein IH, Rush AJ, Suppes T, Kyotoku Y, Warden D. The Quick Inventory of Depressive Symptomatology (clinician and self-report versions) in patients with bipolar disorder. *CNS spectrums.* 2010;15(6):367-373.
27. Miller CJ, Johnson SL, Eisner L. Assessment tools for adult bipolar disorder. *Clinical Psychology: Science and Practice.* 2009;16(2):188-201.
28. Knapen SE. *Rhythm & Blues: Chronobiology in the Pathophysiology and Treatment of Mood Disorders*, Rijksuniversiteit Groningen; 2019.
29. van der Krieke L, Jeronimus BF, Blaauw FJ, Wanders RBK, Emerencia AC, Schenk HM, De Vos S, Snippe E, Wichers M, Wigman JTW, Bos EH, Wardenaar KJ, De Jonge P. HowNutsAreTheDutch (HoeGeklsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research.* 2016;25(2):123.
30. Tyler E, Jones S, Black N, Carter L-A, Barrowclough C. The relationship between bipolar disorder and cannabis use in daily life: An experience sampling study. *PLoS One.* 2015;10(3).
31. de Casterle BD, Gastmans C, Bryon E, Denier Y. QUAGOL: A guide for qualitative data analysis. *IntJNursStud.* 2012;49(3):360.
32. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental cisorders: Meta-analysis and recommendations. *JMedInternet Res.* 2019;21(12):e14475.
33. van Genugten CR, Schuurmans J, Lamers F, Riese H, Penninx BW, Schoevers RA, Riper HM, Smit JH. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *Journal of Clinical Medicine.* 2020;9(2):322.
34. Conner TS, Reid KA. Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science.* 2012;3(3):315-323.
35. Schermer M. Telecare and self-management: Opportunity to change the paradigm? *JMedEthics.* 2009;35(11):688.
36. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:636.
37. Perez Arribas I, Goodwin GM, Geddes JR, Lyons T, Saunders KE. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Transl Psychiatry.* 2018;8(1):1-7.
38. Smit AC, Snippe E, Wichers M. Increasing restlessness signals impending increase in depressive symptoms more than 2 months before it happens in individual patients. *PsychotherPsychosom.*

2019;88(4):249.
39. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, Viechtbauer W, Delespaul P, Myin-Germeys I, Wichers M. Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology.* 2016;35(5):433.
40. Wichers M, Snippe E, Riese H, Bos FM. The network approach to depression: Hype or Holy Grail? under review.
41. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. *International Journal of Bipolar Disorders.* 2018;6(1):1-7.
42. Bastiaansen JA, Kunkels YK, Blaauw FJ, Boker SM, Ceulemans E, Chen M, Chow S-M, de Jonge P, Emerencia AC, Epskamp S. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *JPsychosomRes.* 2020:110211.
43. 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.
44. Fisher AJ, Boswell JF. Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment.* 2016;23(4):496.
45. Brandt, P. and J.T. Williams, *Multiple time series models.* 2007, Thousand Oaks, CA, US: Sage Publications Inc.
46. Emerencia, A.C., et al., *Automating Vector Autoregression on Electronic Patient Diary Data.* IEEE Journal Of Biomedical And Health Informatics, 2016. 20(2): p. 631.
47. Brose, A. and N. Ram, *Within-person factor analysis: Modeling how the individual fluctuates and changes across time*, in *Handbook of Research Methods for Studying Daily Life*, M.R. Mehl and T.S. Conner, Editors. 2012, Guilford Press: New York, NY US. p. 459.

Appendix

1. Experience sampling method diary items

The item list can be viewed in the appendix of Chapter 9 and was constructed in several steps. First, we identified relevant concepts for symptoms of bipolar disorder and searched the literature for ESM studies in patients with bipolar disorder. This yielded a first item list of 71 potentially relevant items. This list was then discussed in individual interviews with three patients and one psychiatrist. The items were finally selected on three criteria: 1) the patients and clinician recognized the item as signaling symptoms of either mania or depression; 2) the patients and the clinician felt comfortable with the formulation of the item (e.g., they could see themselves saying the sentence in daily life); 3) both the patients, the clinician, and the research team believed the item would vary meaningfully within participants. Items that were deemed relevant but too person-specific were put on a list for participants so they might select them for their personal question.

All items were obligatory, with the exception of the comment field at the end of the questionnaire (item 29). All items were assessed five times per day, with the exceptions of the items regarding sleep (item 2- 3), and the item regarding appointments (item 27). These items were only shown if participants answered 'yes' on item 1 or item 26. This way, we ensured that participants could still answer questions about their sleep, even though they might have skipped the first (few) assessments.

2. Data security

ESM assessments, as well as the weekly ASRM and QIDS questionnaires, were administered and secured via RoQua (www.roqua.nl). RoQua is a web-based application that is fully integrated in the personal health record systems used by several mental health care institutions in the Northern Netherlands that are part of the Rob Giel Research Center (www.rgoc.nl). The two institutions that participated in the present study are part of this collective and use the RoQua application for

Routine Outcome Monitoring (ROM).

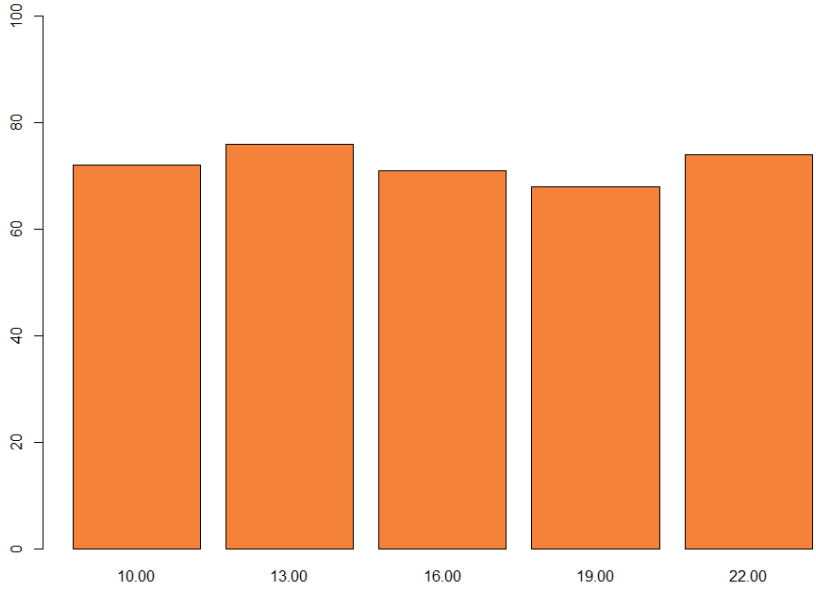
RoQua links all assessment data to a personal identifier, removing all information that may be traceable to the individual, thereby ensuring that the data is stored anonymously and securely. Given that RoQua is used in clinical practice, the application meets the stringent criteria regarding privacy and data security. During the study period, only the patients' clinicians and the researcher FB could access the patients' assessment data. Patients gave consent to both their clinicians and researcher FB to view the ESM and weekly questionnaire data.

3. Personal feedback report

The personal feedback report consisted mostly of descriptive information. Below, we provide an overview of the feedback that was given to each participant, based on simulated data. This way, we are able to show all graphs and explanations offered to participants, as well as explain our data analytic choices herein. All feedback reports contained graphs, as well as text to explain how the graphs should be read. The report did not give advice on the interpretation of the graphs (e.g., 'this graph shows exercising is good for your health'); this was left to the patient and clinician to decide.

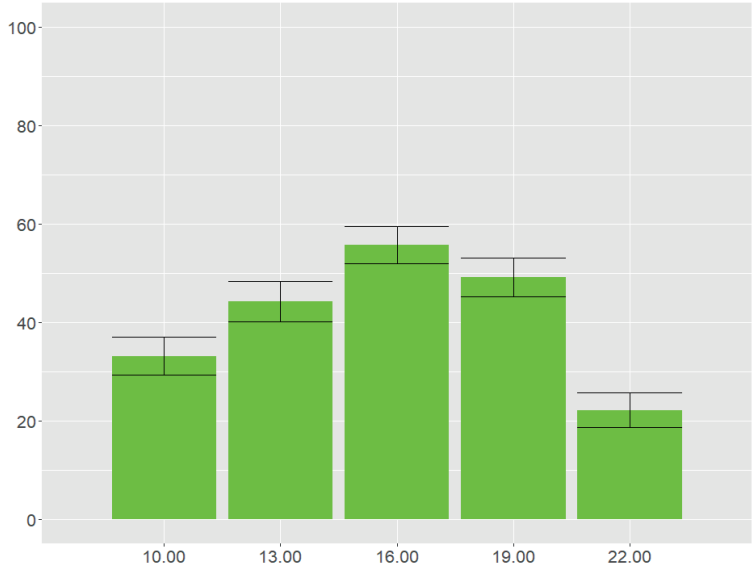
Missing data. First, participants were shown their percentage of missing data across all the five time points in a bar plot (see Figure 8.3). They were presented with their overall percentage of completed assessments (e.g., 76%), and whether sufficient assessments were completed to give reliable information about their data.

Figure 8.3. Percentage of completed assessments at each of the assessment times.



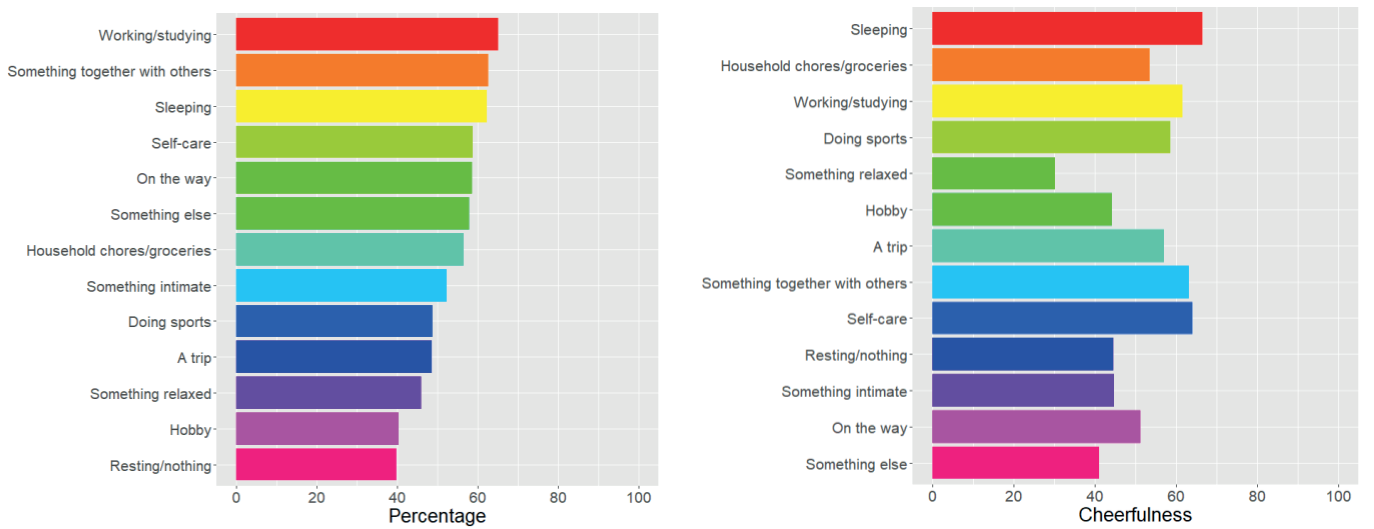
Time of day. Next, mood and symptoms in relation to time of day was depicted in bar plots with error bars (see figure 8.4). Participants were explained that, when the error bars did *not* overlap one another, there was, on average, a difference in how they felt over the course of the day. These plots were shown for every diary item that significantly differed across the five time points. If there were significant differences, participants were told how their symptoms varied (e.g., 'during the monitoring period, you generally felt more down in the morning than in the afternoon and evening'). If no significant differences arose, no plots were shown, and participants were explained that we could not find indications of diurnal variation in mood and symptoms.

Figure 8.4. Diurnal variation in the item "I feel down".



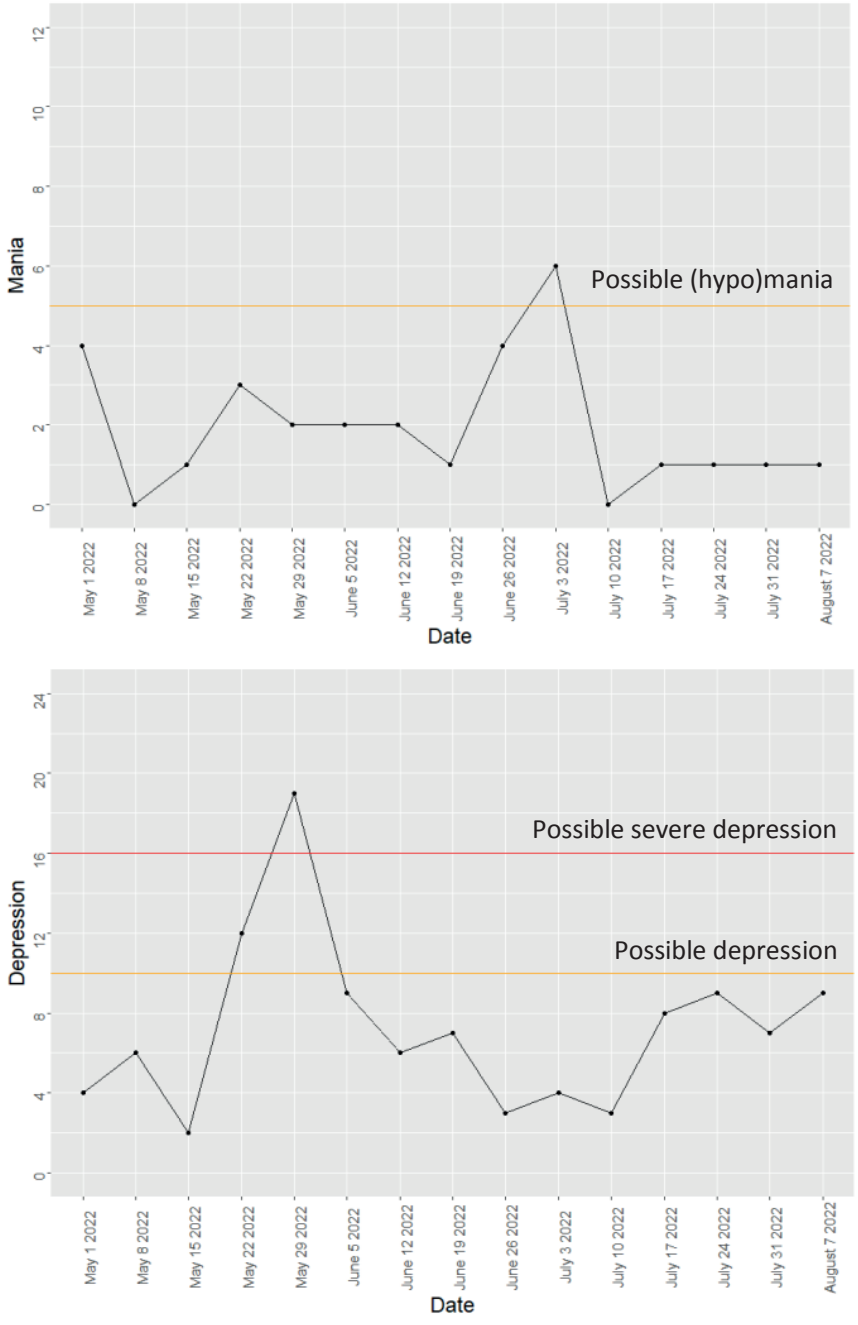
Activities. Furthermore, participants were shown the percentage of time they spent on different activities and their average mood during these activities (cheerful and down) in bar plots (see Figure 8.5). Participants were explained that they could not draw causal inferences from these plots, mainly due to two limitations. First, during the monitoring period, participants had retrospectively indicated every activity they had done during the previous three hours. This means that the relationship between mood and a particular activity could be confounded by function of time (e.g., the activity was done at the beginning of the three hour time block but not at the time of the assessment, when mood was assessed). Second, the relationships between mood and activities was bidirectional. A certain mood could influence the likelihood of an activity, whereas activities could also induce a certain mood.

Figure 8.5. Percentage of time spent on activities and mood during activities.



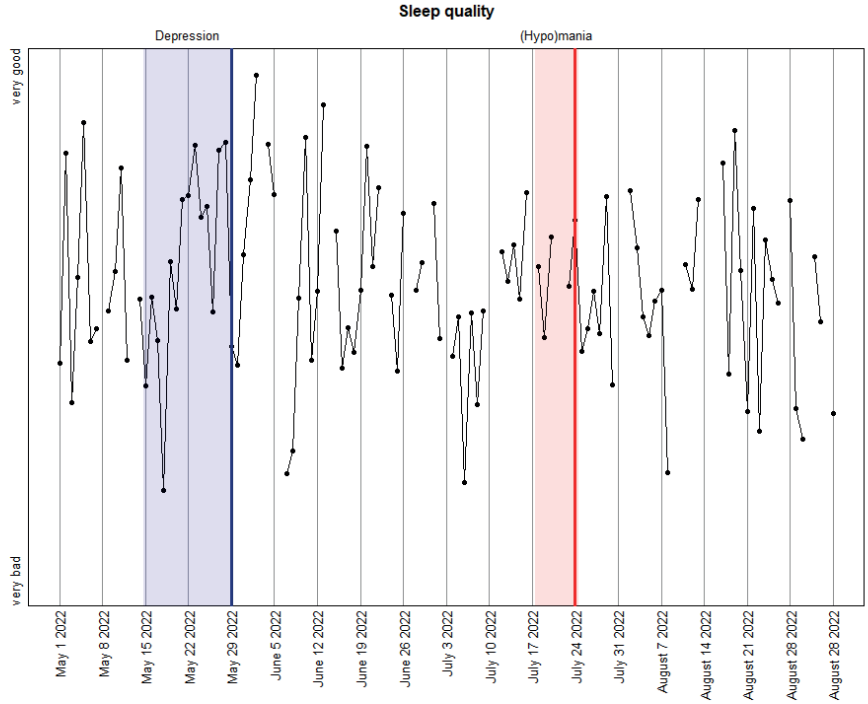
Episode occurrence. Then, participants were shown the weekly mood questionnaires (QIDS and ASRM) in line graph (see Figure 8.6). Participants were explained that previous research had established cut-offs[25-27] (see the orange and red lines), above which there could be indications of a (hypo)manic or depressive episode.

Figure 8.6. Weekly scores on manic (ASRM) and depressive (QIDS) symptoms.



Sleep and appointments. Next, line graphs depicting the once per day variables (having made appointments, sleep duration, and sleep quality) were shown (Figure 8.7). Vertical lines depicted potential episodes of (hypo)mania and depression, to visualize potential associations between the weekly questionnaires and the ESM items. Given that the items of the weekly questionnaires pertained to the previous week, we highlighted both the day of completing the questionnaire (bold blue/ red line) and the previous week (transparent bar). Missing data was depicted as a broken line.

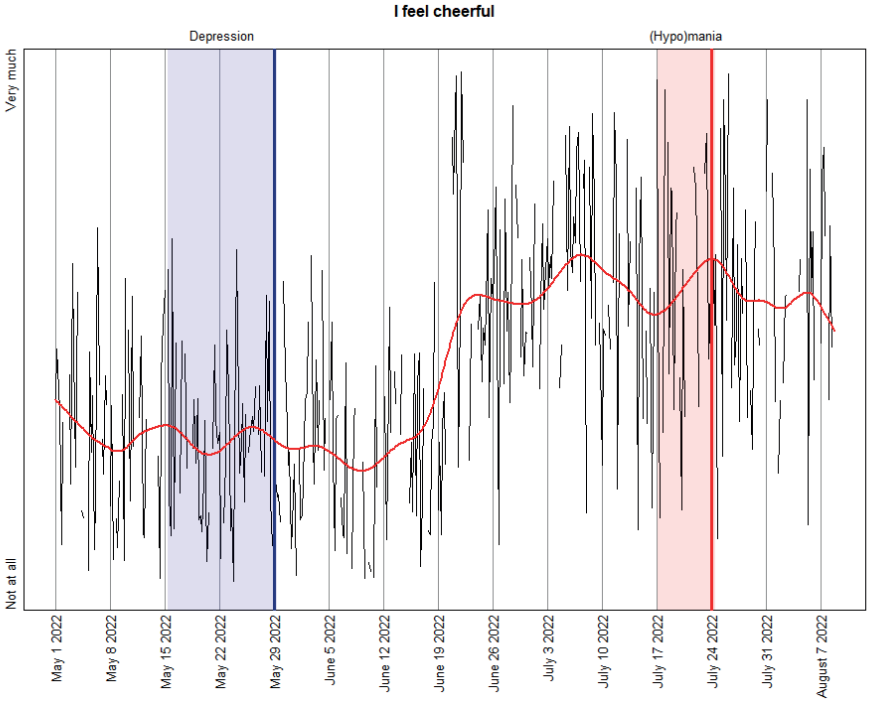
Figure 8.7. Variation in sleep quality, plotted against weeks with increased (hypo)manic and depressive symptoms.



Mood and symptom variation. Similar plots were constructed for all other continuous ESM variables (22 in total; see Figure 8.8). To facilitate visual inspection

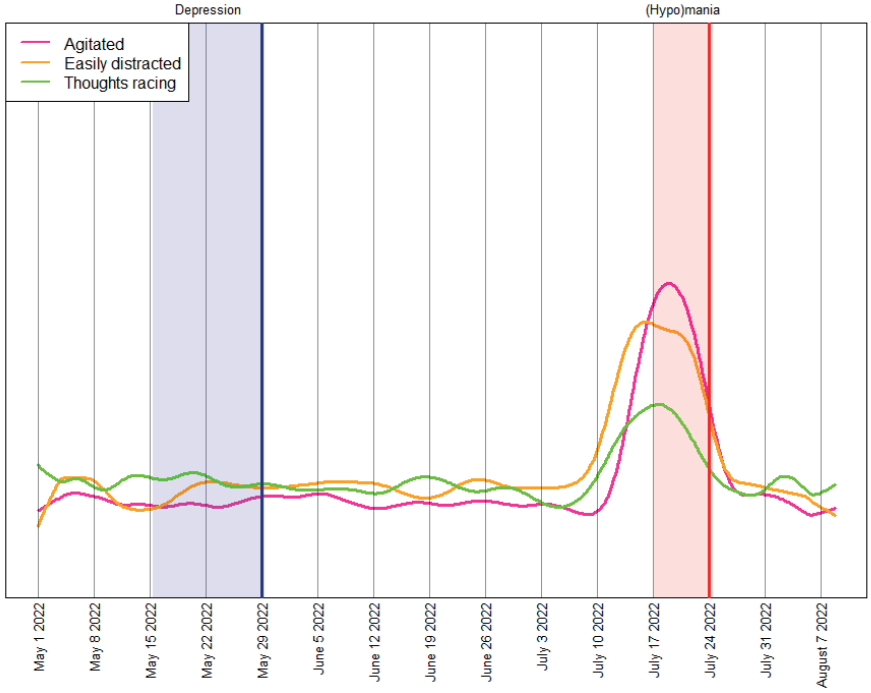
of the graphs, we fitted Kernel smoothing lines over the plots. This helped patients and clinicians to view weekly trends in the ESM data (e.g., has this item decreased or increased in a particular week).

Figure 8.8. Mood variation for the item "I feel cheerful".



Episode indicators. To facilitate interpretation of all 22 line graphs, the researcher examined the Kernel smoothing lines of each graph to determine which variables uniquely increased or decreased during (hypo)manic or depressive episodes (and not during euthymic periods). When such unique variables were found, these were summarized in another line graph (see Figure 7). Participants were explained that they might monitor themselves for these variables specifically to recognize impending episodes. If no indicator variables were found, participants were explained that we could not find specific variables that might signal a depression or (hypo)mania.

Figure 8.9. Indicator plot for (hypo)mania.



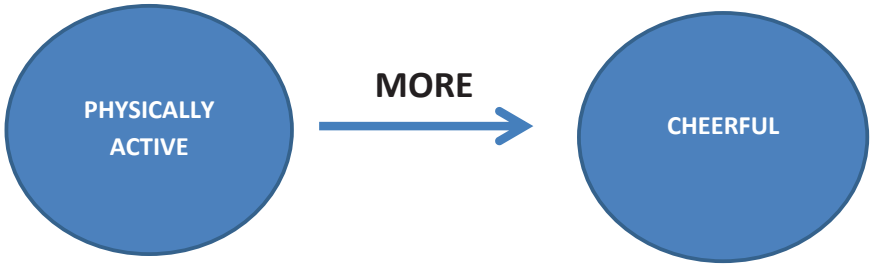
Associations between variables (optional). Some participants wished to see information on the associations between specific variables (e.g., between mood and physical activity or time spent with loved ones). This was tested using lag-1 vector autoregressive models[45] (VAR) through the AutovarCore package[46]. This R-package automatically estimates individual VAR-models. A VAR-model was only constructed if 1) the participant had completed at least 60% of all assessments, 2) the items of interest showed sufficient variation (>10%[47]), and 3) the items were stationary. Results were only deemed significant if all four autovarCore models were statistically significant. Together with the participant, we operationalized the research question, thereby attempting to minimize the number of variables so as to reduce the risk of Type I error.

If a significant association was found, participants were told that we found indications that, in general, variable A resulted in less/more variable B three hours later (e,g, being more physically active led to feeling more cheerful three hours later; see Figure 8). To caution overinterpretation, we stated that this does not necessarily

mean that this is always the case, and that these results might be specific to the monitoring period.

If no significant associations were found, participants were told that we could not find indications that variable A significantly influenced variable B three hours later or vice versa. Participants were explained that this did not mean that the association could never exist, but that this could be due to the constraints of the data and the model (e.g., exercising could have a beneficial effect on other variables that were not tested, or exercising could have a beneficial effect at other lags than lag-1).

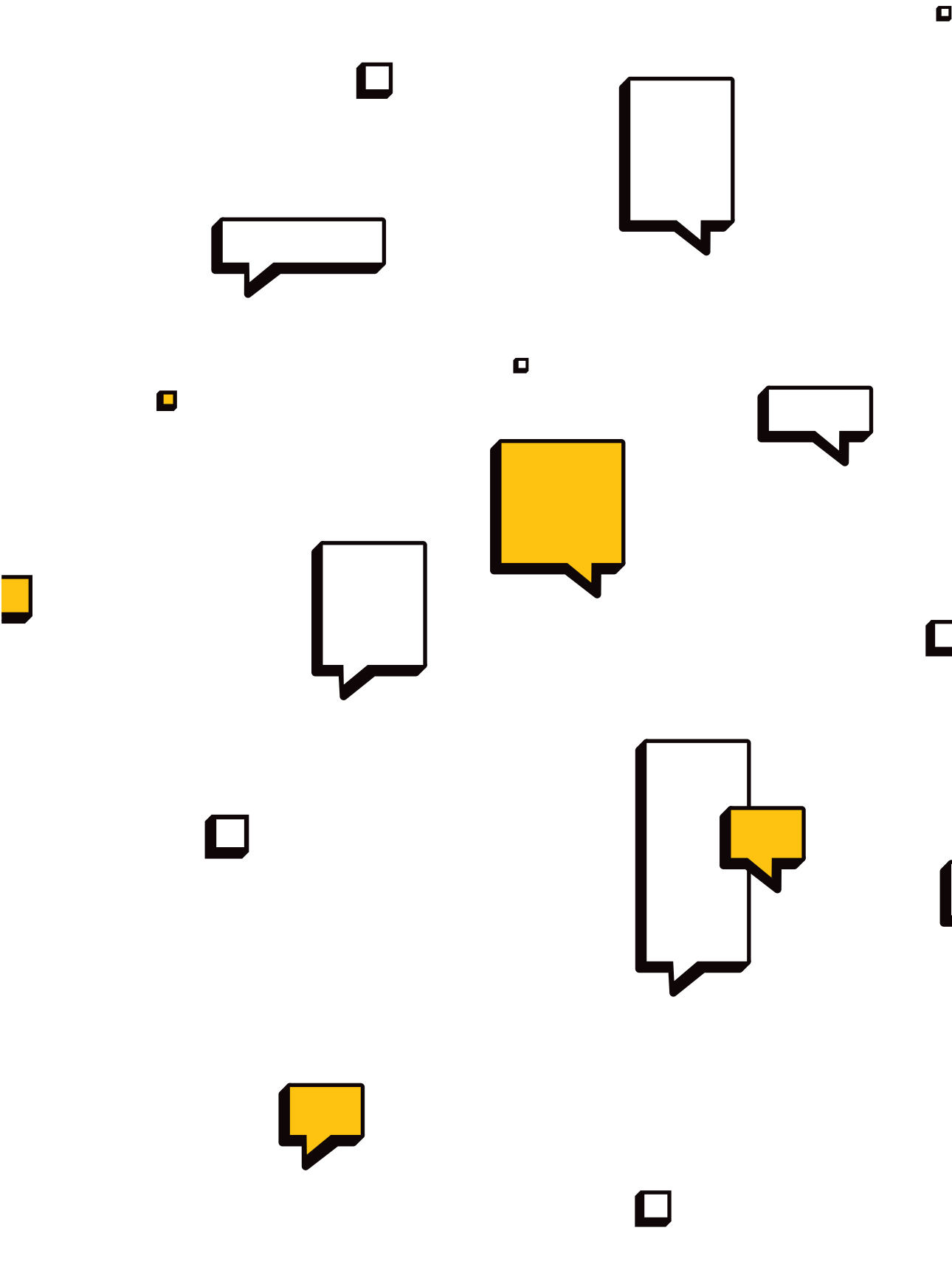
Figure 8.10. Association between being physically active and cheerfulness. This graph shows that, if the participant was physically active, he/she would generally feel more cheerful three hours later.



Comment field. Finally, participants were presented with a table containing all their time-stamped responses to the comment field (see Figure 8.11). Patients and clinicians were encouraged to use them to interpret peaks and valleys in the line graphs.

Figure 8.11. Table containing (fictional) responses in the comment field.

Date	Diary entry
12-7-2022 9:47	e.g. talked to person A about X.
13-7-2022 13:01	e.g. headache
13-7-2022 21:46	e.g. felt down because of Y.



Chapter 9

Anticipating manic and depressive shifts in patients with bipolar disorder using early warning signals

Published as:

Bos FM*, Schreuder MJ*, George SV, Doornbos B, Bruggeman R, van der Krieke L, Haarman BCM, Wichers M, Snippe E. Anticipating manic and depressive shifts in patients with bipolar disorder using early warning signals. Submitted.

*authors contributed equally

Abstract

Objectives. Patients with bipolar disorder experience rapid transitions to manic and depressive episodes, which remain difficult to predict. Smartphone monitoring of affect and symptoms using ecological momentary assessment (EMA) enables the study of early warning signals (EWS) to anticipate ensuing mood shifts. The present study investigated whether EWS can anticipate mood shifts in individual patients, and whether EWS could have clinical utility.

Methods. Twenty bipolar type I/II patients completed EMA questionnaires five times a day for four months (*Mean*=491 observations per person). Weekly completed symptom questionnaires on depressive (Quick Inventory for Depressive Symptomatology Self-Report) and manic (Altman Self-Rating Mania Scale) symptoms were used to determine transitions. EWS (rises in autocorrelation at lag-1 and standard deviation) were calculated in moving windows over 17 affective and symptomatic EMA states. Kendall's tau was calculated to detect significant EWS prior to transitions. Positive and negative predictive values were calculated to determine clinical utility.

Results. Eleven patients reported 1-2 manic or depressive transitions. The presence of EWS increased the probability of detecting impending transitions, but their absence could not be taken as a sign that no transition would occur in the near future. Predictive values point towards several momentary states as accurate indicators of nearby depressive and manic transitions (predictive values ranging from 65-100%): cheerfulness, focusing ability, full of ideas, worry, racing thoughts, agitation, energy, and tiredness. Large individual differences in the utility of EWS were found.

Conclusions. EWS show promise in anticipating manic and depressive transitions in bipolar disorder. Further confirmatory research is warranted.

Introduction

A major challenge in psychiatry is to timely identify impending psychopathological episodes for individual patients. Until now, research has mostly focused on group-level retrospective risk factors[1], which unfortunately say little about *which* individual patient will relapse *when.* However, rapid technological advances have enabled patients to easily monitor their mood and symptoms in real-time, opening the door to prospective and personalized anticipation of clinically relevant symptom changes in the near future[2]. Such early identification of episodes might be especially relevant for patients with bipolar disorder (BD), who experience frequent and disruptive depressive and manic episodes, and whose treatment is strongly focused on episode recognition[3].

Now that intensive longitudinal monitoring through smartphones has become increasingly feasible[4], the field is in need of tools to utilize these data to anticipate future increases in psychopathology. Evidence in the fields of physics and ecology is accumulating that many processes can be characterized as complex dynamical systems, in which abrupt transitions to alternative states are anticipated by increasing instability of the system. This instability is reflected by a process termed critical slowing down[5], which means that, as a system approaches a transition, it gets increasingly slow in recovering from minor perturbations[6]. Critical slowing down manifests in patterns in the dynamics of time series data, including rising autocorrelation (i.e., the system's current state increasingly predicts its next state) and rising variance (i.e., the system's current state shows increasing fluctuations). Because critical slowing down may occur prior to transitions in a system, rising autocorrelation and variance have been termed early warning signals (EWS). Interestingly, EWS have been shown to precede transitions in a wide variety of systems, such as climate change[7], starlight shifts[8], and animal extinction[9].

Researchers increasingly recognize that psychopathology might also behave as a complex dynamical system[10,11]. If this is the case, transitions to psychopathological episodes might be anticipated by EWS in momentary affective and symptomatic states[12,13]. These states can be assessed through ecological

momentary assessment (EMA), with which patients frequently monitor their affect and symptoms in daily life on their smartphones. In psychopathology, rising autocorrelations in momentary states (e.g., affect, stress, worry) indicate higher carryover of one's affective state from one moment to the next. This means that the effects of perturbations (e.g., stressful events) linger longer[14]. Rising trends in variance mean that perturbations have an increasingly strong impact on one's momentary state. Preliminary research suggests that EWS in momentary states can indeed anticipate prospective shifts from a healthy to a depressed state[15-17]. This raises the question whether EWS could also signal impending mood shifts in BD patients.

For BD patients, a smart personalized prediction tool based on smartphone monitoring could revolutionize treatment. Patients and clinicians describe this as one of the greatest promises of e-mental health[18,19]. EMA might fulfill a dual purpose in this regard: the monitoring itself might already increase awareness of mood shifts, whereas the gathered data can be analyzed to provide a personalized EWS-based alert system[20]. Indeed, previous research has demonstrated that daily and weekly mood fluctuations in BD patients follow principles of complex dynamical systems[21-23], and that EWS can be detected prior to mood transitions in simulated actigraphy data[24]. A logical next step would be to examine in empirical data whether EWS in momentary states indeed anticipate mood shifts in BD patients.

Therefore, the present empirical and exploratory study is the first to investigate whether EWS precede mood shifts in BD patients and whether they might have clinical utility. To that end, we employed an exploratory replicated single-subject design in twenty BD patients who participated in EMA for four months. First, we investigated whether EWS improved the detection of impending manic as well as depressive transitions, and whether the absence of EWS could be taken to indicate a lack of impending transitions. Second, we studied which momentary states constituted the best EWS for manic and depressive transitions, and finally, we examined individual differences in the utility of EWS. Taken together, these findings might provide a comprehensive investigation into whether EWS may indeed be used to signal upcoming transitions in BD.

Methods

Participants

For the present prospective observational cohort study, twenty patients with BD type I or II were included (see Table 9.1). To be included, patients had to: 1) be ≥18 years, 2) be diagnosed with and currently in treatment for BD type I/II, and 3) demonstrate high occurrence of manic and/or depressive episodes (≥2 in the previous year). Clinicians of two Dutch tertiary care institutions invited patients for the study until the intended cap of twenty participants was reached. Interested patients were invited to the research facility, where the study was explained in detail and informed consent was signed.

Twenty-eight patients were referred, two of whom were unreachable. Six patients declined participation during the first telephone call, expecting that study participation and the focus on mood would be too burdensome. This left twenty patients that started and finished the study. Although a seemingly small sample, it is relatively large for idiographic studies. Here, power depends not on the number of individuals, but rather on the number of assessments per individual[25]. As such, a sample of twenty patients allows us to examine the relative robustness of EWS in improving the detection of transitions. The study was approved by the University Medical Center Groningen medical ethics committee (no 201501161).

Study design

Participants engaged in EMA and completed weekly questionnaires for four months. Patients received five assessments per day: every three hours (time-contingent schedule), patients received a text message with a link to the EMA on their smartphone, which took approximately 1-2 minutes to complete. Assessments were securely administered and stored via RoQua (www.roqua.nl) in patients' personal health records. Patients chose their own start and end time and had one hour to complete each assessment. No reminder prompts were given. Researchers contacted patients after the first three days of monitoring and if compliance was

low, if regular contact was preferred, or if the weekly questionnaires indicated above-threshold symptoms. Participants were not offered financial compensation.

Measurements

Manic and depressive transitions

Weekly, patients completed the Dutch versions of the Altman Self-Rating Mania Scale (ASRM[26]) and the Quick Inventory for Depressive Symptomatology Self-Report (QIDS-SR[27]). A transition was defined as a clinically relevant abrupt (i.e., within one week) increase in manic (ASRM≥6[28]) or depressive (QIDS-SR≥6[27]) symptoms, without such increases in the two weeks prior to the transition (see the Appendix for all transitions per individual). Our analytic approach further required at least three weeks (≥105 observations) of EMAs prior to the transition.

EMA items

The questionnaire (see the Appendix) consisted of 26 items pertaining to momentary mood, symptoms, sleep, and activities. These items were based on previous EMA research[29-31] and interviews with three patients and a psychiatrist on relevant constructs for people with BD. For the calculation of EWS, we selected the 17 EMA items that assessed affect or symptoms on a continuous scale (i.e., ranging from 0 ("not at all") to 100 ("very much").

Compliance

Two patients were part of a pilot and therefore completed three months of EMA monitoring. The eighteen other patients completed on average 18 weeks of monitoring (*range*=16-32). Average compliance, calculated as the number of completed assessments divided by the number of assessments participants received, was 76% (491 assessments, *SD* = 137.8).

Data analysis

For each EMA momentary state, two EWS indicators were examined: rises in the autocorrelation at lag-1 and rises in the standard deviation as indicator of the variance. These EWS were estimated using a moving windows approach[15,32].

Briefly, this involves iterative computation of EWS within segments (or windows) of the time series, for each individual and transition separately. With each iteration, the window slides one time point ahead until the transition point. Within every window, we computed the autocorrelation and standard deviation of each momentary state. This yielded a new time series for each EWS indicator, allowing us to examine whether rises in the indicator actually preceded an abrupt manic or depressive transition. Pre-processing and analysis of the data were performed in the statistical programming language R[33]. Results were visualized using the R-packages *ggplot2*[34] and *gridExtra*[35]. Our code will become available in the Supplementary Materials accompanying this article.

Pre-processing steps

Outliers were winsorized to minimize their influence on the results. To ensure there were no trends in the mean over time (i.e., stationarity)[32], and reduce the risk of false positives[36], the data were detrended by applying a Gaussian kernel smoothing function over the whole pre-transition period[7,37]. Missing data were not imputed because this might result in spurious correlations[32].

Window size

The window size is related to the timescale at which the system dynamics evolve and may affect results[32]. Therefore, sensitivity analyses were conducted to investigate the effects of windows of one, two or three weeks. We ensured that each window included an equal number of weekend days, thereby negating their effects on the results[24]. The sensitivity analyses indicated small differences for different window sizes (see the Appendix), but results were largely robust. Therefore, based on suggestions on how bipolar dynamics develop in patients with high mood instability[38], we opted for a window size of two weeks (i.e., 70 observations), yielding on average 168 windows (*SD*=99, range=35-415) per transition.

Calculation of EWS

First, within each window, the autoregressive coefficient was calculated at lag-1 over the residuals obtained after detrending[32]. The autocorrelation thus indicates how well a momentary state (e.g., feeling cheerful) predicts itself three hours later.

Overnight lags were prevented by deleting the lagged observation of each day's first observation. Second, variance was estimated from the standard deviation[2] over the residuals within each window[32]. A rising standard deviation indicates that someone's momentary state varies more widely around the mean over time.

Significance testing

The above approach resulted in a new time series data set for every individual, with separate estimates of the autocorrelation and standard deviation. To test whether each EWS indicator significantly increased, we calculated the Kendall correlation coefficient (τK) over this data set[32], in the two weeks prior to the transition (using the R-package *Kendall*[39]). A significant positive correlation indicates that the EWS indicator significantly rose prior to a transition. The value of Kendall's tau was taken to reflect the strength of the EWS. Since the moving windows use overlapping data, we corrected for this dependency between nearby windows by applying the Hamed-Rao correction[40]. Furthermore, we corrected for multiple testing by using the false discovery rate (FDR) approach as proposed by Benjamini and Hochberg[41]. This ensures that, across all tests performed within a single patient, the probability of false positives is 5%.

Predictive value

To gain insight into the predictive utility of EWS for detecting mood transitions, positive and negative predictive values (PPV and NPV respectively) were calculated[42]. The PPV and NPV are based on the sensitivity (proportion of true positives, where EWS preceded transitions) and specificity (proportions of true negatives, where the absence of EWS indicates the absences of transitions). Predictive values indicate to what extent (i) the presence of EWS increases the likelihood of detecting a transition (PPV) and (ii) their absence increases the likelihood of detecting that *no* transition will take place (NPV). PPV and NPV were calculated for both EWS indicators averaged across momentary states, as well as for each momentary state separately. Predictive values follow the below formulas, using the sensitivity and specificity of a particular momentary state (i) in a specific individual (j) for detecting

[2] The standard deviation is the square root of the variance. This transformation does not affect the rank correlation coefficient Kendall's tau.

transitions (t) towards either mania or depression:

$$PPV_{i,t,j} = \frac{sensitivity_{i,t,j} * prevalence_t}{sensitivity_{i,t,j} * prevalence_t + \left(1 - specificity_{i,j}\right) * (1 - prevalence_t)}$$

$$NPV_{i,t,j} = \frac{specificity_{i,j} * (1 - prevalence_t)}{(1 - sensitivity_{i,t,j}) * prevalence_t + specificity_{i,j} * (1 - prevalence_t)}$$

Higher predictive values indicate higher accuracy of EWS in detecting transitions. A PPV of 100% indicates that an EWS has no false positives, whereas a NPV of 100% indicates an EWS has no false negatives. The PPV and NPV demonstrate whether EWS improve the accuracy of detecting transitions above the general prevalence of transitions in our sample, and should thus be higher than the average transition prevalence (i.e., the proportion of manic (32%), depressive (36%), or no transitions (68% for mania and 64% for depression)). See the Appendix for a step-by-step calculation of predictive values and for sensitivity and specificity for each EWS.

To estimate false positives, EWS were also computed for individuals without a transition. For them, we selected a period of approximately 245 observations (i.e., mean number of observations in the pre-transition period of transitioning patients), without any changes of ≥4 on the ASRM and QIDS-SR. This was possible for seven out of nine non-transitioning patients.

Results

Sample characteristics

Demographic and clinical characteristics are depicted in Table 9.1. Of the twenty patients, eleven reported an abrupt manic or depressive mood shift, four of whom reported two transitions. Notably, during transitions, all patients reported subthreshold mania scores. During manic transitions, all but two patients reported above threshold scores for depression.

A clinical illustration

We will first present a clinical case example to illustrate EWS in an individual patient (ID6): a 27-year old woman diagnosed with BD type II. She reported both a manic and a depressive transition on the weekly ASRM and QIDS-SR (Figure 9.1A). Her weekly symptom scores had been stable in the previous weeks. Note that this would imply similar stability in her EMA momentary states until the transitions.

Figure 9.1B shows her EMA observations for the state *'extremely well'*. Higher scores indicate she is feeling more euphoric. It is difficult to distill clear patterns from these data; we see a lot of moment-to-moment variation, and more missed assessments after the manic transition. Figure 9.1C shows significant EWS in the autocorrelation for '*extremely well*' prior to both transitions. This means that, within the two weeks before reporting a manic and depressive shift, her euphoria increasingly lingered over time. Figure 9.1D shows an EWS in the standard deviation for the depressive transition, but not for the manic transition, indicating that her euphoria varied more widely prior to the depressive transition.

Table 9.1. Demographic and clinical characteristics.

	Full sample (N=20)	Patients with transition (N=11)	Patients without transition (N=9)
Gender (N)			
Male	4	3	1
Female	16	8	8
Age (N)			
20-35 years	9	4	5
36-50 years	8	5	3
51-65 years	3	2	1
Education level (N)			
Higher education	9	6	3
Secondary education	5	2	3
Secondary vocational education	3	1	2
Pre-vocational education	3	2	1
Years since bipolar disorder diagnosis (*M, SD*)	6.4 (6.3)	5.0 (5.8)	8.2 (6.8)
Years in treatment (*M, SD*)	10.6 (8.8)	10.1 (8.5)	11.27 (9.7)
Bipolar disorder diagnosis (N)			
Bipolar disorder type I	9	6	5
Bipolar disorder type II	11	5	4
Comorbid diagnoses (N)			
No comorbid Axis I/II disorder	12	5	7
Attention Deficit/Hyperactivity Disorder	1	1	0
Autism Spectrum Disorder	1	1	0
Sleep disorder	1	1	0
Alcohol/drug dependence	1	0	1
Personality disorder	6	5	1
Medication use (N)			
None	2	1	1
Amphetamine	1	1	0
Anti-epileptic	10	8	2
Atypical antipsychotic	10	5	5
Benzodiazepine	9	6	3
Thyreomimetica	2	0	2
Lithium	5	0	5
Monoamine oxidase inhibitor	3	3	0

	Full sample (N=20)	Patients with transition (N=11)	Patients without transition (N=9)
Selective serotonin reuptake inhibitor	4	2	2
Tricyclic antidepressant	1	1	0
Transitions[1] (N)			
To a manic episode		7	
To a depressive episode		8	
Symptom increase in week of transition (*M, SD*)			
Transition to a manic episode (ASRM)		6.7 (1.5)	
Transition to a depressive episode (QIDS-SR)		11.3 (5.8)	
Episode duration after transition in weeks (*M, SD*)			
Manic episode		1.9 (0.9)	
Depressive episode		2.6 (3.1)	
Compliance to EMA (%)	75.9	74.6	77.6

[1] Four patients reported two transitions, the remaining seven patients reported one transition.
Note: ASRM = Altman Self-Rating Mania Scale, M = mean, N = number, QIDS-SR = Quick Inventory of Depressive Symptomatology Self-Report, SD = standard deviation.

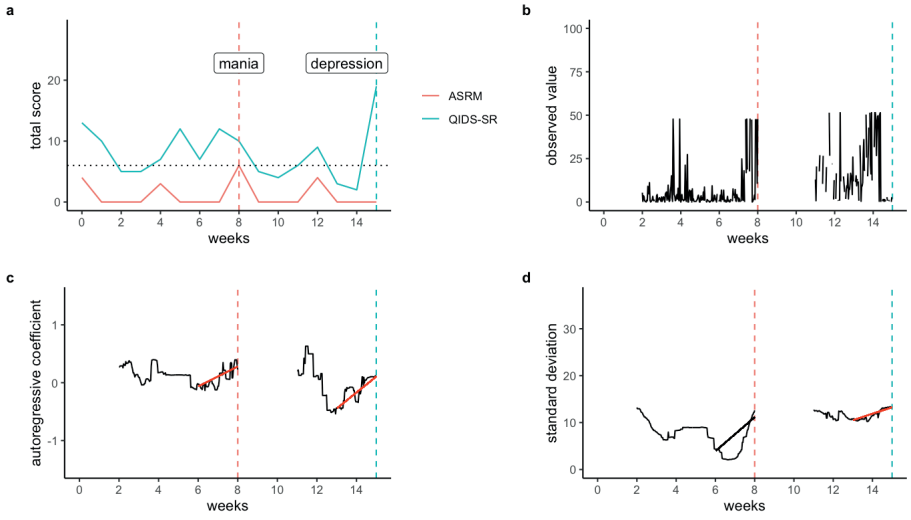
Predictive value of early warning signals

All transitions, both depressive and manic, were preceded by at least one EWS (i.e., a significant rise in the autocorrelation or standard deviation) in at least one of the EMA momentary states. Regarding the autocorrelation, on average, most EWS were found prior to manic transitions (*M*=5.6) versus depressive transitions (*M*=4.5) and non-transitions (*M*=4.2). The standard deviation also yielded the most EWS prior to manic transitions (*M*=6.7), versus depressive transitions (*M*=2.8) and non-transitions (*M*=5.3). In general, average PPVs and NPVs indicate that EWS were better in signaling the presence of transitions than their absence (see Figure 9.2).

Regarding depressive transitions, the average PPV indicates that if at least one EWS in the autocorrelation was detected, the probability of anticipating a depressive transition increased from 36% (prevalence) to 46%. However, if EWS were absent, the probability of correctly inferring *no* depressive transition did not improve (NPV=63% versus 64% prevalence). The standard deviation was not an accurate EWS for depressive transitions, with a PPV of 29% and a NPV of 58%.

For manic transitions, the autocorrelation was again slightly more accurate in signaling transitions. Here, EWS improved the probability of correctly inferring a manic transition from 32% (prevalence) to 48% (autocorrelation) or 41% (standard deviation). However, the NPV demonstrates that the probability of inferring *no* manic transition in the absence of EWS improved only slightly from 68% (prevalence) to 74% (autocorrelation) or 75% (standard deviation).

Figure 9.1. An illustration of early warning signals in one individual (ID6) in the EMA momentary state 'I feel extremely well'.



Note. Figure 9.1A depicts weekly manic (Altman Self-Rating Scale, ASRM, red) and depressive (Quick Inventory of Depressive Symptomatology, QIDS-SR, blue) symptom scores. At week 8 and 15, she reports an abrupt transition to a manic and depressive episode, respectively. Figure 9.1B visualizes her EMA observations for the item "I feel extremely well". Higher scores indicate she is feeling more down. Figure 9.1C shows significant early warning signals (EWS) in the autocorrelation for "I feel extremely well" for the transition to the manic (Kendall's Tau = .54, *corrected* p<.001) as well as the depressive episode (Kendall's Tau = .68, *corrected* p<0.001). Figure 9.1D shows an EWS in the standard deviation prior to the depressive transition (Kendall's Tau = .75, *corrected* p<0.001), but not prior to the manic transition (Kendall's Tau = .50, *corrected* p=0.07).

Predictive value of specific EMA momentary states

Figure 9.2 demonstrates the PPVs and NPVs for all momentary states separately. Five momentary states had a PPV of 100% for both depressive and manic transitions: '*cheerful'* (autocorrelation), *'ability to focus/switch'* (autocorrelation), '*full of ideas'* (autocorrelation), and *'worry'* (standard deviation), indicating they were never found for patients without a transition. The NPVs of these momentary states ranged from 70-83%. These momentary states might thus signal an impending transition without specifying its nature (depressive or manic).

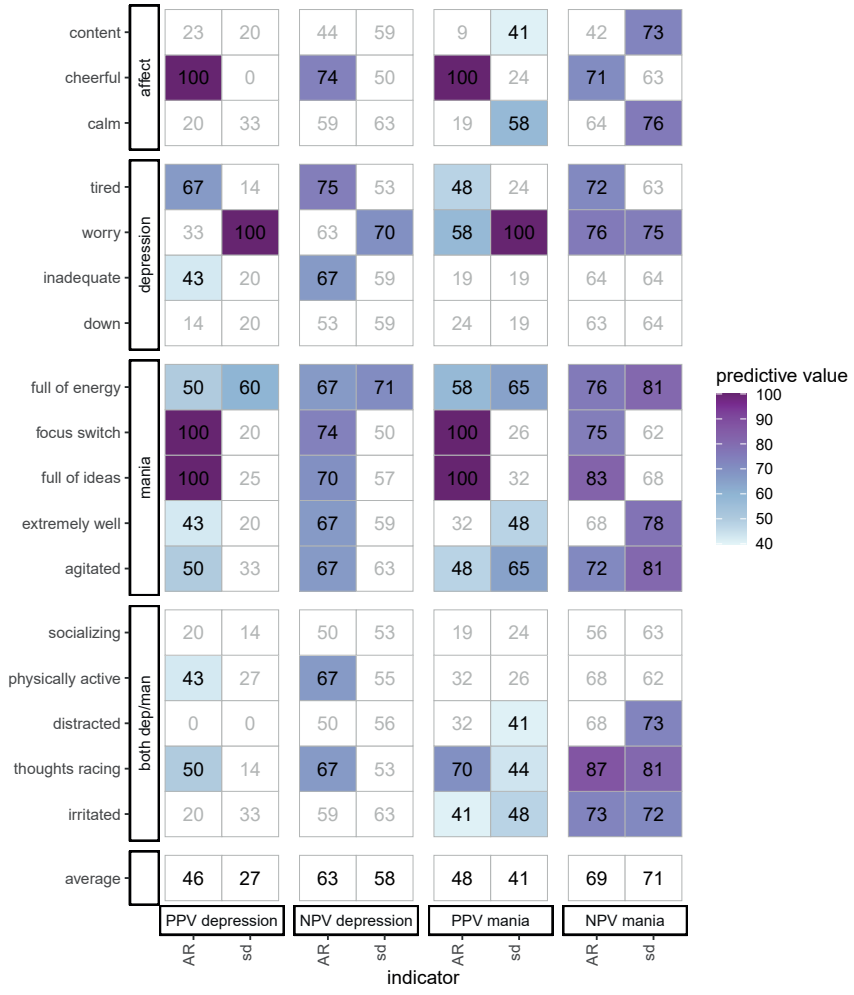
Other momentary states were found to specifically improve the detection of either depressive or manic transitions, as indicated by an average PPV and NPV above 70%. For depressive transitions, the only momentary state yielding such EWS was *'tired'* (autocorrelation). For manic transitions, the best EWS were found in '*racing thoughts*' (autocorrelation), '*agitated'* (standard deviation), and '*full of energy'* (standard deviation).

'Socializing' and '*down'* never improved the detection of either manic or depressive transitions. A particularly inaccurate EWS for manic transitions was *'content'* (autocorrelation). For depressive transitions, three momentary states had a PPV of 0% because EWS were never detected prior to transitions: *'distracted'* (autocorrelation and standard deviation) and *'cheerful'* (standard deviation).

Individual differences

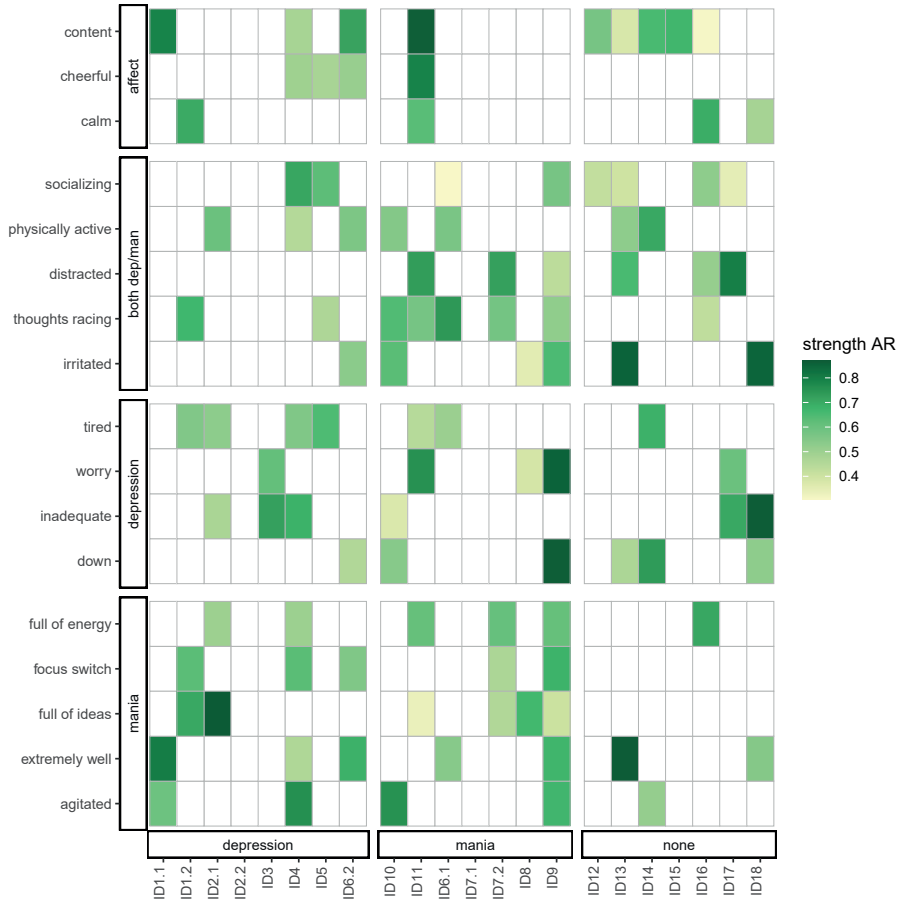
Large individual differences were found in the presence, type, and strength of EWS momentary states (see Figure 9.3). For example, ID4 showed EWS in the autocorrelation for ten momentary states prior to a depressive transition, whereas ID2 reported none prior to their second depressive transition. Whereas the autocorrelation of '*down'* was a particularly strong depressive EWS for ID9, this EWS was not found in any of the seven other transitions. For the four patients that reported two transitions, EWS often did not replicate. Exceptions include the autocorrelations of *'extremely well'* (ID6) and *'physically active'* (ID6), and the standard deviations of '*physically active'* (ID1 and ID7) and *'extremely well'*, *'full of energy'*, *'physically active'*, and *'racing thoughts'* (ID7).

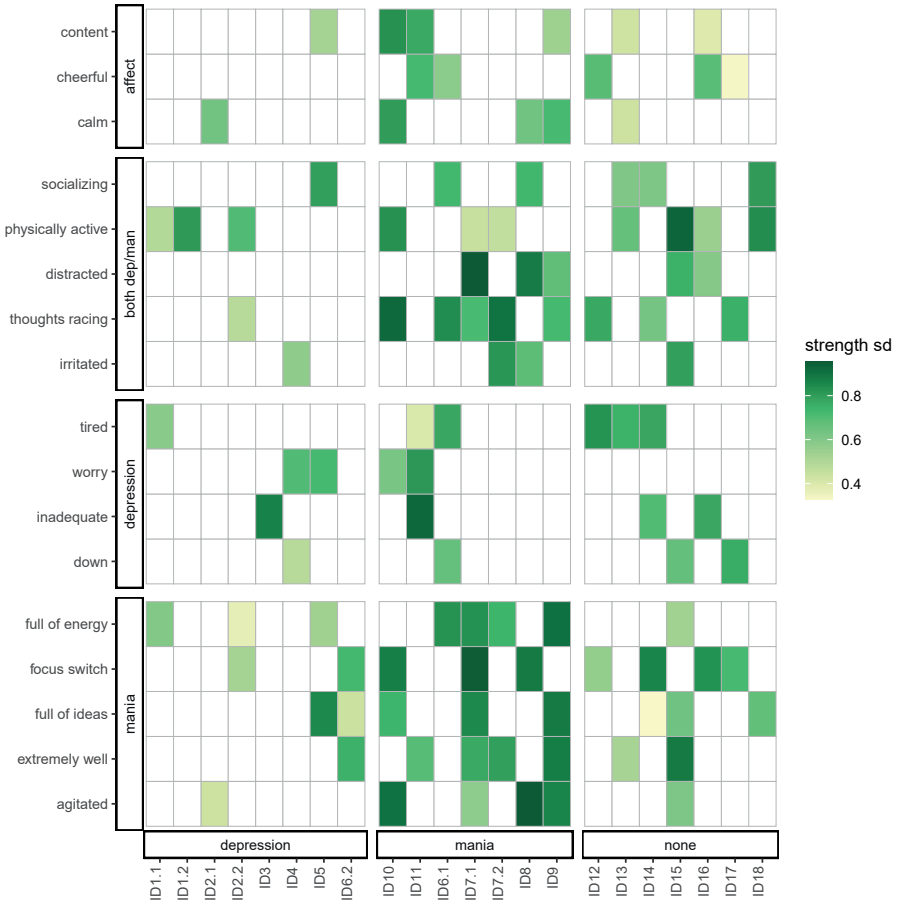
Figure 9.2. Positive and negative predictive values for each early warning signal.



Note. The y-axis represents each momentary state, the x-axis the positive (PPV) and negative (NPV) predictive value, separated for manic and depressive transitions and for the two early warning signals (EWS) indicators: the autocorrelation (AR) and standard deviation (SD). The predictive values can be compared against the prevalence of the transition: the proportion of manic (32%), depressive (36%), or no transitions (68% for mania and 64% for depression). White tiles indicate that this EWS did not improve the detection of a transition above the prevalence of that transition. The color indicates the magnitude of the predictive value for that EWS: the more intense the color, the higher the predictive value. To facilitate interpretation, the EMA momentary states were assigned to summary categories based on hypothesized underlying constructs.

Figure 9.3. Individual differences in the type and strength of the early warning signal.





Note. The y-axis represents each EMA momentary state, the x-axis each transition (note that four individuals had two transitions). To facilitate interpretation, the EMA momentary states were assigned to summary categories based on hypothesized underlying constructs. A colored block indicates that the EWS was significant for that transition. The color intensity indicates the strength of the EWS. Strength of the EWS was operationalized as the value of Kendall's tau. *Abbreviations:* AR = autocorrelation at lag-1, EMA = ecological momentary assessment, EWS = early warning signal, sd = standard deviation.

Discussion

The present exploratory study investigated whether EWS in momentary affective and symptomatic states anticipate abrupt mood transitions in BD patients, and whether EWS might have clinical utility. Results provide preliminary support that EWS can indeed be detected in momentary states that were collected by longitudinal smartphone monitoring, and that EWS could improve the detection of mood transitions. Notably, the presence of EWS increased the probability of detecting impending transitions, but their absence could not be taken as a sign that no transition would occur in the near future. That is, although several momentary states maximized the PPV, indicating that no false positives were found, no momentary state maximized the NPV, indicating the presence of false negatives (i.e., EWS were not always detected prior to transitions). Furthermore, momentary states differed in their predictive utility, and we found large inter-individual differences in the predictive capacity of EWS. Finally, the autocorrelation outperformed the standard deviation as an EWS.

The predictive values point towards several momentary states as promising indicators of nearby depressive and manic transitions. Momentary states that signaled both manic and depressive transitions were *cheerfulness*, *focusing ability*, *full of ideas*, and *worrying*. Manic transitions were most often anticipated by EWS in *racing thoughts*, *agitation,* and *full of energy,* whereas depressive transitions were most often preceded by EWS in *tiredness*. This supports the hypothesis that transitions are best anticipated by EWS in momentary states matching the underlying psychopathology[43]. Importantly, not all momentary states constituted as accurate EWS, highlighting the importance of exploring individual items.

Contrasting to group-level risk factors that offer little guidance on the timing of relapses for individual patients, our study provides preliminary support that EWS might be used to establish *when* individual patients will relapse[44]. If our results are replicated, personalized EWS as generated by smartphone gathered data could be used to alert patients and clinicians to nearby mood shifts. Timely identification of mood shifts is paramount to BD treatment, but is complicated because patients

usually only recognize them when episodes have already started[18]. Early detection using EWS may enable early intervention that could mitigate the severity and impact of ensuing episodes[45]. However, the large heterogeneity we found highlights the necessity of a personalized approach. Furthermore, the finding that EWS had limited replicability within the four individuals with two transitions tentatively suggests that continuous monitoring of various momentary states may be necessary for accurate mood shift detection, as opposed to a less burdensome approach where personalized EWS have to be identified only once prior to a transition. As such, before EWS can find their way to clinical practice, more confirmatory research is needed.

On a critical note, even for the most accurate momentary states, EWS were not always present prior to mood shifts, and were sometimes also found in individuals without actual shifts. False positives may occur when the data show more noise or variability[46]. Patients in our sample indeed showed a high frequency of mood shifts, similar to (ultra) rapid-cycling patients that experience stark fluctuations of mood most of the time [47]. However, failing to anticipate actual mood shifts (false negatives) might be more problematic for clinical applications than the occasional false alarm. The absence of EWS may be explained by the fact that most transitions are often governed by only a few variables[48]; indeed the momentary states that contained EWS differed between the individuals in our sample. False negatives may also have occurred because of the lack of stable mood episodes in the patients in our sample. That is, critical slowing down assumes transitions take place from one stable state to another, whereas the 'stable' state of patients in our sample may well be characterized by large mood instability. Finally, false negatives could occur because the timescale of the EMA assessments (five/day) may not match the timescale at which critical slowing down takes place[49]. Future research may investigate how the bipolar mood system should be characterized and at what timescale it unfolds: should we attempt to anticipate distinct mood episodes, or rather mood instability in general[50]. The latter approach may be more promising given that many BD patients have mixed episodes, as also demonstrated by the patients in our sample. Thus, an interesting follow-up study would be to investigate EWS in more stable BD patients, who transition from stable euthymic states to

manic or depressed states.

The present study is the first to investigate whether EWS anticipate transitions in BD patients in empirical data. Notably, previous studies have only examined autocorrelation and standard deviation as static indicators of future shifts[21-23], whereas we have prospectively examined increases in these indicators to anticipate nearby mood shifts. Further, we used a personalized and idiographic approach that has been advocated to study within-person processes[25]. Other strengths include the relatively large sample (N=20) for idiographic studies, in which each patient could be viewed as a replication of results in other patients, and the large number of observations per person. Furthermore, the diversity in momentary states under investigation might provide new suggestions for confirmatory research.

The present results should be viewed in light of several limitations. First, our sample was diverse, consisting of BD patients with different treatment regimens and comorbid (personality) diagnoses characterized by high mood variability, which might have obscured the relation between EWS and transitions. Second, EWS studies in other fields suggest that results may be dependent on the analytical decisions regarding window size, data detrending, and the period over which the rise is calculated (two weeks)[32,37]. Third, the question remains whether the above-threshold scores on weekly self-report symptom questionnaires adequately reflect abrupt and clinically meaningful transitions. Fourth, our estimates of the prevalence of transitions should be interpreted tentatively given our study design. However, the prevalence of transitions was quite high in this sample, rendering our approach to compare the predictive values against the prevalence rather conservative. Finally, given that our study was exploratory in nature, results may not generalize to other samples or data sets in which different methods to study EWS are employed.

To conclude, EWS show promise in signaling impending transitions to manic and depressive mood shifts in BD. Future confirmatory research may focus on examining EWS in cheerfulness, focusing ability, full of ideas, worry, energy, agitation, racing thoughts, and tiredness, as these items had the best predictive value. Further investigation into the individual characteristics that determine the clinical utility of EWS is warranted.

References

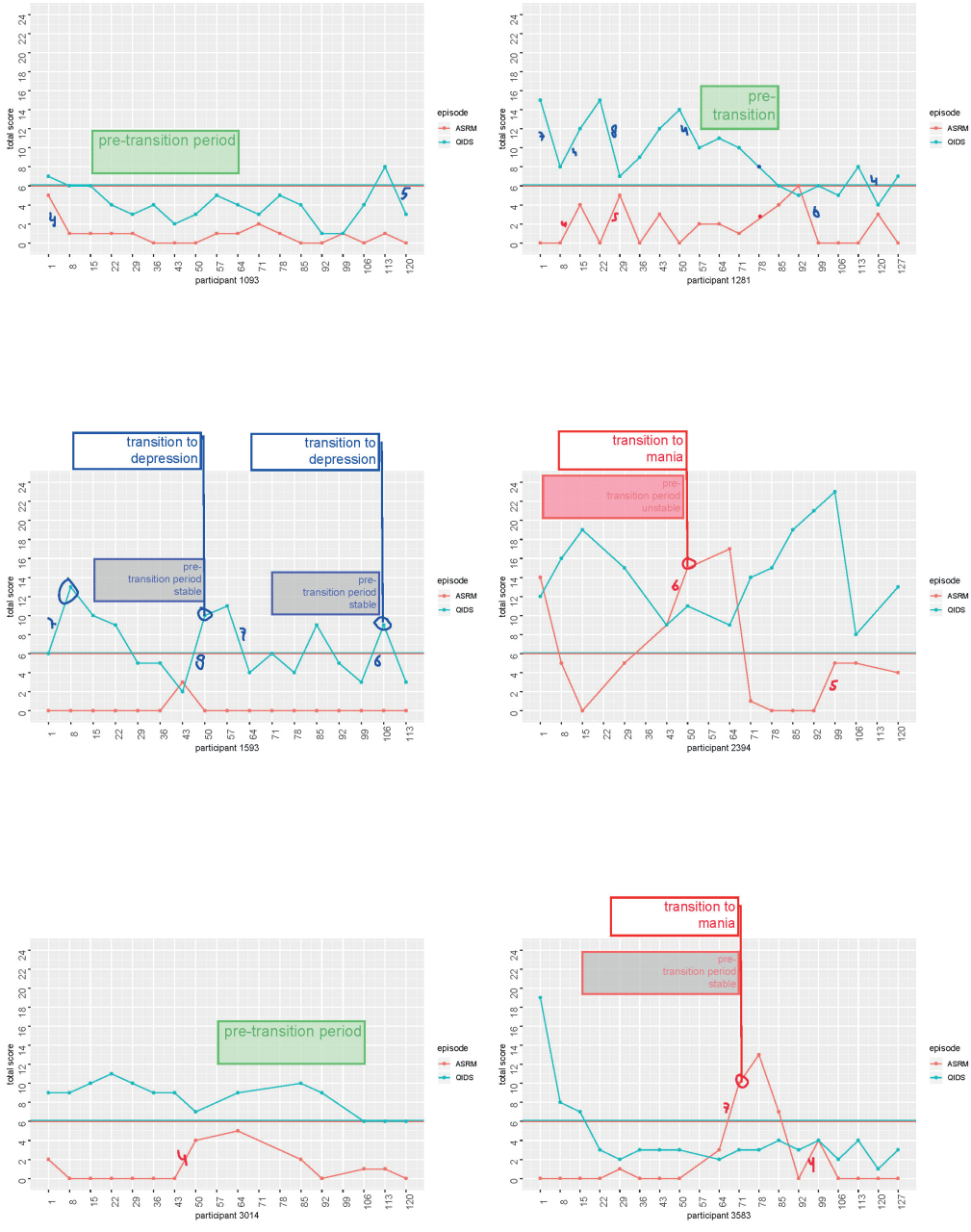
1. van Meter AR, Burke C, Youngstrom EA, Faedda GL, Correll CU. The bipolar prodrome: Meta-analysis of symptom prevalence prior to initial or recurrent mood episodes. *Journal of the American Academy of Child & Adolescent Psychiatry.* 2016;55(7):543-555.
2. Dunster G, Swendsen J, Merikangas KR. Real-time mobile monitoring of bipolar disorder: A review of evidence and future directions. *Neuropsychopharmacology.* 2020:1-13.
3. Michalak EE, Yatham LN, Kolesar S, Lam RW. Bipolar disorder and quality of life: A patient-centered perspective. *Quality of Life Research.* 2006;15(1):25-37.
4. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *JMedInternet Res.* 2019;21(12):e14475.
5. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, Van Nes EH, Rietkerk M, Sugihara G. Early-warning signals for critical transitions. *Nature.* 2009;461(7260):53.
6. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J. Anticipating critical transitions. *Science.* 2012;338(6105):344.
7. Dakos V, Scheffer M, van Nes EH, Brovkin V, Petoukhov V, Held H. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences.* 2008;105(38):14308-14312.
8. George SV, Kachhara S, Misra R, Ambika G. Early warning signals indicate a critical transition in Betelgeuse. *Astronomy & Astrophysics.* 2020;640:L21.
9. Drake JM, Griffen BD. Early warning signals of extinction in deteriorating environments. *Nature.* 2010;467(7314):456-459.
10. Cramer AOJ, van Borkulo CD, Giltay EJ, van der Maas HLJ, Kendler KS, Scheffer M, Borsboom D. Major depression as a complex dynamic system. *PLoS One.* 2016;11(12):e0167490.
11. Hayes AM, Andrews LA. A complex systems approach to the study of change in psychotherapy. *BMC Med.* 2020;18(1):1-13.
12. 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.
13. Olthof M, Hasselman F, Strunk G, van Rooij M, Aas B, Helmich MA, Schiepek G, Lichtwarck-Aschoff A. Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science.* 2020;8(1):25-35.
14. Kuppens P, Allen NB, Sheeber LB. Emotional inertia and psychological maladjustment. *Psychological Science.* 2010;21(7):984.
15. Wichers M, Groot PC, Psychosystems, Group ESM, Group EWS. Critical slowing down as a personalized early warning signal for depression. *PsychotherPsychosom.* 2016;85(2):114.
16. Wichers M, Smit AC, Snippe E. Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: A confirmatory single-subject time-series study. *Journal for Person-Oriented Research.* 2020;6(1):1-15.
17. Cabrieto J, Adolf J, Tuerlinckx F, Kuppens P, Ceulemans E. Detecting long-lived autodependency changes in a multivariate system via change point detection and regime switching models. *Sci Rep.* 2018;8(1):1-15.
18. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, Matthews M. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: Implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association.* 2016;23(3):477-484.
19. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes J, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: A qualitative study. *European Psychiatry.* 2017;41:115-121.
20. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: A qualitative study of patient

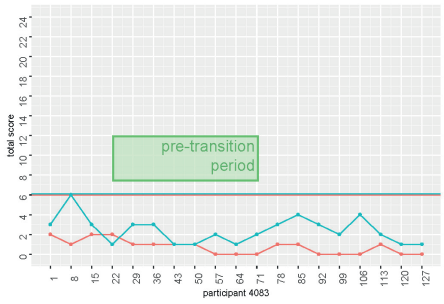
and clinician experiences. *International Journal of Bipolar Disorders.* 2020;8(1):1-14.
21. Bonsall MB, Geddes JR, Goodwin GM, Holmes EA. Bipolar disorder dynamics: affective instabilities, relaxation oscillations and noise. *Journal of the Royal Society Interface.* 2015;12(112):20150670.
22. Ortiz A, Alda M. The perils of being too stable: Mood regulation in bipolar disorder. *Journal of Psychiatry & Neuroscience.* 2018;43(6):363.
23. Curtiss J, Fulford D, Hofmann SG, Gershon A. Network dynamics of positive and negative affect in bipolar disorder. *Journal of Affective Disorders.* 2019;249:270-277.
24. Bayani A, Hadaeghi F, Jafari S, Murray G. Critical slowing down as an early warning of transitions in episodes of bipolar disorder: A simulation study based on a computational model of circadian activity rhythms. *Chronobiology International.* 2017;34(2):235-245.
25. Zuidersma M, Riese H, Snippe E, Booij SH, Wichers M, Bos EH. Single-subject research in psychiatry: Facts and fictions. *Frontiers in Psychiatry.* 2020;11:1174.
26. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman self-rating mania scale. *BiolPsychiatry.* 1997;42(10):948.
27. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *BiolPsychiatry.* 2003;54(5):573.
28. Miller CJ, Johnson SL, Eisner L. Assessment tools for adult bipolar disorder. *Clinical Psychology: Science and Practice.* 2009;16(2):188-201.
29. Knapen SE. *Rhythm & Blues: Chronobiology in the Pathophysiology and Treatment of Mood Disorders*, Rijksuniversiteit Groningen; 2019.
30. van der Krieke L, Jeronimus BF, Blaauw FJ, Wanders RBK, Emerencia AC, Schenk HM, De Vos S, Snippe E, Wichers M, Wigman JTW, Bos EH, Wardenaar KJ, De Jonge P. HowNutsAreTheDutch (HoeGeklsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research.* 2016;25(2):123.
31. Tyler E, Jones S, Black N, Carter L-A, Barrowclough C. The relationship between bipolar disorder and cannabis use in daily life: An experience sampling study. *PLoS One.* 2015;10(3).
32. Dakos V, Carpenter SR, Brock WA, Ellison AM, Guttal V, Ives AR, Kefi S, Livina V, Seekell DA, van Nes EH. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One.* 2012;7(7):e41010.
33. Team RDC. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing; 2014.
34. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016.
35. Auguie B. gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra. 2017.
36. Jäger G, Füllsack M. Systematically false positives in early warning signal analysis. *PLoS One.* 2019;14(2):e0211072.
37. Lenton T, Livina V, Dakos V, Van Nes E, Scheffer M. Early warning of climate tipping points from critical slowing down: Comparing methods to improve robustness. *Philosophical Transactions of the Royal Society A: Mathematical, Physical Engineering Sciences.* 2012;370(1962):1185-1204.
38. Bonsall MB, Wallace-Hadrill SM, Geddes JR, Goodwin GM, Holmes EA. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. *Proceedings of the Royal Society B: Biological Sciences.* 2012;279(1730):916-924.
39. McLeod AI. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. https://CRAN.R-project.org/package=Kendall. 2011.
40. Hamed KH, Rao AR. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology.* 1998;204(1-4):182-196.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B.* 1995;57(1):289-300.
42. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: Predictive values. *Bmj.* 1994;309:102.
43. Wichers M, Schreuder MJ, Goekoop R, Groen RN. Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology.

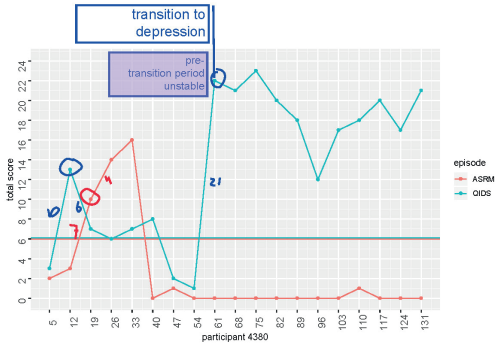
PsycholMed. 2019;49(3):380-387.
44. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences.* 2018;115(27):E6106-E6115.
45. Morriss R. Clinical importance of inter-episode symptoms in patients with bipolar affective disorder. *Journal of Affective Disorders.* 2002;72:S3-S13.
46. Boettiger C, Ross N, Hastings A. Early warning signals: The charted and uncharted territories. *Theoretical Ecology.* 2013;6(3):255-264.
47. Kramlinger KG, Post RM. Ultra-rapid and ultradian cycling in bipolar affective illness. *The British Journal of Psychiatry.* 1996;168(3):314-323.
48. van der Maas HL, Molenaar PC. Stagewise cognitive development: An application of catastrophe theory. *PsycholRev.* 1992;99(3):395.
49. Haslbeck J, Ryan O. Recovering within-person dynamics from psychological time series. *https://psyarxivcom/dymhw/.* 2020.
50. Hadaeghi F, Hashemi Golpayegani MR, Moradi K. Does "crisis-induced intermittency" explain bipolar disorder dynamics? *Frontiers in Computational Neuroscience.* 2013;7:116.

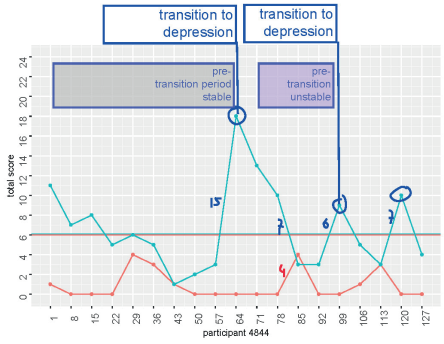
Appendix

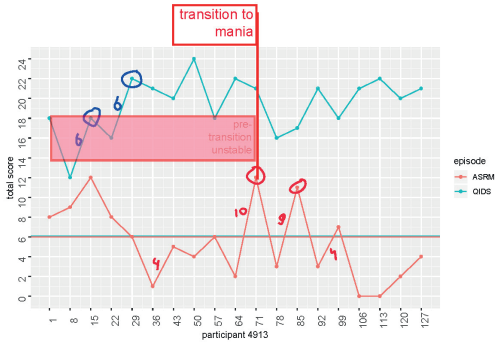
Figure 9.4. All transitions per individual.

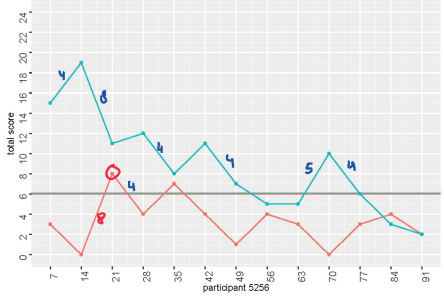


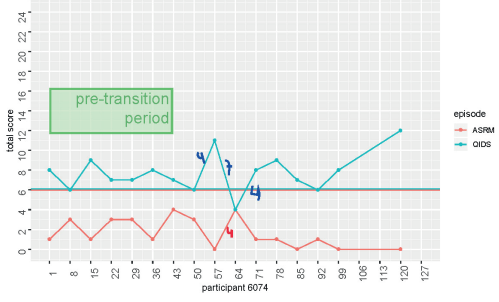

pre-transition period
total score
participant 4083
episode
ASRM
QIDS

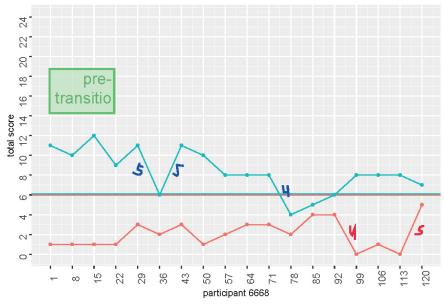

transition to depression
pre-transition period unstable
total score
participant 4380
episode
ASRM
QIDS

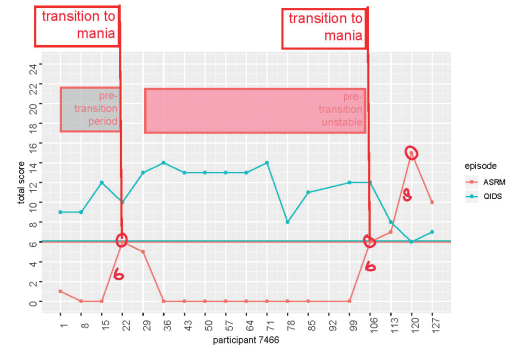

transition to depression
transition to depression
pre-transition period stable
pre-transition unstable
total score
participant 4844
episode
ASRM
QIDS

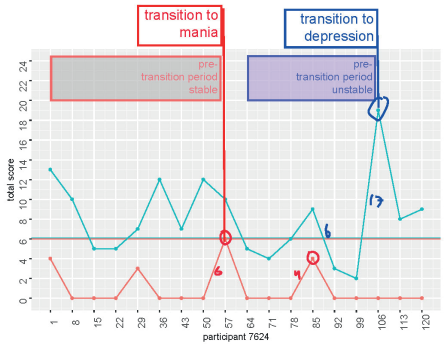

transition to mania
pre-transition unstable
total score
participant 4913
episode
ASRM
QIDS

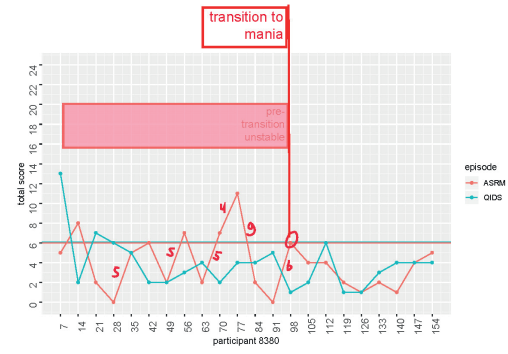

total score
participant 5256
episode
ASRM
QIDS

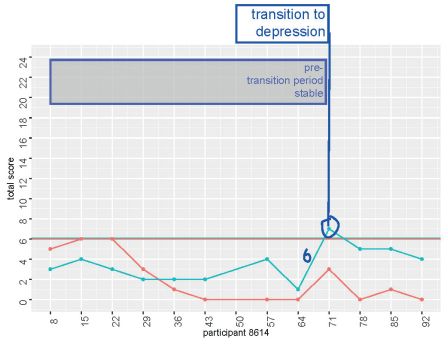

pre-transition period
total score
participant 6074
episode
ASRM
QIDS

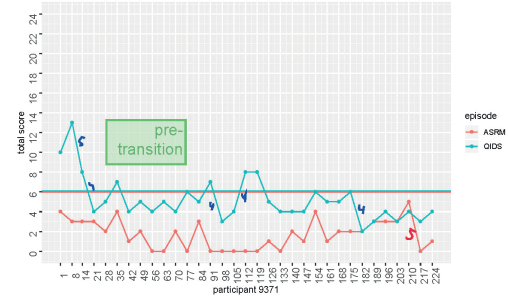

pre-
transitio
total score
participant 6668
episode
ASRM
QIDS

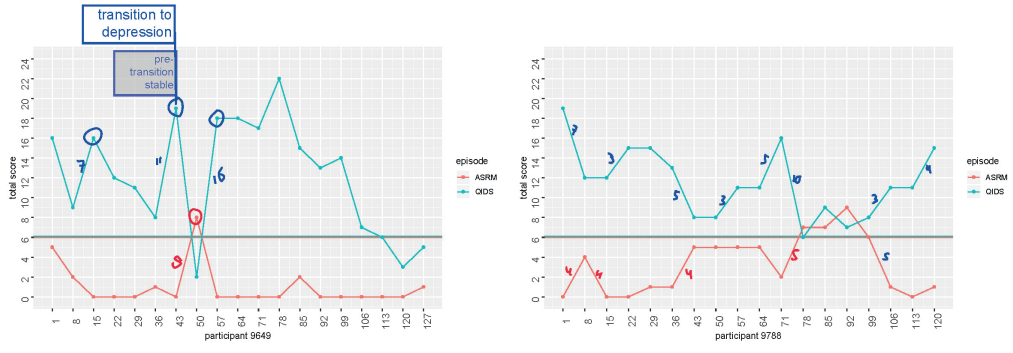

transition to
mania
transition to
mania
pre-
transition
period
pre-
transition
unstable
total score
participant 7466
episode
ASRM
QIDS


transition to
mania
transition to
depression
pre-
transition period
stable
pre-
transition period
unstable
total score
participant 7624
episode
ASRM
QIDS


transition to
mania
pre-
transition
unstable
total score
participant 8380
episode
ASRM
QIDS


transition to
depression
pre-
transition period
stable
total score
participant 8614
episode
ASRM
QIDS


pre-
transition
total score
participant 9371
episode
ASRM
QIDS



2. Experience sampling method diary items

The item list (see Table 9.2) was constructed in several steps. First, we identified relevant concepts for symptoms of bipolar disorder and searched the literature for EMA studies in patients with bipolar disorder. This yielded a first item list of 71 potentially relevant items. This list was then discussed in individual interviews with three patients and one psychiatrist. The items were finally selected on three criteria: 1) the patients and clinician recognized the item as signaling symptoms of either mania or depression; 2) the patients and the clinician felt comfortable with the formulation of the item (e.g., they could see themselves saying the sentence in daily life); 3) both the patients, the clinician, and the research team believed the item would vary meaningfully within participants. Items that were deemed relevant but too person-specific were put on a list for participants so they might select them for their personal question.

All items were obligatory, with the exception of the comment field at the end of the questionnaire (item 29). All items were assessed five times per day, with the exceptions of the items regarding sleep (item 2- 3), and the item regarding appointments (item 27). These items were only shown if participants answered 'yes' on item 1 or item 26. This way, we ensured that participants could still answer questions about their sleep, even though they might have skipped the first (few) assessments.

Table 9.2. Ecological momentary assessment (EMA) diary items.

Dutch	English translation	Anchors (far left, middle, far right)	Which pompt(s)	Reason for exclusion
Is dit de eerste meting die u invult vandaag?	Is this the first assessment you complete today?	- Yes - No	All five prompt	
1. Hoe lang heeft u geslapen?	How long did you sleep?	0 – 12 hours	If answered 'yes' on item 1	Assessed on different timescale, yielding too few data points for analysi
2. De kwaliteit van mijn slaap was…	The quality of my sleep was…	Very bad – reasonably – very good	If answered 'yes' on item 1	Assessed on different timescale, yielding too few data points for analysis
Ik voel me opgewekt	I feel cheerful	Not at all – reasonably – very much	All five prompts	
Ik voel me neerslachtig	I feel down	Not at all – reasonably – very much	All five prompts	
Ik voel me kalm	I feel calm	Not at all – reasonably – very much	All five prompts	
Ik voel me gejaagd	I feel agitated	Not at all – reasonably – very much	All five prompts	
Ik voel me bijzonder goed	I feel extremely well	Not at all – reasonably – very much	All five prompts	
Ik voel me moe	I feel tired	Not at all – reasonably – very much	All five prompts	
Ik voel me tevreden	I feel content	Not at all – reasonably – very much	All five prompts	
Ik voel me geïrriteerd	I feel irritated	Not at all – reasonably – very much	All five prompts	
Ik zit vol energie	I am full of energy	Not at all – reasonably – very much	All five prompts	
Ik zie op tegen de rest van de dag	I dread the rest of the day	Not at all – reasonably – very much	All five prompts	Responses invalid for one participant who accidentally misinterpreted the scale
Ik zit vol goede ideeën	I am full of good ideas	Not at all – reasonably – very much	All five prompts	
Ik heb het gevoel te kort te schieten	I feel inadequate	Not at all – reasonably – very much	All five prompts	
Mijn gedachten gaan snel	My thoughts are racing	Not at all – reasonably – very much	All five prompts	
Ik kan snel schakelen	I am able to focus and switch easily	Not at all – reasonably – very much	All five prompts	
Ik ben snel afgeleid	I'm distracted easily	Not at all – reasonably – very much	All five prompts	
Ik heb zin om met anderen af te spreken	I feel like socializing	Not at all – reasonably – very much	All five prompts	
Eigen vraag	Personal question	Not at all – reasonably – very much	All five prompts	Item differs across participants

Dutch	English translation	Anchors (far left, middle, far right)	Which pompt(s)	Reason for exclusion
Ik heb meer gegeten dan gewoonlijk	I have eaten more than usual	Not at all – reasonably – a lot more than usual	All five prompts	Too dependent on the occurrence of events
Sinds het vorige meetmoment heb ik gepiekerd	Since the previous prompt, I have worried	Not at all – reasonably – very much	All five prompts	
Sinds het vorige meetmoment heb ik veel gecommuniceerd	Since the previous prompt, I have communicated a lot	Not at all – reasonably – very much	All five prompts	Too dependent on the occurrence of events
Sinds het vorige meetmoment heb ik me lichamelijk ingespannen	Since the previous prompt, I have been physically active	Not at all – reasonably – very much	All five prompts	
Denk aan de meest opvallende gebeurtenis sinds het vorige meetmoment. Hoe heftig was deze gebeurtenis?	Think back on the most notable event since the previous prompt. How intense was this event?	Not at all – reasonably – very much	All five prompts	Too dependent on the occurrence of events
Is dit de laatste meting die u invult vandaag?	Is this the last assessment you complete today?	- Yes - No	All five prompt	
3. Ik heb vandaag veel afspraken gemaakt	I have made a lot of appointments today	Not at all – reasonably – a lot	If answered 'yes' on the previous item	Assessed on different timescale, yielding too few data points for analysis
Sinds het vorige meetmoment, wat deed ik?	Since the previous prompt, what have I been doing? (multiple options possible)	- Sleeping - Household chores/groceries - Working/studying - Doing sports/walking/cycling - Something relaxed (e.g., reading, TV) - Hobby (e.g., gardening, music) - A trip (e.g., into town, concert) - Something together with others - Something intimate (e.g., cuddling, sex) - Engaging in self-care - Resting/nothing - On the way - Something else	All five prompts	Not a continuous variable
Noteer hier eventuele opmerkingen. Noteer het ook als er iets gebeurd is dat invloed heeft op uw stemming.	Note observations here if any. Also note anything that may have influenced your mood.	Open entry	All five prompts (optional)	Not a continuous variable

4. Results of the sensitivity analyses for the window size

All early warning signal (EWS) analyses were run for window sizes of 1, 2, and 3 weeks. Table 9.3 shows the predictive values, sensitivity, and specificity of the autocorrelation and standard deviation, averaged across all EMA diary items. The differences between the window sizes are quite small. One notable finding is that a three-week window is preferable when examining the autocorrelation prior to transitions to depression and mania. For the standard deviation, the two-week window was preferable.

Table 9.3. Predictive values, sensitivity, and specificity of early warning signals, for three different window sizes, averaged across all EMA diary items.

	Depression			Mania			
	PPV	**NPV**	**Sensitivity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**
Indicator: AR							
1 week window	43	65	25	45	70	25	78
2 week window	**46**	**63**	**27**	**48**	**69**	**33**	**75**
3 week window	59	69	36	58	73	30	84
Indicator: SD							
1 week window	32	61	18	42	70	30	72
2 week window	**27**	**58**	**23**	**41**	**71**	**38**	**67**
3 week window	18	38	15	38	68	30	67

Abbreviations: AR = autocorrelation, PPV = positive predictive value, NPV = negative predictive value, SD = standard deviation.

5. Calculation of positive and negative predictive values, sensitivity, and specificity

See Table 9.4 for all positive and negative predictive values, sensitivity, and specificity. The sensitivity of EWS reflects the probability of EWS, given that a transition is present. Sensitivity was calculated per indicator (autocorrelation, standard deviation), momentary state (*i*) and transition type (*t*; depression, mania) by dividing the number of EWS in momentary state *i* that anticipated transition *t* by the total number of transitions of type *t*. As an example, the autocorrelation in feeling tired anticipated four out of eight transitions towards depression, resulting in a sensitivity of .50 (i.e., 50%).

$$sensitivity_{i,t} = \frac{EWS_{i,t}}{N_t}$$

The specificity of EWS reflects the probability of *no* EWS, given that there was no transition. This statistic was calculated by inspecting EWS in those individuals who did not experience a transition (N = 7). A specificity of 1 (i.e., 100%) indicates that EWS in this particular item were never found in individuals without transitions.

$$specificity_i = 1 - \frac{EWS_i}{N}$$

Positive predictive values (PPVs) indicate the probability of a transition towards either mania or depression, given that EWS are detected. These values were calculated as follows:

$$PPV_{i,t} = \frac{sensitivity_{i,t} * prevalence_t}{sensitivity_{i,t} * prevalence_t + (1 - specificity_i) * (1 - prevalence_t)}$$

In this formula, the prevalence of a transition of type *t* reflects the number of transitions of type *t* divided by the total number of transitions. In total, we identified 22 transitions, of which 8 (36%) involved a sudden increase in symptoms of depression, 7 (32%) involved an increase in manic symptoms, and 7 (32%) were simulated in individuals without transitions. The latter 'non-transitions' were used to estimate the specificity of EWS.

Negative predictive values (NPVs) indicate the probability that *no* transition will occur, given that EWS are *not* detected. These values were calculated as follows:

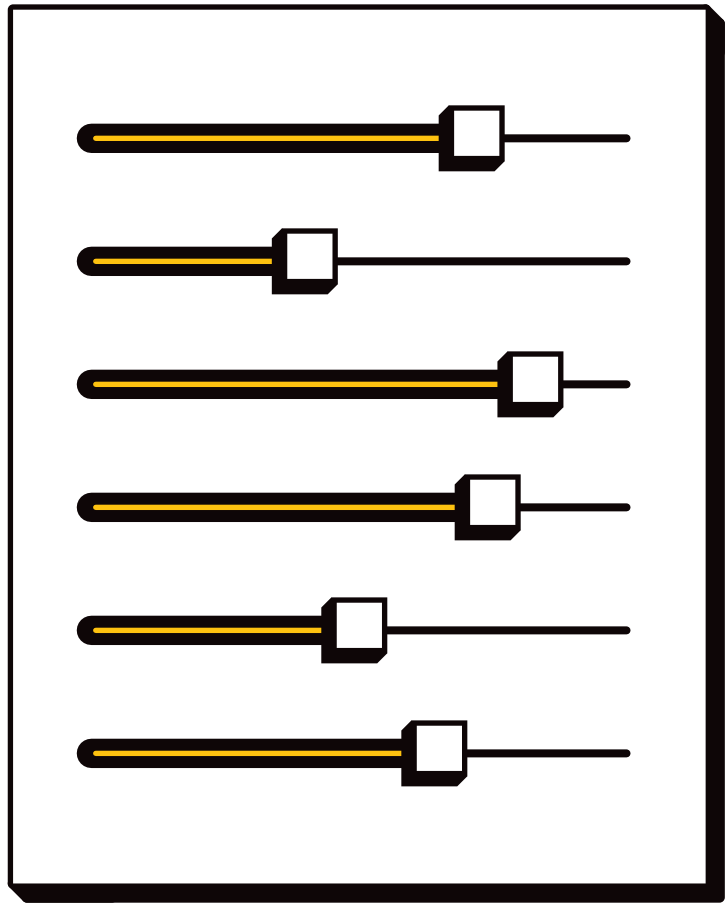
$$NPV_{i,t} = \frac{specificity_i * (1 - prevalence_t)}{(1 - sensitivity_{i,t}) * prevalence_t + specificity_i * (1 - prevalence_t)}$$

Note that the probability of *no* transition is defined by 1 minus the prevalence of a particular type of transition, resulting in 64% (depression) and 68% (mania), respectively. The prevalence of a transition can be thought of as a set-point: in absence of any insight in EWS, the probability that a specific individual will (not) experience a transition equals the average probability of (no) transitions. EWS can be considered informative if (1) their presence considerably heightens the risk of a future transition and (2) their absence lowers the probability that a transition will occur. Therefore, in the below table, we printed PPV and NPV values that exceeded the expected probabilities in bold ($prevalence_t$ or $1 - prevalence_t$.)

Table 9.4. Predictive values, sensitivity, and specificity of early warning signals for transitions towards depression and mania.

	Depression			Mania			
	PPV	**NPV**	**Sensitivity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**
Indicator: AR							
calm	20	59	12	19	64	14	71
cheerful	**100**	**74**	38	**100**	**71**	14	100
content	23	44	38	9	42	14	29
irritated	20	59	12	**41**	**73**	43	71
thoughts racing	**50**	**67**	25	**70**	**87**	71	86
distracted	0	50	0	32	68	43	57
physically active	**43**	**67**	38	32	68	29	71
socializing	20	50	25	19	56	29	43
down	14	53	12	24	63	29	57
inadequate	**43**	**67**	38	19	64	14	71
worry	33	63	12	**58**	**76**	43	86
tired	**67**	**75**	50	**48**	**72**	29	86
agitated	**50**	**67**	25	**48**	**72**	29	86
extremely well	**43**	**67**	38	32	68	29	71
full of ideas	**100**	**70**	25	**100**	**83**	57	100
focus switch	**100**	**74**	38	**100**	**75**	29	100
full of energy	**50**	**67**	25	**58**	**76**	43	86
Indicator: SD							
calm	33	63	12	**58**	**76**	43	86
cheerful	0	50	0	24	63	29	57
content	20	59	12	**41**	**73**	43	71
irritated	33	63	12	**48**	**72**	29	86
thoughts racing	14	53	12	**44**	**81**	71	57
distracted	0	56	0	**41**	**73**	43	71
physically active	27	55	38	26	62	43	43
socializing	14	53	12	24	63	29	57
down	20	59	12	19	64	14	71
inadequate	20	59	12	19	64	14	71
worry	**100**	**70**	25	**100**	**75**	29	100
tired	14	53	12	24	63	29	57
agitated	33	63	12	**65**	**81**	57	86
extremely well	20	59	12	**48**	**78**	57	71
full of ideas	25	57	25	32	68	43	57
focus switch	20	50	25	26	62	43	43
full of energy	**60**	**71**	38	**65**	**81**	57	86

Abbreviations. AR = autocorrelation, NPV = negative predictive value, PPV = positive predictive value, SD = standard deviation.



Chapter 10

General summary and discussion

Preface

In this thesis, I have explored the utility of ecological momentary assessment (EMA) for psychiatric care. In the first part, I have investigated the promise of the network theory for providing personalized models based on EMA data. In the second part, I have more broadly examined the clinical utility of EMA and EMA feedback. In this final Chapter, I will place our findings in the perspective of other recent scientific advances on EMA. First, I will summarize the main findings that originated from this thesis. Next, I will postulate the three main lessons learned regarding the clinical utility of EMA. Finally, I will discuss methodological challenges and recommend focus points for future research. I will end with some concluding remarks.

Summary of main findings

In the first Part of this thesis, I have focused on investigating the potential of personalized models, and of network analysis specifically, to inform on clinical practice.

In **Chapter 2**, we used cross-sectional network analysis to attempt to uncover the working mechanisms of antidepressant medication treatment, selective serotonin reuptake inhibitors (SSRIs) in particular. Results demonstrated that depressive symptoms significantly decreased over time and that they became increasingly interconnected. Furthermore, we found that the clustering of depressive symptoms and their centrality differed at baseline versus after eight weeks of antidepressant treatment. The low sample size and the lack of a control group limited our ability to draw clear conclusions on the working mechanisms of SSRIs. Furthermore, since cross-sectional network associations reflect between-person associations, we were unable to generalize our findings to individual patients. Given these limitations, we concluded that the cross-sectional network approach is at most suggestive of hypotheses that need to be tested in confirmatory research.

Therefore, in **Chapter 3**, we turned to dynamic network analysis. We conducted impulse response function analysis (IRF) on EMA data to attempt to

understand the poorer prognosis of individuals reporting anhedonia (here defined as a loss of interest). We hypothesized that this poorer prognosis might be reflected in the emotional dynamics of individuals with subclinical depression. Therefore, we compared the emotional dynamics of subclinically depressed individuals with anhedonia to those without. Interestingly, although the two groups were matched and the clinical characteristics of both groups were thus the same, their emotional dynamics diverged. Those with anhedonia showed a diminished beneficial impact of positive affect, whereas those without anhedonia showed a stronger impact of stress. We therefore speculated that different affective pathways may underlie the development of depressive symptoms and argued that such personalized models of emotional dynamics may provide relevant information for treatment.

To understand the relationship between cross-sectional and dynamic networks better, in **Chapter 4** we directly compared these two types of networks in empirical data. Were both methodological approaches to yield similar conclusions, we expected that between-person findings as generated by cross-sectional networks were informative of within-person associations. However, each network demonstrated different associations and indicated different targets for interventions. We therefore concluded that between-person associations identified in cross-sectional networks cannot be taken to reflect how symptoms trigger each other over time, and should not be interpreted as such.

In the final Chapter of Part I, we reviewed the network literature in the field of depression research **(Chapter 5)**. We also added to the existing network theory by introducing the micro-level affect dynamics (MAD) network theory. This review demonstrated that such micro-level network studies yielded different conclusions than the more traditionally used macro-level network approach regarding several important network hypotheses, for example the relative importance of depressive symptoms (i.e., centrality), and the hypothesis that stronger network connectivity is related to higher severity of depression. Therefore, we recommended researchers to clearly differentiate between the two network theories when discussing their results. Furthermore, given the reviewed state of the evidence yielded by network studies so far, we concluded that empirical support is unfortunately not yet suggestive of

the use of networks in clinical practice.

In Part II of this thesis I have explored the clinical utility of EMA in diverse settings. First, in **Chapter 6**, we reviewed the literature on the use of EMA in the field of psychopharmacology. This review identified several potential applications of EMA in multiple phases of psychopharmacological treatment: for example, using baseline affective dynamics to predict treatment outcome, or assessing the duration of medication effects. We concluded that EMA could have interesting possibilities for personalizing psychopharmacological treatment and understanding the effects of medication in daily life.

In **Chapter 7**, we assessed the expectations of patients and clinicians about the clinical potential of EMA using interviews and focus groups. Patients and clinicians with diverse backgrounds believed that EMA could be relevant in every phase of care, and that EMA may improve the patient-clinician relationship and the self-management of patients. However, they were also apprehensive of EMA's high assessment burden and potential worsening of symptoms due to the constant focus on them. Patients and clinicians further offered several recommendations for implementation, for example, the importance of the ability to personalize EMA to fit the treatment goals and needs of individual patients.

These findings were further emphasized in **Chapter 8**. Here, we assessed the experiences of patients with bipolar disorder and their clinicians with the addition of four months of EMA and personalized EMA feedback to treatment. Many patients and clinicians confirmed the aforementioned positive and negative effects from EMA monitoring and feedback. Although EMA monitoring was generally seen as beneficial to patients' self-management, the personalized feedback was viewed less favorably. Indeed, there seemed to be a mismatch between what patients expected from the EMA feedback and what they received, highlighting the importance of making EMA feedback clinically relevant. Thus, we concluded that patients and clinicians should collaborate to ensure EMA achieves its intended goal, and made recommendations for discussion points to aid patients and clinicians herein.

Finally, in **Chapter 9**, we investigated the potential of EMA to form a personalized alert system, to warn patients with bipolar disorder and their clinicians

of upcoming manic or depressive mood shifts. We used four months of EMA data to investigate whether, prior to such mood shifts, we could detect early warning signals could be detected in diverse momentary affective states. Positive and negative predictive values were calculated to examine the clinical utility of early warning signals. Indeed, for several EMA affective states, predictive values of detecting transitions were improved by using early warning signals (i.e., rises in autocorrelation and variance). However, we also found large individual differences in the utility of early warning signals. Therefore, we concluded that early warning signals in some EMA affective states show promise to use as a basis for an alert system, but much more confirmatory research is necessary before this can be used as a clinical application.

Lessons learned

This thesis aimed to explore the clinical utility of EMA and personalized models based on EMA data. In the following sections, I will discuss the lessons learned regarding the implementation of EMA, and place the findings generated by this thesis in a broader context.

Lesson 1. Simplicity over complexity: Personalized networks based on EMA data are not yet ready for use in clinical practice

Early on, the clinical utility of EMA has been tied to the potential of personalized models based on EMA data[2]. Personalized networks in particular are proposed as a way to inform patients and clinicians on associations between affect and symptoms relevant to the etiology and maintenance of psychopathology. Network theory appeals intuitively to many clinicians, as they often construct functional analyses in which they attempt to understand the (causal) relationships between the patient's symptoms, emotions, cognitions, and contexts[3]. Personalized networks that reveal such relationships for individual patients might therefore add to the clinician's toolbox. For this reason, several studies have started to experiment with providing personalized networks as feedback to patients[4-9]. In this section, I will first evaluate

whether findings generated by network studies can have direct clinical implications, and secondly, whether personalized network models may indeed be useful for clinical practice.

In Chapter 2 and 3, we used cross-sectional and dynamic network methodologies to advance our knowledge on the working mechanisms of SSRIs and the poorer prognosis of people with anhedonia, respectively. The findings of these studies may leave both clinicians and clinical researchers inclined to make inferences about individual patients. For example, in Chapter 2, we speculated that the high centrality of anhedonia after antidepressant treatment may suggest that patients who still exhibit this symptom might be vulnerable to relapse. However, we are yet far from such direct clinical implications. As we will explore further below, network associations cannot be taken as causal. Furthermore, associations in cross-sectional networks such as the one in Chapter 2 are between-person, or group-level, indicating that they reveal associations averaged across individuals. Chapter 4 and other publications since then[10-13] have demonstrated that such between-person associations do not appear to be generalizable to within-person associations. Thus, we cannot conclude from cross-sectional networks that indeed, for individual patients, anhedonia implies relapse vulnerability. Another problem is that researchers in the field are still debating the question whether the uncovered network structures can be considered stable, generalizable, and replicable[14-18]. Furthermore, in the dynamic networks of Chapter 3, we have seen large individual heterogeneity in the associations, which has been highlighted by other publications as well[11,19-21]. Findings of group-level networks may therefore not always have relevance for individual patients. However, they can identify promising areas for future confirmatory study.

Problems with group-level networks notwithstanding, the main promise of the network theory for clinical practice lay in the construction of personalized network models based on EMA data. A wide variety of analytical models have been developed to estimate such personalized models, mostly variations of vector autoregressive (VAR) models[6,20,22-25]. Unfortunately, it is yet unclear which can be considered the golden standard. Even seemingly trivial methodological

choices, such as handling missing data or variable selection, can greatly affect the resulting network structure, and thus, the resulting conclusions drawn[26]. Why this is problematic is convincingly demonstrated by Bastiaansen and colleagues[27]. They invited twelve research teams to analyze the same EMA data of an individual patient and to offer recommendations for treatment based on their analysis. Although not a prerequisite, many teams employed network (VAR) analysis to analyze the EMA data. Strikingly, there was little overlap in i) the methods used to analyze the data, ii) the uncovered associations, and iii) the recommendations suggested, even when relying on a similar analytic approach. For this reason, one of the teams emphasized that final decisions on treatment targets yielded by the network should only be made in dialogue with the patient[27]. Nonetheless, this study shows that, with so many researcher degrees of freedom in the analysis process, we currently cannot be certain that networks analysis yields valid and reliable associations that have relevance for treatment.

Another issue related to personalized network models for clinical practice relates to the EMA data that is used to construct networks. Whereas affect and symptoms are typically assessed on dimensional scales (e.g., visual analogue scales (0-100) or Likert scales (1-7)), context (e.g., activities, company, location) is assessed by providing patients with lists of categories. For example, for activity, such categories may include working, exercising, and relaxing. Furthermore, other relevant experiences might be assessed on a different timescale: the past night's sleep quality can be assessed only once daily, whereas affect is assessed multiple times per day. Unfortunately, networks based on VAR models cannot deal with categorical variables or variables assessed on different timescales[28]. This is a missed opportunity, because it is precisely the causal interplay between patients' (social) contexts and his or her symptomatology that is often a central focus in treatment. Relatedly, network associations are dependent on the EMA assessment schedule; usually they are at least 90 minutes to even 6 hours apart[29]. This means that rapid (causal) chains of events, such as those expected to take place when experiencing a panic attack, are difficult to show in personalized networks[6,30]. As such, networks based on VAR models currently are likely to miss relevant clinical information.

Even if these issues were to be resolved, networks based on VAR analyses face the limitation that the associations uncovered are temporal, not causal[31]. One of the most appealing promises of network theory lay in the identification of central symptoms: these were suggested to be highly influential symptoms that may maintain or even trigger other symptoms, making them potentially effective treatment targets[32-36]. This would have direct clinical implications: if such a central symptom were present in an individual patient, they may make this person vulnerable to the development of more symptoms. However, this hypothesis is now considered controversial. It is questionable whether network associations are indeed causal in nature[31,37,38], and whether an intervention on the most central node is both feasible and effective[39]. Unfortunately, associations in VAR networks may falsely imply causality to patients and clinicians, whereas they are at most temporally related. For example, networks may reveal a temporal association between worrying and sadness. Depending on the structure of the EMA data, this means that worrying at one point in time *may* precede sadness, say, three hours later. It does not mean that sadness is *caused* by worrying, or that worrying *always* precedes sadness. Furthermore, if the network model did not demonstrate a significant association between two network nodes – for example, between physical activity and affect (in Chapter 3) – this does not mean that this association *does not exist* for this person. It could be that physical activity *is* beneficial to a patient's mood, but that we focused on the wrong variables to assess physical activity or mood, or that we used the wrong timescale in the network[40]. In Chapter 3, we examined the relationship between physical activity and affect over six hours. It might be that the beneficial effects of physical activity are not visible when examining such a short time span, and that a long-term intervention promoting activity does improve mood. This nuance is, of course, easily lost when discussing networks in the treatment room. Indeed, as we have seen in Chapter 8, patients may attach a (too) great importance to scientific analysis of their data. Furthermore, in the rare occasions where we provided personalized network models to patients and clinicians, we noticed that they were perceived as difficult to understand (Chapter 7, 8, and Box 10.1). The findings of Part I of this thesis thus all point to the fact that we should exercise caution when bringing such personalized network models to clinical practice.

To conclude, the highly anticipated promise of EMA to illuminate relevant associations in personalized network models is currently not sufficiently supported by empirical evidence. Some researchers have argued that networks could still have clinical utility in facilitating discussion between patient and clinician[28]. However, more simple visualizations and representations of the EMA data could be equally capable of inspiring such discussion[41]. For example, bar graphs displaying the level of positive mood during multiple activities also visualize the relationship between mood and behavior without directly implying causality. By not testing these associations for significance, patients and clinicians have more room to consider their own interpretations of the findings. Therefore, I recommend that EMA feedback consists of simple, descriptive feedback until personalized statistical models are supported by empirical research.

Lesson 2. Seizing the moment: The promises of EMA are recognized by patients and clinicians – under some conditions

In this thesis, we aimed to investigate *if, how,* and *when* EMA might be useful for clinical practice. To that end, we have explored several promising clinical applications of EMA. In our qualitative studies, patients and clinicians agreed with the promise of EMA for diverse clinical applications that largely overlap with those previously suggested by research (e.g., Chapter 6 and [1,2,42-45]). They suggested EMA might provide relevant information for diagnostics, treatment, and relapse prevention, and may improve patients' insight, self-awareness, and behavioral change (Chapter 7 and 8). Notably, we have learned that EMA monitoring in itself may already be considered an intervention, whereas the promise of EMA feedback was viewed somewhat more cautiously. These findings add to the results of other quantitative and qualitative studies, where patients reviewed EMA as useful and feasible in research settings[46-48]. Furthermore, it suggests that EMA might indeed, as expected, benefit the self-management of patients and facilitate communication between patient and clinician[2,45,49-51]. This was suggested to occur especially because EMA may reduce memory biases[52], and therefore help the patient explain how he or she fared in between treatment sessions. Information relevant to treatment is therefore less likely to be missed[53]. Although the possibility of real-time advice and therapy

exercises – termed ecological momentary interventions (EMI)[54] – was not explicitly mentioned by patients and clinicians, research suggests this might further increase the clinical utility of EMA[55,56]. This could also entail real-time feedback on deviating symptom patterns that may alert patients and clinicians to impending relapses (as explored in Chapter 9), although such an alert system is still in very a preliminary stage[57]. The eagerness of both patients and clinicians to use EMA in treatment provides a solid basis to further develop several clinical applications of EMA and to test out its effectivity in diverse clinical settings.

Regarding the question *for whom* EMA can be expected to be most beneficial, patients and clinicians argued against excluding certain patient groups a priori. This is corroborated by early studies suggesting that EMA is feasible for individuals with a diverse range of mental disorders[58-60], from depressive disorders[50,61,62], bipolar disorder[46,63], anxiety disorders[64], psychotic disorders[65-67], autism spectrum disorder[68,69], and borderline personality disorder[70-72], to eating disorders [73,74]. However, rather than asking *for whom* EMA is beneficial, the question should be rephrased to *when*: EMA was considered to be more or less feasible and useful in diverse stages of care. During acute phases of psychopathology, for example, EMA was considered to be too burdensome. This corresponds to a recent study demonstrating that EMA was perceived as significantly more burdensome during mood episodes[75]. Interestingly, in this study, this did not negatively impact adherence, which may in part explain why compliance rates in EMA studies are usually high, even in populations with severe mental disorders[76]. In daily clinical practice, without the motivation of helping advance scientific knowledge, such compliance rates are likely considerably lower[1]. In our qualitative study, patients with bipolar disorder indicated that EMA is especially useful in the early phases after receiving the diagnosis, to learn effective coping strategies that benefit self-management. EMA thus might be especially interesting for people in the early phases of the disorder[77]. As such, EMA may be considered differentially efficacious and feasible in different stages of care.

Furthermore, not all patients and clinicians perceived EMA as beneficial to patient self-management. Our qualitative findings brought to light that a

substantial subgroup of patients, mostly diagnosed with mood disorders, felt that the heightened focus on mood worsened their symptoms and prevented them from engaging in helpful coping strategies (Chapter 7 and 8). Indeed, for several patients this was sufficient reason not to participate in EMA. Such mood reactivity is often overlooked in quantitative studies[78,79], but consistently mentioned by patients in qualitative research[74,80]. It relates to a tension recognized by e-health research, where e-health applications can have both beneficial and harmful effects to patient self-management[81]. Practically, such negative effects could be prevented to some extent by ensuring the EMA diaries place equal emphasis on both positive and negative experiences. More importantly, clinicians should be encouraged to recognize such negative effects and discuss them with patients.

Ultimately, this means that patients and clinicians have to balance such negative effects and the burden placed upon EMA participants against EMA's perceived benefits. EMA can therefore not be a useful clinical tool for everyone at any time: some patients may decide they already have sufficient self-management skills, or would rather monitor themselves only once-daily or weekly. As such, an important requirement for the implementation of EMA is that both patients and clinicians are made central actors in decisions regarding EMA (Chapter 7 and 8). Both should be engaged in every step of the process, from deciding on the goal of EMA and compiling an EMA diary, to interpreting the feedback together. This will put patients in control and let them decide what kind of information they need to benefit their recovery process[49]. Furthermore, both should have access to the gathered EMA data at any time and place. It will only further enhance the knowledge asymmetry if patients only provide information and clinicians only receive it[49]. And finally, as researchers, we have a responsibility to develop EMA tools that maximize the potential benefits and minimize the potential negative effects. This also entails training clinicians to provide clear guidelines when and how they can use EMA in care, and which potential negative effects they can expect.

To conclude, the promises of EMA for clinical practice are recognized by patients and clinicians, but only under the condition that patients and clinicians are actively involved.

Lesson 3. One size does not fit all: To be effective, EMA needs to be tailored to clinical settings

The final lesson learned pertains to the *how*: how should we implement EMA in clinical practice? It now seems likely that a one-size-fits-all approach is not useful and we will have to adapt EMA as a research methodology to fit better in clinical settings. In Chapter 7 and 8, patients and clinicians have advocated for a more personalized approach to EMA, in which the EMA diary as well as the EMA feedback can be flexibly adapted depending on the patient's care needs. Furthermore, they highlighted the necessity of transdiagnostic EMA diaries, which is consistent with the accumulating evidence that many patients experience symptoms from multiple diagnostic categories[82-84]. This need for personalization and transdiagnostic diaries has been highlighted by research on monitoring in general as well[80,85,86]. However, in both research as well as clinical applications, EMA diaries have so far been mostly standardized. This may result in increased burden for patients, who have to frequently answer questions that are not relevant to their current situation, and may insufficiently relate the findings to their treatment goals. Being able to personalize the EMA diary seems therefore to be a core requirement for successful implementation of EMA.

A second core requirement brought forward by this thesis is the tailoring of EMA to clinical settings. Indeed, this has also been highlighted by studies on other e-health applications[87-89]. Until now, EMA has hardly been adapted from its use in research. For example, in Chapter 8, we provided feedback after four months of treatment, which covered topics deemed relevant at face value: the variability in mood and symptoms and bar graphs displaying the relationship between activities and mood. Such feedback can be interesting, but does not directly relate to the way patients and clinicians talk about these topics in clinical practice. Indeed, in Chapter 8, patients and clinicians described that it was difficult to tie the feedback to clinically relevant conclusions. Instead, they would have liked to have received information on their diagnosis, as well as effects of medications or life style adjustments. They further argued that the feedback would have been more relevant if discussed briefly at every treatment session, instead of after a few months. This provides us

with specific recommendations about the timing and content of feedback when implementing EMA in the treatment of bipolar disorder. It further offers suggestions of how EMA could complement or improve the longitudinal mood monitoring that is already a core component of treatment[80]. As such, in-depth qualitative investigations in specific populations and treatment phases are an essential starting point when attempting to implement EMA in clinical settings.

Recent (qualitative) implementation experiments have confirmed the importance of integrating EMA in the workflow of clinicians. Box 10.1 describes an implementation experiment that we conducted as part of this thesis to explore whether EMA could improve the process of diagnosing first psychosis. We found that, even when considerable effort was spent to tailor EMA to the specific clinical setting, clinicians still had difficulty applying EMA in their workflow. These findings are confirmed by Daniels and colleagues, who attempted to integrate EMA in family medicine[90]. They observed that clinicians felt unequipped to use EMA feedback as a tool to inform functional analysis during consultations. Other recent studies have also suggested that although patients are enthusiastic about the possibilities of EMA, clinicians are not yet convinced of its clinical utility[7,91]. These studies demonstrate the necessity of involving clinicians when developing EMA applications for clinical use. Their suggestions can help integrate EMA in their current treatment strategies. Furthermore, they are illustrative of the importance of training clinicians in the use of EMA. Such training should not only involve the basic characteristics of EMA, but also relate it to existing treatment techniques already employed by clinicians. For example, in cognitive behavioral therapy (CBT), EMA could expand on the self-monitoring techniques that are central to CBT[92].

To conclude, in order to effectively implement EMA in clinical settings, we need to develop personalized, tailored, and targeted EMA interventions for diverse clinical populations and treatment phases.

Box 10.1. An illustrative EMA implementation experiment: Using EMA to inform the diagnosis of first psychosis.

We conducted an implementation experiment at the Department of Psychiatry at the University Medical Center Groningen, to test whether EMA could inform on the diagnosis of first psychosis. We expected EMA to provide relevant contextual information to aid the clinicians in their diagnostic process. To make the implementation as realistic as possible, focus groups were held with the clinicians to i) construct a relevant (yet standardized) EMA diary, ii) tie the EMA data to clinically relevant feedback, and iii) integrate EMA in the departmental meeting structure and diagnostic process. Psychiatric nurses were trained to instruct patients in the use of EMA, and a researcher was present at the weekly department meeting to remind the clinicians to consider EMA. Prior to the start of the project, clinicians voiced being enthusiastic about the possibilities of EMA. They were interviewed at diverse stages of the project to understand their experiences.

The project was conducted from December 2017 to September 2018. Of the 36 patients deemed suitable for EMA, twenty patients were included. Clinicians indicated that the EMA feedback was largely helpful in confirming and visualizing findings from more traditional diagnostic tools. However, in our implementation, we failed to help clinicians in incorporating EMA in their workflow, who often forgot to discuss the EMA feedback with their patients. In part, this was a technical problem: the researcher e-mailed the EMA feedback to the clinicians; whereas all other diagnostic information is stored in patients' personal health records. Furthermore, many patients completed very few assessments (compliance, on average, was 40%, in line with another implementation study in psychosis1), limiting the utility of the EMA feedback. This may be due to the standardized diary, and our decision to not appoint the responsibility of discussing EMA to a specific clinician, who therefore did not feel compelled to frequently remind patients of the importance of completing the EMA diaries.

This implementation experiment demonstrates that actual embedding of EMA in routine care is difficult. Even when considerable effort is spent to tailor EMA to the clinical setting, implementation may fail because patients and clinicians are insufficiently involved in the design of the EMA tool. We learned from these findings that appropriate training is essential, not only in what EMA is, but also in how EMA can be most effectively used in clinical settings.

Developing a clinical EMA tool: PETRA

The above lessons have demonstrated that EMA indeed shows promise as a clinical tool in psychiatry, but have also highlighted several pitfalls. In the following sections, I will describe how our findings have resulted in the development of a clinical EMA tool.

Based on the findings of the present thesis, we considered a clinical EMA tool to be most promising if it meets four core demands. First, it should offer extensive possibilities for personalization. EMA diaries should be tailored to assess only relevant experiences, which are likely to transcend diagnostic boundaries; as such, the diary should be transdiagnostic. Most EMA applications used in research are standardized or offer limited options for personalization; and even if they do offer some personalization options, clinicians are required to invest considerable time and efforts making large-scale enrollment in clinical care not feasible. Furthermore, many EMA tools are designed with specific populations in mind (e.g.,[56,93,94], but see for exceptions[77,95]) and therefore do not appeal to patients with symptoms belonging to multiple diagnoses. In addition to personalization, a second requirement is therefore that the EMA tool is be user-friendly and tailored to fit diverse clinical settings. This means that the tool should be largely automated and easy to use for clinicians. Many e-health applications fail because they inadequately take into account the experiences of users: here, patients and clinicians[96]. Patients and clinicians should therefore be involved early on in the development of an EMA tool to ensure its usability in clinical practice. Third, the tool should be supported by scientific research. Many apps provide monitoring options that do not meet criteria for the construction of valid EMA diaries or are not evidence-based[97,98]. We have seen above that EMA applications that are not sufficiently grounded in empirical evidence may give rise to false conclusions, or may simply cost clinicians a lot of time because they construct diaries that do not provide interpretable data. Fourth and finally, a clinical EMA tool should meet demands regarding privacy and safe data storage. Many applications store data on third-party servers, which is problematic under the General Data Protection Regulation law of the European Union. In my view, a tool that meets these four demands is necessary to be able to fully explore the efficacy of EMA in clinical practice.

Although several companies provide research applications of EMA, and mobile phone app stores are overflowing with e-health applications, none meets all the criteria stated above. Therefore, we started the PETRA project. PETRA is an acronym for PErsonalized Treatment by Real-time Assessment. PETRA is a web-based tool integrated in patients' electronic health records via the RoQua system

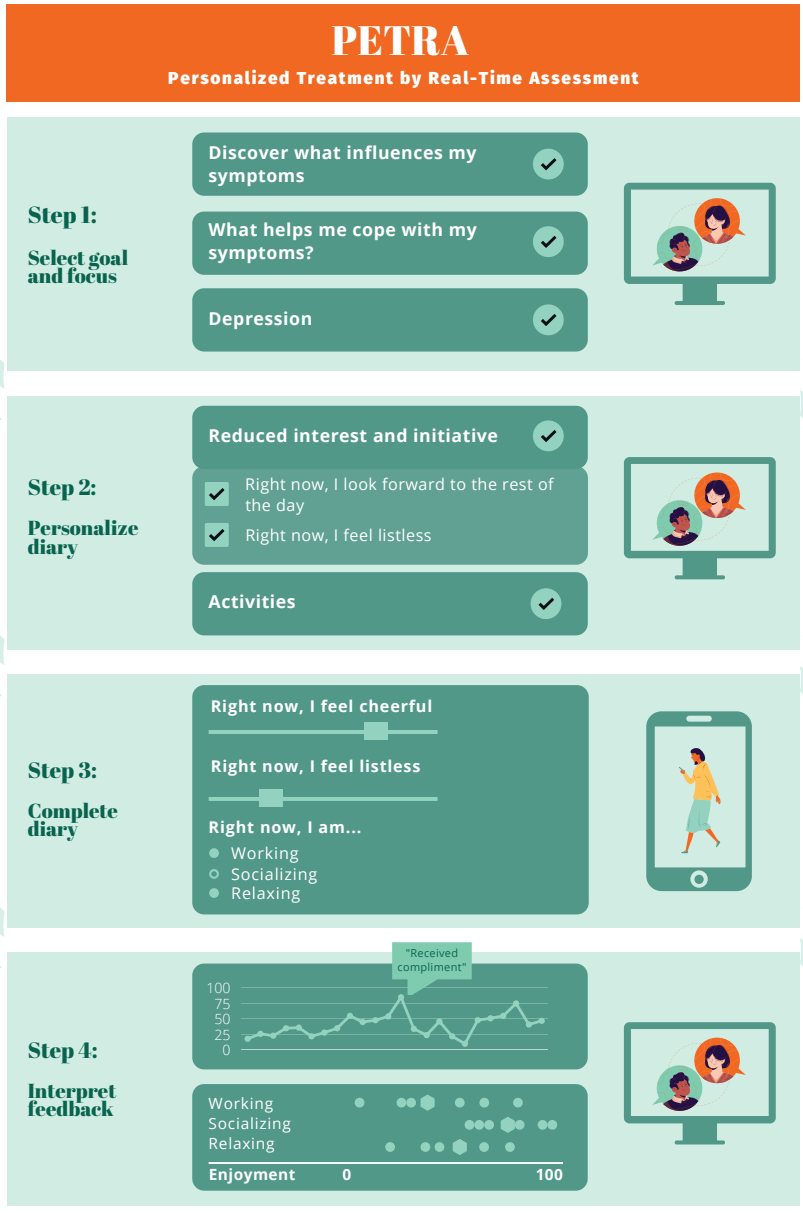
(see www.roqua.nl) to facilitate easy adoption in clinicians' workflow[99]. It is currently being developed and scheduled to go live February 2021 (www.petrapsy.nl/en).

PETRA is developed based on principles of design thinking and the roadmap for the development of e-health technologies of the Center for e-Health Research (CeHRes)[89,100]. Both methodologies advocate for involving relevant users early in the development of e-health tools. The first phases of e-health development consist of gaining a thorough understanding of the contexts in which the e-health tool is used, by conducting literature reviews (e.g., Chapter 6), interviews, focus groups (e.g., Chapter 7 and 8), and pilots (in our clinic, but also e.g.,[4,5,7]). The next phase constitutes an iterative design process, in which prototypes are developed, evaluated, and adapted frequently based on feedback of users. This differs from the usual modus of operandi in science, where interventions are usually only pilot-tested once before their efficacy is evaluated in randomized controlled trials (RCTs). Such an approach may result in e-health technologies that are of limited use to patients and clinicians. A final core part of the CeHRes roadmap involves establishing a multidisciplinary project team that can oversee the various (practical) components of the development. Therefore, the PETRA project team consists of multidisciplinary members: the core scientific team, a scientific programmer, the RoQua programmers to ensure a sustainable and future-proof embedding in care, a user-experience designer to integrate the needs of all stakeholders, patients, and clinicians. All project members critically review the prototypes, which are then continuously adapted and reevaluated.

Based on our findings of the first investigative phases, PETRA consists of a decision aid and a feedback module (see Figure 10.1). In four easy steps, patients and clinicians can together decide on the goal of the personalized EMA diary, tweak the proposed diary, complete the diaries, and interpret the personalized feedback. Given that both patients and clinicians were seen as central actors in the clinical use of EMA, and EMA is expected to benefit the patient-clinician relationship, PETRA is designed to be used collaboratively. The decision aid was developed because patients and clinicians highlighted the need for personalized EMA diaries, but at the same time emphasized they would have little time for making such diaries.

Building a diary from the ground up would be too demanding and requires extensive knowledge on validity of EMA diaries that is not available to the typical clinician or patient. Therefore, PETRA was designed such that users only have to indicate the goal they intend to achieve with EMA and the main symptoms they would like to focus on. PETRA then automatically preselects several relevant diary items, which can be adapted to optimally fit with the patient's situation. For example, a patient treated for depression might deselect the diary questions on anhedonia and instead select diary questions on anxiety. Patients and clinicians also have the option to formulate EMA items themselves; this could be relevant in cases where patients have highly person-specific experiences (for example, in psychosis, it may be relevant to specify the content of delusions). Each diary automatically comes with an optional comment field, so patients can reflect on their experiences and how those influenced mood and symptoms. Furthermore, based on the selected goal, PETRA proposes an EMA schedule: the preferred number of assessments per day, as well as the timing of assessments (e.g., at fixed moments or within semi-random intervals during the day), and the number of days necessary to result in valid and interpretable feedback. This preselection can then be further tweaked according to personal needs and preferences, while remaining within valid boundaries (e.g., it is not possible to select a too short assessment period).

Figure 10.1. PETRA: A clinical EMA tool.



The second part of PETRA consists of a feedback module that visualizes the gathered EMA data. Based on our reviews of the literature (e.g., Chapter 5), and our qualitative studies (Chapter 7 and 8), we concluded that PETRA should first focus on simple, descriptive feedback rather than complex analytical models such as networks. We believed EMA to be especially interesting for providing a less biased overview of how the patient fared in between sessions, and by bringing to light those moments with especially high or low symptoms. By zooming into those moments, and analyzing situational information that may have brought about these highs and lows, patients can become more able to cope with these situations. Therefore, the PETRA feedback module is centered around three major themes in psychiatric treatment: the variability of mood and symptoms, associations between (social) contexts and mood and symptoms, and the impact of pleasant and unpleasant events. Furthermore, patients are actively encouraged to provide qualitative descriptions, and the feedback module summarizes these evaluations in multiple ways, for example, through word clouds and displaying contextual descriptions next to moments of especially high or low mood. PETRA is designed to be used iteratively and during every treatment session. At any point during the four steps (see Figure 10.1) alterations can be made. If patients and clinicians discuss the feedback and decide that they want to gather additional information, diary items can be removed and added, or the goal and focus can be changed to result in a new diary. Finally, the modular structure of the feedback module ensures that novel feedback can be easily added in later stages. For example, if scientific research can provide a solid evidence basis for personalized models, PETRA can offer this information directly in its feedback module (see, e.g., [33,101,102]).

To conclude, the PETRA tool is being developed to meet the demands of patients and clinicians for a personalized, user-friendly, scientifically valid, and privacy-safe EMA tool that is useful in psychiatric care. Future research into its efficacy will have to demonstrate whether PETRA indeed has beneficial effects on patients' well-being, self-management, and treatment process.

Methodological considerations on qualitative research

The findings of the present thesis should be viewed in light of its methodological designs. In this discussion, I elaborate mostly on our qualitative findings brought forward by Chapter 7 and 8, and Box 10.1. Qualitative research is a crucial complement to quantitative studies in this field, to understand the clinical context in which EMA is expected to maximize its promises and minimize its pitfalls[87,103]. Such information is difficult to uncover in quantitative study designs. However, in qualitative research, data is gathered until saturation is reached, meaning that no new information is heard. This inherently means that sample sizes are smaller than quantitative researchers are used to. This makes qualitative findings more difficult to generalize to settings other than Dutch tertiary care institutions and patients with mood disorders. Furthermore, participation bias is not unlikely: we may have recruited mostly patients and clinicians that are already enthusiastic about using EMA in treatment, thereby overestimating its eventual uptake in clinical practice. Nonetheless, qualitative research does not claim to offer universal conclusions about the topic under study. Rather, it gives insight into the processes perceived to form the working mechanisms of EMA[104], and pinpoints promising directions for further (quantitative) study. Our qualitative findings are at the very least suggestive of the merits of further exploration of specific clinical applications of EMA. This is where we will turn to in the final section of this Chapter.

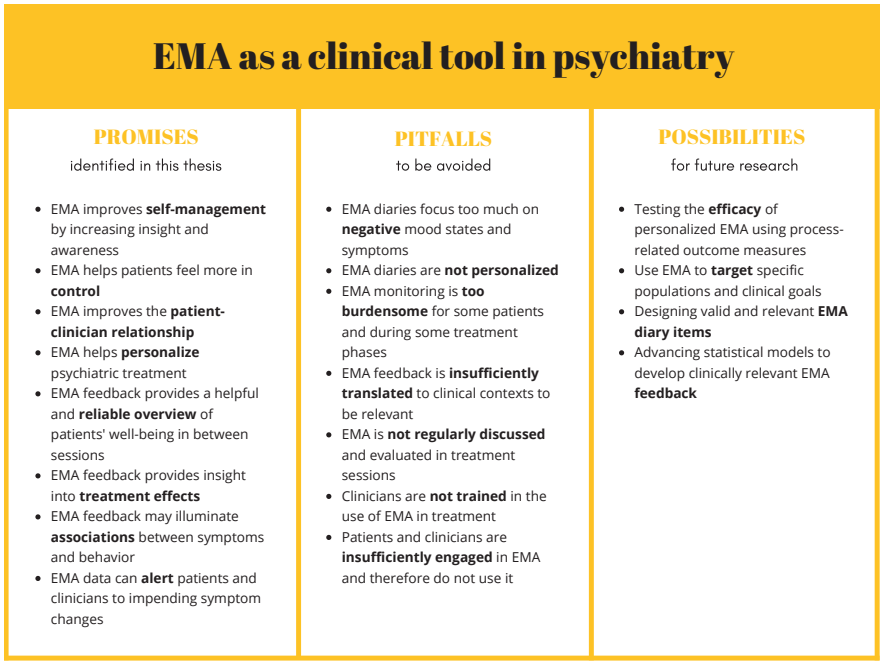
A research agenda for the clinical implementation of EMA: recommendations

Now that we have reviewed the promises and pitfalls of EMA, it is time to turn towards possibilities for future research (for an overview, see Figure 10.2). In this section, I will propose three directions for future research that may further enhance our understanding of EMA as a clinical tool.

First, EMA as a clinical intervention needs to receive more empirical

investigation. The consolidation of the evidence-base of EMA is crucial in the further development of EMA as a clinical tool. This will reveal whether the promise of EMA for psychiatric treatment is supported by research.

Figure 10.2. Overview of the promises, pitfalls, and possibilities of EMA as a clinical tool in psychiatry.



It will also set EMA as an e-health technology apart from the myriad of apps available to clinicians, who will then be more inclined to use EMA[105]. This is, therefore, a crucial next step in future research on EMA as a clinical tool. So, how to proceed? Specifically, future RCTs should test the efficacy of personalized EMA, in diverse or even transdiagnostic populations[77], using process-related outcome measures. So far, relatively few RCTs have investigated the efficacy of EMA as a clinical tool[9,62,93]. Other studies so far are less systematic and are limited by the lack of control groups or small sample sizes (e.g., [61,63,64,67,106,107]). These studies evaluated standardized EMA diaries, whereas we have now seen the importance of personalizing EMA and integrating it in specific clinical contexts. Moreover, given that many RCTs have

focused on patients with depression, future RCTs might turn to other populations, such as patients with bipolar disorder or psychosis. Here, especially, EMA might be an ideal match, because treatments of severe psychopathology often already aim to improve self-management with self-monitoring[80]. And finally, future RCTs should focus less on symptom improvement and instead turn to process-related outcome measures. Studies that have so far evaluated the efficacy of EMA do not always report a beneficial effect on symptoms, distress, or quality of life. However, the intervention is without exception rated highly feasible and acceptable by patients. This mismatch between patients' experiences and reported evidence may lie in the type of outcome measures used in RCTs and observational trials. As we have seen in our qualitative research, patients and clinicians expect EMA to mostly improve the patient's self-management, the patient-clinician relationship, and shared-decision making; aspects usually not assessed in clinical trials. Indeed, one RCT that did assess empowerment and behavioral change reported beneficial effects of EMA[19,50]. Future trials may therefore want to include quantitative assessments of these aspects as outcome measures, as for example in the currently ongoing Therap-i study, where EMA is used to support case—conceptualization in the treatment of depression[108].

A second step entails further development of EMA diary items. Many EMA diary items originate from the Positive and Negative Affect Scale (PANAS)[109] and are carefully designed to capture momentary fluctuations. However, two problems arise here: i) current EMA diary items are not specifically developed to have clinical relevance, and ii) there is currently little consensus among researchers which EMA diary items measure which underlying construct (e.g., energy, mania, or affect). Regarding the first problem, we have seen earlier that networks currently cannot incorporate contextual information and therefore have difficulty showing the relationship between symptoms and behavior. If EMA feedback is to be relevant to psychiatric care, it is important that we first identify what patients and clinicians deem relevant constructs to be assessed[110]. In Chapter 8, for example, our qualitative interviews on relevant constructs to patients with bipolar disorder yielded several EMA items relevant to their treatment. Researchers are therefore encouraged to give patients and clinicians a stronger voice in their EMA study designs. The second

problem relates to the construct consistency of EMA diary items. For example, it remains unclear whether EMA is affected by a phenomenon termed response shift, which occurs when patients start to interpret the diary items differently over the course of the assessment period[111]. In my experience, patients need several days learning to interpret the diary items and develop consistency in their responses. Furthermore, researchers have differed in their preferred way to analyze EMA items: some use the separate items (as we did in Chapter 4 and 9), whereas others prefer to use factor analysis to group items of different constructs together (e.g., [64,112-114]). The latter makes a more confirmatory approach possible (because fewer variables are analyzed), but can be difficult because many participants show different factor loadings. This was the reason to use the separate items in Chapter 9. Because research has so far used a myriad of formulations to assess similar constructs, it is difficult to generalize study findings. An important step forward is the EMA item repository, with which researchers can register the EMA items used and to which constructs they were considered to belong[115]. This enables more systematic research into EMA diary items and their construct validity.

A final step will involve the tailoring of specific EMA applications to diverse clinical settings. This means we will have to systematically investigate in which parts of diagnosis and treatment phases, and in which clinical populations, EMA can play an informative role. For example, EMA might be specifically investigated as an intensive monitoring tool to alert patients to impending manic and depressive episodes in bipolar disorder. As is now possible with the PETRA tool, the EMA diary content, timing, and frequency of assessments should be adapted to answer such clinical goals. Furthermore, research should further explore diverse types of feedback based on EMA data. This means we should explore the possibilities of descriptive feedback (e.g.,[41]) in addition to theoretically and empirically advancing statistical methods to analyze EMA data, such as network models, early warning signals, or machine learning. For example, machine learning could be helpful as a diagnostic tool, to see which patterns in EMA data are more suggestive of one diagnosis (e.g., bipolar disorder) versus the other (e.g., borderline personality disorder)[116]. Thus, we need to tie characteristics of EMA data to treatment response and outcome so patients and clinicians will be better able to use this information in treatment. For

example, an early promise of EMA was the suggestion that certain patterns in EMA data (e.g., the variability or extent of affect) could be related to worse or favorable outcomes (e.g., [53,112]). In our studies, many patients and clinicians have repeatedly asked: *what does this pattern in EMA data mean*? As of yet, this question is very difficult to answer. We do not yet know when variability is sufficiently high to be wary of impending symptom changes, or whether certain patterns are specific to one disorder as opposed to a differential diagnosis. Therefore, more research into how EMA data can be analyzed to provide valid personalized feedback is warranted. Patients and clinicians need to be continuously involved in this development, to make sure the feedback is usable and relevant to clinical practice.

Concluding remarks

The present thesis aimed to explore the clinical utility of EMA and personalized feedback on EMA data. We have seen that patients and clinicians are enthusiastic about the promise of EMA to improve psychiatric care. Such promising applications include enhancing the self-management of patients, the personalization of psychiatric treatment, improving the patient-clinician relationship, and providing a reliable overview of the patients' well-being in between treatment sessions. However, this thesis has also identified several pitfalls to be avoided. Problems arise when EMA diaries are focused too much on negative mood states and symptoms, when EMA diaries are not personalized and tailored to its clinical goal, and when patients and clinicians are insufficiently engaged in EMA. Finally, we have examined possibilities for future research on the clinical use of EMA. In the coming years, we will need to systematically evaluate the efficacy of EMA in RCTs, which should focus on personalized and transdiagnostic EMA diaries that are tailored to specific clinical settings. Furthermore, we need to further develop clinically relevant EMA diaries and advance statistical methods to analyze EMA data to provide clinically relevant feedback. In doing so, we may further consolidate EMA as a clinical tool for psychiatric care.

References

1. Verhagen SJ, Berben JA, Leue C, Marsman A, Delespaul PA, van Os J, Lousberg R. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. *PLoS One.* 2017;12(10):e0186294.
2. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, Drukker M, Reininghaus U, Jacobs N, Lataster T, Simons C, Lousberg R, Gülöksüz S, Leue C, Groot PC, Viechtbauer W, Delespaul P. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *DepressAnxiety.* 2017;34(6):481.
3. Beck JS. *Cognitive therapy: Basics and beyond.* New York: Guilford Press; 1995.
4. 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.
5. Kroeze R, van der Veen DC, Servaas MN, Bastiaansen JA, Oude Voshaar R, Borsboom D, Ruhe E, Schoevers R, Riese H. Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research.* 2017;3(1):1.
6. Epskamp S, van Borkulo CD, van der Veen DC, Servaas MN, Isvoranu A-M, Riese H, Cramer AOJ. Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science.* 2018;6(3):416.
7. Frumkin MR, Piccirillo ML, Beck ED, Grossman JT, Rodebaugh TL. Feasibility and utility of idiographic models in the clinic: A pilot study. *Psychotherapy Research.* 2020:1-15.
8. van der Krieke L, Blaauw FJ, Emerencia AC, Schenk HM, Slaets JPJ, Bos EH, de Jonge P, Jeronimus BF. Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *PsychosomMed.* 2016;79(2):213-223.
9. 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. *Behavior Therapy.* 2017;48(1):76.
10. Bos EH, Wanders RBK. Group-level symptom networks in depression. *JAMA Psychiatry.* 2016;73(4):411.
11. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences.* 2018;115(27):E6106-E6115.
12. Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives.* 2004;2(4):201.
13. Hamaker EL. Why researchers should think 'within-person': A paradigmatic rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:43.
14. Forbes MK, Wright AG, Markon KE, Krueger RF. Further evidence that psychopathology networks have limited replicability and utility: Response to Borsboom et al. (2017) and Steinley et al. (2017). *JAbnormPsychol.* 2017;126(7):1011–1016.
15. Forbes MK, Wright AGC, Markon KE, Krueger RF. The network approach to psychopathology: promise versus reality. *World Psychiatry.* 2019;18(3):272.
16. Forbes MK, Wright AGC, Markon KE, Krueger RF. Evidence that psychopathology symptom networks have limited replicability. *JAbnormPsychol.* 2017;126(7):969.
17. Borsboom D, Fried EI, Epskamp S, Waldorp LJ, van Borkulo CD, van der Maas HLJ, Cramer AOJ. False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *JAbnormPsychol.* 2017;126(7):989-999.
18. Fried EI, Eidhof MB, Palic S, Constantini G, Huisman-van Dijk HM, Bockting CLH, Engelhard I, Armour C, Nielsen ABS, Karstoft KI. Replicability and generalizability of PTSD networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological*

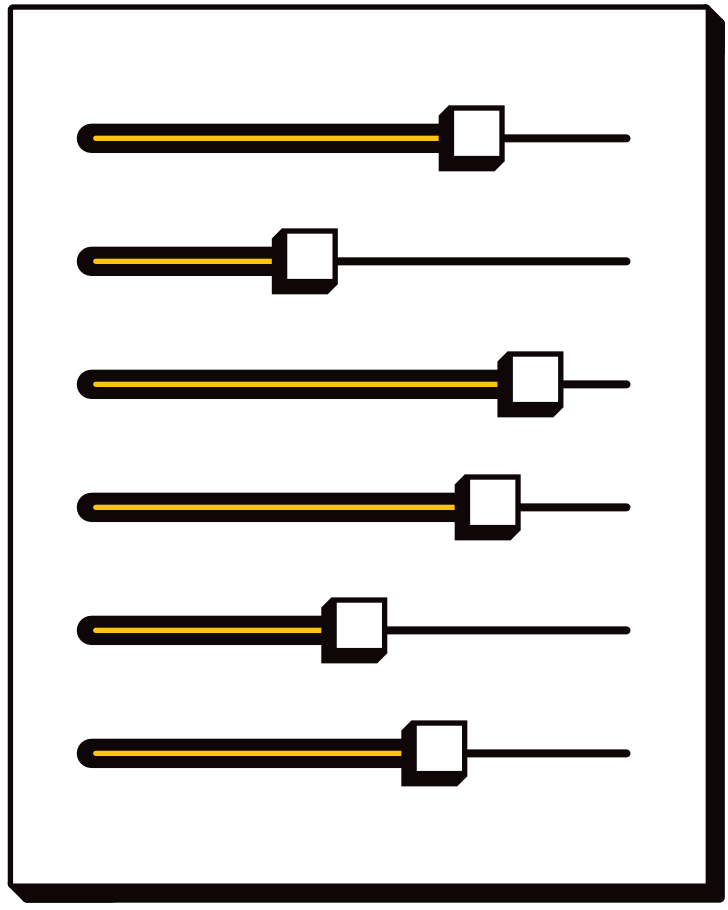
Science. 2018;6(3):335-351.
19. Snippe E, Simons CJP, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, Viechtbauer W, Delespaul P, Myin-Germeys I, Wichers M. Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology.* 2016;35(5):433.
20. Rosmalen JGM, Wenting AMG, Roest AM, de Jonge P, Bos EH. Revealing causal heterogeneity using time series analysis of ambulatory assessments: Application to the association between depression and physical activity after myocardial infarction. *PsychosomMed.* 2012;74(4):377.
21. Stavrakakis N, Booij SH, Roest AM, de Jonge P, Oldehinkel AJ, Bos EH. Temporal dynamics of physical activity and affect in depressed and nondepressed individuals. *Health Psychology.* 2015;34:1268.
22. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, Borsboom D, Tuerlinckx F. A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One.* 2013;8(4):e60188.
23. Wild B, Eichler M, Friederich H-C, Hartmann M, Zipfel S, Herzog W. A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology.* 2010;10(1):28.
24. Abegaz F, Wit E. Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics.* 2013;14(3):586.
25. Brandt P, Williams JT. *Multiple time series models.* Thousand Oaks, CA, US: Sage Publications Inc; 2007.
26. de Vos S, Wardenaar KJ, Bos EH, Wit EC, Bouwmans MEJ, de Jonge P. An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *PLoS One.* 2017;12(6):e0178586.
27. Bastiaansen JA, Kunkels YK, Blaauw FJ, Boker SM, Ceulemans E, Chen M, Chow S-M, de Jonge P, Emerencia AC, Epskamp S. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *JPsychosomRes.* 2020:110211.
28. von Klipstein L, Riese H, van der Veen DC, Servaas MN, Schoevers RA. Using person-specific networks in psychotherapy: Challenges, limitations, and how we could use them anyway. *BMC Med.* 2020;18.
29. Janssens KA, Bos EH, Rosmalen JG, Wichers MC, Riese H. A qualitative approach to guide choices for designing a diary study. *BMC Medical Research and Methodology.* 2018;18(1):1-12.
30. Robinaugh D, Haslbeck J, Waldorp L, Kossakowski J, Fried EI, Millner A, McNally RJ, van Nes EH, Scheffer M, Kendler KS. Advancing the network theory of mental disorders: A computational model of panic disorder. *PsyArXiv.* 2019.
31. Bringmann LF, Eronen MI. Don't blame the model: Reconsidering the network approach to psychopathology. *PsycholRev.* 2018;125(4):606.
32. David SJ, Marshall AJ, Evanovich EK, Mumma GH. Intraindividual dynamic network analysis–implications for clinical assessment. *Journal of Psychopathology and Behavioral Assessment.* 2018;40(2):235.
33. Rubel JA, Fisher AJ, Husen K, Lutz W. Translating person-specific network models into personalized treatments. Development and demonstration of the Dynamic Assessment Treatment Algorithm for Individual Networks (DATA-IN). *Psychotherapy Psychosomatics.* 2018;87(4):249-251.
34. Borsboom D, Cramer AO. Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology.* 2013;9:91-121.
35. Fried EI, Epskamp S, Nesse RM, Tuerlinckx F, Borsboom D. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders.* 2016;189:314.
36. Robinaugh DJ, LeBlanc NJ, Vuletich HA, McNally RJ. Network analysis of persistent complex bereavement disorder in conjugally bereaved adults. *JAbnormPsychol.* 2014;123(3):510.
37. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, Wigman J, Snippe E. What do centrality measures measure in psychological networks. *JAbnormPsychol.* 2019;128(8):892.
38. Dablander F, Hinne M. Node centrality measures are a poor substitute for causal inference. *Sci Rep.* 2019;9(1):1-13.

39. Eronen MI. Causal discovery and the problem of psychological interventions. *New Ideas in Psychology.* 2020;59:100785.
40. Haslbeck J, Ryan O. Recovering within-person dynamics from psychological time series. *https://psyarxivcom/dymhw/.* 2020.
41. Bringmann LF, Wichers M, Riese H, Stulp G. ESMvis: A tool for visualizing individual Experience Sampling Method (ESM) data. *Quality of Life Research.* 2020:1-10.
42. Wichers M, Hartmann JA, Kramer IMA, Lothman C, Peeters F, van Bemmel L, Myin-Germeys I, Delespaul P, van Os J, Simons CJP. Translating assessments of the film of daily life into person-tailored feedback interventions in depression. *Acta PsychiatrScand.* 2011;123(5):402.
43. 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.
44. 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.
45. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, van Bemmel AL, Peeters F, Delespaul P, van Os J. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta PsychiatrScand.* 2011;124(4):262.
46. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes J, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: A qualitative study. *European Psychiatry.* 2017;41:115-121.
47. Palmier-Claus J, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, Barkus E, Kapur S, Wykes T, Lewis SW. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: A qualitative study. *BMC Psychiatry.* 2013;13.
48. Hanssen E, Balvert S, Oorschot M, Borkelmans K, van Os J, Delespaul P, Fett A-K. An ecological momentary intervention incorporating personalised feedback to improve symptoms and social functioning in schizophrenia spectrum disorders. *Psychiatry Res.* 2020;284:112695.
49. Alpay L, van der Boog P, Dumaij A. An empowerment-based approach to developing innovative e-health tools for self-management. *Health Informatics Journal.* 2011;17(4):247-255.
50. Simons CJP, Hartmann JA, Kramer I, Menne-Lothmann C, Höhn P, Van Bemmel AL, Myin-Germeys I, Delespaul P, Van Os J, Wichers M. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. *European Psychiatry.* 2015;30(8):900.
51. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, Sanci L, Patton G. Self-monitoring using mobile phones in the early stages of adolescent depression: Randomized controlled trial. *JMedInternet Res.* 2012;14(3):e67.
52. Reis HT. Why researchers should think 'real-world': A conceptual rationale. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:3.
53. Barge-Schaapveld D, Nicolson NA. Effects of antidepressant treatment on the quality of daily life: An experience sampling study. *JClinPsychiatry.* 2002;63(6):477.
54. Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology.* 2010;15(1):1.
55. Wenze SJ, Armey MF, Miller IW. Feasibility and acceptability of a mobile intervention to improve treatment adherence in bipolar disorder: a pilot study. *Behavior Modification.* 2014;38(4):497-515.
56. Wenze SJ, Armey MF, Weinstock LM, Gaudiano BA, Miller IW. An open trial of a smartphone-assisted, adjunctive intervention to improve treatment adherence in bipolar disorder. *Journal of Psychiatric Practice.* 2016;22(6):492.
57. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: Status and considerations. *International Journal of Bipolar Disorders.* 2018;6(1):1-7.
58. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY US: Guilford Press; 2012:636.
59. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, Reininghaus U. Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry.* 2018;17(2):123-132.

60. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul P, Lewis SW, Dunn G. Experience sampling research in individuals with mental illness: Reflections and guidance. *Acta PsychiatrScand.* 2011;123(1):12.
61. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, Mohr DC. Harnessing context sensing to develop a mobile intervention for depression. *JMedInternet Res.* 2011;13(3):e55.
62. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, Schruers K, van Bemmel AL, Myin-Germeys I, Delespaul P, van Os J, Wichers M. A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry.* 2014;13(1):68.
63. Depp CA, Ceglowski J, Wang VC, Yaghouti F, Mausbach BT, Thompson WK, Granholm EL. Augmenting psychoeducation with a mobile intervention for bipolar disorder: A randomized controlled trial. *Journal of Affective Disorders.* 2015;174:23-30.
64. Fisher AJ, Bosley HG, Fernandez KC, Reeves JW, Soyster PD, Diamond AE, Barkin J. Open trial of a personalized modular treatment for mood and anxiety. *Behaviour Research and Therapy.* 2019;116:69-79.
65. Vaessen T, Steinhart H, Batink T, Klippel A, Van Nierop M, Reininghaus U, Myin-Germeys I. ACT in daily life in early psychosis: an ecological momentary intervention approach. *Psychosis.* 2019;11(2):93-104.
66. Reininghaus U, Klippel A, Steinhart H, Vaessen T, Van Nierop M, Viechtbauer W, Batink T, Kasanova Z, Van Aubel E, Van Winkel R. Efficacy of Acceptance and Commitment Therapy in Daily Life (ACT-DL) in early psychosis: Study protocol for a multi-centre randomized controlled trial. *Trials.* 2019;20(1):1-12.
67. Bell IH, Rossell SL, Farhall J, Hayward M, Lim MH, Fielding-Smith SF, Thomas N. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): Feasibility, acceptability and preliminary clinical outcomes. *SchizophrRes.* 2020;216:479-487.
68. Chen Y-W, Cordier R, Brown N. A preliminary study on the reliability and validity of using experience sampling method in children with autism spectrum disorders. *Developmental Neurorehabilitation.* 2015;18(6):383.
69. Hare DJ, Gracey C, Wood C. Anxiety in high-functioning autism: A pilot study of experience sampling using a mobile platform. *Autism.* 2016;20(6):730-743.
70. Helweg-Jørgensen S, Lichtenstein MB, Fruzzetti AE, Dahl CM, Pedersen SS. Daily Self-Monitoring of Symptoms and Skills Learning in Patients With Borderline Personality Disorder Through a Mobile Phone App: Protocol for a Pragmatic Randomized Controlled Trial. *JMIR Research Protocols.* 2020;9(5):e17737.
71. Nica EI, Links PS. Affective instability in borderline personality disorder: Experience sampling findings. *CurrPsychiatry Rep.* 2009;11(1):74.
72. Tsanas A, Saunders K, Bilderbeck A, Palmius N, Osipov M, Clifford G, Goodwin G, De Vos M. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders.* 2016;205:225-233.
73. Lindgreen P, Clausen L, Lomborg K. Clinicians' perspective on an app for patient self-monitoring in eating disorder treatment. *IntJEatDisord.* 2018;51(4):314-321.
74. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: Qualitative study. *JMIR mHealth & uHealth.* 2018;6(6):e10253.
75. van Genugten CR, Schuurmans J, Lamers F, Riese H, Penninx BW, Schoevers RA, Riper HM, Smit JH. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *Journal of Clinical Medicine.* 2020;9(2):322.
76. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *JMedInternet Res.* 2019;21(12):e14475.
77. Schick A, Pätzold I, Rauschenberg C, Hirjak D, Banaschewski T, Meyer-Lindenberg A, Wellek S, Boecking B, Reininghaus U. The effects of a novel, accessible, transdiagnostic ecological momentary intervention for improving resilience in youth (EMIcompass): Study protocol for a randomized controlled trial. *PsyArXiv.* 2020, October 30.

78. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain.* 2003;104(1-2):343.
79. Sobell MB, Bogardis J, Schuller R, Leo GI. Is self-monitoring of alcohol consumption reactive? *Behav Assess.* 1989;11(4):447.
80. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, Matthews M. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: Implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association.* 2016;23(3):477-484.
81. Schermer M. Telecare and self-management: Opportunity to change the paradigm? *JMedEthics.* 2009;35(11):688.
82. McGorry PD, Hartmann JA, Spooner R, Nelson B. Beyond the "at risk mental state" concept: Transitioning to transdiagnostic psychiatry. *World Psychiatry.* 2018;17(2):133.
83. Reininghaus U, Böhnke JR, Chavez-Baldini U, Gibbons R, Ivleva E, Clementz BA, Pearlson GD, Keshavan MS, Sweeney JA, Tamminga CA. Transdiagnostic dimensions of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *World Psychiatry.* 2019;18(1):67-76.
84. Wichers M, Schreuder MJ, Goekoop R, Groen RN. Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology. *PsycholMed.* 2019;49(3):380-387.
85. Rosmalen JG, van Gils A, Mesa MAA, Schoevers RA, Monden R, Hanssen DJ. Development of Grip self-help: An online patient-tailored self-help intervention for functional somatic symptoms in primary care. *Internet Interventions.* 2020;19:100297.
86. Wright AG, Woods WC. Personalized models of psychopathology. *Annual Review of Clinical Psychology.* 2020;16.
87. Grol R, Grimshaw J. From best evidence to best practice: Effective implementation of change in patients' care. *The Lancet.* 2003;362(9391):1225.
88. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy research.* 2015;25(1):6.
89. Van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, Seydel ER. A holistic framework to improve the uptake and impact of eHealth technologies. *JMedInternet Res.* 2011;13(4):e111.
90. Daniëls NE, Hochstenbach LM, van Bokhoven MA, Beurskens AJ, Delespaul PA. Implementing experience sampling technology for functional analysis in family medicine–a design thinking approach. *Front Psychol.* 2019;10:2782.
91. Ellison WD. An initial study of practicing psychologists' views of the utility of ecological momentary assessment for difficult psychotherapy cases. *Administration and Policy in Mental Health and Mental Health Services Research.* 2020:1-11.
92. Cohen JS, Edmunds JM, Brodman DM, Benjamin CL, Kendall PC. Using self-monitoring: Implementation of collaborative empiricism in cognitive-behavioral therapy. *Cognitive Behavioral Practice.* 2013;20(4):419-428.
93. Bastiaansen JA, Ornee DA, Meurs M, Oldehinkel A. An evaluation of the efficacy of two add-on ecological momentary intervention modules for depression in a pragmatic randomized controlled trial (ZELF-i). *PsycholMed.* 2020:1-10.
94. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, Simmons MB, Alvarez-Jimenez M, Goodrich S, Davey CG. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. *JMIR Mental Health.* 2018;5(1):e9.
95. Mestdagh M, Verdonck S. Creating a user-friendly and statistically sound EMA tool for blended care: www.m-path.io. *Conference presentation.*
96. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder M, Seydel E. Evaluation of internet-based technology for supporting self-care: Problems encountered by patients and caregivers when using self-care applications. *JMedInternet Res.* 2008;10(2):e13.
97. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for

smartphone health apps. *Telemedicine and e-Health.* 2015;21(2):97-104.
98. Marshall JM, Dunstan DA, Bartik W. The digital psychiatrist: In search of evidence-based apps for anxiety and depression. *Frontiers in Psychiatry.* 2019;10:831.
99. Bos FM, von Klipstein L, Emerencia AC, Veermans E, Wichers M, Riese H. PErsonalized Treatment Real-time Assessment (PETRA): A flexible interface for personalized diaries in psychiatry. In preparation.
100. van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R. *EHealth Research, Theory and Development: A Multi-Disciplinary Approach.* Routledge; 2018.
101. Lutz W, Rubel J, Schwartz B, Schilling V, Deisenhofer A-K. Towards integrating personalized feedback research into clinical practice: Development of the Trier Treatment Navigator (TTN). *Behaviour Research and Therapy.* 2019;120.
102. Lutz W, Schwartz B, Hofmann SG, Fisher AJ, Husen K, Rubel JA. Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Sci Rep.* 2018;8(1):7819.
103. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: What it is and how to do it. *Bmj.* 2013;347.
104. Hammarberg K, Kirkman M, de Lacey S. Qualitative research methods: When to use them and how to judge them. *Human Reproduction.* 2016;31(3):498-501.
105. Giota KG, Kleftaras G. Mental health apps: Innovations, risks and ethical considerations. *E-Health Telecommunication Systems Networks.* 2014;2014.
106. Barakat S, Maguire S, Surgenor L, Donnelly B, Miceska B, Fromholtz K, Russell J, Hay P, Touyz S. The role of regular eating and self-monitoring in the treatment of bulimia nervosa: A pilot study of an online guided self-help CBT program. *Behavioral Sciences.* 2017;7(3):39.
107. Depp CA, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson TL, Lebowitz BD, Jeste DV. Mobile interventions for severe mental illness: Design and preliminary data from three approaches. *JNervMentDis.* 2010;198(10):715.
108. Riese H, von Klipstein L, Schoevers RA, van der Veen DC, Servaas MN. Personalized ESM monitoring and feedback to support psychological treatment for depression: A pragmatic randomized controlled trial (Therap-i). *BMC Psychiatry.* 2021;21(1):1-11.
109. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *JPersSocPsychol.* 1988;54(6):1063.
110. Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, Fried EI, Gartlehner G, Cipriani A, Tran VT. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *The Lancet Psychiatry.* 2020;7(8):692-702.
111. Barclay-Goddard R, Epstein JD, Mayo NE. Response shift: A brief overview and proposed research priorities. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation.* 2009;18(3):335.
112. 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. *EurNeuropsychopharmacol.* 2011;21(3):241.
113. Wichers M, Lothmann C, Simons CJP, Nicolson NA, Peeters F. The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: A momentary assessment study. *The British Journal of Clinical Psychology.* 2012;51(2):206.
114. Fisher AJ, Boswell JF. Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment.* 2016;23(4):496-506.
115. Kirtley OJ, Hiekkaranta AP, Kunkels YK, Verhoeven D, Van Nierop M, Myin-Germeys I. The Experience Sampling Method (ESM) Item Repository. . *Retrieved from osfio/kg376.* 2019, April 2.
116. Perez Arribas I, Goodwin GM, Geddes JR, Lyons T, Saunders KE. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Transl Psychiatry.* 2018;8(1):1-7.



Nederlandse samenvatting

Achtergrond

Ongeveer een derde van de wereldbevolking wordt ooit in zijn of haar leven gediagnosticeerd met een psychische stoornis. Dit kan variëren van een depressie of een angststoornis, tot minder voorkomende diagnoses als een bipolaire stoornis of schizofrenie. Een psychische stoornis kan zich op veel verschillende manieren uiten; de ene persoon met een depressie is de andere niet. Dat maakt het lastig te bepalen welke behandeling het beste werkt voor wie. Daarom gaan er in de wetenschap de laatste jaren steeds meer stemmen op voor een meer gepersonaliseerde, idiografische manier van onderzoek doen. In tegenstelling tot traditioneel groepsonderzoek bestudeert deze tak het individu: bijvoorbeeld hoe psychische klachten zich uiten, met elkaar samenhangen of zich ontwikkelen binnen individuen.

Een manier om dergelijke individuele klachtenpatronen in kaart te brengen is *ecological momentary assessment* (EMA), in het Nederlands ook wel dagboekmetingen genaamd. Met EMA kan men zijn of haar ervaringen meerdere keren per dag bijhouden via de smartphone (zie Figuur 1 voor een voorbeeld). Er worden dan vragen gesteld over de stemming, eventuele psychische klachten, gedachten en de context (activiteiten, gebeurtenissen en gezelschap). Het aantal dagboekjes varieert meestal van drie tot soms wel tien keer per dag. In onderzoek wordt EMA veelvuldig gebruikt om beter te begrijpen hoe psychische klachten zich uiten in het dagelijks leven. In de afgelopen jaren werd echter ook vaak gesproken over de belofte van EMA voor de klinische praktijk.

Veel onderzoekers verwachten dat EMA zou kunnen helpen om mensen met psychische klachten meer grip op hun welzijn te geven. EMA zou mensen helpen om regelmatig even stil te staan bij hun welzijn en hoe ze deze kunnen verbeteren. Een andere veelgenoemde belofte is de mogelijkheid om de ingevulde dagboekgegevens te visualiseren als ‘dagboekfeedback’. Deze feedback kan een behandelaar vervolgens bespreken met de patiënt: dit zou relevante inzichten

kunnen bieden voor diagnostiek of vervolgbehandeling.

Er zijn echter nog veel vragen over *hoe* EMA precies in de behandelpraktijk toegepast dient te worden. Tot voor kort werd EMA vooral in onderzoekscontext toegepast, waarbij men kort (maximaal twee weken) een gestandaardiseerd dagboekje invulde. Er is nog een brug te slaan naar de toepassing in de behandelpraktijk. Op het eerste gezicht lijkt EMA goed te passen in de traditie van registratie die in veel behandelingen, met name cognitieve gedragstherapie, al wordt toegepast. Maar hoe zien behandelaren en patiënten het gebruik van EMA precies voor zich in de praktijk? Wat zijn mogelijke toepassingen die we verder kunnen exploreren? Is EMA voor iedere patiënt geschikt, of zijn er ook personen of momenten waarop EMA juist nadelig kan werken? En tot slot: wat voor dagboekfeedback is nuttig?

Figuur 1. Voorbeeld van een EMA-dagboek en EMA-dagboekfeedback.

Dit laatste punt behoeft nog enige uitleg. Dagboekfeedback kan bestaan uit relatief eenvoudige grafieken. Daarin worden bijvoorbeeld de schommelingen in de stemming in de afgelopen periode weergegeven, of de gemiddelde stemming

gedurende activiteiten (zie Figuur 1). Er kunnen echter ook complexere statistische methoden worden toegepast, waarmee de verbanden tussen verschillende psychische klachten in kaart worden gebracht. Dit wordt gedaan door middel van netwerkanalyse. Hierbij wordt getoetst of een bepaalde klacht voorafgaat een andere: is het bijvoorbeeld zo dat, bij een bepaalde persoon, piekeren vaak wordt gevolgd door somberheid? Een persoonlijk netwerkmodel zou mogelijk inzichtelijk kunnen maken hoe diverse klachten met elkaar samenhangen, en op die manier richting kunnen geven aan de behandeling. Klachten die sterk samenhangen met veel andere klachten, zogenaamde centrale symptomen, zouden bijvoorbeeld een veelbelovend startpunt kunnen vormen van de behandeling.

Netwerkanalyse vloeit voort uit de netwerktheorie, die postuleert dat stoornissen ontstaan doordat symptomen elkaar aansteken over de tijd. De analyses die de theorie in praktijk brengen, hebben zich in de afgelopen jaren snel ontwikkeld. Maar ook hier zijn nog veel onduidelijkheden. We zouden verwachten dat een behandeling invloed heeft op de netwerkstructuur, is dat ook zo? In hoeverre komen verschillende typen netwerkanalyses met elkaar overeen? Laten zij dezelfde verbanden zien?

In dit proefschrift is de belofte van EMA voor de klinische praktijk onderzocht. In het eerste deel heb ik gekeken naar de belofte van netwerkanalyse voor de behandelpraktijk. In het tweede deel heb ik onderzocht of, hoe en wanneer EMA toegepast kan worden in de behandeling van diverse psychische stoornissen.

Belangrijkste bevindingen

Netwerkanalyse kan worden toegepast op EMA-data, maar ook op symptoomvragenlijsten. Dergelijke symptoomvragenlijsten worden meestal eens in de zoveel weken afgenomen in de context van een behandeling of in onderzoek om te zien of symptomen zijn afgenomen. Cross-sectionele netwerkanalyse, toegepast op symptoomvragenlijsten van een grote groep mensen, kan inzichtelijk maken hoe klachten gemiddeld genomen met elkaar samenhangen. In **hoofdstuk 2** hebben we cross-sectionele netwerkanalyse toegepast om de werkingsmechanismen van antidepressiva te onderzoeken. Dit onderzochten we in een groep van 170 depressieve patiënten die gerandomiseerd werden naar behandeling met een

selectieve serotonine-heropnameremmer (SSRI). Voorafgaand aan de behandeling en na acht weken antidepressivagebruik vulden zij een uitgebreide vragenlijst in over depressieve symptomen. De netwerkanalyse liet zien dat symptomen gemiddeld genomen sterker met elkaar samenhingen na acht weken antidepressivagebruik. Ook zagen we dat symptomen die eerder sterk met elkaar samenhingen, later minder centraal werden, en vice versa. Omdat dit onderzoek geen controlegroep bevatte, zijn deze verschillen echter niet met zekerheid aan de antidepressiva toe te schrijven. Daarnaast heeft cross-sectionele netwerkanalyse als beperking dat de resultaten lastig te generaliseren zijn naar individuele patiënten. Het lijkt er dus op dat dit type analyse vooral geschikt is om hypotheses op te werpen die getest dienen te worden in vervolgonderzoek.

Daarom hebben we in **hoofdstuk 3** dynamische netwerkanalyse toegepast, in het bijzonder *impulse response function* (IRF) *analysis*. Dynamische netwerkanalyse kan worden toegepast op individuele EMA-data. Dat heeft twee voordelen boven cross-sectionele netwerkanalyse: ten eerste kan er per patiënt inzichtelijk kan worden gemaakt welke emoties met elkaar in verband staan, en ten tweede kan er gekeken worden naar hoe klachten op elkaar volgen over de tijd. In deze studie hebben we ons gericht op 40 mensen met subklinische depressieve klachten. Zij werden opgedeeld in twee groepen: mensen met het kernsymptoom anhedonie (verlies van interesse) en mensen zonder dit symptoom. We waren benieuwd of de slechtere prognose van mensen met anhedonie verklaard zou kunnen worden door een andere dynamiek tussen emoties van moment tot moment. En inderdaad: hoewel de twee groepen klinisch gezien vergelijkbaar waren, zagen we verschillende emotionele dynamiek. Terwijl mensen zonder anhedonie vooral een sterke impact ervoeren van stress, hadden mensen met anhedonie vooral minder baat bij positief affect. Het laat zien dat de ontwikkeling van depressieve klachten mogelijk op verschillende manieren kan verlopen, en dat persoonlijke modellen gebaseerd op EMA-data wellicht een waardevolle aanvulling zouden kunnen zijn voor de behandeling.

In **hoofdstuk 4** zijn we dieper gedoken in de relatie tussen cross-sectionele en dynamische netwerkanalyse. Tot dusver was het namelijk nog niet bekend in

hoeverre de resultaten van beide type analyses tot dezelfde conclusies zouden leiden. Als dat het geval zou zijn, zouden we de resultaten van cross-sectionele analyses ook kunnen toepassen op individuele patiënten. Een groot voordeel van cross-sectionele analyses is dat er al veel datasets beschikbaar zijn, en de dataverzameling van dergelijke data een stuk minder tijdrovend is dan de benodigde data voor dynamische analyses. Daarom construeerden we cross-sectionele en dynamische netwerken van zeven depressieve symptomen in een groep van 104 depressieve patiënten. Helaas zagen we in de cross-sectionele en dynamische netwerken verschillende verbanden tussen de symptomen, en duidde iedere netwerkvorm andere symptomen aan als centraal. Dat betekent dat de verbanden die we zien in cross-sectionele netwerken niet zomaar aanduiden hoe symptomen over de tijd met elkaar in verband staan.

In het laatste hoofdstuk van deel I, **hoofdstuk 5**, hebben we de netwerkliteratuur op het gebied van depressie onder de loop genomen. We maakten hierbij onderscheid tussen vier typen netwerkanalyses: netwerken gebaseerd op symptomen (macroniveau) versus op emoties (microniveau), en netwerken gebaseerd op cross-sectionele gegevens versus dagboekdata. In totaal hebben we 56 studies beschreven. We zagen in de review dat studies op microniveau andere conclusies opleverden dan de studies op macroniveau. Wel lieten veel studies overtuigend zien dat positief affect vaak een belangrijke rol speelt in netwerken van symptomen of emoties. Daarnaast introduceerden we een nieuwe netwerktheorie, de microlevel affect dynamics (MAD) theorie, die de huidige (cross-sectionele) netwerktheorie aanvult. Gezien de verschillen tussen de typen netwerkanalyses is het belangrijk voor onderzoekers om hier duidelijk onderscheid in te maken bij het bespreken van resultaten. Ook komen we in dit hoofdstuk tot de conclusie dat netwerkanalyse in zijn huidige vorm nog niet klaar is voor gebruik in de klinische praktijk.

In deel twee van dit proefschrift heb ik gekeken naar de toepassing van EMA in verschillende behandelcontexten.

Hoofdstuk 6 richtte zich op mogelijke toepassingen van EMA op het gebied van psychofarmacologie. We hebben in totaal 18 studies beschreven die EMA hebben gebruikt om de effecten van medicatie te bestuderen in diverse psychiatrische

populaties. Zo werd EMA bijvoorbeeld gebruikt om behandeluitkomsten te voorspellen of de duur van medicatie-effecten te bepalen. EMA biedt dus interessante mogelijkheden om farmacologische behandeling verder te personaliseren en te begrijpen hoe medicatie werkt in het dagelijks leven van individuele patiënten.

In **hoofdstuk 7** onderzochten we de verwachtingen van patiënten en behandelaren omtrent het gebruik van EMA in de behandeling. Zo interviewden we in totaal 22 patiënten met verschillende diagnoses en 21 behandelaren (psychiaters, psychologen en verpleegkundigen). Velen van hen waren enthousiast over het nut van EMA. Zo verwachtten zij dat EMA relevant kon zijn in verschillende fasen van zorg, en dat het de relatie tussen patiënt en behandelaar zou kunnen verbeteren. Tegelijk noemden zij ook nadelen die tot dusverre nog niet in eerder onderzoek waren voorgekomen: de hoge belasting en mogelijke verergering van klachten door er steeds bij stil te staan. Patiënten en behandelaren benadrukten het belang van het personaliseren van het EMA-dagboek zodat het goed aansluit bij de behandeling.

Deze bevindingen werden bekrachtigd door de resultaten in **hoofdstuk 8**. Hier lieten we 20 patiënten met een bipolaire stoornis vier maanden lang vijf EMA-dagboekjes per dag invullen. Na afloop maakten we een persoonlijk rapport met EMA-feedback die tussen patiënt en behandelaar werd besproken. Veel patiënten vonden het invullen van de dagboekjes het waardevolst: zo beschreven ze dat ze zich bewuster werden van hun stemming en daardoor meer grip ervoeren op hun klachten. De EMA-feedback sloot minder goed aan bij de verwachting van patiënten en behandelaren. Het is daarom van groot belang om de EMA-feedback klinisch relevant te maken en daarin samen te werken met patiënten en behandelaren.

Tot slot hebben we in **hoofdstuk 9** de mogelijkheid onderzocht van een gepersonaliseerd alarmsysteem op basis van EMA. Zo bekeken we of we de EMA-data verzameld in hoofdstuk 8 konden gebruiken om een plotselinge toename in manische of depressieve klachten te voorspellen. Eerder onderzoek bood namelijk aanwijzingen dat een verminderde veerkracht in emoties (ook wel ‘vroege waarschuwingssignalen’ genoemd) voorafgaat aan het toenemen van psychische klachten. Dit onderzochten we in 17 verschillende momentane emoties. We vonden inderdaad dergelijke waarschuwingssignalen in bepaalde emoties, met name in

opgewektheid, snel kunnen schakelen, vol ideeën zitten, piekeren, gedachtevlucht, agitatie, vol energie zitten, en vermoeidheid. Desalniettemin vonden we niet dezelfde waarschuwingssignalen bij iedereen, vonden we ook vals positieven en verschilden de hoeveelheid en het type waarschuwingssignaal sterk per persoon. EMA biedt dus interessante mogelijkheden voor een alarmsysteem, maar meer vervolgonderzoek is zeker nog nodig voordat het in de praktijk toegepast kan worden.

Conclusie

Dit proefschrift heeft laten zien dat de klinische toepassing van EMA veelbelovend is. EMA lijkt patiënten meer grip te kunnen bieden op hun klachten, behandelingen persoonlijker te maken, en het gesprek tussen patiënt en behandelaar te verbeteren. Aan de andere kant hebben we gezien dat het gebruik van EMA spaak kan lopen als de methode onvoldoende relevant is gemaakt voor de praktijk. Zo is het juist in de praktijk van belang dat de dagboekjes zich niet alleen richten op klachten maar juist ook positief worden ingestoken, en dat zij op maat gemaakt kunnen worden voor de patiënt en het behandeldoel. Het is daarom essentieel dat patiënten en behandelaren worden betrokken bij de verdere ontwikkeling van EMA als behandeltool. Het enthousiasme waarmee behandelaren en patiënten het gebruik van EMA tegemoetzien, heeft geleid tot de ontwikkeling van PETRA, een online platform waarmee behandelaren en patiënten eenvoudig een gepersonaliseerd dagboekje kunnen maken, invullen en feedback bekijken. Gezien de huidige stand van onderzoek rondom netwerkanalyse, zal de praktijk het voorlopig nog moeten doen met eenvoudige, beschrijvende dagboekfeedback. Dat is jammer, omdat dit een van de grootste beloften was van EMA voor de behandelpraktijk. Dit proefschrift heeft echter laten zien dat we de meerwaarde van EMA vooral kunnen vinden in het versterken van de eigen regie van de patiënt. De komende jaren zal de wetenschap zich verder moeten buigen over de ontwikkeling van klinisch relevante visualisaties. Ook zal gerandomiseerd onderzoek moeten uitwijzen of EMA inderdaad een positief effect heeft op het functioneren en de veerkracht van patiënten. Op die manier kan EMA het ideaal van gepersonaliseerde zorg nog weer een stap dichterbij brengen.

Dankwoord

Nou tou moar, daar issie dan! Wat heb ik genoten en wat ben ik dankbaar voor degenen die dit avontuur samen met mij hebben beleefd.

Te beginnen bij mijn geweldige promotieteam. Samen hebben we er echt een feestje van gemaakt. **Marieke**, wat was ik blij toen ik als PhD bij jou aan de slag mocht! Jij was voor mij een ideale promotor: open, betrokken, leergierig en met een scherpe blik op wat het veld en je promovendus nodig heeft. Jij wist altijd de juiste vragen te stellen om een paper of project verder te brengen. Ik heb jouw adviezen erg gemist in het staartje van mijn traject, en misschien onze discussies over onze bevindingen nog wel meer. Hopelijk krijgen we die kans nog eens.

Richard, jij was de kritische behandelaar van ons team en hield ons scherp op de toegevoegde waarde voor de kliniek. Ook was je voor mij een geweldige mentor en coach. Luisteren is een kunst, en jij hebt die als geen ander onder de knie. Zo heb je mij altijd aangemoedigd en me het vertrouwen gegeven dat ik op de goede weg zit. Als onze *designated survivor* heb je in de laatste jaren steeds meer begeleiding op je genomen en ik ben je dankbaar voor hoe je dat hebt opgepakt.

Evelien, ik weet nog goed hoe wij elkaar voor het eerst ontmoeten in *sunny Philadelphia*. Ik heb jou de oren van het hoofd gekletst, iets wat ik in de jaren die erop volgden nog heel vaak zou doen. Wat ben ik blij met jou als copromotor. Ik heb heel veel van jou geleerd op het gebied van ESM, statistiek, en goede wetenschap beoefenen – jouw kennis van ons veld is ongeëvenaard. Maar bovenal was jij vaak de persoon met wie ik mijn hoogte- en dieptepunten wilde delen. Ik ben blij en dankbaar dat we naast collega's ook goede vrienden zijn geworden.

Lian, als er iemand is die ik moet bedanken ben jij het; jij hebt dit project gepitcht. Ik heb enorme bewondering voor hoe jij wetenschap en praktijk weet te combineren. Jij leerde mij veel over hoe ik onderzoek dichtbij de mensen kan brengen. Ook leerde jij mij om met beide benen op de grond te blijven staan, en om ervoor te zorgen dat mijn werk mijn privéleven niet teveel naar de achtergrond verdrijft.

Mijn proefschrift is onlosmakelijk verbonden aan het **PETRA-project**, waarbinnen ik al mijn ideeën over EMA als klinische tool in de praktijk kon brengen. Ik viel

met m'n neus in de boter. Een implementatieproject is spannend en heeft van veel mensen visie, tijd en energie gevraagd. Graag sta ik hier dan ook stil bij het PETRA-projectteam. **Harriëtte**, wij gingen dit project samen aan en wat hebben we veel voor elkaar gebokst in de afgelopen jaren! Ik ben je enorm dankbaar voor het vertrouwen dat je mij hebt gegeven. Jouw vermogen om op alle niveaus in een organisatie mensen bij elkaar te brengen is weergaloos. Ik kijk heel erg uit naar de komende tijd, als we PETRA eindelijk in de praktijk kunnen gaan zien, en lekker kunnen gaan sparren over het vervolg. **Ando**, de magie die jij elke week weer hebt gecodeerd tovert steevast een glimlach op mijn gezicht. **Tom**, jij leerde mij dat klein en simpel beginnen groot resultaat oplevert, een belangrijke vaardigheid voor een perfectionist als ik. Ook bedankt voor je hulp met dit proefschrift. **Lino**, als 'ESM-in-de-praktijk-onderzoekers' kunnen wij uren filosoferen over de ideale toepassing in de kliniek. Jij hebt altijd nieuwe en innovatieve ideeën voor PETRA. Dat maakt het ook zo leuk als we die samen in de praktijk kunnen brengen. **Erwin**, en ook **Dennis**, **Henk** en **Inge**, bedankt voor alle mogelijkheden die jullie bieden, en het geduld en de precisie waarmee jullie PETRA in RoQua een plaats geven. Ook dank aan alle RoQua-teamleden, in het bijzonder **Jorn**, **Inge** en **Ellen** voor alle hulp omtrent de dagboekstudies van dit proefschrift. **Jannie** en **Roan**, ook jullie bedankt voor het fijne samenwerken.

Dit proefschrift is enorm geholpen door behandelaar-onderzoekers die met ons meedachten. **Bennard**, jij zit altijd vol goede ideeën en gelukkig klopte je daarmee bij ons aan. Je hebt veel invloed gehad op mijn vorming als onderzoeker: jij deed me inzien dat ik het liefst praktisch toepasbaar onderzoek wilde doen, als spil tussen onderzoekers, behandelaren en patiënten. Daar ben ik je enorm dankbaar voor. Nog steeds tillen we samen het onderzoek op het snijvlak tussen wetenschap en praktijk naar een hoger niveau. Daar geniet ik enorm van. **Wim**, ook jij gaf ons de kans om op jouw afdeling te experimenteren met EMA. Dank voor alle tijd en mogelijkheden die je ons daarvoor hebt geboden. Ook dank ik graag de behandelteams van de **bipolaire poli** en de **psychosepoli** van het UCP en GGZ Drenthe voor hun inspanningen rondom de dataverzameling.

Aan alle **deelnemers** van mijn onderzoek, als patiënt of als behandelaar: ik ben

jullie bijzonder dankbaar voor alle –vaak persoonlijke– inzichten die jullie met mij wilden delen. Jullie veerkracht is voor mij een grote inspiratiebron.

Also many thanks to my other amazing co-authors and collaborators, with whom conducting research has been a great pleasure and learning experience: **Frank, Marieke S., Sandip**, **Eiko, Laura, Sona, Claudi, Stijn, Peter, Jessica, Claudia, Marije, Robert,** and **Taylor**. In particular, **Frank, Marieke,** and **Sandip**: your statistical expertise and the enthusiasm with which you talked about it made our collaboration so much fun. Our meetings often felt like the highlight of the week.

I'd also like to thank my international collaborators, **Steve**, **Rob**, and **Aaron**, for offering me a place at their labs. Your work on improving treatments of psychopathology is inspiring and has encouraged me to conduct research that, in the end, will benefit patients.

Veel dank aan de leden van de beoordelingscommissie: **prof. dr. Judith Rosmalen, prof. dr. Inez Myin-Germeys,** en **prof. dr. Philippe Delespaul.**

Ook bedank ik graag mijn mede-onderzoekers van het **RGOc**, **ICPE** en het **UCP** van wie ik veel steun heb ervaren tijdens mijn onderzoek. **Meike,** jij brengt cliënt en onderzoeker nader bij elkaar, en hebt ons enorm geholpen bij het betrekken van cliënten bij ons onderzoek. **Kaying**, wij bewandelden samen het implementatiepad en dat was zowel inspirerend als erg gezellig. **Margo** en **Esther**, jullie stonden altijd voor mij klaar, en ik voelde me dankzij jullie erg welkom toen ik weer terugkeerde. Bedankt daarvoor. **Maurits**, ik ging ooit koffie leren drinken zodat ik als nieuwe collega met jullie mee koffie kon halen. Zes jaar vol koffiewandelingen later zijn we een mooie vriendschap – en koffieverslaving – rijker. We voelen elkaar vaak goed aan.

Lieve **YAM(F)ARM**, wat had ik een geluk toen jullie erbij kwamen. Ik geniet ontzettend van het sparren over ons onderzoek, de wetenschap, of gewoon het leven. Stuk voor stuk zijn jullie fantastische onderzoekers van wie ik veel leer en tegen wie ik erg opkijk. Vele schrijfweken, borrels, vakanties, liedjesschrijverij en ape-indexes later zijn we daarnaast ook een hechte vriendengroep geworden. Bij jullie voel ik me altijd een winnaar, ook als het allemaal even wat moeilijker gaat. **Yoram**, jij bent onze stille kracht, onze handyman en professioneel cocktailbarman.

Bij jou hoef ik nooit veel te zeggen om me toch begrepen te voelen. **Arnout**, ik zal nooit vergeten dat jij mij vijfsterrendiners kwam brengen in de laatste fase van mijn proefschrift. Jij had altijd een rotsvast vertrouwen in mij. Die gedachte heeft me erg gesterkt in moeilijke perioden. **Marmar**, wij lijken uit hetzelfde hout gesneden. Bij jou heb ik vaak aan één woord al genoeg. Samen *nerden*, even ventileren of gewoon beppen – het is altijd even leuk. Samen hebben we enorm naar de grote dag toegeleefd. Dat jij dan naast me staat geeft me veel rust en maakt dat ik er zelfs een beetje naar uitkijk. Enne, deze tekst zal het enige onderdeel zijn van dit proefschrift dat je níet bent nagelopen: ook daarvoor duizendmaal dank. **FARM**, als kamergenootjes hebben we heel wat gelachen en gehuild samen. Jullie sprongen voor mij op de bres toen ik terugkeerde. Met jullie support kon ik alles aan. **Anouk**, jij kan me overal enthousiast voor maken. Je leerde me om voor mezelf op te komen en daar blijf ik je dankbaar voor. **Robin**, wij hebben de online shops veel klandizie bezorgd in de luwe uurtjes op kantoor. Jij stond altijd voor me klaar, of het nou was met een taartje aan de voordeur of door op het juiste moment een scherpe vraag te stellen. **Marieke**, ik ken denk ik niemand die zo vrolijk, warm en hartelijk is als jij. Ik heb de foute Nederlandse hits (en je zittende dansmoves) moeten missen het afgelopen jaar, maar gelukkig zijn daar heel veel fijne wandelingen voor in de plaats gekomen. Jongens, bedankt dat jullie er waren.

Ook bedank ik hier heel graag mijn lieve vrienden, **Yfke, Julia, Jasper**, **Femke**, **Robin A.** en **Leonie**. Bij jullie hoef ik niet dr. Boss te zijn, maar mag ik als mezelf komen. **Yfke**, ook al woon je 'ver' weg, je voelt altijd fijn dichtbij. Er gaat vaak nauwelijks een dag voorbij zonder een berichtje aan elkaar. Ik ben blij dat we al onze belevenissen zo fijn met elkaar kunnen delen. Niet alleen lijken we in veel dingen op elkaar, ook houden we van dezelfde kneuterige hobbies en ik geniet er enorm van om die samen met jou te beoefenen op onze vakanties en weekendjes weg. In veel dingen lijken wij erg op elkaar. **Julia**, onze gedachten zitten vaak op dezelfde golflengte. Of het nou gaat om #phdlife, vriendschap, of de recentste shopuitspatting: het is altijd gezellig. Ik voel me altijd heel begrepen en vertrouwd bij jou. **Jasper en Femke** (en Tjibbe), samen bevochten we al een virus nog voor het werkelijkheid werd. Mijn weekenden bij jullie zijn steevast ontspannen en gezellig. Enne, misschien toch ooit een vlogkanaal beginnen samen? **Robin,** jij maakt me

altijd aan het lachen, bedankt voor het meeleven. **Leonie**, wij moeten onze angsten soms allebei even flink laten schrikken, maar we doen het toch maar mooi. Ik ben heel trots op ons.

Rob, we kennen elkaar al bijna twintig jaar en hebben al een hoop meegemaakt samen. Jij bent degene die ik bel als er feest gevierd moet worden of dieptepunten moeten worden verwerkt. Dan luister je, leef je mee, en help je me een oplossing te zoeken of te relativeren. Met je heerlijke humor moedig je me aan om uit mijn comfortzone te stappen. Niet zelden zeg je tegen me: wat er ook gebeurt, wat er ook misgaat, een dag later gaan we ook gewoon weer koffie drinken hoor. Daarmee geef je me heel veel rust: ik voel me 100% mezelf bij jou. Ik ben dan ook heel blij en dankbaar dat je ook als paranimf met me hebt meegeleefd en naast me zult staan tijdens de verdediging. Samen maken we er echt een mooi avontuur van, zoals we dat met zoveel kleine en grote momenten doen. Op nog vele jaren mooie vriendschap!

Lieve **oma Willy**, jij was altijd met mij en mijn onderzoek begaan en leefde echt met me toe naar de grote dag. Wat is het zuur dat je er dan niet bij kunt zijn. Ik hoop dat ik je toch een beetje trots heb gemaakt. Lieve **oma Joy**, ook jij leeft altijd mee, dankjewel daarvoor.

Lieve **Jasper en Maaike**, ik grijp hier even m'n kans om de sentimentele zus te zijn (in een dankwoord mag dat!). Wat ben ik trots op hoe jullie een leven opbouwen daar in Leiden. Jullie staan altijd voor anderen klaar, zo ook voor mij tijdens de moeilijkere momenten in de afgelopen jaren. Of het nou was met snickerijs, een scooterritje, of gewoon een luisterend oor. Bij jullie voel ik me altijd thuis. Ik kan niet wachten op toekomstige avonturen.

Lieve **papa en mama**, dankzij jullie zat de liefde voor de universiteit er al vroeg in. Jullie moedigen me altijd aan om te doen waar ik gelukkig van word en geven me het vertrouwen dat ik veel meer kan dan ik zelf vaak denk. Tijdens alle pieken en dalen in de afgelopen jaren hebben jullie naast mij gestaan - ik kon altijd op jullie rekenen. Ik voel me heel erg rustig en begrepen door jullie. Ook leerden jullie mij dat de mooiste, meest waardevolle momenten altijd ontstaan in de verbinding met anderen. Dat is iets waar ik in mijn werk en in mijn leven het meest van geniet.

Curriculum Vitae

Fionneke Bos was born on May 2, 1992 in Groningen, The Netherlands. After finishing her secondary education (Praedinius Gymnasium) in 2010, she completed her bachelor Psychology and the extracurricular program Honours College at the University of Groningen in 2013. After that, she started the Behavioural and Social Sciences Research Master, specialization Clinical Psychology, at the University of Groningen. As part of this program, she embarked on a six-month research visit to Vanderbilt University and the University of Pennsylvania in the United States, where she performed research for her master's thesis on network analysis of depressive symptoms. She graduated cum laude.

In 2015, she started her PhD project at the Rob Giel Research Center at the University Medical Center in Groningen, under supervision of prof. dr. Marieke Wichers, prof. dr. Richard Bruggeman, dr. Evelien Snippe, and dr. Lian van der Krieke. From 2018 onwards, she combined her PhD research with the coordination of the PETRA project, focusing on the development of a tool for personalized diaries. Currently, Fionneke works as a postdoctoral researcher at the Rob Giel Research Center and the iLab of the University Center Psychiatry in Groningen, continuing her research on the application of ecological momentary assessment in clinical practice.

List of peer-reviewed publications

Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: A qualitative study of patient and clinician experiences. *International Journal of Bipolar Disorders.* 2020;8(1):1-14.

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;70(11):983-991.

Wichers M, Snippe E, Riese H, Bos FM. De netwerkbenadering bij depressie: Veel noten op de zang of heilige graal? *Tijdschrift voor Gedragstherapie.* 2019;52(1):53-68.

Bos FM, Blaauw FJ, Snippe E, van der Krieke L, de Jonge P, Wichers M. Exploring the emotional dynamics of subclinically depressed individuals with and without anhedonia: An experience sampling study. *Journal of Affective Disorders.* 2018;228:186-193.

Bos FM, Fried EI, Hollon SD, Bringmann LF, Dimidjian S, DeRubeis RJ, Bockting CLH. Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *SocPsychiatry PsychiatrEpidemiol.* 2018;53(6):617-627.

Bos FM, Snippe E, de Vos S, Hartmann JA, Simons CJP, van der Krieke L, de Jonge P, Wichers M. Can we jump from cross-sectional to dynamic interpretations of networks? Implications for the network perspective in psychiatry. *PsychotherPsychosom.* 2017;86(3):175-177.

Bos FM, Schoevers RA, aan het Rot M. Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review. *European Neuropsychopharmacology.* 2015;25(11):1853-1864.

Submitted for publication

Bos FM, Schreuder MJ, George SV, Doornbos B, Bruggeman R, van der Krieke L, Haarman BCM, Wichers M, Snippe E. Anticipating manic and depressive shifts in patients with bipolar disorder using early warning signals. Submitted.

Wichers M, Riese H, Snippe E, Bos FM. A narrative review of network studies in depression: What different network methodologies tell us about depression. Submitted.