University of Groningen

What's in a mood?

Helmich, Marieke A.

DOI:
[10.33612/diss.200101721](https://doi.org/10.33612/diss.200101721)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Helmich, M. A. (2022). *What's in a mood? looking for dynamic predictors of individual improvement in depression*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.200101721>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

What's in a mood?

Looking for dynamic predictors of
individual improvement in depression

Marieke A. Helmich

The research in this thesis was conducted as part of the Transitions in Depression (TRANS-ID) research project. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (ERC-CoG-2015; No 681466 to M. Wichers). We are grateful to all who contributed to making this project possible.

Layout and cover design: Marieke A. Helmich
Printed by: Ipskamp

© Marieke A. Helmich, 2021. All rights reserved.

university of
groningen

What's in a mood?

Looking for dynamic predictors of
individual improvement in depression

PhD thesis

to obtain the degree of PhD at the
university of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Wednesday 2 February 2022 at 12.45 hours

by

Maria Agatha Helmich

born on 13 February 1992
in Haaren

Supervisors

Prof. dr. M. Wichers

Prof. dr. A.J. Oldehinkel

Co-supervisors

Dr. E. Snippe

Dr. L. F. Bringmann

Assessment Committee

Prof. dr. A. M. Hayes

Prof. dr. N. Jacobs

Prof. dr. R. C. Oude Voshaar

Chapter overview

Chapter 1	General introduction	6
Chapter 2	Sudden gains in day-to-day change: revealing nonlinear patterns of individual improvement in depression	26
Chapter 3	Daily dynamics of negative affect: indicators of rate of response to treatment and remission from depression?	48
Chapter 4	Early warning signals and critical transitions in psychopathology: challenges and recommendations	68
Chapter 5	Transitions in Depression (TRANS-ID) Recovery: study protocol for a repeated intensive longitudinal $n = 1$ study design to search for personalized early warning signals of critical transitions towards improvement in depression	88
Chapter 6	The Duration-adjusted Reliable Change Index (DaRCI): defining clinically relevant symptom changes of varying durations	114
Chapter 7	Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression	136
Chapter 8	Personalized detection of impending symptom transitions in depression using early warning signals during and shortly after antidepressant discontinuation	180
Chapter 9	Summary & General discussion	206
&	Nederlandse samenvatting	228
&	Dankwoord / a word of thanks	236
&	The course of my PhD in 7000+ mood measurements	243
&	Curriculum Vitae	247
&	List of publications SHARE	248

1.

General introduction

With parts adapted from:

Bringmann, L.F., **Helmich, M.A.**, Eronen, M.I., Völkle, M.C. (In press) Complex systems approaches to psychopathology. In Krueger, R.F. & Blaney, P.H. (Eds.), *Oxford Textbook of Psychopathology*. Oxford University Press.

The process by which depression arises and evolves over time is complex. It is not yet well understood who will experience improvement or recurrence of symptoms, and what predicts such symptom change. Neither antidepressant medication, nor psychological therapy are always successful treatments [1,2]. Moreover, even for individuals who successfully recover, depressive symptoms tend to recur with time [3–5]. Therefore, understanding how individuals recover from depression, and how treatment can help reduce depressive symptoms are important clinical goals.

In recent years, researchers have started questioning whether the moment-to-moment dynamics and intra-individual changes in disorders like depression may contain important information about the clinical prognosis for individual patients [6–9]. Particularly, we have witnessed a sharp increase in studies that approach psychopathology from the perspective of complexity science and dynamical systems theory to help explain the layered processes of within-person symptom change [10–18].

In this general introduction, I will explain why approaching psychopathology as a dynamic process is relevant and describe several approaches for capturing individual change in time series data of symptoms and affect. The chapters in this thesis will focus mostly on the importance of considering change in depression by investigating treatment response from a complex dynamical systems perspective. I aim to answer whether particular change patterns as well as the dynamics of momentary mood can be predictive of depressive symptom improvement.

The complex dynamical systems perspective on psychopathology

Within the complex dynamical systems framework, psychopathology is studied as a process, at the level of the individual. A person is viewed as a system of interdependent psychological processes, such as cognitions, behaviours, and emotions that interact and change over time [19–21]. The dynamic nature of psychological processes can be seen in everyday experience: most people will notice that their feelings generally fluctuate throughout the day, from hour to hour or even minute to minute [22–26]. Similarly, depressed patients will experience more symptoms in some weeks, while at other times the symptoms seem to wane. Yet, traditionally, researchers have studied psychopathology as a trait-like phenomenon [27]: measuring individuals cross-sectionally (once), and asking

them to retrospectively evaluate their symptoms over a period of several weeks or even a whole life span [28]. Such studies provide an important group-level understanding of depressive symptomatology – the heterogeneity in symptoms, the severity, the duration –, but oversimplify the changeability of disorders within people's life spans [29].

By contrast, intervention studies do look at diagnostic change: they assess individuals before and after an intervention to determine which form of psychological treatment or medication works best for most people, providing important information for therapists and policy makers [30]. Moreover, researchers are developing methods to use those group-level data to improve personalized treatment selection, and within-person predictions of treatment response, and symptom recurrence [31–34]. Yet, a lack of precise insight in individual-level changes remains: in these studies, the manner of change itself is not considered relevant, only the amount of symptom reduction; the time between pre- and post-assessment is left unregarded. Therefore, to gain a better understanding of the diversity of treatment trajectories, as well as the timing and antecedents of individual change during therapy, using a dynamic approach is important [35].

It is relevant to take into account that the dynamics and change processes in complex dynamical systems can occur at several different time scales, which are often not distinct, but hierarchically related [36,37]. For example, an episode of depression unfolds over a time scale of weeks or months. However, the components that make-up the depressive episode, such as insomnia, pessimistic thoughts, and rumination, take place at a faster time scale of weeks or days. These symptoms, in turn, can be seen as consisting of moment-to-moment changes and interactions among affect states [14,38]. Because complex dynamical systems are composed of interacting variables that can change at different rates, zooming in on the moment-to-moment changes in one variable (e.g., affect) can be informative of an upcoming change in another, slower changing variable (e.g., symptoms) [9,39]. Specifically, experiences of negative affect may not be disabling if they occur every now and then, but become problematic and disabling when they occur repeatedly or are constantly present, turning into a symptom of depression. Thus, studying mood in daily life at the time scale of minutes or hours can give a more fine-grained picture of the development of symptoms related to depression [9].

To capture depressive symptom changes as a process over time, (intensive) longitudinal data or time series are needed, with observations occurring frequently enough that meaningful fluctuations and changes in mood and symptoms are sampled, and not lost in between measurements [40]. Such data consist of many repeated assessments, for example over days, weeks, or months, sometimes also multiple times per day, typically gathered using methods known as the experience sampling method (ESM), ecological momentary assessment (EMA) or ambulatory assessment (AA) [41–45]. Although ESM, EMA and AA have different historical backgrounds, their goals and applications are overlapping, and the terms are increasingly used interchangeably.

These methods capture the individual in the course of daily life, thereby taking intra-daily dynamics and variability in emotions, cognition and behaviour into account [46,47]. Assessing the state of a person's feelings, thoughts, and behaviours at a given moment, within the context of their normal activities, reduces biases of more traditional methods, such as retrospective recall bias. Take as an example a study of Mokros [48], in which patients were asked to report their symptoms during the week and also to recall the symptoms they had experienced at the end of the week. This study vividly showed that the symptoms that were reported on a daily basis during the week were strikingly different from the symptoms that were recalled at the end of the week (see also Ref. [41]).

Early warning signals

With EMA studies becoming more commonplace [35], researchers have begun to examine this rich time series data for new predictors of within-person symptom change and are drawing on dynamical systems theory for methods to improve timely intervention [17,49]. In fact, one particular reason that the complex dynamical systems framework has generated interest in the field of psychopathology research is the hypothesis that indicators of change in the underlying system known as early warning signals (EWS) may be detectable in momentary affective states before a sudden shift in symptoms, such as a depressive relapse [50], or sudden improvements in the context of therapy [51]. These EWS have been successfully used to predict impending system reorganizations known as “critical transitions” in a wide range of complex systems, including in ecology [39,52–56], climate

data [57], financial markets [58], human development [19,59,60], epidemiology [61], and EEG-measured sleep states [62].

To understand how EWS work, it can be helpful to visualize the system in the form of a stability landscape. **Figure 1.1** provides a simplified visualization of such a stability landscape, which represents the person as a psychological system. The two basins depict a current and alternative attractor in the system (e.g., a depressed and non-depressed state), and the ball symbolizes the individual's psychological state at a given time (e.g., current emotion). The shape of the landscape, and the depth of the basin in particular, reveals how resilient a person is to external shocks (e.g., events that strongly influence one's feelings, like having a fight with a friend, or winning the lottery).

In **Figure 1.1A**, the system appears relatively stable, the basin is 'deep' – the current attractor is strong – and even though the ball can move around in response to external perturbations – pushing the ball away from the equilibrium position – with time, the ball will be pulled back to the resting position at the bottom of the basin. Note that, while a potential alternative attractor appears on the right side of **Figure 1.1A**, only a very strong impulse would cause the ball to be pushed into that basin. Translated to psychology: if we view the left side of the figure as an individual's mental state, which is currently relatively stable and healthy, it would require a strong trigger – like losing a loved one – for the system to switch to the right side, which can be understood as a more negative, possibly depressed, mental state.

In **Figure 1.1B**, the system has lost resilience: the current attractor has become weaker, and the basin shallower. As a result of this change in the stability landscape, the system in **Figure 1.1B** is more vulnerable to 'tip over' with a sudden shift to the alternative stable state (i.e., attractor) on the right side of the figure (**Figure 1.1C**): no longer does the ball need a large impulse to be pushed over the edge of the left-hand basin, even a small push (e.g., a conflict at work, or a restful holiday) could effectively cause a critical transition to a depressed state or another psychopathological state.

When a system changes from **Figure 1.1A** to **Figure 1.1B**, a phenomenon known as 'critical slowing down' occurs, as the basin of attraction loses strength and becomes shallower, the dynamical behaviour of the ball starts to change: when pushed away from the equilibrium position, it goes farther and returns more slowly.

Figure 1.1*Visual representation of stability landscapes.*

Note. The stability landscape represents a person's psychological system, where the shape of the system reflects how resilient a person is to external shocks (e.g., psychologically impactful events). The ball represents the current state of the system (e.g., current emotion), and the attractors the possible stable states that the system can be in (e.g., depressed state). In **Figure 1.1A**, the system shows a strong ('deep') attractor. In **Figure 1.1B**, the system is destabilizing and the attractor has weakened. In **Figure 1.1C**, the ball has 'tipped over' into the new attractor, and the old attractor disappears.

Adapted from "Early warning signals and critical transitions in psychopathology: challenges and recommendations," by Helmich et al. (2021), *Current Opinion in Psychology*, 41, p. 51-58. Copyright 2021 by M. A. Helmich.

This change in dynamics, in turn, can be observed as altered temporal dynamics of the system's variables, which serve as EWS. For instance, an increase in lag-1 autocorrelation can be an indicator of critical slowing down: as the ball spends more time away from its stable point (the mean), consecutive states are more similar and more highly correlated with previous states. Moreover, as the distance from the equilibrium position also increases, critical slowing down is expected to

correspond to EWS in the form of increased variance over time. Translated to the individual's psychological viewpoint, and to momentary affect in specific, the effect of critical slowing down would show as emotions that become increasingly volatile as the system is destabilizing (increased variance), and once something triggers an emotional reaction, this feeling is likely to linger longer over time (increased autocorrelation).

As one would perhaps expect from the EWS-hypothesis, studies on momentary affect dynamics have related higher levels of autocorrelation and variance in daily affective experiences to psychological well-being and depression [25,63–66]. Specifically, they found that – on average – indications of destabilization in EMA-measured positive and negative affect (statistically defined as autocorrelation at lag 1, the standard deviation, and the mean square of successive differences) were associated with stronger depressive symptoms concurrently [63,67,68] and later symptom change [69–72]. However, most studies have not examined how these affect dynamics link to the trajectory of symptoms over time, with only one such study using positive and negative affect dynamics to predict the rate of treatment response [73].

Furthermore, while these studies on affect dynamics provide some evidence for the notion that (developing) depression is related to the temporal dynamics associated with EWS, this evidence is indirect. In group-level studies, higher levels of autocorrelation of variance compared to others (other systems) are taken as indicators of EWS [51]. However, true indications of EWS occur within-system, as a change in dynamics compared to a person's own stable state, for instance, as a rising trend in autocorrelation and variance [50,74,75]. Thus, a true test of this hypothesis requires studies capturing change in autocorrelation and variance at the individual level (see also Ref. [76]).

In the first study to examine EWS within-person, Wichers and colleagues [50] monitored one individual with a history of major depression episodes while he was tapering his antidepressant medication. The participant completed ten EMA-questionnaires each day, which were sent at semi-randomly determined times across the day, and his depressive symptoms were assessed every week, over a period of about eight months (239 days). While the participant's symptoms were stable at the start of the study period, reducing medication dosage increases the risk of a depressive relapse, and he did relapse after 18 weeks. Upon examination

of the temporal dynamics in the data, the researchers found EWS in the form of significant rises in autocorrelation and variance in the weeks preceding the symptom transition. Thus, this study provided important first empirical evidence that EWS could be found for a specific individual, prior to a transition in depressive symptoms.

If it is found that EWS are indeed generic indicators of future symptoms shifts, individual patients, or at-risk individuals could be monitored using EMA and EWS could be used to make personalized prediction and timely intervention possible. For instance, if EWS are found to be consistently effective at predicting impending changes in individuals, their application in clinical practice could include early detection of episodes of mental illness [71,77], identifying sensitive periods in which psychological treatment may be more likely to effectuate positive symptom change [16,78], as well as improving timely intervention in cases of depressive relapse and suicide attempts [50,79]. In sum, the promise of EWS is that clinically important symptom changes could be anticipated by monitoring individuals' moment-to-moment (affective) experiences in the course of daily life.

Applying dynamical systems principles to depressive symptom improvement

Based on the complex dynamical systems framework, we may expect that studying the individual *over time*, in periods where they are likely to show symptom improvement, may provide an important new way to expand our understanding of change processes in depression. Particularly, this within-person approach allows us to question whether specific overall change trajectories, discontinuities or dynamics may be telling of a higher chance of symptom improvement.

An important dearth in our understanding of depression as a dynamical process is what kind of symptom changes can be classified as transitions between stable states (i.e., critical transitions, or phase shifts [39,80,81]) in the context of recovery. A 'critical transition' appears as a sizeable change in states (e.g., towards lower levels of symptoms), which appears relatively suddenly, and is not readily reversed (see **Figure 1.1**) [55,81–83]. However, a clear analogue of such a transition is not directly found in the clinical literature. For instance, what is considered

‘rapid’ in the context of depressive recovery; definitions vary between a substantial symptom improvement over the course of days, to as long as four weeks [84,85]. And while discontinuous phenomena such as sudden gains, depression spikes, and rapid early response have been well-described in weekly therapy data and typically contain a criterion of persistence of the improvement, these patterns have rarely been studied explicitly from a complex systems approach, nor with fine-grained data [59,86–91]. Thus, more research is needed to explore what types of shifts toward symptom improvement may represent potential critical transitions.

Currently, there is insufficient knowledge of what kind of trajectory depressive symptom improvement typically follows, and whether critical transitions can be identified in time series of within-person symptom assessments. Most studies on therapeutic change in depression have focused on identifying predictors of treatment outcomes that can be assessed at baseline, and can be linked to stronger pre-post intervention symptom improvement on average (i.e., at the group-level) [92–96]. Implicit in those studies is the assumption that the manner in which people’s symptoms change over time, from pre- to post-treatment, is similar between persons and likely linear [97]. However, nonlinearity is an essential quality of dynamical systems in general [60,98,99]. Based on the idea that some patients may experience a change to an alternative stable state as they recover, we may expect discontinuous change patterns to be a common occurrence in the overall trajectory of patients in treatment. Indeed, the handful of studies that have examined the form of change trajectories at the individual level have shown that overall improvement patterns are often nonlinear [79,86,93,100]. Furthermore, there is one type of symptom shift that clinicians have frequently observed within the course of therapy, known as “sudden gains” [101–103]. A sudden gain is defined as a large decrease in symptom scores between therapy sessions, which, drawing on dynamical systems theory, could be hypothesized to be a potential critical transition between attractor states (see **Figure 1.1**).

Notably, the manner in which individuals respond to treatment, and the timing, rate and overall shape of symptom improvement appear to be potentially relevant predictors of better outcomes. So far, a few studies have examined the trajectory of depressive symptom improvement itself – whether treatment response occurred gradually, in an early steep curve, or by a sudden mean-shift –, and related it to the likelihood of better outcomes (stronger symptom reduction, or

reaching remission status) [84,89,104–108]. Additionally, the presence of a sudden gain has been shown to be an important predictor of stronger symptom reduction after psychotherapeutic treatment and follow-up in many studies [101,109–117]. Clinically, it seems logical that showing a sudden, stable improvement in symptoms means that a patient will have a more positive treatment outcome than a patient without such a sudden gain. However, it remains an open question how sudden gains fit into the larger overall improvement pattern, and whether they represent critical transitions as described by dynamical systems theory, observed in clinical practice.

Finally, a rigorous, repeated study of the notion that individuals and change in depressive symptoms function according to the principles of complex systems theory at the psychological level is needed. To that end, the study by Wichers [50] warrants replication to confirm that EWS reliably precede depressive symptom recurrence. Additionally, testing whether EWS occur prior to transitions in the opposite direction, toward symptom improvement rather than recurrence, is highly relevant. No study has examined this thus far, yet this has great clinical relevance, with the promise of EWS as personalised indicators of ‘sensitive periods’ in which therapy may be particularly effective. Furthermore, this would provide an important strengthening of the theoretical basis for applying complex dynamical systems principles to psychopathology.

To empirically test these hypotheses, intensive longitudinal within-person datasets are needed, consisting of many (several hundred) observations per patient, covering both symptoms and affect, to ensure that the EWS (i.e., changes in dynamics over time) can be calculated reliably. Such a dataset must include a baseline period in which the system can be expected to be ‘stable’, and must then enter a condition – such as starting therapy – that increases the chance of a transition in symptoms that may be preceded by destabilization and rises in EWS. Moving beyond the narrative metaphor of individuals and their mental states as complex dynamical systems requires a test of these theoretical principles in empirical data, to see if this approach can live up to its promise to help anticipate and explain clinical change phenomena in a way that is relevant for patients. To that end, the chapters in this thesis examine depression as a dynamic process, studying idiographic change over the course of therapy and the role of affect dynamics and EWS as predictors of symptom improvement and treatment response.

Outline of this thesis

In **Chapter 2** we investigate the occurrence of sudden gains, and the shape of individual trajectories in daily self-rated problem severity during treatment, and the relationship between these patterns and treatment response. **Chapter 3** aims to uncover whether within-person affect dynamics can predict the rate of symptom improvement during treatment, and the remission status of participants over eighteen months. In **Chapter 4** we review some of the conceptual and practical challenges involved in studying EWS and critical transitions in psychopathology and list various important considerations for those designing empirical studies to study these phenomena. **Chapter 5** describes the study procedures we used to have an optimal chance of detecting EWS before sudden depressive symptom improvements: by collecting four months of intensive longitudinal individual time series of depressed patients who are starting psychological treatment. In **Chapter 6**, I propose a novel method for identifying reliable symptom changes of varying duration in repeated symptom assessments, while maintaining a given confidence level. Next, **Chapter 7** investigates the hypothesis of whether EWS can be detected prior to transitions toward improvement in depressed patients receiving treatment. Similarly, **Chapter 8** examines the hypothesis that EWS may occur prior to recurrence of depressive symptoms in a sample of formerly depressed individuals who are tapering their antidepressant medication. Finally, **Chapter 9** provides a summary and general discussion of the chapters in this thesis, and outlines several challenges and recommendations for researchers studying EWS and transitions in psychopathology.

References

1. Keller, M. B. (2004). Remission versus response: The new gold standard of antidepressant care. *Journal of Clinical Psychiatry*, 65([suppl 4]), 53–59.
2. Riedel, M., Möller, H.-J., Obermeier, M., Schennach-Wolff, R., Bauer, M., Adli, M., Kronmüller, K., Nickel, T., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W., & Seemüller, F. (2010). Response and remission criteria in major depression – A validation of current practice. *Journal of Psychiatric Research*, 44(15), 1063–1068. <https://doi.org/10.1016/j.jpsychires.2010.03.006>
3. Buckman, J. E. J., Underwood, A., Clarke, K., Saunders, R., Hollon, S. D., Fearon, P., & Pilling, S. (2018). Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clinical Psychology Review*, 64, 13–38. <https://doi.org/10.1016/j.cpr.2018.07.005>
4. Burcusa, S. L., & Iacono, W. G. (2007). Risk for recurrence in depression. *Clinical Psychology Review*, 27(8), 959–985. <https://doi.org/10.1016/j.cpr.2007.02.005>
5. Fava, G. A. (1996). The concept of recovery in affective disorders. *Psychotherapy and Psychosomatics*, 65(1), 2–13. <https://doi.org/10.1159/000289025>
6. Delignières, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences*, 8(4), 479–510.
7. Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322. <https://doi.org/10.1177/1754073915590619>
8. Kazdin, A. E. (2001). Progression of therapy research and clinical application of treatment require better understanding of the change process. *Clinical Psychology: Science and Practice*, 7(4), 143–151. <https://doi.org/10.1093/clipsy/8.2.143>
9. Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-Level Affect Dynamics in Psychopathology Viewed From Complex Dynamical System Theory. *Emotion Review*, 7(4), 362–367. <https://doi.org/10.1177/1754073915590623>
10. Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9(1), 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
11. Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of Counseling Psychology*, 63(4), 379–395. <https://doi.org/10.1037/cou0000150>
12. Granic, I. (2005). Timing is everything: Developmental psychopathology from a dynamic systems perspective. *Developmental Review*, 25(3–4), 386–407. <https://doi.org/10.1016/j.dr.2005.10.005>
13. Hayes, A. M., & Strauss, J. L. (1998). Dynamic systems theory as a paradigm for the study of change in psychotherapy: An application to cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 66(6), 939–947. <https://doi.org/10.1037/0022-006X.66.6.939>
14. Jeronimus, B. F. (2019). Dynamic system perspectives on anxiety and depression. In E. S. Kunnen, N. M. P. de Ruiter, B. F. Jeronimus, & M. A. van der Gaag (Eds.), *Psychosocial Development in Adolescence: Insights from the Dynamic Systems Approach* (1st ed.). Routledge Psychology.
15. Richardson, M. J., Dale, R., & Marsh, K. L. (2014). Complex dynamical systems in social and personality psychology: Theory, modeling, and analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 253–282). Cambridge University Press.

16. Schiepek, G. (2009). Complexity and nonlinear dynamics in psychotherapy. *European Review*, 17(02), 331. <https://doi.org/10.1017/S1062798709000763>
17. Schiepek, G., & Tschacher, W. (1992). Application of synergetics to clinical psychology. In W. Tschacher, G. Schiepek, & E. J. Brunner (Eds.), *Self-organization and clinical psychology. Empirical approaches to synergetics in psychology* (pp. 3–31). Springer. https://doi.org/10.1007/978-3-642-77534-5_1
18. Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, 16, 49–74. <https://doi.org/10.1146/annurev-clinpsy-102419>
19. Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to development and action*. MIT Press.
20. Van Geert, P. L. C. (2020). Dynamic systems, process and development. *Human Development*, 63(3–4), 153–179. <https://doi.org/10.1159/000503825>
21. Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, 53(6), 782–805. <https://doi.org/10.1080/00273171.2018.1496813>
22. Brans, K., Koval, P., Ku, P. V., Yan, L., Lim, L., & Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion*. <https://doi.org/10.1037/a0032400>
23. Chan, D. K. C., Zhang, X., Fung, H. H., & Hagger, M. S. (2015). Does emotion and its daily fluctuation correlate with depression? A cross-cultural analysis among six developing countries. *Journal of Epidemiology and Global Health*, 5(1), 65–74. <https://doi.org/10.1016/j.jegh.2014.09.001>
24. Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99(6), 1042–1060. <https://doi.org/10.1037/a0020962>
25. Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, 6(3), 383–391. <https://doi.org/10.1037/1528-3542.6.3.383>
26. Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., Thiery, E., Delespaul, P. H., & van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, 124(1–2), 191–195. <https://doi.org/10.1016/j.jad.2009.11.010>
27. Hamaker, E. L. (2012). Why researchers should think ‘within-person’: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 43–61). Guilford.
28. Kruijsaar, M. E., Barendregt, J., Vos, T., de Graaf, R., Spijker, J., & Andrews, G. (2005). Lifetime prevalence estimates of major depression: An indirect estimation method and a quantification of recall bias. *European Journal of Epidemiology*, 20(1), 103–111. <https://doi.org/10.1007/s10654-004-1009-0>
29. Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., & Harrington, H. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. *JAMA Network Open*, 3(4), e203221. <https://doi.org/10.1001/jamanetworkopen.2020.3221>
30. Lichtwarck-Aschoff, A., van Geert, P., Bosma, H., & Kunnen, S. (2008). Time and identity: A framework for research and theory formation. *Developmental Review*, 28(3), 370–400. <https://doi.org/10.1016/j.dr.2008.04.001>
31. van Loo, H. M., Bigdeli, T. B., Milaneschi, Y., Aggen, S. H., & Kendler, K. S. (2020). Data mining algorithm predicts a range of adverse outcomes in major depression. *Journal of Affective Disorders*, 276, 945–953. <https://doi.org/10.1016/j.jad.2020.07.098>
32. DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment

- Recommendations. A Demonstration. *PLOS ONE*, 9(1), e83875. <https://doi.org/10.1371/journal.pone.0083875>
33. Klein, N. S., Holtman, G. A., Bockting, C. L. H., Heymans, M. W., & Burger, H. (2018). Development and validation of a clinical prediction tool to estimate the individual risk of depressive relapse or recurrence in individuals with recurrent depression. *Journal of Psychiatric Research*, 104, 1-7. <https://doi.org/10.1016/j.jpsychires.2018.06.006>
 34. Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3), 404-413. <https://doi.org/10.1038/nn.4238>
 35. Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, 26(1), 10-15. <https://doi.org/10.1177/0963721416666518>
 36. Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, 1-7. <https://doi.org/10.1017/S0033291719002514>
 37. Riese, H., & Wichers, M. (2019). Comment on: Eronen MI (2019). The levels problem in psychopathology. *Psychological Medicine*, 1-2. <https://doi.org/10.1017/S0033291719003623>
 38. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(7), 1349-1360. <https://doi.org/10.1017/S0033291713001979>
 39. Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press. <https://press.princeton.edu/titles/8950.html>
 40. Haslbeck, J. M. B., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*. <https://doi.org/10.1080/00273171.2021.1896353>
 41. aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*, 32(6), 510-523. <https://doi.org/10.1016/j.cpr.2012.05.007>
 42. Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, 175(9), 526-536. <https://doi.org/10.1097/00005053-198709000-00004>
 43. Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life: The experience sampling method*. Datawyse / Maastricht University Press.
 44. Ebner-Priemer, U. W., Kubiak, T., & Pawlik, K. (2009). Ambulatory Assessment. *European Psychologist*, 14(2), 95-97. <https://doi.org/10.1027/1016-9040.14.2.95>
 45. Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4(1), 1-32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
 46. Devries, M. W. (1987). Investigating mental disorders in their natural settings. *Journal of Nervous and Mental Disease*, 175(9), 509-513. <https://doi.org/10.1097/00005053-198709000-00001>
 47. Myin-Germeyns, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine*, 39(9), 1533-1547. <https://doi.org/10.1017/S0033291708004947>
 48. Mokros, H. B. (1993). Communication and Psychiatric Diagnosis: Tales of Depressive Moods From Two Contexts. *Health Communication*, 5(2), 113-127. https://doi.org/10.1207/s15327027hc0502_4
 49. Nelson, B., & Hartmann, J. A. (2018). Prediction in mental health research and its limits (or why life can only be understood backwards but must be lived forwards). *Early Intervention in Psychiatry*, 12(5), 767-770. <https://doi.org/10.1111/eip.12530>

50. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>
51. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
52. Carpenter, S. R., & Brock, W. A. (2006). Rising variance: A leading indicator of ecological transition. *Ecology Letters*, 9(3), 311–318. <https://doi.org/10.1111/j.1461-0248.2005.00877.x>
53. Dakos, V., van Nes, E. H., Donangelo, R., Fort, H., & Scheffer, M. (2010). Spatial correlation as leading indicator of catastrophic shifts. *Theoretical Ecology*, 3(3), 163–174. <https://doi.org/10.1007/s12080-009-0060-6>
54. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59. <https://doi.org/10.1038/nature08227>
55. Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I. A., Levin, S. A., & Van Nes, E. H. (2012). Anticipating critical transitions. *Science*, 338(6105), 344–348. <https://doi.org/10.1126/science.1225244>
56. Van Nes, E. H., & Scheffer, M. (2005). Implications of spatial heterogeneity for catastrophic regime shifts in ecosystems. *Source: Ecology Ecology*, 868723(867), 1797–1807.
57. Lenton, T. M., Livina, V. N., Dakos, V., Van Nes, E. H., & Scheffer, M. (2012). Early warning of climate tipping points from critical slowing down: Comparing methods to improve robustness. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1962), 1185–1204. <https://doi.org/10.1098/rsta.2011.0304>
58. Wen, H., Ciamarra, M. P., & Cheong, S. A. (2018). How one might miss early warning signals of critical transitions in time series data: A systematic study of two major currency pairs. *PLoS One*, 13(3), e0191439. <https://doi.org/10.1371/journal.pone.0191439>
59. Van Der Maas, H. L. J., & Hopkins, B. (1998). Developmental transitions: So what's new? *British Journal of Developmental Psychology*, 16(1), 1–13. <https://doi.org/10.1111/j.2044-835x.1998.tb00747.x>
60. Van Der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review*, 99(3), 395–417. <https://doi.org/10.1037/0033-295X.99.3.395>
61. Brett, T., Ajelli, M., Liu, Q. H., Krauland, M. G., Grefenstette, J. J., Van Panhuis, W. G., Vespignani, A., Drake, J. M., & Rohani, P. (2020). Detecting critical slowing down in highdimensional epidemiological systems. *PLoS Computational Biology*, 16(3). <https://doi.org/10.1371/journal.pcbi.1007679>
62. de Mooij, S. M. M., Blanken, T. F., Grasman, R. P. P., Ramautar, J. R., Van Someren, E. J. W., & van der Maas, H. L. J. (2020). Dynamics of sleep: Exploring critical transitions and early warning signals. *Computer Methods and Programs in Biomedicine*, 193, 105448. <https://doi.org/10.1016/j.cmpb.2020.105448>
63. Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13(6), 1132–1141. <https://doi.org/10.1037/a0033579>
64. Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, 141(4), 901–930. <https://doi.org/10.1037/a0038822>

65. Bos, F. M., Blaauw, F. J., Snippe, E., van der Krieke, L., de Jonge, P., & Wichers, M. (2018). Exploring the emotional dynamics of subclinically depressed individuals with and without anhedonia: An experience sampling study. *Journal of Affective Disorders*, 228, 186–193. <https://doi.org/10.1016/j.jad.2017.12.017>
66. Crowe, E., Daly, M., Delaney, L., Carroll, S., & Malone, K. M. (2018). The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in Major Depression. *Psychiatry Research*. <https://doi.org/10.1016/j.PSYCHRES.2018.02.032>
67. Sperry, S. H., Walsh, M. A., & Kwapil, T. R. (2020). Emotion dynamics concurrently and prospectively predict mood psychopathology. *Journal of Affective Disorders*, 261, 67–75. <https://doi.org/10.1016/j.jad.2019.09.076>
68. Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, 121(4), 819–829. <https://doi.org/10.1037/a0027978>
69. Curtiss, J., Fulford, D., Hofmann, S. G., & Gershon, A. (2019). Network dynamics of positive and negative affect in bipolar disorder. *Journal of Affective Disorders*, 249, 270–277. <https://doi.org/10.1016/j.jad.2019.02.017>
70. Eldesouky, L., Thompson, R. J., Oltmanns, T. F., & English, T. (2018). Affective instability predicts the course of depression in late middle-age and older adulthood. *Journal of Affective Disorders*, 239, 72–78. <https://doi.org/10.1016/j.JAD.2018.06.038>
71. Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, 12(2), 283–289. <https://doi.org/10.2337/db10-1371>
72. Kuranova, A., Booij, S. H., Menne-Lothmann, C., Decoster, J., Van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., Van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. *BMC Medicine*, 18(1), 1–11. <https://doi.org/10.1186/s12916-020-1500-9>
73. Husen, K., Rafaeli, E., Rubel, J. A., Bar-Kalifa, E., & Lutz, W. (2016). Daily affect dynamics predict early response in CBT: Feasibility and predictive validity of EMA for outpatient psychotherapy. *Journal of Affective Disorders*, 206, 305–314. <https://doi.org/10.1016/j.JAD.2016.08.025>
74. Cabrieto, J., Tuerlinckx, F., Kuppens, P., Hunyadi, B., & Ceulemans, E. (2018). Testing for the presence of correlation changes in a multivariate time series: A permutation based approach. *Scientific Reports*, 8(769), 1–20. <https://doi.org/10.1038/s41598-017-19067-2>
75. Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H., & Scheffer, M. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One*, 7(7). <https://doi.org/10.1371/journal.pone.0041010>
76. Bos, E. H., & De Jonge, P. (2014). ‘Critical slowing down in depression’ is a great idea that still needs empirical proof. *Proceedings of the National Academy of Sciences*, 111(10), E878–E878. <https://doi.org/10.1073/pnas.1323672111>
77. Wichers, M., Schreuder, M. J., Goekoop, R., & Groen, R. N. (2019). Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology. *Psychological Medicine*, 49(3), 380–387. <https://doi.org/10.1017/S0033291718002064>
78. Hayes, A. M., & Yasinski, C. (2015). Pattern destabilization and emotional processing in cognitive therapy for personality disorders. *Frontiers in Psychology*, 6(FEB), 107. <https://doi.org/10.3389/fpsyg.2015.00107>

79. Bryan, C. J., & Rudd, M. D. (2018). Nonlinear change processes during psychotherapy characterize patients who have made multiple suicide attempts. *Suicide and Life-Threatening Behavior*, 48(4), 386–400. <https://doi.org/10.1111/sltb.12361>
80. Kelso, J. A. S., Ding, M., & Schöner, G. (1993). *Dynamic pattern formation: A primer* (p. 50). The MIT Press.
81. Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, 234(4), 65–83.
82. Dutilh, G., Wagenmakers, E. J., Visser, I., & Van Der Maas, H. L. J. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, 35(2), 211–250. <https://doi.org/10.1111/j.1551-6709.2010.01147.x>
83. Litzow, M. A., & Hunsicker, M. E. (2016). Early warning signals, nonlinearity, and signs of hysteresis in real ecosystems. *Ecosphere*, 7(12). <https://doi.org/10.1002/ecs2.1614>
84. Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, 1(2), 138–155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>
85. Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2015). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, 27(2), 478–488. <https://doi.org/10.1037/pas0000060>
86. Hayes, A. M., Feldman, G. C., Beevers, C. G., Laurenceau, J. P., Cardaciotto, L. A., & Lewis-Smith, J. (2007). Discontinuities and cognitive changes in an exposure-based cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 75(3), 409–421. <https://doi.org/10.1037/0022-006X.75.3.409>
87. Yasinski, C., Hayes, A. M., Ready, C. B., Abel, A., Görg, N., & Kuyken, W. (2019). Processes of change in cognitive behavioral therapy for treatment-resistant depression: Psychological flexibility, rumination, avoidance, and emotional processing. *Psychotherapy Research*, 1–15. <https://doi.org/10.1080/10503307.2019.1699972>
88. Laurenceau, J.-P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, 27(6), 682–695. <https://doi.org/10.1016/j.cpr.2007.01.007>
89. Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, 118(1–3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>
90. Schiepek, G., Heinzel, S., Karch, S., Plöderl, M., & Strunk, G. (2016). Synergetics in psychology: Patterns and pattern transitions in human change processes. In G. Wunner & A. Pelster (Eds.), *Selforganization in Complex Systems: The Past, Present, and Future of Synergetics* (pp. 181–208). Springer International Publishing. http://link.springer.com/10.1007/978-3-319-27635-9_12
91. Schiepek, G., Viol, K., Aichhorn, W., Hütt, M.-T., Sungler, K., Pincus, D., & Schöller, H. J. (2017). Psychotherapy is chaotic – (Not only) in a computational world. *Frontiers in Psychology*, 8, 379. <https://doi.org/10.3389/fpsyg.2017.00379>
92. Dunn, T. W., Vittengl, J. R., Clark, L. A., Carmody, T., Thase, M. E., & Jarrett, R. B. (2012). Change in psychosocial functioning and depressive symptoms during acute-phase cognitive therapy for depression. *Psychological Medicine*, 42(2), 317–326. <https://doi.org/10.1017/S0033291711001279>
93. Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, 27(6), 715–723. <https://doi.org/10.1016/j.cpr.2007.01.008>

94. Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
95. Keller, M. B. (2003). Past, present, and future directions for defining optimal treatment outcome in depression. *JAMA*, 289(23), 3152–3160. <https://doi.org/10.1001/jama.289.23.3152>
96. Rabin, A. S., Kaslow, N. J., & Rehm, L. P. (1984). Changes in symptoms of depression during the course of therapy. *Cognitive Therapy and Research*, 8(5), 479–487. <https://doi.org/10.1007/BF01173285>
97. Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, 5(4), 439–458. <https://doi.org/10.1111/j.1468-2850.1998.tb00166.x>
98. Globus, G. C., & Arpaia, J. P. (1994). Psychiatry and the new dynamics. *Biological Psychiatry*, 35(5), 352–364. [https://doi.org/10.1016/0006-3223\(94\)90039-6](https://doi.org/10.1016/0006-3223(94)90039-6)
99. Strunk, G., & Lichtwarck-Aschoff, A. (2019). Therapeutic chaos. *Journal for Person-Oriented Research*, 5(2), 81–100. <https://doi.org/10.17505/jpor.2019.08>
100. Heinzel, S., Tominschek, I., & Schiepek, G. (2014). Dynamic patterns in psychotherapy—Discontinuous changes and critical instabilities during the treatment of obsessive compulsive disorder. *Nonlinear Dynamics, Psychology, and Life Sciences*, 18(2), 155–176.
101. Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(1), 93–101. <https://doi.org/10.1037/a0026455>
102. Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444–447. <https://doi.org/10.1037/0022-006X.70.2.444>
103. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
104. Barkham, M., Stiles, W. B., & Shapiro, D. A. (1993). The shape of change in psychotherapy: Longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology*, 61(4), 667–677. <https://doi.org/10.1037/0022-006X.61.4.667>
105. Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75(6), 864–874. <https://doi.org/10.1037/0022-006X.75.6.864>
106. Tadić, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., & Hegerl, U. (2010). Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders*, 120(1–3), 86–93. <https://doi.org/10.1016/j.jad.2009.04.014>
107. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2013). Nomothetic and idiographic symptom change trajectories in acute-phase cognitive therapy for recurrent depression. *Journal of Consulting and Clinical Psychology*, 81(4), 615–626. <https://doi.org/10.1037/a0032879>
108. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2016). Defined symptom-change trajectories during acute-phase cognitive therapy for depression predict better longitudinal outcomes. *Behaviour Research and Therapy*, 87, 48–57. <https://doi.org/10.1016/j.brat.2016.08.008>
109. Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive-behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology*, 84(8), 726–736. <https://doi.org/10.1037/ccp0000101>

110. Greenfield, M. F., Gunthert, K. C., & Haaga, D. A. F. (2011). Sudden gains versus gradual gains in a psychotherapy training clinic. *Journal of Clinical Psychology*, 67(1), 17–30. <https://doi.org/10.1002/jclp.20748>
111. Hardy, G. E., Stiles, W. B., Cahill, J., Ispan, C., Macaskill, N., & Barkham, M. (2005). Sudden gains in cognitive therapy for depression: A replication and extension. *Journal of Consulting and Clinical Psychology*, 73(1), 59–67. <https://doi.org/10.1037/0022-006X.73.1.59>
112. Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy*, 43(6), 703–714. <https://doi.org/10.1016/j.brat.2004.06.002>
113. Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasch, C., Mocanu, S., Vocks, S., Schulte, D., & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research*, 23(1), 14–24. <https://doi.org/10.1080/10503307.2012.693837>
114. Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., Iveson, M., & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology*, 71(1), 14–21. <https://doi.org/10.1037/0022-006X.71.1.14>
115. Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of sudden gains in acute phase treatment of depression. *Journal of Consulting and Clinical Psychology*, 73(1), 173–182. <https://doi.org/10.1037/0022-006X.73.1.173>
116. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2015). Detecting sudden gains during treatment of major depressive disorder: Cautions from a Monte Carlo analysis. *Current Psychiatry Reviews*, 11(1), 19–31.
117. Warren, R., Burrow, J., Conroy, D., Lukela, J., & Kahn, D. A. (2014). ‘I Didn’t know cognitive therapy was deep’: A case study of sudden and lasting gains in cognitive-supportive therapy of depression. *Journal of Psychiatric Practice*, 20(5), 379–388. <https://doi.org/10.1097/01.pra.0000454785.97976.4d>

2

Sudden gains in day-to-day change: revealing nonlinear patterns of individual improvement in depression

Published as:

Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, 88(2), 119-127. <https://doi.org/10.1037/ccp0000469>

Abstract

Objective: We examined individual overall trajectories of change and the occurrence of sudden gains in daily self-rated problem severity, and the relation of these patterns to treatment response.

Method: Mood disorder patients ($N = 329$, mean age = 44, 55% women) completed daily self-ratings about the severity of their complaints as a standard part of treatment, using the Therapy Process Questionnaire (TPQ). Per individual, the best-fitting defined (linear, log-linear, one-step) trajectory was tested for significance: for change over time, and for specificity of the best-fitting trajectory. 203 cases had ICD-10 Symptom Rating (ISR) depression scores post-treatment: a score ≤ 1 identified 114 treatment responders. Relation to response was examined for sudden gains and type of change trajectory.

Results: 138 cases (42%) had a significant decrease in problem severity, of which 54 cases (16%) had a defined trajectory: 50 cases with one-step improvement, and 4 with a linear improvement in daily problem severity. Sudden gains occurred in 28% of the total sample, and within 58% of improvement patterns. Specifically, sudden gains occurred in 68% of significant one-step trajectories, and 25% of the linear cases. Sudden gains and non-specific change trajectories were significantly more frequent for treatment responders.

Conclusions: At the day-level, patterns of improvement are nonlinear for most patients. Sudden gains occur within various forms of overall change and are associated with treatment response. Clinically relevant improvements in depression occur both gradually and abruptly, and this finding allows for the possibility that the remission process functions according to dynamical systems principles.

Keywords: depression; treatment response; idiographic change pattern; sudden gains; daily assessments

Introduction

The course of depressive symptom change over time has been at the core of many psychotherapy studies aimed at gaining insight into how people get better and whether a particular pattern of improvement is indicative of better long-term outcomes [1–4]. Depression research generally holds the assumption that remission and recovery of symptoms is a linear process, based on group-level studies with pre- and post-measurements [5,6]. This assumed gradual improvement has been challenged by investigations of the therapy process at the individual level, showing that idiographic patterns of change can also be nonlinear [7–10]. A study using weekly symptom assessments revealed that a steady, linear shape of change was only present in about 20% of patients [11]. Around 30% of people showed a log-linear trajectory of change, with strong improvements at the beginning of therapy, followed by slower, less steep progress thenceforth [11–13].

Another frequently found pattern is a sudden, large, clinically relevant decrease in symptoms in the course of treatment. [14]. Sudden gains have been reported to occur for as many as 23% to 46% of patients in various types of psychological treatment, at various time points in the treatment process [8,15–18]. However, even when sudden gains occur, they do not always define the overall trajectory: they have been found within the context of an overall gradual course of change in a few studies [8,19,20], and the only study that examined individual shapes of change showed that a trajectory-defining mean-shift in symptoms (i.e., one-step change) was the best fitting model for only 16% of their sample [11].

Whereas traditional approaches to mental disorders cannot explain this wide variety of change patterns, a dynamical system conceptualisation can explain both the presence of gradual change patterns and the occurrence of abrupt shifts [21,22]. From this perspective, mental disorders are conceptualised as a complex system of interacting symptoms, behaviour, cognition and emotions, which is capable of taking on different dynamically stable states [22–25]. While change within and from such dynamically stable states often appears gradual, shifts between two states may be abrupt [21,26]. Psychopathology researchers have previously shown that mood systems exhibit generic ‘early warning signals’ that occur when a relevant change is imminent [22,27–31]. Just as a dynamical system becomes less stable after exposure to a large or a repeating stimulus, and is more likely to ‘tip over’ from one state to another [23,32–34], psychotherapy may be

an influence on the mood system that increases a patient's likelihood for change and brings them closer to a sudden transition toward improvement of symptoms [8,30,35]. Thus, one reason to examine how often sudden gains happen within the overall trajectory of change is that this may provide important clues about whether treatment response in depression can be viewed as a nonlinear, dynamical systems process, in which changes occur both gradually and abruptly.

A second reason to study individual patterns of change during treatment is that they may be associated with later mental health outcome. Patterns of early improvement during treatment have been linked to better long-term outcomes in several studies [13,36–38] and defined change trajectories (linear, log-linear or one-step overall patterns) have been found to have a long-term advantage over less orderly change trajectories [39]. Sudden gains have also generally been found to be indicative of a better degree of improvement [4,40]. Yet, it remains unclear to what extent the overall trajectory of improvement itself is predictive of outcome (cf. Ref. [39]), and to what extent sudden gains within these trajectories contribute to a stronger degree of person-specific improvement in depressed patients.

In order to differentiate the relative contributions to better outcomes of a sudden gain or the shape of an individual's overall trajectory, we require a more detailed measurement of change during treatment. The way a change pattern looks over time is highly dependent on the measurement frequency [15,41] and the few studies that examined individual change patterns during treatment used weekly assessments. Although sudden gains theoretically occur between adjacent therapy sessions, without a finer-grained study using daily assessments we remain blind to the momentum, magnitude, and stability of the changes in the days between therapy sessions. Therefore, we examine the individual courses of change in perceived problem severity with naturalistic, daily data gathered from a clinical sample with mood disorders during their treatment for depression.

The aims of the current study are: 1) to examine the relative frequency of defined (linear, log-linear and one-step) overall trajectories of improvement in problem severity scores of patients with mood disorders during therapy; 2) to determine how often sudden gains occur within these trajectories; and 3) to examine whether response to therapy relates to a) a defined trajectory of overall change, b) having a sudden gain, and c) the combination of a specific trajectory of change and a gain.

Method

Sample and procedures

The dataset was derived from four clinics in Austria and Germany between June 2008 and August 2014. Therapists used the internet-based Synergetic Navigation System (SNS) to monitor the therapy process in real-time through a daily questionnaire [42] with the aim to optimize treatment [22,43]. As a part of care as usual, the sample of 329 patients with an ICD-10 diagnosis of mood disorder [44] filled out questionnaires each evening starting from the first day of treatment. Ethical approval for the application of the SNS for patient monitoring and the usage of the retrieved data was given by the ethical committee of the state of Salzburg, and all patients signed an informed consent confirming that their anonymised data could be used for empirical purposes and scientific publication [41].

Treatment

The intensive daily treatment program in all clinics consisted of multiple integrative components, including individual and group therapy, mostly cognitive behavioural therapy, physiotherapy, psychomotor therapy, psychoeducation, and creative therapy. Most people were at the clinic as inpatients, though some went home in evenings or weekends. Standard duration of treatment was one to three months, with the potential to be extended by one additional month.

Materials

Therapy Process Questionnaire

Patients completed daily ratings on the 47 items of the Therapy Process Questionnaire (TPQ) [42]. The TPQ is divided into a five-factor structure, of which we focused on factor II: Problem Severity [42]. The Problem Severity factor is comprised of five items on which the current degree of hindrance due to complaints and symptoms is rated by the patient. For example, “*Today I felt helpless and at the mercy of my problems*”, and “*Today my problems affected my daily life*”, with the response scale ranging from 0 (not at all) to 6 (very much).

ISR-depression scale

Patients completed the ICD-10 Symptom Rating (ISR) at the start and end of therapy to assess the extent to which they suffered from specific symptoms in the past two weeks [45]. The depression scale assesses four constructs: ‘depressed mood’, ‘lack of joy’, ‘lack of energy’ and ‘low self-esteem’, each rated on a five-point Likert scale, ranging from 0 (not applicable) to 4 (extremely applicable). The instrument has good internal validity, is sensitive to change in depression and has shown convergent validity with other instruments [46–48].

Analysis

Pre-processing steps

To examine the period over which most patients received treatment (i.e., between one and three months) any observations after 100 days were dropped. Missing observations were deleted list-wise.

Overall trajectories of change

To determine which theoretical overall change trajectory best typified the pattern of daily problem severity ratings over time for each individual, we modelled the following defined trajectories: a) Linear change — gradual improvement, b) log-linear change — early, fast improvement that levels out with time, c) a one-step change — modelled as a shift in means, d) null-model — no change over time, for comparison (see also Refs. [11,39]). Per individual, problem severity factor scores were regressed on a) a linear function of time (1, 2, 3..., n), b) a log-linear function of time ($\ln(1, 2, 3..., n)$), c) a one-step model of time (time until the largest shift = 1, time after shift = 0), d) no change over time (intercept-only).

To define the point of the ‘step’ in the one-step model, we used the `e.divisive` function (`ecp` package; [49]) to detect the single largest change point for each individual, by specifying $k = 1$. For this, and the sudden gains analysis (next section), the following settings were kept consistent: the `alpha`-argument was set to the default of 1, so that any distributional change contributed to the detection of a relevant change point. The significance level was set to $p < .01$, and the maximum number of random permutations to $R = 4999$. The number of observations

between potential change points was set to seven days: `min.size = 7`. This ensures that a change point can only be detected after a period of at least one week, in accordance with previous studies (e.g., [14,17]).

To identify the best-fitting trajectory per individual (i.e., linear, log-linear, one-step change, and for comparison: a null-model), we used leave-one-out cross-validation (LOOCV) with the `caret` package [50]. This method iteratively uses $n - 1$ observations to predict one left-out observation for the specified model shape until it has tested all data points once. The absolute error reflects the difference between the omitted actual observation and the predicted observation over all iterations; the model with the lowest mean absolute error (MAE) describes the data best. Next, to examine how often the model with the lowest MAE represented significant change over time, we tested the best model against the null-model for all individuals. To further differentiate whether the significant change trajectory over time took a defined (linear, log-linear or one-step) shape, or was less specific, we tested the best model for each person against their second-best trajectory. Specifically, the comparisons consisted of permutation-based one-way tests of independence on the absolute error time series of the different models, paired on time point, using 500,000 Monte Carlo resampling iterations. This permutation-based method accounts for the small sample sizes, outliers and non-normal distributions (`coin` package; [51]).

Sudden gains

To formulate decision rules to test whether the largest shift in the distribution of problem severity scores met the criteria of a sudden gain, we used the definition of sudden gains from Tang & DeRubeis [14]: an improvement between two time points that is large in: 1) absolute magnitude, 2) relative magnitude, and 3) magnitude relative to symptom fluctuation.

To pinpoint the moment at which the largest shift in problem severity scores took place, we used the `e.divisive` function again. First, we did not restrict the number of possible change points ($k = \text{null}$) and selected only those cases that had at least one significant change point. Second, to keep only the largest shift in the distribution of scores for those cases, the analysis was run again, now specifying maximum one change point with $k = 1$.

To determine if the absolute magnitude of the identified shift was sufficiently large (criterion 1), we tested whether the difference between the mean of the week before the change point and the mean of the week after the change point was larger than or equal to an absolute value of 0.8. This value reflects the average within-person standard deviation of the problem severity scores of the first two weeks. We adjusted Tang and DeRubeis' [14] original first criterion (i.e., 1 *SD* on the Beck Depression Inventory (BDI); [52]) as our study used a different measurement instrument with no established clinical cut-off.

The relative magnitude of change (criterion 2) was tested identically to Tang and DeRubeis [14]: by calculating whether there was at least 25% relative difference in means between the seven days before and the seven days after the identified sudden transition.

The magnitude of the shift relative to symptom fluctuation (criterion 3) was tested as part of the change point analysis: the e.divisive function bisects the data and detects a point at which the relative difference in distributional characteristics (most prominently the means and variance) between the two sections is largest. This method might be less prone to detecting false positive sudden transitions than the original criteria (see Ref. [20]), as e.divisive explicitly tests whether the random fluctuations in scores are significantly different between distributions of scores before and after the change point using permutation testing procedures [49,53].

Relation to outcome

Response was defined as an absolute low symptom score on the ISR-depression scale of ≤ 1 post-treatment, indicating minimal depressive severity [48]. Scores were only available for part of the sample ($n = 203$), and splitting the group on the ≤ 1 cut-off resulted in 114 responders, and 89 non-responders.

We examined the association between treatment responder status (0 = non-responder, 1 = responder) and the frequency of patterns of overall change and sudden gains using χ^2 tests of independence with an α -level of .05. For the overall trajectories, we first examined whether responder status was related to having a defined trajectory of change over time (significant trajectory = 1, non-specific significant change = 0). Then, depending on the resulting group

sizes, we tested whether responder status was related to a certain trajectory group, (categorical variable: linear = 1, log-linear = 2, one-step = 3). Finally, the relative proportion of sudden gains (no gain = 0, gain = 1) in responders and non-responders was examined.

Results

Descriptives

The sample ($N = 329$) included 181 females and 148 males, ages between 18 and 69 years old ($M = 43.8$, $SD = 11$). All patients had an ICD-10 diagnosis of mood disorder [50–54]: Bipolar disorder ($n = 23$); Major depressive disorder (MDD) — single episode ($n = 149$); MDD — recurrent ($n = 155$); MDD — persistent ($n = 2$). On average, missing data was low: 3.05% ($SD = 3.94$). With a minimum of 28 and a maximum of 100 data points, the mean number of observations was 62.9 ($SD = 22.3$). Ratings on the problem severity scale had a mean of 3.03 ($SD = 1.14$) in the first seven days, and a mean of 2.43 ($SD = 1.33$) in the last seven days. The mean ISR score at intake was 2.27 ($SD = 1.05$), and 1.11 ($SD = 0.99$) post-treatment.

Frequency of change patterns

Overall trajectories

The best-fitting model was indicative of significant change over time (fitted better than the null-model) for 176 cases (53.5%). Of those, 84 individuals (25.5%) had a significant defined trajectory – linear, log-linear or one-step (i.e., the best model outperformed the second-best). Examples of individual trajectories of problem severity over time are illustrated in **Figure 2.1**. Focusing specifically on improvement during treatment, **Table 2.1** shows results for the 138 (78%) cases with significant decrease in problem severity over time (i.e., had a negative beta-coefficient). Of those cases, 54 (39%) had significant, defined change trajectories over time, which most often took the shape of a one-step pattern (50 cases; 93%), followed by a linear trajectory (4 cases; 7%). Log-linear trajectories never outperformed the second-best model.

Table 2.1

Change pattern frequencies for cases with a significant improvement in problem severity scores over time.

		Overall trajectories			Sudden gain		
		<i>n</i>	%	% of total	<i>n</i>	%	% of total
Significant improvement over time		138	100%	41.9%	81	58.7%	24.6%
Defined trajectory over time	Linear	4	2.9%	1.2%	1	0.7%	0.3%
	Log-linear	0	0%	0%	0	0%	0%
	One-step	50	36.2%	15.2%	34	24.6%	10.3%
	Total	54	39.1%	16.4%	35	25.4%	10.6%

Note. $N = 138$, total sample $N = 329$. Significant improvement: the best model fit better than the null-model at $p < .05$ and the beta-coefficient of the best model was negative. Defined trajectory: the best model fit better than the second-best at $p < .05$

Sudden gains

In the total sample of 329 participants, we identified 189 (57%) cases that met the criterion of a significant shift in the distribution of their problem severity scores relative to symptom fluctuations. The criterion of absolute size of the gain being 0.8 or larger between the week before and after the change point narrowed the sample to 146 cases (44%). Applying the criterion of a relative change of 25% mean difference between the weeks before and after the change point, yielded a final number of 93 (28%) sudden gains, and 41 (12%) sudden losses (i.e., problem severity increased). Results for the full sample ($N = 329$) are reported separately to allow proportions of change patterns to be compared to the broader existing literature, see **Table S2.1** in the **Supplementary materials**.

Occurrence of sudden gains within significant improvement trajectories

A sudden gain was identified for 81 (59%) of the 138 cases with significant change toward improvement. In 34 cases (68%) of the defined one-step change group a sudden gain occurred, in the linear group, 1 case (25%) had a sudden gain. Notably, 16 cases had a trajectory-defining one-step shift that did not meet sudden gain criteria.

Figure 2.1*Three different individual overall trajectories of improvement.*

Note. TPQ = Therapy Process Questionnaire.

Panels A and B: a defined change trajectory indicates that the best-fitting model fitted significantly better than the second-best model. **Panel C:** a non-specific change indicates that there was significant change over time (compared to the null-model), but that the best-fitting defined (here: one-step) trajectory could not be distinguished statistically from the second-best alternative (linear, in this case).

Associations between patterns of change and treatment response

Non-specific and defined trajectories

To examine whether having a defined trajectory of overall change was related to responder status, we compared the proportions of responders in the group with a significant trajectory of improvement over time ($n = 33$, responder $n = 14$) to the cases that had a significant improvement over time but not a defined trajectory ($n = 52$, responder $n = 41$). We found a significant association between responder status and (un)defined change trajectories, $\chi^2(1, N = 85) = 4.40, p = .036$. Specifically, responders were more likely to have a non-specific trajectory, and non-responders more frequently had a defined trajectory, than would be expected by chance.

Specific trajectories

Because the vast majority of cases was categorised as a one-step change, we could not test for associations between the different theoretical change trajectory groups and treatment response.

Sudden gains

Responders ($n = 114$; sudden gain $n = 42$) were found to have significantly more sudden gains relative to non-responders ($n = 89$; sudden gain $n = 17$), $\chi^2 (1, N = 203) = 7.63, p = .006$.

Note: we provide a table of the relative frequencies of the significant change trajectories and sudden gains and losses for the (non-)responder subsample in **Supplementary Table S2.2**.

Post hoc: relationship between one-step trajectories and sudden gains

Having found that both sudden gains and undefined change trajectories are related to treatment response, we considered it relevant to examine how often sudden gains shaped the overall trajectory: i.e., do they occur more frequently in one-step improvement? Given the dominance of one-step trajectories in the defined change trajectory group (50 out of 54 cases), we were able to extend our analyses with an additional chi-square test, to compare the proportions of sudden gains in the group with a significant one-step improvement over time ($n = 50$, sudden gain $n = 34$) to the group with non-specific significant change over time ($n = 84$, sudden gain $n = 46$). We found that having defined one-step trajectory was unrelated to having a sudden gain or not: $\chi^2 (1, N = 134) = 2.28, p = .131$. Thus, sudden gains need not be trajectory-defining, as they occurred equally often in one-step trajectories and other forms of overall improvement, where they were part of a larger gradual or nonlinear change pattern.

Discussion

In this study, we mapped different trajectories of improvement in depressive problem severity at a fine-grained daily level, which adds a new layer of detail to our knowledge of sudden gains and overall change processes in depression. We found that more than half of our sample followed non-specific improvement trajectories, and one-step trajectories were the most common defined trajectory. Our results

indicate that responders were more likely to have an undefined shape of overall change, and that sudden gains did not occur more frequently in defined one-step trajectories than in less specific improvement trajectories over time. Moreover, we replicated the finding that sudden gains are a frequent phenomenon and that they are predictive of treatment response.

A defined one-step overall trajectory was clearly most prevalent in our sample. This contrasts findings from studies on week-level individual change, which showed that log-linear shape of change was most prevalent, and one-step trajectories were the least frequent [5,11,39]. Group-level studies have most often found linear trajectories in depressive scores during therapy, and our results suggest that this pattern is not applicable to many individuals at a day-to-day level (see also Refs. [6,8,54]). In fact, 61% of patients with a significant improvement could not be categorised by a specific trajectory — this is more than the 34% undefined change that has been found at week-level [11,39]. Moreover, undefined trajectories of change were related to being a treatment responder, which is in direct contrast to the finding that defined change patterns are related to better treatment outcomes [39].

It could be that our results contrast those studies because daily measurements capture the fluctuating, dynamic nature of depressive problems and the variability and heterogeneity of therapeutic change over time more closely [55]. The theoretical trajectories of change employed in this study are derived from studies on weekly data, which typically have fewer observations per person (e.g., 8-20 treatment sessions in various studies on sudden gains, see Ref. [4]), making modelling more complex patterns over time statistically challenging for week-level studies, and possibly of lesser interest. Having more detail over the days creates its own challenge, as we found that the simple theoretical models fit our data less well, resulting in larger errors [56]. This may partly explain the relative dominance of the one-step model in the defined group: modelling two means over time allows for more statistical flexibility in finding an optimal fit (and lower errors), than the single line of the linear or log-linear models. The rich, daily data in this study thus shows that the explanatory model of linear improvement during therapy is inadequate for describing patterns of treatment response for most patients in our sample.

Sudden gains occurred frequently within significant improvement trajectories (59% of cases), irrespective of it having a defined one-step change or

undefined change. This shows that even when large jumps in symptom reduction occur, they need not define the overall course of change. A previous simulation study already showed that this finding is likely: our study now confirms the occurrence of sudden gains within heterogeneous, gradual overall improvement trajectories with empirical data [19]. The prevalence of sudden gains, and the individual variation in overall change during therapy can be taken as encouragement to look for new avenues of conceptualising depression and mood systems to account for nonlinearity and individuality.

Although not the only avenue of interest, having found these continuous and discontinuous changes means we should not overlook a complex dynamical systems explanation for understanding patterns of treatment response in depression [25,29]. Some empirical research has supported the idea that sudden changes in symptoms may be indications of a critical shift in the mood system, where a move toward a more adaptive state is reached after the positive influence of therapy [22,28,31,43,57]. Examining whether sudden gains in depression indicate tipping points is relevant, as dynamical systems theory may provide us with methods to anticipate changes in symptoms or depressive burden. By looking for the generic signs of imminent change that typify complex dynamical systems (e.g., increases in autocorrelation, variance, dynamic complexity, and connectivity), we may learn to anticipate and protect against changes toward maladaptive states, or encourage positive change when the system is particularly susceptible [29,31,58,59]. Clearly, both clinicians and patients would benefit from knowing when and how changes are taking place, and may even adapt the therapy accordingly [13,32,60,61]. However, these explanations for depressive remission remain tentative for now and require further testing.

Strengths of our study include the high level of detail gained from daily measurements, along with a large sample size and statistically conservative methods throughout – cross-validation and permutation tests were used in assessing overall change trajectory model fit and significance, and in the estimation of the optimal change point. Our analyses also accounted for individual differences in how treatment response in depression develops over time. By focusing on a group of patients who improved over the course of treatment, our findings give a closer description of the day-to-day patterns of improvement in depression. Clinically, this has allowed us to say that if a sudden gain occurs for a patient who

keeps taking steps towards improvement – no matter the exact trajectory – their outlook is more promising than for patients who improve without a sudden gain. We chose a different measurement instrument to study the presence of sudden gains (the TPQ rather than the BDI; [42,62]) than the original authors [14], which allowed us to get a daily pattern of experienced problem severity that serves as a more direct gauge of functioning and response to therapy than the presence of symptoms alone [54]. Furthermore, using change point analysis to automate testing of the third sudden gains criterion (relevant change relative to symptom fluctuation) made our method an objective and conservative way to observe the conditions of this debated criterion [39].

A limitation of our study is that we were unable to compare specific change trajectory groups among one another on treatment response as the group sizes were too unbalanced to test the differences. We also did not investigate the occurrence of more than one, or cascades of gains [15,22]. We were also limited by our use of the theoretical change trajectories from the weekly literature, as our results suggest that more complex, individualised models may be needed to capture the dynamic nature of change in depressive complaints in daily data. Finally, due to the observational design of this study, we are unable to identify the processes and determinants that underlie the change patterns we identified.

To conclude, this study examined individual trajectories of improvement during therapy with fine-grained daily ratings, and highlights the importance of looking beyond the existing linear theoretical explanations of how depression changes over time. The presence of gradual and abrupt shifts in problem severity can be taken as a first indication that the process of depressive remission may behave according to the principles of complex dynamical systems, in which nonlinear change is common, and this warrants further investigation. Clinically, this is important, as this study shows that nonlinear, variable patterns of change, including sudden gains, can be expected for many depressed patients during treatment and are associated with treatment response.

References

1. Kopta, S. M. (2003). The dose—Effect relationship in psychotherapy: A defining achievement for Dr. Kenneth Howard. *Journal of Clinical Psychology, 59*(7), 727–733. <https://doi.org/10.1002/jclp.10167>
2. Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist, 41*(2), 159–164. <https://doi.org/10.1037/0003-066X.41.2.159>
3. Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology, 58*(10), 1291–1304. <https://doi.org/10.1002/jclp.10070>
4. Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80*(1), 93–101. <https://doi.org/10.1037/a0026455>
5. Hayes, A. M., & Hayes, S. (2007). Towards an understanding of the process and mechanisms of change in cognitive behavioral therapy: Linking innovative methodology with fundamental questions. *Clinical Psychology Review, 27*(6), 679–681. <https://doi.org/10.1016/j.cpr.2007.01.006>
6. Laurenceau, J.-P., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review, 27*(6), 682–695. <https://doi.org/10.1016/j.cpr.2007.01.007>
7. Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., Petrovic, A., Zobel, A., Henigsberg, N., Rietschel, M., Aitchison, K. J., Farmer, A., & McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine, 40*(08), 1367–1377. <https://doi.org/10.1017/S0033291709991528>
8. Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review, 27*(6), 715–723. <https://doi.org/10.1016/j.cpr.2007.01.008>
9. Dunn, T. W., Vittengl, J. R., Clark, L. A., Carmody, T., Thase, M. E., & Jarrett, R. B. (2012). Change in psychosocial functioning and depressive symptoms during acute-phase cognitive therapy for depression. *Psychological Medicine, 42*(2), 317–326. <https://doi.org/10.1017/S0033291711001279>
10. Rabin, A. S., Kaslow, N. J., & Rehm, L. P. (1984). Changes in symptoms of depression during the course of therapy. *Cognitive Therapy and Research, 8*(5), 479–487. <https://doi.org/10.1007/BF01173285>
11. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2013). Nomothetic and idiographic symptom change trajectories in acute-phase cognitive therapy for recurrent depression. *Journal of Consulting and Clinical Psychology, 81*(4), 615–626. <https://doi.org/10.1037/a0032879>
12. Ilardi, S. S., & Craighead, W. E. (1994). The Role of Nonspecific Factors in Cognitive-Behavior Therapy for Depression. *Clinical Psychology: Science and Practice, 1*(2), 138–155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>
13. Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders, 118*(1–3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>
14. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
15. Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasch, C., Mocanu, S., Vocks, S., Schulte, D., & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14–24. <https://doi.org/10.1080/10503307.2012.693837>

16. Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy*, 43(6), 703–714. <https://doi.org/10.1016/j.brat.2004.06.002>
17. Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 73(1), 168–172. <https://doi.org/10.1037/0022-006X.73.1.168>
18. Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444–447. <https://doi.org/10.1037/0022-006X.70.2.444>
19. Thomas, C., & Persons, J. B. (2013). Sudden gains can occur in psychotherapy even when the pattern of change is gradual. *Clinical Psychology: Science and Practice*, 20(2), 127–142. <https://doi.org/10.1111/cpsp.12029>
20. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2015). Detecting sudden gains during treatment of major depressive disorder: cautions from a monte carlo analysis. *Current Psychiatry Reviews*, 11(1), 19–31. <https://doi.org/10.2174/1573400510666140929195441>
21. Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of Counseling Psychology*, 63(4), 379–395. <https://doi.org/10.1037/cou0000150>
22. Schiepek, G. (2009). Complexity and Nonlinear Dynamics in Psychotherapy. *European Review*, 17(02), 331. <https://doi.org/10.1017/S1062798709000763>
23. Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major Depression As a Complex Dynamic System. *PLoS One*, 11(12), [e0167490]. <https://doi.org/10.1371/journal.pone.0167490>
24. Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L. J., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: Bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry*, 15(1), 222. <https://doi.org/10.1186/s12888-015-0596-5>
25. Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive-behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology*, 84(8), 726–736. <https://doi.org/10.1037/ccp0000101>
26. Thelen, E., & Smith, L. B. (1994). *A Dynamic Systems Approach to Development and Action*. MIT Press.
27. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59. <https://doi.org/10.1038/nature08227>
28. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
29. Hayes, A. M., & Strauss, J. L. (1998). Dynamic systems theory as a paradigm for the study of change in psychotherapy: An application to cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 66(6), 939–947. <https://doi.org/10.1037/0022-006X.66.6.939>
30. Schiepek, G., Heinzel, S., Karch, S., Plöderl, M., & Strunk, G. (2016). Synergetics in Psychology: Patterns and Pattern Transitions in Human Change Processes. In G. Wunner & A. Pelster (Eds.), *Selforganization in Complex Systems: The Past, Present, and Future of Synergetics* (pp. 181–208). Springer International Publishing. https://doi.org/10.1007/978-3-319-27635-9_12

31. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical Slowing Down as a Personalized Early Warning Signal for Depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>
32. Schiepek, G., Tominschek, I., & Heinzel, S. (2014). Self-organization in psychotherapy: Testing the synergetic model of change processes. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01089>
33. Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press.
34. Scheffer, M. (2010). Foreseeing tipping points. *Nature*, 467(7314), 411–412. <https://doi.org/10.1038/467411a>
35. Haken, H. (1992). Synergetics in psychology. In W. Tschacher, G. Schiepek, & E. J. Brunner (Eds.), *Self-organization and clinical psychology. Empirical approaches to synergetics in psychology* (pp. 32–54). Springer. <https://doi.org/10.1007/978-3-642-77534-5>
36. Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2015). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, 27(2), 478–488. <https://doi.org/10.1037/pas0000060>
37. Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of Early Change in Psychotherapy Under Routine Outpatient Conditions. *Journal of Consulting and Clinical Psychology*, 75(6), 864–674. <https://doi.org/10.1037/0022-006X.75.6.864>
38. Tadić, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., & Hegerl, U. (2010). Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders*, 120(1–3), 86–93. <https://doi.org/10.1016/j.jad.2009.04.014>
39. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2016). Defined symptom-change trajectories during acute-phase cognitive therapy for depression predict better longitudinal outcomes. *Behaviour Research and Therapy*, 87, 48–57. <https://doi.org/10.1016/j.brat.2016.08.008>
40. Greenfield, M. F., Gunthert, K. C., & Haaga, D. A. F. (2011). Sudden gains versus gradual gains in a psychotherapy training clinic. *Journal of Clinical Psychology*, 67(1), 17–30. <https://doi.org/10.1002/jclp.20748>
41. Schiepek, G., Aichhorn, W., Gruber, M., Strunk, G., Bachler, E., & Aas, B. (2016). Real-Time Monitoring of Psychotherapeutic Processes: Concept and Compliance. *Frontiers in Psychology*, 7, 604. <https://doi.org/10.3389/fpsyg.2016.00604>
42. Schiepek, G., Aichhorn, W., & Strunk, G. (2012). Der Therapie-Prozessbogen (TPB) – Faktorenstruktur und psychometrische Daten. *Zeitschrift Für Psychosomatische Medizin Und Psychotherapie*, 58(3), 257–266. <https://doi.org/10.13109/zptm.2012.58.3.257>
43. Schiepek, G., Aichhorn, W., & Schöller, H. (2018). Monitoring change dynamics – a nonlinear approach to psychotherapy feedback. *Chaos and Complexity Letters*, 11(3), 355–375.
44. World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization.
45. Tritt, K., Von Heymann, F., Zaudig, M., Zacharias, I., Söllner, W., & Loew, T. (2008). Development of the 'ICD-10-Symptom-Rating' (ISR) questionnaire. *Zeitschrift Für Psychosomatische Medizin Und Psychotherapie*, 54(4), 409–418. <https://doi.org/10.13109/zptm.2008.54.4.409>
46. Fischer, H. F., Tritt, K., Klapp, B. F., & Fliege, H. (2011). How to compare scores from different depression scales: Equating the Patient Health Questionnaire (PHQ) and the ICD-10-Symptom Rating (ISR) using Item Response Theory. *International Journal of Methods in Psychiatric Research*, 20(4), 203–214. <https://doi.org/10.1002/mp.350>

47. Fischer, H. F., Tritt, K., Klapp, B. F., & Fliege, H. (2010). Faktorstruktur und psychometrische Eigenschaften des ICD-10 Symptom-Rating (ISR) an Stichproben psychosomatischer Patienten. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie*, 60(8), 307–315. <https://doi.org/10.1055/s-0029-1214419>
48. Brandt, W. A., Loew, T., Von Heymann, F., Stadtmüller, G., Tischinger, M., Strom, F., Molfenter, J., Georgi, A., & Tritt, K. (2015). How does the ICD-10 symptom rating (ISR) with four items assess depression compared to the BDI-II? A validation study. *Journal of Affective Disorders*. <https://doi.org/10.1016/j.jad.2014.10.070>
49. James, N. A., & Matteson, D. S. (2014). ecp: An R Package for Nonparametric Multiple ChangePoint Analysis of Multivariate Data. *Journal of Statistical Software*, 62(7), 1–25.
50. Kuhn, M. (2008). The caret Package: Classification and Regression Training. *Journal of Statistical Software*, 28(5), 1–26.
51. Hothorn, T., Hornik, K., van de Wiel, M., & Zeileis, A. (2008). Implementing a Class of Permutation Tests: The {coin} Package. *Journal of Statistical Software*, 28(8), 1–23.
52. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4(6), 561–571.
53. Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods*, 49(3), 988–1005. <https://doi.org/10.3758/s13428-016-0754-9>
54. Barkham, M., Stiles, W. B., & Shapiro, D. A. (1993). The shape of change in psychotherapy: Longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology*, 61(4), 667–677. <https://doi.org/10.1037/0022-006X.61.4.667>
55. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(07), 1349–1360. <https://doi.org/10.1017/S0033291713001979>
56. Delignières, D., Fortes, M., & Ninot, C. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences*, 8(4), 479–510.
57. Schiepek, G., Viol, K., Aichhorn, W., Hütt, M.-T., Sungler, K., Pincus, D., & Schöller, H. J. (2017). Psychotherapy Is Chaotic – (Not Only) in a Computational World. *Frontiers in Psychology*, 8, 379. <https://doi.org/10.3389/fpsyg.2017.00379>
58. Hayes, A. M., Yasinski, C., Barnes, J. Ben, & Bockting, C. L. H. (2015). Network destabilization and transition in depression: New methods for studying the dynamics of therapeutic change. *Clinical Psychology Review*, 41, 27–39. <https://doi.org/10.1016/j.cpr.2015.06.007>
59. Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2019). Critical Fluctuations as an Early-Warning Signal for Sudden Gains and Losses in Patients Receiving Psychotherapy for Mood Disorders. *Clinical Psychological Science*, 216770261986596. <https://doi.org/10.1177/2167702619865969>
60. Schiepek, G., & Tschacher, W. (1992). *Application of Synergetics to Clinical Psychology* (pp. 3–31). Springer. https://doi.org/10.1007/978-3-642-77534-5_1
61. Krause, M. S., & Lutz, W. (2009). Process Transforms Inputs to Determine Outcomes: Therapists Are Responsible for Managing Process. *Clinical Psychology: Science and Practice*, 16(1), 73–81. <https://doi.org/10.1111/j.1468-2850.2009.01146.x>
62. Beck, A. T., Steer, R. A., & Carbin, M. C. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77–100.

Supplementary material

Table S2.1

Frequencies of significant change over time and sudden gains and losses in the full sample.

		Change trajectory		Sudden gain		Sudden loss	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Significant change over time		176	53.5%	81	24.6%	28	8.5%
Defined trajectory over time	Null-model	3	0.9%	0	0%	0	0%
	Linear	4	1.2%	1	0.3%	0	0%
	Log-linear	0	0%	0	0%	0	0%
	One-step	77	23.4%	34	10.3%	18	5.5%
Total		84	25.5%	35	10.6%	18	5.5%

Note. Percentages represent the proportion of the total sample ($N = 329$). Significant change: the best model fitted better than the null-model at $p < .05$. Defined trajectory: the best model fitted better than the second-best at $p < .05$.

Table S2.2

Frequencies of significant change and defined trajectories over time and sudden gains and losses, for the 203 patients that could be defined as responders and non-responders.

	Change trajectory						Sudden gain						Sudden loss						
	Overall			non-responders			Overall			non-responders			Overall			non-responders			
	n	%		n	%		n	%		n	%		n	%		n	%		
Total	203	100%	114	56.2%	89	43.8%	59	29.1%	42	20.7%	17	8.4%	27	13.3%	12	5.9%	15	7.4%	
Significant change over time	108	53.2%	69	34%	39	19.2%	54	26.6%	38	18.7%	16	7.9%	18	8.9%	6	3%	12	5.9%	
Defined trajectory over time	Null-model	2	1%	1	0.5%	1	0.5%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
	Linear	1	0.5%	0	0%	1	0.5%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
	Log-linear	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
	One-step	48	23.6%	27	13.3%	21	10.3%	24	11.8%	15	7.4%	9	4.4%	11	5.4%	5	2.5%	6	3%
	Total	51	25.1%	28	13.8%	23	11.3%	24	11.8%	15	7.4%	9	4.4%	11	5.4%	5	2.5%	6	3%

Note. Response was defined as a score of ≤ 1 on the ICD-10 Symptom Rating depression scale post-treatment. Percentages represent the proportion of the cases with a post-treatment score ($N = 203$). Significant change: the best model fitted better than the null-model at $p < .05$. Defined trajectory: the best model fitted better than the second-best at $p < .05$.

3

Daily dynamics of negative affect: indicators of rate of response to treatment and remission from depression?

Based on:

Helmich, M.A., Wichers, M., Peeters, F., & Snippe, E. (Under review). Daily dynamics of negative affect: indicators of rate of response to treatment and remission from depression? *Manuscript under review*. <https://doi.org/10.31234/osf.io/hv65n>

Abstract

Background: More instability (MSSD) and variability (SD) of negative affect (NA) have been related to current and future depressive symptoms. We investigated whether MSSD and SD of NA were predictive of the rate of symptom change during treatment and of reaching remission status during versus post-treatment/not remitting, beyond mean-levels of NA.

Method: Forty-six participants with a primary diagnosis of major depressive disorder (assessed with the Structured Clinical Interview for DSM-IV) completed six days of experience sampling, with ten semi-random beeps per day before starting a combination of pharmacotherapy and supportive therapy. During and after treatment, the Hamilton Depression Rating Scale (HDRS) diagnostic interview was performed monthly for 18 months. Rate of depressive symptom change was modelled as a linear decrease in HDRS scores in a multilevel model, predicted by the mean, SD and MSSD of NA, as well as relevant baseline predictors. Logistic regression was used to predict remission status during versus post-treatment/no remission (HDRS of ≤ 7 within or after five months) with the same variables.

Results: The mean of NA, but not the SD and MSSD of NA predicted the rate of depressive symptom reduction over five months. The odds of being remitting during treatment were not associated with any of the predictors.

Conclusions: Our results suggest that assessing NA instability and variability before the start of treatment for depression may not give an indication of the treatment response over time or remission status. Clinically, the mean of NA may be promising as an indicator of response potential before treatment onset.

Keywords: depression; treatment response; negative affect dynamics; remission; ecological momentary assessment

Introduction

The likelihood that a patient responds to treatment for depression is hard to predict. The process of remission is heterogeneous and even those patients that respond during therapy may not maintain the symptom improvements over the long term [1]. A more unfavorable course of depression is often characterized by symptom instability and recurrences of symptoms [2]. Because such an unstable course is debilitating, it is important to identify predictors of treatment response and change in depressive symptoms at an early stage.

Drawing on dynamical systems theory, potential candidates for determining the likelihood of treatment response and long-term presence of depressive symptoms may be found in the dynamics of moment-to-moment negative affect (NA). In a complex dynamical system, large state changes are often preceded by destabilization in the underlying components of the system [3–5]. Analogously, a large ‘state change’ like treatment response could therefore show early signs of destabilization in the experiences underlying depressive symptoms, like NA (e.g., “I feel down”) within the course of daily life [6,7]. Moreover, from a clinical point of view, the destabilization may indicate a depressed patient’s ‘readiness for change’ [8], and therapy may further contribute to a positive direction of the change [2,9–11].

Researchers have posited that a less stable system may show more *variability* and *instability* in NA scores [12]. A higher standard deviation (SD) of NA scores reflects more variability and a higher mean square of successive differences (MSSD) of NA scores captures instability of NA or how strongly NA changes from one moment to the next. Specifically, researchers have been using ecological momentary assessments (EMA) to calculate instability and variability in NA using these representative time series of people’s NA throughout the day, over multiple days [13]. Based on such measures, it could be hypothesized that higher variability and instability in EMA assessments of NA at the start of treatment would predict more favorable treatment outcomes.

However, research suggests that more instability and variability of NA scores may not be favorable indicators of the (current) state of a patient. Cross-sectional studies have shown NA variability to be related to stronger depressive symptoms [14] and borderline personality disorder [15,16], and affective instability to be

predictive of depressive symptoms up to five years later, as well as the likelihood of having lifetime major depressive episodes [17]. Moreover, when measured with EMA in the context of daily life, depressed individuals appear to have higher levels of variability and instability in NA than healthy individuals [18,19], and these indices of NA dynamics have been found to be related to current depression and psychological well-being [20–22]. Thus, instability and variability of NA scores may as well reflect mechanisms underlying psychopathology [12,23,24] instead of being indicators of ‘readiness for change’ and may therefore as well predict less favorable treatment outcomes.

So far, only two studies have examined how EMA-measured NA dynamics relate to depressive symptom improvement during treatment. First, Husen *et al.* [25] studied a sample of 39 outpatients with mixed primary diagnoses of affective and anxiety disorder, who were set to receive cognitive behavioral therapy (CBT). On average 2.3 months before treatment onset, patients completed a two-week EMA-period of four measurements a day, to capture real-time affective states (including NA). They examined the relationship between the steepness of the decrease in symptoms over five weeks of weekly therapy sessions, and the mean and MSSD of NA, while controlling for baseline symptom severity. They found that lower NA instability was associated with a stronger decrease in symptoms over time. While the mean of NA was predictive of treatment response in a single-predictor model, this effect turned non-significant in their combined multi-predictor model.

Second, Bosley *et al.* [26] studied a sample of 32 patients with mood and anxiety disorders, who completed four daily assessments of NA for 30 days before starting CBT treatment. They found that the MSSD and SD of NA were unrelated to pre-post treatment depression symptom change, although the MSSD of NA did predict a decrease in anxiety symptoms in response to treatment beyond mean-levels of NA. The evidence regarding the role of NA dynamics predicting the course of depression treatment is thus inconclusive, and no study has yet investigated whether NA dynamics captured in daily life can inform us on long-term remission likelihood in depression.

Furthermore, more evidence is needed about the added value of NA instability and variability above and beyond the mean of NA in predicting treatment response in depression. For instance, Funkhouser *et al.* [27] found mean NA to be related to previous MDD episodes, while instability of NA was unrelated to the

MDD risk factors in their study. Bos *et al.* [28] examined the relationship between NA dynamic measures and depression and found that when including both the SD and MSSD in their model, only the SD remained a significant predictor, but this, in turn, became non-significant after adjusting for mean NA levels. Similarly, a meta-analysis of EMA-studies by Dejonckheere *et al.* [29] showed that most complex measures of affect dynamics had little added value over the mean of NA in the prediction of life satisfaction and depressive symptoms. These studies underscore the importance of accounting for statistical overlap between NA measures, which is why studying both variability and instability in the same sample is relevant and investigating their explanatory power beyond mean-levels of NA particularly so. Therefore, the current study aims to investigate the predictive value not only of instability and variability of NA, but also of mean NA levels.

If dynamic indices or mean levels of NA can be informative of a patient's potential for depressive symptom improvement, monitoring their affect before and during treatment could be of great clinical importance. Our dataset uniquely combines an EMA baseline period immediately before treatment onset with eighteen monthly clinical interviews that allow us to assess both the change in symptoms during five months of treatment, as well as long-term remission incidence. Therefore, the aims of this study are 1) to test whether within-person instability (MSSD), variability (SD) and the mean of NA are predictive of the rate of depressive symptom change during five months of treatment for depression, and 2) to test whether these NA measures are predictive of remission status over eighteen months. Specifically, we examine whether NA dynamics are predictive of the rate of depressive symptom change and remission status above and beyond the mean-level of NA and baseline symptoms and patient characteristics.

Method

Participants

Forty-seven participants with a primary diagnosis of Major Depressive Disorder (MDD) on the Structured Clinical Interview for DSM-IV (SCID) [30] as assessed by a research psychiatrist (F.P.), were recruited among individuals seeking treatment at a university-affiliated mental health center or the outpatient department of a psychiatric hospital in Maastricht, the Netherlands. Inclusion criteria were: 18–65 years of age, at least moderate depressive symptoms as indicated by a score of ≥ 18 on the Hamilton Depression Rating Scale (HDRS) [31], and no antidepressant treatment at the study start – although the use of low-dose anxiolytic drugs was permitted ($n = 8$). Exclusion criteria were current substance abuse or psychotic symptoms and insufficient command of the Dutch language. Participants were also required to have a minimum of 30% valid (i.e., delay in answering ≤ 25 minutes) EMA observations [32]. This led to one exclusion, and a final sample of 46 participants. Participants gave written informed consent, and received \$30 after completing the EMA period. The study design was approved by the Medical Ethics Committee of the academic hospital (METC azM/UM).

Procedure

At baseline, before treatment onset, repeated self-reports on participants' momentary affect, activities and contexts in daily life were gathered by EMA during a period of six consecutive days. Ten times a day, between 7:30 a.m. and 10:30 p.m., at semi-random intervals of circa 90 minutes, a beep from a digital wristwatch prompted participants to fill in the pencil-and-paper self-report form [33]. Immediately after the EMA period, all participants entered a naturalistic treatment phase. The HDRS was used to measure depressive symptom severity at baseline, and every following month over the course of 18 months.

Treatment

All patients received a combination of pharmacotherapy and supportive psychotherapy, in accordance with American Psychiatric Association

guidelines (APA)[34]. Serotonergic antidepressants were prescribed in flexible dosage depending on participants' response and tolerance. In case of non-response, different anti-depressant treatments were prescribed: venlafaxine or a tricyclic agent, and in case of continued non-response, lithium was added [35]. Supportive therapy in this mental health center focused on improving depressive symptoms and associated problems in participants' daily lives. The therapy consisted of a broad combination of psychoeducation, providing advice and encouragement, problem solving, behavioral activation, and stimulating participants to improve the quality of interpersonal relations. Typically, treatment lasted between 15 and 20 therapy sessions, with sessions occurring weekly and decreasing in frequency following agreement between therapist and participant.

Measurements

Negative affect

Momentary mood states were assessed with nine adjectives rated on 7-point scales (1 – not at all, 7 – very), e.g., “I feel down. To create the NA-variable, the following items were averaged per measurement occasion: down, tense, lonely, irritated, anxious, guilty, agitated, tired, irritable (for more detail, see Ref. [36]).

Depressive symptom severity

The HDRS-17 is a 17-item clinician-rated scale [31], which was used in this study to measure the severity of depressive symptoms at baseline, and at 18 monthly follow-up occasions. A trained research assistant performed the ratings of the HDRS via telephone.

Baseline patient characteristics

In the baseline interview, information was gathered concerning the patients' family history of depression (yes or no), duration of the depressive complaints in months, and whether the current episode was their first or not.

Analysis

Variability and instability in NA

The level of within-person variability in NA was estimated by calculating the standard deviation (SD) over all the available EMA-observations for each person.

To determine the level of NA instability across the EMA-measurements for each person, we calculated a time-adjusted mean square successive difference (MSSD) as suggested by Jahng *et al.* [37]:

$$\text{time-adjusted MSSD} = \frac{\text{median}(\Delta t)}{n - 1} \sum_{i=1}^{n-1} \frac{(X_{i+1} - X_i)^2}{(t_{i+1} - t_i)}$$

The time-adjusted MSSD corrects for the irregular intervals between observations inherent in our semi-random EMA-design by subtracting the median lag-duration per individual from the observed lags for that person. Overnight lags were omitted from the MSSD-calculations by inserting a missing observation between the last observation of the day and the first observation of the next day.

Analysis 1: Predicting rate of symptom change during treatment

In accordance with Husen *et al.* [25], monthly HDRS measurements were nested within patients in a multilevel model. We examined whether within-person variability and instability in NA were associated with a steeper linear¹ decline of HDRS symptom scores. We also examined the individual predictive value of mean momentary NA, the baseline HDRS score, episode duration, family history, and first or recurrent episode. Only the HDRS observations during treatment were used (i.e., the first five months), which is the same number of measurements used by Husen *et al.* (i.e., first five weeks). Maintaining a minimum of four measurements for inclusion in the analysis to ensure a reliable estimation of the response slope led to six exclusions, and 40 participants entered in the analysis.

Each single-predictor model included the variable Time (1, 2, 3, 4, 5), the candidate predictor and its interaction with the Time slope. Predictors that were significantly ($p < .05$) associated with a steeper linear rate of symptom change were included in a multi-predictor model.

¹ In accordance with the method described in Husen *et al.* [25], we conducted an AIC comparison between two unconditional growth models: a linear model (AIC = 1158.38) and a log-linear model (change in HDRS over time is modeled as a $\log_{10}$ function of time; AIC = 1163.21). For us, the linear model fit better, whereas Husen *et al.* ended up modeling symptoms as a logarithmic curve.

Analysis 2: Predicting remission status

We calculated time to remission for each patient ($n = 46$) by counting the number of monthly HDRS measurements (maximum eighteen) before a HDRS score ≤ 7 was reached (q.v. Ref. [38]). A HDRS score ≤ 7 is regarded as an indication of symptom remission [39,40]. Because the outcome variable was not normally distributed and appeared more bi-modal, we applied a median split to create a binary variable with 1 indicating *remission during treatment*, within five months ($n = 28$), and 0 indicating *post-treatment or no remission* ($n = 18$, 5 post-treatment remitters, and 13 non-remitters; see **Figure 3.1**).

To examine whether the candidate predictors were predictive of remission during treatment or post-treatment/no remission, we first estimated single-predictor logistic regression models using likelihood ratio testing. Significant predictors ($p < .05$) were then tested in a multi-predictor logistic regression model.

Figure 3.1

Histogram of the time to remission in months.

Note. Remission is indicated when a patient obtains a score of ≤ 7 on the monthly Hamilton Depression Rating Scale interview. The vertical dashed line indicates the applied median split at five months. Non-responders are indicated at ‘...’ on the x-axis, as it is unclear whether and when those patients experienced symptom improvement.

Results

Descriptives

Of the 46 participants, 27 were female (59%), and mean age was 40.3 ($SD = 11.1$). Secondary Axis I diagnoses were indicated for 21 (46%) of the patients, of which four participants had three or more secondary diagnoses: generalized social phobia ($n = 9$), dysthymic disorder ($n = 5$), panic disorder with agoraphobia ($n = 4$), posttraumatic stress disorder ($n = 4$), specific phobia ($n = 4$), bulimia nervosa ($n = 1$).

Thirty-three people (72%) reached the remission-threshold within the 18 months of study ($M = 4.3$, $SD = 3.6$, see also **Figure 3.1**). On average, 45 ($SD = 10.1$) out of 60 possible EMA questionnaires were completed (75%), and 14.1 ($SD = 3.3$) of the maximum 18 monthly HDRS assessments (78%). The overall Cronbach's alpha for the NA scale was 0.84, with the reliability of individual items ranging from 0.51 to 0.80.

Table 3.1

Descriptive statistics and Pearson's correlations of baseline symptom severity and sample characteristics and negative affect dynamic measures.

Measure	Descriptive statistics			Correlations						
	Mean (SD)	Min	Max	HDRS0	Duration	Episode	Family history	Negative affect		
								Mean	SD	MSSD
Baseline characteristics										
HDRS0	23.9 (3.9)	19	34	–						
Duration (months)	20.2 (22.8)	2	120	–0.07	–					
Previous episode	$n = 19$ (41%)	–	–	–0.11	–0.21	–				
Family History	$n = 25$ (54%)	–	–	–0.07	0.07	–0.03	–			
Negative affect										
Mean level (M)	2.98 (0.97)	1.46	5.23	0.25	0.33	–0.04	0.01	–		
Variability (SD)	0.70 (0.27)	0.21	1.44	0.25	0.01	0.05	0.07	0.29	–	
Instability (MSSD)	0.56 (0.47)	0.06	2.19	0.26	0.01	0.03	0.03	0.15	0.90	–

Note. HDRS0 = Hamilton Depression Rating Scale, baseline assessment; SD = Standard Deviation; MSSD = Mean Square Successive Difference. Significant correlations ($p < .05$) are marked in **bold** and p -values reported in text. Negative affect measures are calculated over the within-persons negative affect variables.

Table 3.1 presents descriptive statistics and Pearson correlations for the baseline predictors and NA variables. Some significant intercorrelations were found: mean-level NA was positively related to SD of NA ($p \approx .050$), and to duration of the complaints ($p = .025$). The SD and MSSD of NA were highly correlated at $p < .001$.

Analysis 1: Predicting rate of symptom change during treatment

Neither instability nor variability in NA were predictive of a steeper rate of symptom response during treatment (see **Table 3.2**, Analysis 1). Only Mean NA was a significant predictor, with higher daily NA relating to a stronger rate of treatment response. Because only one single-predictor model gave a significant result², no multi-predictor model was run.

The fixed effects for baseline HDRS score, and mean levels of NA were significant, indicating that, on average, a higher baseline HDRS and higher mean of NA were associated with higher symptom levels during the five months of treatment.

Analysis 2: Predicting remission status

In the single-predictor models, no significant associations were found between remission status and NA instability, NA variability or any of the other predictors (see **Table 3.2**, Analysis 2). Again, no multi-predictor was warranted.

² To control for the potential confounding effect of baseline symptom severity, the models were also run with the baseline HDRS score as a covariate. This did not result in any different conclusions.

Table 3.2

The predictive value of negative affect dynamics and baseline variables on the linear rate of change in HDRS scores over five months (Analysis 1), and on remission during treatment versus post-treatment/no remission (Analysis 2).

Analysis 1: HDRS over time

Single-predictor models	Fixed effects		Predictor interaction with time slope	
	Estimate (SE)	<i>p</i>	Estimate (SE)	<i>p</i>
Model 1 HDRS0	0.93 (0.33)	.007	−0.19 (0.10)	.063
Model 2 SD NA	−0.06 (5.67)	.991	−1.22 (1.69)	.471
Model 3 MSSD NA	−2.51 (3.74)	.506	0.02 (1.14)	.987
Model 4 Mean NA	4.45 (1.20)	<.001	−0.81 (0.40)	.042
Model 5 Duration	0.10 (0.05)	.065	−0.01 (0.02)	.467
Model 6 Episode	0.63 (2.74)	.820	−0.24 (0.81)	.771
Model 7 Family history	−0.52 (2.73)	.850	−0.69 (0.80)	.393

Analysis 2: Remission during treatment vs post-treatment/no remission

Single-predictor models	Log odds	OR	95% CI	<i>p</i>
Model 1 HDRS0	−0.06	0.94	[0.80, 1.09]	.414
Model 2 SD NA	1.55	4.71	[0.46, 68.56]	.198
Model 3 MSSD NA	1.18	3.26	[0.77, 21.96]	.114
Model 4 Mean NA	−0.57	0.56	[0.28, 1.06]	.076
Model 5 Duration	−0.01	0.99	[0.96, 1.02]	.418
Model 6 Episode	−0.21	0.81	[0.24, 2.72]	.729
Model 7 Family history	0.66	1.93	[0.59, 6.60]	.279

Note. HDRS = Hamilton Depression Rating Scale; HDRS0 = baseline HDRS; NA = Negative Affect; SD = Standard Deviation; MSSD = Mean Square Successive Difference; SE = Standard Error; OR = Odds Ratio; CI = Confidence Interval. Significant *p*-values (< .05) are marked in **bold**.

Analysis 1: Level 1: *N* = 179, Level 2: *N* = 40. Estimates are the fixed and interaction effects of single-predictor models. In each model, Time as the HDRS-observation number was entered as a level 1 predictor. All level 2 predictors were grand-mean centered.

Analysis 2: Remission during treatment (*n* = 28) = HDRS score ≤ 7 within five months; Post-treatment/No remission (*n* = 18) = remission after five months or no remission. The reference group (0) is post-treatment/no remission. OR > 1 indicates the odds are higher to reach remission status during treatment per one unit increase on a given predictor. OR < 1 indicate an inverse relationship (lower odds of remission during treatment), the inverse odds can be calculated as 1/OR. All predictors were mean-centered.

Discussion

The aim of this study was to examine whether instability and variability in daily NA dynamics were indicative of a stronger rate of treatment response, and of remission during treatment versus post-treatment/no remission, beyond mean levels of NA. We did not find instability or variability to be predictive of the rate of symptom change over five months, only a higher mean NA significantly predicted a steeper decrease in HDRS-scores over time. Moreover, none of the NA dynamics or baseline predictors were related to remission occurring during treatment compared to post-treatment/no remission.

We hypothesized that NA dynamics would be associated with depressive symptom change during treatment and with reaching remission during eighteen months follow-up, yet neither variability nor instability of NA showed the expected effect. Rather than the fluctuations in momentary NA captured by the SD and MSSD, it appears that the mean-levels of NA may be more promising as a predictor of depressive change. In addition to the positive association to the rate of symptom improvement over five months of treatment, higher NA mean-levels were, on average, predictive of higher symptom HDRS symptom levels over five months and the mean of NA was positively correlated with a longer duration of complaints at baseline. It seems plausible that experiencing stronger NA in daily life is an indicator that concurrent symptoms are more severe [22,26], and there is, consequently, more room for improvement during treatment. This is also supported by Husen *et al.* [25] and Bosley *et al.* [26] who both found baseline symptom severity to be related to stronger treatment response. Therefore, measuring NA levels within the context of daily life before treatment onset could be a helpful gauge of a patient's potential for improvement during therapy.

Our findings add to the debate on the importance of accounting for mean-levels of NA when relating momentary-measured mood indices to depression [13,29]. In their study on the association between NA dynamics and current depressive symptom levels, Bos *et al.* [28] found the predictive effect of the MSSD to be rendered non-significant by the addition of the SD, and the effect of the SD to be reduced to non-significance by the addition of the mean in their model. Similarly, the SD and MSSD of NA were highly correlated in our sample, and the SD – but not the MSSD – was also significantly correlated with the mean of NA. Indices

like the SD and MSSD are often dependent on the mean, which can complicate their interpretation [20,28,29,41]. While we could not test this in a multi-predictor model, it could be that we did not find the MSSD and SD to be predictors of change in depression symptoms because they were not (as strongly) correlated with the mean as in previous studies.

At this time, the evidence regarding NA dynamics and treatment response is sparse, and inconclusive. We found no associations of NA variability and NA instability with symptom decreases over five months of treatment, nor with remission status over 18 months. Similarly, Bosley *et al.* [26] found no relationship between NA dynamics and change in pre-to-post treatment depression symptom scores, but Husen *et al.* [25] found lower instability of daily NA predicted stronger early response rates to CBT treatment. Two out of three empirical studies show no association between NA dynamics and treatment response, yet with so little evidence it is not yet possible to draw any firm conclusions about the utility of NA dynamics in predicting change in depression symptoms.

Moreover, design differences between studies hinder a more general interpretation of these findings. First, the number of measurements within the day will affect a time-dependent variable such as the MSSD, and there is insufficient knowledge about what the optimal time between measurements is to capture affect instability in depression [20,42]. In our study, EMA measurements were ten times daily, every 90 minutes on average, versus four measurements a day, at 240-minute intervals for Bosley *et al.* [26] and Husen *et al.* [25]. Shorter intervals and higher sampling frequency may capture moment-to-moment changes in NA more closely, and may give more reliable estimates of the spread of scores throughout the day expressed in the SD and MSSD [13,37,43].

Second, it seems reasonable that the relationship between NA dynamics and symptom response is sensitive to *when* in the treatment process NA is measured. Whereas Bosley *et al.* [26] and our study measured NA dynamics directly before treatment, Husen *et al.* [25] measured it on average 2.3 months before treatment onset. Capturing daily NA dynamics directly before treatment versus while a patient is on the waitlist several months before treatment, may have been sufficiently different contexts that it could have affected the study findings. It is noteworthy that the two studies in which the EMA period was immediately

followed by treatment did not find a relationship between NA dynamics and depression outcomes.

Third, an important reason our design may have had limited ability to replicate the findings of Husen *et al.* [25] is that we examined the rate of symptom change over five *months*, whereas they investigated the first five *weeks* of treatment. While the theoretical rationale of instability and variability representing the kind of destabilization that precedes state transitions in dynamical systems has shown promise in some studies e.g., Refs. [6,7], it may be that the time span over which such dynamics are predictive of change is much shorter than the five to eighteen months studied in our sample [44]. To know whether instability and variability are indeed indicators of an upcoming symptom transition, the NA dynamics should ideally be measured immediately before the period of change occurs. While we measured NA instability and variability directly before treatment onset and most of our participants did remit during the treatment phase — within the first five months — this nevertheless meant that NA dynamics were measured several months before a transition towards lower symptom levels could be identified. Moreover, the monthly symptom interviews may have been too far apart in time to adequately capture any early symptom changes in our model of the HDRS over time. A study design with more frequent assessments of symptoms and NA dynamics over a longer period of time during treatment is necessary to determine whether NA instability and variability precede symptom improvements [5,45].

Strengths of the study are that we tested the predictive capacity of dynamic variables independently first, to account for multi-collinearity. We also extended the studies of Husen *et al.* [25] and Bosley *et al.* [26] by examining remission status over the course of eighteen months. With our sample size and model specification being consistent with Husen *et al.*, we expected the power to detect an effect to be similar in our multilevel analyses. For the logistic regression, a sensitivity analysis in G*Power 3.1 [46] suggested that with the current sample size ($N = 46$) we had 80% power to find an OR of 2.75 or larger. However, our sample size is modest and could have limited our ability to find smaller effects. Moreover, while it would have been interesting to compare whether NA dynamics were different for those who remitted during treatment, post-treatment and non-remitters, the groups were too small and unequally sized to make this comparison.

In conclusion, future studies should aim to elucidate to what extent the relationship between NA dynamics, mean-level NA and later improvement in depression is sensitive to various measurement timings: how far in advance the EMA measurements are collected, the number of and time interval between NA observations within the day, as well as between symptom assessments. Moreover, the mean-level of NA measured before treatment onset appears to be a simple yet informative indicator of potential for symptom improvement during therapy. It would therefore be clinically relevant for researchers to further study the mean of NA as a way to predict depressive improvement during early and later phases of treatment.

References

1. Buckman, J. E. J., Underwood, A., Clarke, K., Saunders, R., Hollon, S. D., Fearon, P., & Pilling, S. (2018). Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clinical Psychology Review*, 64, 13–38. <https://doi.org/10.1016/j.cpr.2018.07.005>
2. Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., Thiery, E., Delespaul, P. H., & van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, 124(1–2), 191–195. <https://doi.org/10.1016/j.jad.2009.11.010>
3. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(07), 1349–1360. <https://doi.org/10.1017/S0033291713001979>
4. Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major Depression As a Complex Dynamic System. *PLoS One*, 11(12), [e0167490]. <https://doi.org/10.1371/journal.pone.0167490>
5. Hayes, A. M., & Andrews, L. A. (2020). A Complex Systems Approach to the Study of change in psychotherapy. *BMC Medicine*, 18(197), 1–46. <https://doi.org/10.1186/s12916-020-01662-2>
6. Kuranova, A., Booij, S. H., Menne-Lothmann, C., Decoster, J., Van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., Van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. *BMC Medicine*, 18(1), 36. <https://doi.org/10.1186/s12916-020-1500-9>
7. Olthof, M., Hasselman, F., Strunk, G., Aas, B., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Destabilization in self-ratings of the psychotherapeutic process is associated with better treatment outcome in patients with mood disorders. *Psychotherapy Research*, 30(4), 520–531. <https://doi.org/10.1080/10503307.2019.1633484>
8. Hayes, A. M., & Strauss, J. L. (1998). Dynamic systems theory as a paradigm for the study of change in psychotherapy: An application to cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 66(6), 939–947. <https://doi.org/10.1037//0022-006X.66.6.939>
9. Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, 88(2), 119–127. <https://doi.org/10.1037/ccp0000469>
10. Schiepek, G., Aichhorn, W., & Schöller, H. J. (2017). Monitoring change dynamics: A nonlinear approach to psychotherapy feedback. *Chaos and Complexity Letters*, 11(3), 355–375.
11. Hayes, A. M., & Yasinski, C. (2015). Pattern destabilization and emotional processing in cognitive therapy for personality disorders. *Frontiers in Psychology*, 6(FEB), 107. <https://doi.org/10.3389/fpsyg.2015.00107>
12. Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-Level Affect Dynamics in Psychopathology Viewed From Complex Dynamical System Theory. *Emotion Review*, 7(4), 362–367. <https://doi.org/10.1177/1754073915590623>
13. Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology*, 118(1), 195–202. <https://doi.org/10.1037/a0014868>
14. Thompson, R. J., Berenbaum, H., & Bredemeier, K. (2011). Cross-sectional and longitudinal relations between affective instability and depression. *Journal of Affective Disorders*, 130(1–2), 53–59. <https://doi.org/10.1016/j.jad.2010.09.021>

15. Houben, M., & Kuppens, P. (2020). Emotion dynamics and the association with depressive features and borderline personality disorder traits: Unique, specific, and prospective relationships. *Clinical Psychological Science*, 8(2), 226–239. <https://doi.org/10.1177/2167702619871962>
16. Köhling, J., Moessner, M., Ehrental, J. C., Bauer, S., Cierpka, M., Kämmerer, A., Schauenburg, H., & Dinger, U. (2016). Affective instability and reactivity in depressed patients with and without borderline pathology. *Journal of Personality Disorders*, 30(6), 776–795. <https://doi.org/10.1521/pedi.2015.29.230>
17. Eldesouky, L., Thompson, R. J., Oltmanns, T. F., & English, T. (2018). Affective instability predicts the course of depression in late middle-age and older adulthood. *Journal of Affective Disorders*, 239, 72–78. <https://doi.org/10.1016/j.jad.2018.06.038>
18. Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, 6(3), 383–391. <https://doi.org/10.1037/1528-3542.6.3.383>
19. Thompson, R. J., Mata, J., Jaeggli, S. M., Buschkuhl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, 121(4), 819–829. <https://doi.org/10.1037/a0027978>
20. Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13(6), 1132–1141. <https://doi.org/10.1037/a0033579>
21. Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, 141(4), 901–930. <https://doi.org/10.1037/a0038822>
22. Sperry, S. H., Walsh, M. A., & Kwapil, T. R. (2020). Emotion dynamics concurrently and prospectively predict mood psychopathology. *Journal of Affective Disorders*, 261, 67–75. <https://doi.org/10.1016/j.jad.2019.09.076>
23. Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99(6), 1042–1060. <https://doi.org/10.1037/a0020962>
24. Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, 7(4), 355–361. <https://doi.org/10.1177/1754073915590617>
25. Husen, K., Rafaeli, E., Rubel, J. A., Bar-Kalifa, E., & Lutz, W. (2016). Daily affect dynamics predict early response in CBT: Feasibility and predictive validity of EMA for outpatient psychotherapy. *Journal of Affective Disorders*, 206, 305–314. <https://doi.org/10.1016/j.jad.2016.08.025>
26. Bosley, H. G., Soyster, P. D., & Fisher, A. J. (2019). Affect dynamics as predictors of symptom severity and treatment response in mood and anxiety disorders: evidence for specificity. *Journal for Person-Oriented Research*, 5(2), 101–113. <https://doi.org/10.17505/jpor.2019.09>
27. Funkhouser, C. J., Kaiser, A. J. E., Alqueza, K. L., Carrillo, V. L., Hoffman, L. M. K., Nabb, C. B., Auerbach, R. P., & Shankman, S. A. (2021). Depression risk factors and affect dynamics: An experience sampling study. *Journal of Psychiatric Research*, 135(January), 68–75. <https://doi.org/10.1016/j.jpsychires.2021.01.007>
28. Bos, E. H., de Jonge, P., & Cox, R. F. A. (2018). Affective variability in depression: Revisiting the inertia-instability paradox. *British Journal of Psychology*. <https://doi.org/10.1111/bjop.12372>
29. Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, 3(5), 478–491. <https://doi.org/10.1038/s41562-019-0555-0>
30. First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1995). *Structured clinical interview for DSM-IV axis I disorders*. New York State Psychiatric Institute.
31. Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23(1), 56–62.

32. Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life: The experience sampling method*. Datawyse / Maastricht University Press.
33. Peeters, F., Nicolson, N. A., Berkhof, J., Delespaul, P., DeVries, M., & De Vries, M. (2003). Effects of daily events on mood states in major depressive disorder. *Journal of Abnormal Psychology*, 112(2), 203–211. <https://doi.org/10.1037/0021-843X.112.2.203>
34. American Psychiatric Association. (2000). Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. *The American Journal of Psychiatry*, 157(4 Suppl), 1–45.
35. Peeters, F., Berkhof, J., Rottenberg, J., & Nicolson, N. A. (2010). Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. *Behaviour Research and Therapy*, 48(8), 754–760. <https://doi.org/10.1016/j.BRAT.2010.04.008>
36. Wichers, M., Lothmann, C., Simons, C. J. P., Nicolson, N. A., & Peeters, F. (2012). The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: A momentary assessment study. *British Journal of Clinical Psychology*, 51(2), 206–222. <https://doi.org/10.1111/j.2044-8260.2011.02021.x>
37. Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, 13(4), 354–375. <https://doi.org/10.1037/a0014173>
38. van Winkel, M., Nicolson, N. A., Wichers, M., Viechtbauer, W., Myin-Germeys, I., & Peeters, F. (2015). Daily life stress reactivity in remitted versus non-remitted depressed individuals. *European Psychiatry*, 30(4), 441–447. <https://doi.org/10.1016/j.eurpsy.2015.02.011>
39. Riedel, M., Möller, H.-J., Obermeier, M., Schennach-Wolff, R., Bauer, M., Adli, M., Kronmüller, K., Nickel, T., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W., & Seemüller, F. (2010). Response and remission criteria in major depression – A validation of current practice. *Journal of Psychiatric Research*, 44(15), 1063–1068. <https://doi.org/10.1016/j.JPSYCHIRES.2010.03.006>
40. Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., Rush, A. J., & Weissman, M. M. (1991). Conceptualization and Rationale for Consensus Definitions of Terms in Major Depressive Disorder. *Archives of General Psychiatry*, 48(9), 851–855. <https://doi.org/10.1001/archpsyc.1991.01810330075011>
41. Mestdag, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelineing the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, 23(4), 690–707. <https://doi.org/10.1037/met0000153>
42. Ebner-Priemer, U. W., & Sawitzki, G. (2007). Ambulatory assessment of affective instability in borderline personality disorder: The effect of the sampling frequency. *European Journal of Psychological Assessment*, 23(4), 238–247. <https://doi.org/10.1027/1015-5759.23.4.238>
43. Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research*, 37(1), 127–160. https://doi.org/10.1207/S15327906MBR3701_06
44. Strunk, G., & Lichtwarck-Aschoff, A. (2019). Therapeutic Chaos. *Journal for Person-Oriented Research*, 5(2), 81–100. <https://doi.org/10.17505/jpor.2019.08>
45. Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Recovery: Study protocol for a repeated intensive longitudinal n = 1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *PsyArXiv*. <https://doi.org/10.31234/osf.io/fertq>
46. Erdfelder, E., Faul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>

4

Early warning signals and critical transitions in psychopathology: challenges and recommendations

Published as:

Helmich, M. A., Olthof, M., Oldehinkel, A. J., Wichers, M., Bringmann, L. F., & Smit, A. C. (2021). Early warning signals and critical transitions in psychopathology: challenges and recommendations. *Current Opinion in Psychology*, 41, 51–58.
<https://doi.org/10.1016/j.copsyc.2021.02.008>

Abstract

Empirical evidence is mounting that monitoring momentary experiences for the presence of early warning signals (EWS) may allow for personalized predictions of meaningful symptom shifts in psychopathology. Studies aiming to detect EWS require intensive longitudinal measurement designs that center on individuals undergoing change. We recommend that researchers: (a) define criteria for relevant symptom shifts *a priori* to allow specific hypothesis testing; (b) balance the observation period length and high-frequency measurements with participant burden by testing ambitious designs with pilot studies; (c) choose variables that are meaningful to their patient group and facilitate replication by others. Thoroughly considered designs are necessary to assess the promise of EWS as a clinical tool to detect, prevent or encourage impending symptom changes in psychopathology.

Keywords: psychopathology; ecological momentary assessment; symptom change; early warning signals; critical transitions

Introduction

In the search for a better understanding of the development and maintenance of mental illness, researchers have been drawing inspiration from the field of complex dynamical systems. Rooted in mathematics and physics, dynamical systems theory can be applied to the study of multi-dimensional processes of change, such as those occurring in ecology, finance, power grids, and, indeed, in psychopathology [1–5]. Although discontinuous patterns of change (e.g., sudden gains and losses, symptom spikes, rapid early response, relapse) have been frequently reported in the field of clinical psychotherapy [6–10], traditional models of mental disorder typically do not provide an integrated conceptualization of the complex, varied patterns of symptom change that different individuals show. The dynamical systems framework has been suggested as a way to incorporate this wide variety of clinical change patterns into a coherent model that can further our understanding, early detection and the treatment of mental illness [11,12*,13].

Within the dynamical systems conceptualization, an individual constitutes a multi-dimensional complex system of interacting components (e.g., behaviors, emotions, cognitions, and somatic experiences), which describe the system by their joint dynamic patterns over time [4,14*,15,16]. Individuals can move between different dynamically stable states (attractors), such as mental disorder or a psychologically healthy state. Psychopathology is theorized to occur when the system becomes more and more ‘attracted’ to maladaptive functional patterns, which are maintained despite the negative effects on an individual’s wellbeing [17–19]. In other words, when moment-to-moment experiences and behaviors start to reinforce each other to create an overall more negative state, over time, the persistent presence of such negative experiences can become problematic to the point that they can be classified as the cluster of symptoms that can be recognized as a disorder [20].

One particular phenomenon has captured the interest of researchers for its potential clinical utility: early warning signals (EWS). In a dynamical system, a shift between states may appear abrupt and discontinuous on a global system level while it is preceded by a gradual destabilization, leading to EWS in the temporal dynamics of intensive repeated measurements of one or multiple variables at underlying (time) scales [21**]. For instance, as the system destabilizes prior to a

Table 4.1

A non-exhaustive list of Early Warning Signals described by the underlying theoretical process and the statistical indicators.

Theoretical process	Description	Statistical indicators
Critical slowing down further detail, see Ref. [21**]	Close to a tipping point, the attractor loses resilience and becomes weaker, perturbations are more likely to push the system farther away from the center and it takes longer for the system to return to the equilibrium	Rises in lag-1 autocorrelation Rises in variance Rises in skewness Rises in connectivity (i.e., variable cross-correlations)
Critical fluctuations for a recent application, see Ref. [25**]	Once the existing attractor is fully destabilized, the system regains all its degrees of freedom, leading to fluctuations between all possible system states until it settles in a stable attractor.	Rises in dynamic complexity Rises in entropy Rises in variance
Flickering further detail, see Ref. [39]	If a system has two stable states and the dominant attractor is becoming weaker, perturbations can cause the system to ‘flicker’ back and forth between alternative states, until one of the attractors becomes strong enough for the system to settle into one state	Regime switching Bimodality Rises in variance Rises in skewness Rises in kurtosis

Note. Definitions of important terms used in the table: **Attractor:** a stable state of the system, or dynamic regime; visible in the interactions and convergence of dynamic processes when observed over time. **Perturbations:** external shocks or stressors. **Tipping point:** the point from which the system undergoes a transition into an alternative stable state.

transition, rises in the autocorrelation and variance of emotions may serve as EWS [22–24] (for more detail, see **Figure 4.1** and **Table 4.1**). For psychopathology, it is hypothesized that EWS can be observed in the dynamics of ecological momentary assessments¹ (EMA) of emotions, cognitions and behavior, and may be detected *before* the system’s ‘tipping point’ is reached and a critical transition occurs at the symptom-level [24,25**,26**]. Consistent evidence of EWS would provide an important empirical basis for adopting the dynamical systems conceptualization

¹ Similar hypotheses have been investigated in clinical psychology using coded recordings of therapy sessions and dyadic interactions [29,30]. Researchers are also exploring whether EWS can be detected in passively collected physical measurements, like activity and heart rate data [57**,62, 63].

Figure 4.1

Visual representation of the system with alternative stable states, destabilizing, and restabilized.

Note. The ball represents the current state of the system, and the basin in which it lies represents its current attractor. Other basins represent alternative attractors. For psychopathology, the ball symbolizes a momentary state (e.g., current emotion), and the attractors the possible states that the system can be in (e.g., equilibria of affect). In **Figure 4.1A**, the system occupies a stable ('deep') attractor: if any perturbations push the ball away from the equilibrium position, it is unlikely to deviate much and will return quickly to the resting state. In **Figure 4.1B**, the system is destabilizing and the temporal dynamics start exhibiting signs of 'critical slowing down' and 'critical fluctuations': because the attractor has weakened, even small perturbations can cause the ball to move farther away and return more slowly. This shows as rising EWS, like lag-1 autocorrelation and variance, and indicates that the system may be more likely to tip over into the alternative attractor. In **Figure 4.1C**, the system reaches a 'tipping point', and restabilizes into the new attractor. The old attractor disappears.

of psychopathology and the notion of critical transitions into the theoretical understanding of the development, persistence and recurrence of mental disorders. Such evidence would also be of great clinical importance, as it promises the possibility to improve early detection of episodes of mental illness [27–29], to anticipate and prevent negative events like depressive relapse and suicide attempts [26*,30,31], and, conversely, encourage positive symptom changes in the context of therapy [5,8,12*,13].

Yet, moving beyond dynamical systems as a narrative metaphor to explain clinical change phenomena involves both theoretical and practical challenges [32], and gathering the right data has been a major hurdle. In particular, it requires study designs that observe individuals *as they change over time* [23,33*], which implies methodological choices that differ from many previous EMA and intensive longitudinal designs where stationarity or ‘the absence of change’ is assumed [34,35**]. Indirect evidence of EWS-phenomena has been found at group level (e.g., Refs. [22,27,29,36–38]), but only a few studies thus far have measured how individual patients change over time and have been able to empirically test the hypothesis that rises in EWS occur before sudden symptom transitions [24,25**,26**]. The aim of this paper is to outline a selection of challenges and considerations that should be heeded when designing studies on EWS in psychopathology (for a checklist overview, see **Table 4.2** at the end of the article).

Challenge 1: Defining transitions

The primary utility of EWS is that they precede upcoming critical transitions, which typically appear as large, sudden state shifts in the global behavior of the system [1]. What does a critical transition between stable states look like in the context of mental disorder? Because mental experience cannot be observed directly and has no absolute quantification, creating a one-size-fits-all definition of a critical transition in psychopathology is not straightforward. Careful consideration should be given to which constructs to measure (see Challenge 3), and to the fact that the idiosyncrasy of self- and clinician-rated symptoms means that potential critical transitions may have different magnitudes for different individuals, and different time spans in different disorders (see Challenge 2). Therefore, the first challenge is to define and identify relevant symptom changes that could constitute a critical transition in psychopathology.

The clearest transition is one between healthy and psychopathological states, which could be identified using existing, relatively well-defined diagnostic thresholds [12*]. However, for some individuals, the step across the diagnostic boundary may be small and not meaningful; or conversely, symptom changes that remain above or beneath the threshold may still be relevant transitions between

alternative stable states for an individual [40]. Therefore, apart from diagnostic shifts, researchers should search for transitions that are meaningful in the context of individuals' trajectories of symptoms over time [33*].

Quantitative methods can detect potential critical transitions as changes that stand out in the context of a particular within-person time series as shifts that are larger in magnitude than the natural variation expected in a stable, or gradually changing system [18,28]. What is considered a large change in scores may be based, for instance, on thresholds established with a person's own baseline variance [41**], criteria for sudden change between therapy sessions [10,25**], a significant change in scores for a particular questionnaire [42], or statistically identified change points or regime switches [34,43,44]. Still, although quantitative approaches should work well on average [45], using only numerical data risks misidentifying some shifts that were experienced as (ir)relevant by the individual. Important changes may not be visible in the data if measurements are too far apart (see Challenge 2), or if the questionnaire lacks the relevant variable (see Challenge 3), and conversely, sometimes numerically large changes are not judged as important by the patient themselves.

This leads to the other end of the spectrum: qualitative identification of critical transitions. Defining relevant symptom shifts by listening to the patient or clinician can improve the ecological validity of a study. However, retrospective reports can be biased and the precise timings and impact of transitions can be hard to recall or put into words. In $n = 1$ case studies or in a clinical setting [46], transition identification can be strengthened by using various sources of information, both qualitative and quantitative. For larger group studies, using in-depth clinical interviews to identify transitions may be less desirable — though not unfeasible, see Ref. [47].

To conclude, there is no standard way to identify potential critical transitions in psychopathology, and there is a need for exploratory and methodological studies to deepen the understanding of what critical transitions look like in psychological systems. Future studies may also include criteria for how long a symptom shift should be maintained to constitute a transition to a new state. Until there is more empirical evidence, researchers who want to conduct a confirmatory test of the EWS-hypothesis must thoroughly consider *a priori* how they will establish transparent, reproducible methods to identify transitions in their sample.

Challenge 2: Timing all measurements appropriately

To capture a rise in EWS before a symptom transition in an individual, researchers must (a) select a population and period in which the transition has a realistic likelihood to occur, (b) estimate the change processes' duration, and (c) collect high-resolution data (i.e., many observations over time).

First, choosing a population in which symptom shifts are a common occurrence increases the chances of observing a critical transition and being able to test for preceding EWS. Clinical knowledge of a particular disorder and change process can help researchers to decide who to measure and when (for what to measure, see Challenge 3). For instance, discontinuous changes in symptoms are common in depression, even more so for patients receiving therapy [10,25**] or tapering antidepressant medication [47]. A practical advantage of studying a population with many transitions is that, in total, a smaller sample may suffice to find consistent evidence of EWS before transitions.

A second issue to consider is the observation period. Symptoms should be assessed often (in most cases, at least weekly), quantitatively and/or qualitatively, and long enough to observe the entire transition, within a time frame appropriate to the rate of change of the disorder: rapid-cycling bipolar patients may shift into manic and depressed states over the course of a few days (or even hours [48]), whereas reaching a state of remission from depression is considered a 'rapid response' to therapy if it takes several weeks [9,49] and usually spans a period of months. Moreover, data collection should start while the system is believed to still be in a relatively stable state (**Figure 4.1A**), gather enough data while the system is destabilizing (**Figure 4.1B**), and continue at least until after the transition (**Figure 4.1C**). Only with a comparatively stable period at the start of the time series can changes in system dynamics (i.e., EWS) be detected and used as indicators that the system is destabilizing and likely to 'tip over' into an alternative state.

The third consideration is the determination of the (EMA) sampling regime. Different experiential processes fluctuate and change at different rates (compare a minute of irritation to feeling down all day), and therefore the variable in which EWS will be calculated should be sampled frequently enough to capture those temporal fluctuations [14*,18,35**]. It is also worth considering that psychological time series are often noisy, and analysis methods typically require many (equidistant)

data points to give robust results (for example, Refs. [50-54]). Methods are being developed that may elucidate optimal sampling frequencies in the future [55], but until we know the temporal resolution at which the fluctuations in momentary variables are best captured, high-frequency measurements (ideally multiple within the day and as many as possible) are the safest choice [56].

However, a tradeoff must be made between the need for high-resolution data and practical feasibility. For instance, while promising results have been obtained with once-daily self-ratings of depressed patients in treatment [25**], fast-changing systems, such as rapid-cycling bipolar cases, may require so many observations a day to get sufficient data between state shifts that it becomes practically or mentally unfeasible for participants. Although ambitious designs have been successful [8,47,57**,58], intensive longitudinal self-reporting can be burdensome and less feasible for some individuals or diagnostic groups [59]. Therefore, we strongly recommend running pilot studies to explore whether gathering sufficient high-resolution data is realistic and feasible for the intended population and change process.

Challenge 3: Selecting relevant variables

Theory can guide the first steps in choosing variables in which to expect EWS. One of the properties of dynamical systems is that processes at different levels are interdependent: ‘zooming in’ on symptoms of a disorder reveals the underlying moment-to-moment experiences (“*I feel...*”, “*I think...*”, “*I am...*”)[20,60,61] in which EWS may be detectable². Indeed, momentary affect may be a logical micro-variable choice, as affective disturbances are involved in virtually all psychiatric disorders [15] and have been studied with EMA for years in many patient groups [59]. Additionally, because variables are expected to become more and more alike (correlated) near the tipping point [21**,36], a few, or as little as one variable could be sufficient to detect EWS and impending transitions. For example, changes in the dynamics of “I feel down” may precede a relapse in depression, as the item reflects a momentary

² Even though, theoretically, EWS would also be expected to occur in symptom time series, symptoms are conventionally measured with retrospective questionnaires that cover periods of multiple days (e.g., the past week), which results in time series that lack the necessary detail to capture the dynamics of the system and the relevant rises in EWS.

experience of a core depressive symptom (i.e., prolonged feelings of sadness) [26**,32,41**]. Still, the number and content of the variables a researcher chooses to include may impact (a) the quality of their data: whether participant responses show variation and change over time in that variable; and (b) how broadly they can draw their conclusions: whether the variables show EWS for multiple people.

Including variables that can be expected to show natural variation at the chosen sampling rate (e.g., within the day) and remain relevant over a longer period is important to ensure that the high-frequency observations actually measure the changing state of the system [32,35**]. Statistically, many EWS cannot be calculated when there is insufficient variance, and although creating aggregate variables from multiple items can make the analysis more robust to low variation and outliers, single items may offer better interpretability and replicability.

Choosing items that participants will find relevant is another important point. In the clinical context, researchers (and clinicians) may personalize items and prioritize the best possible signal for an individual to monitor whether treatment is effective [8,25**], or if the risk of relapse is rising [41**]. On the other hand, group studies (multiple within-person studies) may prefer to draw generalizable conclusions and choose a set of items that are expected to work well by showing variation and change over time for most people [33*].

To conclude, selecting variables in which one can theoretically expect and practically detect EWS deserves further study in and of itself. Therefore, including as many EMA-items as possible may be of limited value and would only needlessly increase the burden on participants. Researchers would do well to create designs with a selection of theory-driven variables, and keep the future use in other studies or clinical practice in mind.

Conclusion

To make the most out of pioneering research on EWS in psychopathology, it is important to consider how to define transitions, time all measurements appropriately, and select theoretically relevant and practically useful variables. Furthermore, sound methodological choices cannot be based on theory alone, and in many cases pilot data will be needed to construct a strong and empirically effective research design. Like a dynamical system, the methodological challenges involved in capturing EWS of upcoming symptom shifts are interdependent, and the success of a study depends on carefully weighing all the aforementioned design choices before data collection starts. Only then can we move beyond *post hoc* reasoning about EWS and test our hypotheses, and hope to bridge the gap between theory and clinical utility.

Table 4.2

A conceptual checklist for designing studies on early warning signals and critical transitions in psychopathology.

Conceptual level	Questions to consider before starting data collection	See Refs. in
In my study population...		
Level of symptoms	☐ What kind of discontinuous symptom changes are known from the clinical literature?	C1
	☐ How can a relevant transition be distinguished from normal variation?	C1
	☐ Could apparent symptom shifts be caused by external life events (e.g., a pet dying, or having the flu)? If so, can these be differentiated from true transitions by gathering contextual information?	C1
	☐ How fast can switches between states take place?	C1
	☐ How much time can a symptom change take to still be considered sudden, as would be expected of a ‘critical transition’? And therefore, be relevant to predict with EWS.	C1
	☐ Over what time period can the system be expected to move from a relatively stable, into a period of destabilization, and finally, into a new stable state? Therefore, how long should the observation period be?	C2
Level of momentary experience	☐ How often do symptoms need to be assessed to ensure the full transition is observed and the timing of this transition can be accurately estimated?	C2
	In my study population...	
	☐ At what sampling interval can the dynamics and EWS be accurately observed?	C2
Level of momentary experience	☐ How often do momentary affect variables need to be assessed to effectively a) capture the EWS dynamics and, b) have enough data and statistical power to find EWS before the transition is expected to occur?	C2

	☐ What momentary experience or emotion is theoretically most related to the symptoms and disorder?	C3
Level of momentary experience (continued)	☐ Are the chosen variables likely to show change as the system approaches a transition?	C3
	☐ Should the variables... <ul style="list-style-type: none">☐ be relevant, reliable and valid for most individuals to allow generalizability, replicability and gather consistent evidence of EWS?☐ emphasize personalized signals, to predict impending symptom changes in the clinical context?	C3
	☐ Does data collection take place during a presumed change process (e.g., treatment, tapering of medication)?	C2
	☐ Can a pilot study be run to test the feasibility? That is, to find the balance between participant burden and compliance and the ideal overall duration and frequency of the measurements.	C2
Overall design choices	☐ Which analysis method can I use to detect EWS in my data?	C2
	☐ What is the goal of the study? For instance... <ul style="list-style-type: none">☐ Exploratory work: methodological studies to improve understanding of what critical transitions in psychopathology look like☐ Hypothesis-testing: draw generalizable conclusions about the occurrence of EWS☐ Clinical utility: find indicators and methods that could easily be implemented in clinical settings	C1, C3

Note. C = Challenge; EWS = Early Warning Signals

References

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

● of special interest

●● of outstanding interest

1. Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press.
 2. Dakos, V., Carpenter, S. R., van Nes, E. H., & Scheffer, M. (2014). Resilience indicators: Prospects and limitations for early warnings of regime shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1659), 20130263–20130263. <https://doi.org/10.1098/rstb.2013.0263>
 3. Wen, H., Ciamarra, M. P., & Cheong, S. A. (2018). How one might miss early warning signals of critical transitions in time series data: A systematic study of two major currency pairs. *PLOS ONE*, 13(3), e0191439. <https://doi.org/10.1371/journal.pone.0191439>
 4. Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9(1), 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
 5. Hayes, A. M., Yasinski, C., Barnes, J. Ben, & Bockting, C. L. H. (2015). Network destabilization and transition in depression: New methods for studying the dynamics of therapeutic change. *Clinical Psychology Review*, 41, 27–39. <https://doi.org/10.1016/j.cpr.2015.06.007>
 6. Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, 27(6), 715–723. <https://doi.org/10.1016/j.cpr.2007.01.008>
 7. Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(1), 93–101. <https://doi.org/10.1037/a0026455>
 8. Schiepek, G., Aichhorn, W., Gruber, M., Strunk, G., Bachler, E., & Aas, B. (2016). Real-time monitoring of psychotherapeutic processes: concept and compliance. *Frontiers in Psychology*, 7, 604. <https://doi.org/10.3389/fpsyg.2016.00604>
 9. Kelley, M. E., Dunlop, B. W., Nemeroff, C. B., Lori, A., Carrillo-Roa, T., Binder, E. B., Kutner, M. H., Rivera, V. A., Craighead, W. E., & Mayberg, H. S. (2018). Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse. *Depression and Anxiety*, 35(10), 992–1000. <https://doi.org/10.1002/da.22832>
 10. Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, 88(2), 119–127. <https://doi.org/10.1037/ccp0000469>
 11. Hofmann, S. G., Curtiss, J. E., & Hayes, S. C. (2020). Beyond linear mediation: Toward a dynamic network approach to study treatment processes. *Clinical Psychology Review*, 76. <https://doi.org/10.1016/j.cpr.2020.101824>
 12. Hayes, A. M., & Andrews, L. A. (2020). A complex systems approach to the study of change in psychotherapy. *BMC Medicine*, 18(197), 1–46. <https://doi.org/10.1186/s12916-020-01662-2>
- Review providing an up-to-date perspective on and comprehensive overview of complex systems theory and applications in psychotherapy research.

13. Rubel, J. A., Fisher, A. J., Husen, K., & Lutz, W. (2018). Translating person-specific network models into personalized treatments: Development and demonstration of the dynamic assessment treatment algorithm for individual networks (DATA-IN). *Psychotherapy and Psychosomatics*, 87(4), 249–251. <https://doi.org/10.1159/000487769>
14. Jeronimus, B. F. (2019). Dynamic system perspectives on Anxiety and Depression. In E. S. Kunnen, N. M. P. de Ruiter, B. F. Jeronimus, & M. A. van der Gaag (Eds.), *Psychosocial Development in Adolescence: Insights from the Dynamic Systems Approach* (1st ed.). Routledge Psychology.
- **Book chapter that reviews the literature of anxiety and depression in the dynamical systems context and discusses dynamical systems perspectives as a way to improve understanding of those disorders across the lifespan.**
15. Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, 7(4), 362–367. <https://doi.org/10.1177/1754073915590623>
16. Olthof, M., Hasselman, F., Maatman, F. O., Bosman, A., & Lichtwarck-Aschoff, A. (2020). Adaptive DynAmic Pattern Theory (ADAPT) of Psychopathology. *Preprint*. <https://doi.org/10.31234/OSF.IO/F68EJ>
17. Fisher, A. J., Newman, M. G., & Molenaar, P. C. M. (2011). A quantitative method for the analysis of nomothetic relationships between idiographic structures: Dynamic patterns create attractor states for sustained posttreatment change. *Journal of Consulting and Clinical Psychology*, 79(4), 552–563. <https://doi.org/10.1037/a0024069>
18. Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99(6), 1042–1060. <https://doi.org/10.1037/a0020962>
19. Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of Counseling Psychology*, 63(4), 379–395. <https://doi.org/10.1037/cou0000150>
20. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(07), 1349–1360. <https://doi.org/10.1017/S0033291713001979>
21. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59. <https://doi.org/10.1038/nature08227>
- **Seminal paper that introduces the notion that early warning signals can occur before sudden shifts between dynamic regimes in a variety of complex dynamical systems.**
22. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
23. Bos, E. H., & De Jonge, P. (2014). ‘Critical slowing down in depression’ is a great idea that still needs empirical proof. *Proceedings of the National Academy of Sciences*, 111(10), E878–E878. <https://doi.org/10.1073/pnas.1323672111>
24. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>

25. Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, 8(1), 25–35. <https://doi.org/10.1177/2167702619865969>

First study to show, in a large sample, that EWS in the form of increased dynamic complexity in daily self-ratings were associated with an increased probability of transitions in symptom severity within individuals.

26. Wichers, M., Smit, A. C., & Snippe, E. (2020). 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*, 6(1), 1–15. <https://doi.org/10.17505/jpor.2020.22042>

Confirmatory study using EMA data of six subjects, collected over a period of 3–6 months, to test empirically if EWS rise before a transition toward higher levels of depressive symptoms. The results replicate findings of a previous case study, with rises in EWS occurring a month before the symptom transitions.

27. Kuranova, A., Booij, S. H., Menne-Lothmann, C., Decoster, J., Van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., Van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. *BMC Medicine*, 18(1), 36. <https://doi.org/10.1186/s12916-020-1500-9>

28. Wichers, M., Schreuder, M. J., Goekoop, R., & Groen, R. N. (2019). Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology. *Psychological Medicine*, 49(3), 380–387. <https://doi.org/10.1017/S0033291718002064>

29. Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, 12(2), 283–289. <https://doi.org/10.2337/db10-1371>

30. Yasinski, C., Hayes, A. M., Ready, C. B., Abel, A., Görg, N., & Kuyken, W. (2019). Processes of change in cognitive behavioral therapy for treatment-resistant depression: Psychological flexibility, rumination, avoidance, and emotional processing. *Psychotherapy Research*, 1–15. <https://doi.org/10.1080/10503307.2019.1699972>

31. Bryan, C. J., & Rudd, M. D. (2018). Nonlinear change processes during psychotherapy characterize patients who have made multiple suicide attempts. *Suicide and Life-Threatening Behavior*, 48(4), 386–400. <https://doi.org/10.1111/sltb.12361>

32. Olthof, M., Hasselman, F., & Lichtwarck-Aschoff, A. (2020). Complexity in psychological self-ratings: Implications for research and practice. *BMC Medicine*, 18(317). <https://doi.org/10.1186/s12916-020-01727-2>

33. Wright, A. G. C., & Woods, W. C. (2020). Personalized Models of Psychopathology. *Annual Review Of Clinical Psychology Personalized*, 16, 49–74. <https://doi.org/10.1146/annurev-clinpsy-102419>

Review of personalized approaches to psychopathology, providing a summary of the advances in idiographic research and a useful discussion of future directions.

34. Albers, C., & Bringmann, L. F. (2020). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *European Journal of Psychological Assessment*, 36(February), 492–499. <https://doi.org/10.1027/1015-5759/a000589>

35. Hamaker, E. L., Ceulemans, E., Grasman, R. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: state of the art and future challenges. *Emotion Review*, 7(4), 316–322. <https://doi.org/10.1177/1754073915590619>

An overview of modeling techniques to study affect dynamics with intensive longitudinal data, including considerations that are relevant to studying individual change over time.

36. Curtiss, J., Fulford, D., Hofmann, S. G., & Gershon, A. (2019). Network dynamics of positive and negative affect in bipolar disorder. *Journal of Affective Disorders*, 249, 270–277. <https://doi.org/10.1016/j.jad.2019.02.017>
37. Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13(6), 1132–1141. <https://doi.org/10.1037/a0033579>
38. Schreuder, M. J., Hartman, C. A., George, S. V., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Early warning signals in psychopathology: What do they tell? *BMC Medicine*, 18(1), 1–12. <https://doi.org/10.1186/s12916-020-01742-3>
39. Dakos, V., van Nes, E. H., & Scheffer, M. (2013). Flickering as an early warning signal. *Theoretical Ecology*, 6(3), 309–317. <https://doi.org/10.1007/s12080-013-0186-4>
40. van Os, J., Guloksuz, S., Vijn, T. W., Hafkenscheid, A., & Delespaul, P. (2019). The evidence-based group-level symptom-reduction model as the organizing principle for mental health care: Time for change? *World Psychiatry*, 18(1), 88–96. <https://doi.org/10.1002/wps.20609>
41. Smit, A. C., Snippe, E., & Wichers, M. (2019). Increasing restlessness signals impending increase in depressive symptoms more than 2 months before it happens in individual patients. *Psychotherapy and Psychosomatics*, 88(4), 249–251. <https://doi.org/10.1159/000500594>

Study showing that impending symptom changes in depression can be predicted in real time by monitoring deviations from a person's own baseline-established mean. Note, this is not an early warning signals study, but important for its prospective and clinical utility.

42. Helmich, M. A. (2020). Duration-adjusted Reliable Change Index (DaRCI): Defining clinically relevant symptom changes of varying durations. *OSF*. <https://psyarxiv.com/q7ch9/>
43. Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L. J., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: Bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry*, 15(1), 222. <https://doi.org/10.1186/s12888-015-0596-5>
44. Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods*, 49(3), 988–1005. <https://doi.org/10.3758/s13428-016-0754-9>
45. Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19–30. <https://doi.org/10.1037/1040-3590.12.1.19>
46. Schiepek, G. K., Stöger-Schmidinger, B., Aichhorn, W., Schöller, H., & Aas, B. (2016). Systemic case formulation, individualized process monitoring, and state dynamics in a case of dissociative identity disorder. *Frontiers in Psychology*, 7(OCT), 1545. <https://doi.org/10.3389/fpsyg.2016.01545>
47. Smit, A. C., Snippe, E., Hoenders, H. J. R., & Wichers, M. (2020). Transitions In Depression: If, how, and when depressive symptoms increase during and after tapering of antidepressant medication. *Submitted*.

48. Wilk, K., & Hegerl, U. (2010). Time of mood switches in ultra-rapid cycling disorder: A brief review. In *Psychiatry Research* (Vol. 180, Issue 1, pp. 1–4). <https://doi.org/10.1016/j.psychres.2009.08.011>
49. Paul, R., Andlauer, T. F. M., Czamara, D., Hoehn, D., Lucae, S., Pütz, B., Lewis, C. M., Uher, R., Müller-Myhsok, B., Ising, M., & Sämann, P. G. (2019). Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. *Translational Psychiatry*, 9(1), 1–15. <https://doi.org/10.1038/s41398-019-0524-4>
50. Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H., & Scheffer, M. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS ONE*, 7(7). <https://doi.org/10.1371/journal.pone.0041010>
51. Cabrieto, J., Adolf, J., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2018). Detecting long-lived autodependency changes in a multivariate system via change point detection and regime switching models. *Scientific Reports*, 8(15637), 1–15. <https://doi.org/10.1038/s41598-018-33819-8>
52. Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425. <https://doi.org/10.1037/met0000085>
53. Liu, R., Chen, P., Aihara, K., & Chen, L. (2015). Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. *Scientific Reports*, 5(1), 17501. <https://doi.org/10.1038/srep17501>
54. Cabrieto, J., Tuerlinckx, F., Kuppens, P., Hunyadi, B., & Ceulemans, E. (2018). Testing for the presence of correlation changes in a multivariate time series: a permutation based approach. *Scientific Reports*, 8(769), 1–20. <https://doi.org/10.1038/s41598-017-19067-2>
55. Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous time approach to intensive longitudinal data: What, Why and How? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (Issue January). Springer. <https://doi.org/10.1007/978-3-319-77219-6>
56. Dablander, F., Pichler, A., Cika, A., & Bacilieri, A. (2020). Anticipating critical transitions in psychological systems using early warning signals: theoretical and practical considerations theory of critical slowing down. *Submitted*, 1–30. <https://doi.org/10.31234/osf.io/5wc28>
57. Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Recovery: Study protocol for a repeated intensive longitudinal n=1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *OSF, February*. <https://doi.org/10.31234/osf.io/fertq>

Detailed EWS study protocol paper of an intensive longitudinal design intended to monitor depressed patients' symptoms weekly and momentary experiences five times daily over four months during which they receive therapy and are expected to be more likely to show (sudden) symptom improvement.

58. Schreuder, M. J., Groen, R. N., Wigman, J. T. W., Hartman, C. A., & Wichers, M. (2020). Measuring psychopathology as it unfolds in daily life: Addressing key assumptions of intensive longitudinal methods in the TRAILS TRANS-ID study. *BMC Psychiatry*, 20(351). <https://doi.org/10.1186/s12888-020-02674-1>
59. Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research*, 21(12), e14475. <https://doi.org/10.2196/14475>
60. Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, 1–7. <https://doi.org/10.1017/S0033291719002514>

61. Riese, H., & Wichers, M. (2019). Comment on: Eronen MI (2019). The levels problem in psychopathology. *Psychological Medicine*, 1-2.. <https://doi.org/10.1017/S0033291719003623>
62. Kunkels, Y. K., Smit, A. C., George, S. V., Snippe, E., van Roon, A. M., Wichers, M., & Riese, H. (2020). Risk Ahead: Behavioral early-warning signals of increases in depressive symptoms during anti-depressant tapering. *OSF*, preregistration. <https://osf.io/dfmw3/>
63. Kunkels, Y. K., Riese, H., Knapen, S. E., Lek, R. F. R. der, George, S. V., Roon, A. M. van, Schoevers, R. A., & Wichers, M. (2021). Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study. *Translational Psychiatry*, 11(1), 1-10. <https://doi.org/10.1038/s41398-021-01465-w>

5

Transitions in Depression (TRANS-ID) Recovery: study protocol for a repeated intensive longitudinal n = 1 study design

Based on:

Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Recovery: study protocol for a repeated intensive longitudinal n = 1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *PsyArXiv*. <https://doi.org/10.31234/osf.io/fertq>

Abstract

Aim: The Transitions in Depression (TRANS-ID) Recovery study has gathered intensive longitudinal data in a group of individuals with depression during psychological therapy. The study was designed to gather high-resolution time series to monitor individual change processes in great detail, to allow for personalized predictions of shifts in depressive symptoms.

Method: The data collection combined experience sampling methods to assess momentary affect and behavior (five 27-item questionnaires a day, for four months), ambulatory assessment of physical activity and heart rate (continuous, for four months), and depression symptom assessments (weekly for six months, and monthly for the six months thereafter, twelve months total). In addition to a baseline diagnostic interview, baseline questionnaires covered a range of constructs, including overall psychopathology symptoms, medication use, psychological treatment history, alexithymia, life events, quality of life, and chronotype. After the four-month ambulatory assessment period, a personal report of the experience sampling data was given to each participant, and a semi-structured qualitative interview was conducted to evaluate participants' own retrospective experience of symptom changes during the research period.

Conclusion: The TRANS-ID Recovery study procedures and materials are described in detail in this document. The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen (reg. number: NL58848.04.16).

Introduction

The Transitions in Depression (TRANS-ID) Recovery study was aimed at gathering intensive longitudinal data in a group of depressed patients during psychological therapy. We expected that relatively swift and large symptom improvements may occur during therapy, and that these transitions may be preceded by generic early warning signals (EWS), such as seen in complex dynamical systems [1–6]. However, the nature of such transitions toward improvement of depressive symptoms, particularly the dynamics within the day and at the within-person level, is largely unexplored and undefined from a complex system framework. One of the aims of this project was thus to gain insight into the process of depressive symptom change during therapy, and create a better understanding of what might constitute a ‘critical transition’ toward fewer symptoms.

Per person, we gathered between 500–600 momentary measurements of mood (five times a day, for four months) using the Experience Sampling Method (ESM), as well as weekly symptom checklists and heart rate and actigraphy measurements. This enabled us to capture changes and improvements in depressive complaints and behavior within persons on several time scales, and to examine whether EWS such as autocorrelation, variance and overall network connectivity increase before large symptom jumps in repeated $N = 1$ investigations.

Study population

Intended sample size

We aimed to include 50 people, with an expected attrition rate of 10% resulting in an expected sample size of 45 participants. We expected sudden shifts in symptoms in about 50% of the group [7–11]. Participants entered the study between May 2017 and May 2019, with the data collection finalising in May 2020 — ending on a total of 41 participants that completed the full study procedures.

Selection criteria

Inclusion criteria were checked during the screening telephone call and in more detail during the baseline interview. To be eligible for participation the following criteria needed to be met: age ≥ 18 ; presence of depressive symptoms (Inventory of Depressive Symptomatology score ≥ 14); being set to start psychological treatment for depressive symptoms (within a month from the participation start date); capable of following the study procedures; and willing to and capable of giving informed consent.

Exclusion was decided on the following criteria: presence of a current manic episode or current psychotic symptoms; chronic depressive symptoms (> 2 years); reported primary diagnosis of a personality disorder; insufficient Dutch language skills to understand the diary questions, inability to work with a smartphone.

Procedure

Recruitment

To recruit study participants, a website was set up that contained information on the study aims, study procedures, and the eligibility criteria – a link to the full information letter was also provided. Potential participants indicated their interest in the study by filling out their phone number, full name, and date of birth on a form on the website, thereby consenting to be called by a member of the research project team. We advertised the study in local and national newspapers as well as via social media, including paid advertisements. In addition, information on the study and a link to the website was distributed among mental health care institutions.

Two paths to inclusion were possible (details below). Procedure (1) flyer only — a participant-led approach, in which participants registered through the study website after seeing the flyer or advertisement for the study. Procedure (2) flyer and information letter — a clinician-led approach, in which the full information package was given to participants by their mental health care provider, and a consent for contact-form replaced the need for participants to register themselves via the website. Procedure 2 was an amended approach (approved by the Medical Ethics Committee) for one mental health care institution, which allowed the steps before the baseline interview to progress more quickly and more patients to be included *before* they started therapy. A graphical representation of the recruitment procedures is provided in **Figure 5.1**.

Inclusion procedure 1: flyer only

Flyers and posters were distributed via mental health clinics and placed in community centers. The flyers contained a brief overview of the study and the link to the website. In the participating clinics, flyers were given to potential participants by a mental health care provider at their intake or sent along with the intake-invitation letter. Interested persons could then get in touch with the researchers by signing up for the study through the website.

Figure 5.1

Inclusion procedures 1 and 2 for TRANS-ID Recovery.

Note. **Dashed outline:** general recruitment approach for that procedure. **Filled boxes:** procedural steps that occur only if applicable.

Procedures 1: participants take the initiative to register via the website and are provided with an information pack after the telephone screening. Researchers await the signed consent form before continuing the inclusion procedures.

Procedures 2: participants are given the flyer and information letter by their mental health care provider. They can indicate with the clinician during intake if they are interested to be contacted by the researchers. Providing the information pack early allowed for a speedier procedure: if participants seemed eligible during the telephone screening, they could immediately schedule a baseline interview.

Screening telephone call

One of the members of the research team called new applicants within a few days after registration. During this telephone call, information about the study was provided and individuals were given the opportunity to ask questions. A first check of all the inclusion and exclusion criteria was performed.

Information letter

Eligible individuals who remained interested in study participation after the screening telephone call, were sent the patient information letter and an informed consent form by post and/or email. Individuals were asked to use the return-envelope to send the informed consent form within two weeks. If the informed consent form was not received after two weeks, they were sent a reminder. After the informed consent was received, an appointment was made via telephone for the introductory baseline interview.

Inclusion procedure 2: flyer and information letter

In some mental health care institutions, the intake procedures allowed the flyer and the full information letter to be distributed together. Individuals who had been referred to the mental-health institution because of depressive symptoms received the flyer and information letter with their intake-invitation. The information letter stipulated that the TRANS-ID Recovery study would be discussed with them during intake. The practitioner asked whether the patient would be interested in study participation and whether they agreed to be contacted by the researchers with more information. If they agreed, patients confirmed this by signing a contact consent form with their contact phone number and signature.

Screening telephone call

Patients who signed the contact consent form during intake, were called by one of the members of the research team. During the telephone call, information about the study was provided and individuals were given the opportunity to ask questions. A first check of all the inclusion and exclusion criteria was performed. As they had previously received the full information letter, they were asked if they

were still interested to participate in the study and if so, an appointment for the introductory baseline interview was set straightaway at the end of the screening telephone call.

Baseline assessments

Pre-interview inclusion check

To capture the period of change during therapy as closely as possible, the baseline interview was planned in the range of one month before a patient's treatment start, to maximum two weeks after the start of therapy. In some cases this meant that patients on a waitlist for treatment also had to wait for the start of the study, and in other cases participants were excluded for being too far into the therapy process.

A week before the baseline interview, an e-mail invitation for the online baseline assessments was sent (see **Table 5.1** for the list of instruments, and **Figure 5.2** for an overview of the entire study timeline). Completing the questionnaires took about 20 minutes, maximum one hour. The score on the Inventory for Depressive Symptomatology—Self Report (IDS-SR)[12,13] was checked by a researcher before the interview took place to determine whether individuals met the inclusion criterion of 'presence of depressive symptoms'. A score lower than 14 on the IDS-SR, meant exclusion from study participation.

Baseline interview

Diagnostic check: mini-SCAN

In the first half of the interview, history of depression, current depressive symptoms, chronic depression, as well as the presence of a psychotic or a bipolar disorder were assessed using the relevant prompts and diagnostic questions of the mini-SCAN [14]. If a diagnosis of current psychosis and/or mania was indicated, or if there was no current depressive episode indicated, further participation in the study was not possible, and the interview was ended at that point. Participants who had not done so yet (after following recruitment Procedures 2) signed the informed consent form.

Figure 5.2

Overview of the participation trajectory: duration of the various assessment periods, timing of the interviews and questionnaires, and the within-day structure for the ambulatory assessments.

Note. To further clarify the visual representation of the ambulatory assessments, participants did the following: they wore the actigraph continuously, and 1) upon waking: pressing the actigraph event marker, and five minutes of heart rate measurement; 2) during the day: five two-minute ESM questionnaires at three-hour intervals (the times in the figure are an example); 3) before sleep: five minutes of heart rate measurement, and pressing the actigraph event marker.

Training for the ambulatory assessments

Next, the physiological measurements were explained. The participant was asked to start wearing the actigraph continuously from that point onward. The use of the event-marker was demonstrated. Moreover, a step-by-step demonstration of the use of the Cortrium heart rate monitor was given, along with a visual instruction sheet, and a detailed trouble-shooting booklet. Participants were asked to repeat the heart rate measurement without help from the researcher to ensure they mastered all steps. They were given the device, a dedicated iPhone for the measurements, and a supply of electrodes to take home. All participants signed a lending-agreement for the electronic devices they were provided with.

After this, participants were given a detailed instruction for the ESM-questions, for which they could use their own smartphone. A test-questionnaire was sent to them, so they could practice filling out the diary questions, and the

weekly symptom questions. Researchers followed a detailed protocol to give concrete examples and prompts in reaction to a participant's intuitive response style. This was intended to train participants to best capture both daily and long-term changes in their mood over four months' time — i.e., training them to use the scale across their depressed and healthy range of feelings. If individuals did not manage to fill out the ESM-questions on the smartphone by themselves after practicing, they were excluded from the study. Furthermore, participants were allowed to add two additional questions to make the list more personalized. Researchers helped them with the phrasing to make the items as informative as possible. As the ESM-text messages were sent at the same times every day, the researchers took time to find the optimal schedule for each individual. At the end of the instruction, the key points for answering the ESM-questions were repeated, and any remaining questions answered. Appointments were made for two follow-up calls.

Practical details

The baseline interview took two to two and a half hours. Most interviews took place face-to-face at the University Centre for Psychiatry in Groningen. Travel costs for participants were reimbursed. However, in some cases, traveling to Groningen was not feasible. In those cases, interviews took place via video call on two occasions: a first appointment for the mini-SCAN diagnostic check (and in some cases, the ESM-explanation), and a second appointment for the (continuation of the) ambulatory assessment training. The break between the two sessions was required to send the research devices to the participant, which would only be done after eligibility was confirmed by the mini-SCAN.

Ambulatory assessment period (months 1–4)

Contact during the ambulatory period

The researcher who included a participant at the baseline interview remained a participant's primary point of contact throughout the ambulatory period. Regular moments for telephone contact were set during the ambulatory period. First, approximately two days after the baseline interview was completed, and again two weeks after the start. After that point, participants were contacted every three to four weeks, depending on preferences and availability.

During the first call, participants were encouraged to discuss any difficulties or questions that had arisen since the initial instructions. The researcher asked whether participants could recall the most important instructions from the ESM-training, and repeated and reinforced the points that were missed or unclear. Additionally, participants were asked if they had removed the actigraph at any time, or had any issues with the heart rate monitor. During the regular 3-week contacts, the researchers asked how the measurements were going, as well as registering any new information about how participants were doing in terms of mood, symptoms, and treatment.

The calls were intended to motivate the participant, answer questions about the study procedures, and to provide technical help. The researcher was also available by telephone and e-mail if participants needed help at other moments. Regular (weekly) data checks were done to verify that participants were completing the daily questionnaires at a rate of about 80%, and that the weekend symptom-questionnaires were completed. If it appeared that a participant had a low completion rate, the researcher called and/or emailed the participant to discuss and resolve any issues or concerns that were preventing the participant from filling out the questions.

About a week before the end of the ambulatory assessment period, participants were contacted to plan the evaluation interview. They were also asked whether any of the momentary mood items in the ESM-questionnaire were particularly interesting for them to have highlighted in their personal reports. They were also asked to bring any agenda or notebooks with them; this was intended as a memory aid during the interview to help them give a more precise timing of important events or psychological changes with more precision.

Evaluation (after month 4)

Evaluation questionnaires

An online evaluation questionnaire containing questions regarding the therapy, changes in medication, and evaluations of the ambulatory assessments (see Materials, and **Table 5.1**) was sent in the week after the end of the ambulatory assessment period. Participants were asked to complete it within seven days, or at least before the evaluation interview.

Evaluation interview

During this second appointment, participants returned the research devices, went through a semi-structured interview about the changes in their depressive complaints during the research period, and received a personal report of their data over time. The interview questions pertained to daily assessments, and events and experiences that might have influenced their depressive symptoms and the transition towards improvement (for more detail, see Materials below). The personal report consisted of descriptive graphs of their daily ESM-data and the weekly symptom scores, as well as a print-out of the comments they wrote in the ESM-questionnaires and in the weekly open questions. Noticeable peaks and shifts in the daily or weekly data were discussed with the participant, as were periods of missing data.

The full evaluation interview took approximately two hours. All interviews were led by one researcher (M.A.H.) to maintain consistency, and she served as a point of contact for all participants for the remainder of their study period. The evaluation interview typically took place face-to-face at the University Center of Psychiatry, Groningen. However, if preferable, the interview could take place via video call, in which case participants were asked to send the electronic devices back via post.

Table 5.1

Overview of the instruments used in the various (online) questionnaires.

Purpose	Instrument	Abbreviation	Method	Moment of assessment			Refs.
				Baseline (pre-AAP)	Evaluation (post-AAP)	Six-months (halfway)	
Descriptive characteristics	* Demographic information	-	SR	x			
	* Medication	-	SR	x	x		
	* History of and current psychological treatment	-	SR/INT	x	x		
	* Evaluation of ambulatory assessment and study procedures	-	SR		x		
Social functioning and well-being	The Self-Rating Scale of Happiness	-	SR	x			[23]
	* Meaning in Life Questionnaire—Short Form	MLQ-SF	SR	x			[24]
	Social Support List—Interactions and Discrepancies (first version)	SSL-I/SSL-D	SR	x			[25,26]
	General Self-Efficacy	GSE	SR	x			[27]
	Toronto Alexithymia Scale—Difficulty Identifying Feelings and Difficulty Describing Feelings subscales (12 items)	TAS-20 DDF and DIF subscales	SR	x	x	x	[28,29]
Personal risk factors	Munich Chronotype Questionnaire	MCTQ	SR	x			[30–32]
	List of Threatening Experiences	LTE	SR	x		x	[33]
	* * Long-term difficulties	-	SR	x			[34–36]
Psychopathology and Depressive Severity	Inventory of Depressive Symptomatology—Self Report	IDS-SR	SR	x			[12,13,37]
	Symptom Checklist-90—Revised	SCL-90-R	SR	x			[22,38,39]
	* * Short version of the Schedules for Clinical Assessment in Neuropsychiatry	mini-SCAN	INT	x			[15]

Note. AAP = Ambulatory Assessment Period; SR = Self-Report; INT = Interview

* = Developed or adapted for TRANS-ID (from other instruments)

* * = Developed or adapted by TRacking Adolescents' Individual Lives Survey (TRAILS)

* * * = Used for screening; inclusion criteria check

Follow-up symptom assessments (months 5–12)

Every Monday, a researcher checked whether any weekend symptom questionnaires (weekly and monthly) were missing and, if so, contacted those participants via telephone or e-mail to request that they complete the questionnaire by the end of the day.

Follow-up contact moments and questionnaires

In the eight months after the ambulatory assessment period, the moments of contact became less frequent, just as the assessments did. The evaluation interview took place two to four weeks after the ambulatory period ended, and after that, participants were contact twice more via telephone: at the halfway point at six months, and at the end of the study period, at twelve months. At the six-month halfway point, the weekend questionnaires were switched from a weekly protocol to monthly, and the participants received an email with a link to the online ‘six-months’ questionnaires (see **Table 5.1**).

Materials

Baseline, evaluation and six-months questionnaires

A selection of questionnaires was sent to participants before the baseline interview, after the four months of ambulatory assessment, and at six months (see **Figure 5.2**). For an overview of all (standard) instruments and the relevant assessment moments, see **Table 5.1**.

Mini-SCAN

The mini-SCAN is a validated semi-structured interview designed to assess psychiatric disorders [15], which automates the scoring and diagnosis via an online interface. Yes-or-no questions appear on screen for the interviewer to read out, followed by further questioning at the researcher’s discretion, until

they are satisfied to mark a symptom as present ‘yes’, or not ‘no’. For the purpose of this study, only the bipolar (mania), psychosis and depression screening prompts were used. The section on mood disorders was always administered, a list of approximately 30 questions, with three additional questions on psychotic characteristics only appearing if indicated by a previous answer.

Additionally, for this study, a relevant period for questioning needed to be determined. In particular, a period over which the complaints had been relatively stable and comparable over time. Thus, from the last noticeable, large shift in symptoms – be it an improvement or worsening of the complaints – until the day of the interview. If this period was too difficult to determine or too lengthy (e.g., a four months or more) to recall easily, the researcher proposed focusing on the past month. In total, the mini-SCAN interview took approximately 20-45 minutes.

Ambulatory assessment protocols

Experience Sampling protocol

The ESM measurements consisted of five measurements a day, at fixed intervals of three hours, over the course of four months (total ca. 600 prompts per participant). The questionnaires were sent via text-message, with a reminder-text after ten minutes and a half hour time window for completion. A measurement took about two minutes to complete and the timings were set according to the sleep-wake habits of a participant, with specificity up to fifteen minutes to be optimally convenient for a participant.

Experience Sampling instrument

The full 27-item ESM-questionnaire is available in the appendix (**Table A5.1**), as well as in the project’s online repository (<https://osf.io/a8572/>). Items were rated on a visual analogue scale, with anchors ‘not at all’, to ‘very much’ and the underlying scale ranging from 0 to 100, at two-decimal point precision. The items list includes ten momentary affect items based on the circumplex model of affect [16], momentary cognition items, levels of physical discomfort, current social environment and the degree of physical activity and types of activities in the past three hours. Participants were also given the opportunity to add a maximum

of two personalized items. Finally, there was a text-field in which any remarkable or influential moments could be reported.

Heart rate measurement protocol

Heart rate measurements were done using the Cortrium C3 device [17,18], with three ECG-channels and a Bluetooth connection to a dedicated research iPhone. The Cortrium C3 device has been tested for validity against the golden standard [19]. In order to capture a stationary heart rate signal, it is preferable that the measurements are preceded by a relatively calm period of activity. Therefore, participants were asked to complete their five-minute measurements every morning just after waking up, and in the evening before going to bed.

To start a measurement, participants attached three electrode stickers to the Cortrium device, turned it on, and attached it to the middle of their chest, on their sternum and ribs. They were instructed to find a comfortable sitting position, in which they could sit still for five minutes. Because of the live connection to the phone, participants could check whether the signal from all three channels seemed in order before they started each recording. If all was in order, they were to ‘freeze’ the screen so they would not be viewing their own heart rate, and were encouraged to place the phone away from them, and set an alarm for five minutes. In the Cortrium app on the iPhone, they then pressed ‘start recording’. Talking, drinking and moving were not allowed during that period. After the five minutes passed, participants ended the recording and, with a click of a button, sent the data to an anonymized and secure UMCG server. All data were also stored locally on the iPhone.

Actigraphy protocol

The MotionWatch 8 by CamNtech [20,21] was used for the actigraphy measurements. The accelerometer measured cumulative uni-axial MotionWatch-counts over 60-second epochs.

Participants were asked to wear the wristwatch-like device continuously during the ambulatory assessment period of four months. Twice a day, the event marker button on the actigraph was to be pressed: when they woke up and when they went to sleep. The MotionWatch 8 is waterproof, so participants could wear the device in the shower and swimming pool as well. Only for sauna visits, certain contact-sports, and in rare cases if severe skin irritation occurred was non-wear warranted. To mark a period of non-wear, participants were asked to press the event marker twice in quick succession, in order to mark the start and end of a non-wear period distinctively from a sleep/wake period.

Weekly and monthly symptom questionnaires

Assessment of symptoms and individual perception of changes and events

To assess the course of depressive symptoms over time, participants completed fourteen items of the SCL-90 depression subscale every week in the first six months, and monthly in the last six months of participation (see **Figure 5.2**). The SCL-90 depression subscale [22] consists of sixteen items, but the questions on suicidal ideation, and self-esteem were removed for this study as they were deemed potentially too burdensome to fill in at a weekly frequency for a vulnerable population such as depressed patients.

In addition to the symptom checklist, participants were given several optional questions, which they could complete if they were applicable. First, three open questions: 1) have they recently experienced a noticeable shift in symptoms, 2) do they expect a noticeable shift in symptoms in the near future, and 3) have people in their close environment remarked on any changes, and *if so* (for 1, 2, and 3), what and when? Finally, they were given room to describe and rate any important or impactful experiences of the past week or month — a maximum of three at week-level, and five at the month-level assessment. The full weekly questionnaire took about five (additional) minutes to complete.

During the ambulatory assessment period (months 1–4)

In the first four months, the weekly questionnaire was sent along with the ESM-prompts during the weekend. From the first measurement time point on Saturday, a participant's ESM-questionnaire was immediately followed by the weekly depressive symptom questionnaire. The weekly questionnaire was presented again after each ESM-prompt — the same 30-minute completion window applied — until completed or until the very last prompt on Sunday evening.

Follow-up symptom assessment periods (weekly months 5–6; monthly months 6–12)

For the two months of weekly, and six months of monthly (every 28 days) assessments, participants received a link to the questionnaire via text or email on the weekend. Timing of the texts was kept consistent with the participant's ESM-schedule, and prompts were sent until the questionnaire was completed. A maximum of ten prompts (no reminder texts after ten minutes) were sent on Saturday and Sunday — five per day, three hours apart. The questionnaire could be completed at the participant's convenience within 72 hours of the first prompt (no 30-minute completion window).

Evaluation interview

The semi-structured evaluation interview was designed specifically for the TRANS-ID Recovery study, with the intent to gather qualitative data about the changes in depression that participants experienced during the first four months of the study and what the process of change looked like for an individual. Sessions were recorded on audio and answers were noted down verbatim during the interview.

The interview contained open questions, probes for deeper questioning, and several questions on a scale from 1 (not at all) — 7 (very much). The overall tone of the interview put emphasis on a participant's subjective experience, and their views and ideas of how, why, and when their complaints changed over time. Questions included: compared to the start of the study period, which depressive symptoms have changed?; what changed first?; was the change rapid or gradual?; what was the most important driver of change?; what symptoms have improved and what has not?; what components of the therapy were helpful?; how strongly

do they feel they have 1) experienced a decrease in depressive symptoms, and 2) recovered from their depressive symptoms? on a scale from 1–7; what were important events or insights that influenced their symptoms? Participants were also asked questions about the added value of monitoring daily experiences during psychological treatment and to comment on the data collection and the study in general. The evaluation interview questions are available in Dutch upon request.

Ethics and dissemination

Participation in this study was voluntary and written informed consent for scientific publication of their anonymised data was obtained from all participants, in accordance with the Declaration of Helsinki. All project-leading researchers undertook a training to increase their sensitivity to detect (legal) incapacity in participants. Questions to test for full understanding of study obligations and motivation (uncoerced willingness) to participate were part of routine protocol. All procedures were approved on December 12, 2016 by the Medical Ethical Committee of the University Medical Center Groningen (reg. number: NL58848.04.16).

Participants were financially compensated for study participation. Participants could receive a maximum of €250,- for study participation, depending on their compliance with completing the daily, weekly, and monthly questionnaires.

Because the TRANS-ID Recovery group concerned a vulnerable group of individuals who were selected for their depressive complaints, a protocol was prepared to detect cases in which (immediate) clinical care was warranted. For instance, if a participant disclosed any suicidal ideation or plans. The protocol was developed together with psychologists and psychiatrists from the University Centre for Psychiatry in Groningen.

ESM and questionnaire data was collected via and stored in a protected online environment (governed by RoQua, www.roqua.nl), to which only selected researchers could gain access via two-step authentication. The safety of this system is guaranteed by, and meets the standards of the University Medical Center Groningen (UMCG). Participants were given a unique identification number,

the key to which was also only available in the protected RoQua environment to project coordinators. Any data with personal identifiers (informed consent forms, therapist information) were kept in paper form in a locked cabinet, and in digital form on a dedicated server with no link to the coded data. Anonymous interview data and electronic written records of communications with participants were stored under participant identification numbers, on a dedicated protected server.

After the study's main results have been published, the data can be shared under certain conditions upon reasonable request and in accordance with data security and legal privacy requirements. The research results will be submitted for publication in (inter)national medical or psychological journals. Public disclosure and publication of the research data will be according to the Central Committee on Research Involving Human Subjects (CCMO) statement on publication policy.

References

1. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical Slowing Down as a Personalized Early Warning Signal for Depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>
2. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
3. Hayes, A. M., & Hayes, S. (2007). Towards an understanding of the process and mechanisms of change in cognitive behavioral therapy: Linking innovative methodology with fundamental questions. *Clinical Psychology Review*, 27(6), 679–681. <https://doi.org/10.1016/j.cpr.2007.01.006>
4. Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2019). Critical Fluctuations as an Early-Warning Signal for Sudden Gains and Losses in Patients Receiving Psychotherapy for Mood Disorders. *Clinical Psychological Science*, 216770261986596. <https://doi.org/10.1177/2167702619865969>
5. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59. <https://doi.org/10.1038/nature08227>
6. Schiepek, G. K., Aichhorn, W., & Schöller, H. J. (2017). Monitoring change dynamics: A nonlinear approach to psychotherapy feedback. *Chaos and Complexity Letters*, 11(3), 355–375.
7. Masterson, C., Ekers, D., Gilbody, S., Richards, D., Toner-Clewes, B., & McMillan, D. (2014). Sudden gains in behavioural activation for depression. *Behaviour Research and Therapy*.
8. Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(1), 93–101. <https://doi.org/10.1037/a0026455>
9. Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, 88(2), 119–127. <https://doi.org/10.1037/ccp0000469>
10. Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444–447. <https://doi.org/10.1037/0022-006X.70.2.444>
11. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
12. Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J., & Burns, C. (1986). The inventory for depressive symptomatology (IDS): Preliminary findings. *Psychiatry Research*, 18(1), 65–87. [https://doi.org/10.1016/0165-1781\(86\)90060-0](https://doi.org/10.1016/0165-1781(86)90060-0)
13. Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26(3), 477–486. <https://doi.org/10.1017/s0033291700035558>
14. Nienhuis, F. J., Van De Willige, G., Rijnders, C. A. T., De Jonge, P., & Wiersma, D. (2010). Validity of a short clinical interview for psychiatric diagnosis: The mini-SCAN. *British Journal of Psychiatry*, 196(1), 64–68. <https://doi.org/10.1192/bjp.bp.109.066563>

15. Nienhuis, F. J., Van De Willige, G., Rijnders, C. A. T., De Jonge, P., & Wiersma, D. (2010). Validity of a short clinical interview for psychiatric diagnosis: The mini-SCAN. *British Journal of Psychiatry*, 196(1), 64–68. <https://doi.org/10.1192/bjp.bp.109.066563>
16. Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734. <https://doi.org/10.1017/S0954579405050340>
17. *Instrumentation—Cortrium C3*. (2020). <https://osf.io/t7cfe/>
18. *Cortrium.com*. (2019). <https://www.cortrium.com/>
19. Kunkels, Y. K., van Roon, A. M., Wichers, M., & Riese, H. (2021). Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring. *Psychophysiology*, 58(10), e13898. <https://doi.org/10.1111/psyp.13898>
20. *MotionWatch 8 Overview—CamNtech*. (2020). <https://www.camntech.com/products/motionwatch/motionwatch-8-overview>
21. *Instrumentation—MotionWatch 8*. (2020). <https://osf.io/7q64p/>
22. Arrindell, W. A., & Ettema, J. H. M. (1981). Dimensional structure, reliability and validity of the Dutch version of the Symptom Checklist (SCL-90): Data based on a phobic and a ‘normal’ population. [Dutch]. *Nederlands Tijdschrift Voor de Psychologie En Haar Grensgebieden*, 36(2), 108.
23. Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. In *Social Behavior and Personality* (Vol. 34, Issue 2, pp. 139–150). <https://doi.org/10.2224/sbp.2006.34.2.139>
24. Steger, M. F., & Samman, E. (2012). Assessing meaning in life on an international scale: Psychometric evidence for the meaning in life questionnaire-short form among Chilean households. *International Journal of Wellbeing*, 2(3), 182–195. <https://doi.org/10.5502/ijw.v2i.3.2>
25. van Sonderen, E. (2012). *Measuring Social Support [Het meten van sociale steun met de Sociale Steun Lijst-Interacties (SSL-I) en Sociale Steun Lijst Discrepanties (SSL-D): Een handleiding]* (2nd ed.). Noordelijk Centrum voor Gezondheidsvraagstukken, NCG.
26. Bridges, K. R., Sanderman, R., & Van Sonderen, E. (2002). An English language version of the social support list: Preliminary reliability. *Psychological Reports*, 90(3 PART 1), 1055–1058. <https://doi.org/10.2466/pr0.90.3.1055-1058>
27. Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and Validation. *Psychological Reports*, 51(2), 663–671. <https://doi.org/10.2466/pr0.1982.51.2.663>
28. Meganck, R., Vanheule, S., & Desmet, M. (2008). Factorial validity and measurement invariance of the 20-item Toronto alexithymia scale in clinical and nonclinical samples. *Assessment*, 15(1), 36–47. <https://doi.org/10.1177/1073191107306140>
29. Meganck, R., Markey, S., & Vanheule, S. (2012). Reliability and factor structure in an adolescent sample of the Dutch 20-item Toronto Alexithymia Scale. *Psychological Reports*, 111(2), 393–404. <https://doi.org/10.2466/02.08.15.PR0.111.5.393-404>
30. Zavada, A., Gordijn, M. C. M., Beersma, D. G. M., Daan, S., & Roenneberg, T. (2005). Comparison of the Munich Chronotype Questionnaire with the Horne-Östberg’s morningness-eveningness score. *Chronobiology International*, 22(2), 267–278. <https://doi.org/10.1081/CBI-200053536>
31. Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2012). Munich Chronotype Questionnaire (MCTQ). In *STOP, THAT and One Hundred Other Sleep Scales* (pp. 245–247). Springer Science+Business Media, LLC. https://doi.org/10.1007/978-1-4419-9893-4_58

32. Roenneberg, T., Wirz-Justice, A., & Mrosovsky, M. (2003). Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms*, 18(1), 80–90. <https://doi.org/10.1177/0748730402239679>
33. Brugha, T., Bebbington, P., Tennant, C., & Hurry, J. (1985). The List of Threatening Experiences: A subset of 12 life event categories with considerable long-term contextual threat. *Psychological Medicine*, 15, 189–194. <https://doi.org/10.1017/S003329170002105X>
34. Van De Willige, G., Ormel, J., & Giel, R. (1995). Etiological meaning of severe life events and long-term difficulties in the onset of depression and anxiety [Etiologische Betekenis Van Ingrijpende Gebeurtenissen En Langdurige Moeilijkheden Voor Het Ontstaan Van Depressie En Angststoornissen]. *Tijdschrift voor Psychiatrie*, 37(9), 689–703.
35. Overview of measures collected in TRAILS. (2020). <https://www.trails.nl/uploads/fckconnector/6b976452-c8e4-4955-875b-d1170814efb8>
36. Amon, P., Olak, K., Ormel, J., Huisman, M., Verhulst, F. C., Oldehinkel, A. J., & Burger, H. (2009). Life Stressors as Mediators of the Relation Between Socioeconomic Position and Mental Health Problems in Early Adolescence: The TRAILS Study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(10), 1031–1038. <https://doi.org/10.1097/CHI.0b013e3181b39595>
37. Rush, A. J., Carmody, T., & Reimnitz, P.-E. (2000). The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research*, 9(2), 45–59. <https://doi.org/10.1002/mpr.79>
38. Arrindell, W. A., & Ettema, J. H. M. (2003). *SCL-90: Manual to a multidimensional psychopathology indicator [SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator]* (Pearson, Ed.; 2nd ed.). Swets & Zeitlinger.
39. Derogatis, L. R. (1977). *SCL-90-R: administration, scoring and procedures manual-I for the R(evised) version*. John Hopkins University School Medicine.

Appendix

Table A5.1

TRANS-ID Recovery Experience Sampling questionnaire items.

#	Item	Answer options and scale anchors
1	I feel content	Not at all — Very much ^o
2	I feel down	Not at all — Very much ^o
3	I feel irritated	Not at all — Very much ^o
4	I feel cheerful	Not at all — Very much ^o
5	I feel listless	Not at all — Very much ^o
6	I feel restless	Not at all — Very much ^o
7	I feel stressed	Not at all — Very much ^o
8	I feel energetic	Not at all — Very much ^o
9	I feel at ease	Not at all — Very much ^o
10	I feel tired	Not at all — Very much ^o
11	I am experiencing physical discomfort (e.g., headache, backache, dizziness, nausea)	Not at all — Very much ^o
12	I can concentrate well	Not at all — Very much ^o
14	I am worrying	Not at all — Very much ^o
15	I want to be with other people	Not at all — Very much ^o
16	My thoughts about myself are currently...	Very negative — Very positive ^o
17	Who am I with at this moment:	☐ Nobody ☐ Partner ☐ House mates ☐ Family ☐ Family (living elsewhere) ☐ Friends ☐ Colleagues/class mates ☐ Caregiver ☐ Acquaintances ☐ Strangers
18a/b	◆ <i>Optional personalized question (momentary or about the past three hours)</i>	
19	The past three hours I was physically active	Not at all — Very much ^o
33	The past three hours I have spoken with others	Not at all — Very much ^o
20	I have had contact with (an)other(s) in the following way:	☐ Spent time with somebody ☐ (Video)calling ☐ WhatsApp-exchange/chat ☐ I had no contact with others

#	Item	Answer options and scale anchors
21	The past three hours I have done the following:	☐ Eating ☐ Household tasks/groceries/admin ☐ Self care (e.g., shower, brushing teeth, clothes) ☐ Working/studying ☐ Taking care of (grand)children ☐ Psychological treatment ☐ Sports/walking/cycling ☐ Something calm (e.g., watching TV, internet) ☐ Hobby (e.g., make music, DIY) ☐ Outing (e.g., visit to town, a museum) ☐ Something with another/others ☐ Making contact with someone ☐ Be on the way (somewhere) ☐ Sleeping ☐ Resting/nothing
22	Think of the most striking event or activity in the past three hours. L How (un)pleasant was this event?	Very unpleasant — Very pleasant [◊]
23	L How impactful was this event?	Not at all — Very much [◊]
24	☾ I am experiencing tension due to something that will occur in the near future	Not at all — Very much [◊]
25a	* I look forward to the rest of the day	Not at all — Very much [◊]
25b	☾ I look forward to tomorrow	Not at all — Very much [◊]
26a/b	◆ <i>Optional personalized question (end of the day)</i>	
27a	Is this the first questionnaire you are completing today? L Yes → I have slept well last night	Yes / No Not at all — Very much [◊]
28	◆ How many cigarettes have you smoked in the past three hours?	0 — 25
29	◆ How many cups of coffee and glasses of caffeinated soda did you drink in the past three hours?	0 — 25
30	☾◆ How many glasses of alcohol did you drink today?	0 — 25
32	Note down any remarks: was there something that influenced your answers?	{text field}

Note. This table, and the original Dutch version of the questionnaire can also be found in the online project repository <https://osf.io/a8572/>. Numbering in this table is kept consistent with the TRANS-ID dataset.

[◊]: Item rated on a Visual Analogue Scale, underlying range of 0.00 to 100.00.

◆: Item added only if applicable for participant.

☾: Item appears only in the last questionnaire of the day.

*: Item occurs only the first four questionnaires of the day.

6

The Duration-adjusted Reliable Change Index (DaRCI): defining clinically relevant symptom changes of varying durations

Based on:

Helmich, M. A. (In preparation). The Duration-adjusted Reliable Change Index (DaRCI): Defining clinically relevant symptom changes of varying durations. *PsyArXiv Preprint*. <https://doi.org/10.31234/osf.io/q7ch9>

Abstract

Aim: Identifying relevant symptom shifts as they unfold can be challenging, as the time period over which they take place is not uniform for all people. This paper proposes an adaptation of the well-established Reliable Change Index (RCI) that allows researchers and clinicians to explore the presence of symptom changes of varying durations in individual patients' time series: the Duration-adjusted RCI (DaRCI).

Method: The DaRCI takes the RCI cut-off score for change between two points as a starting point, and proportionally extends this measure over multiple observations, while maintaining reliability at a given confidence level. Researchers must choose the relevant time period between two observations, and additional increments are added accordingly. To illustrate the ability of this method to detect changes of various durations, simulated depressive symptom time series with varying degrees of discontinuity and overall mean change in scores were used.

Results: The results show that the DaRCI thresholds over two, three and four observations were effective at identifying the simulated change periods over multiple time points, starting from relatively gradual change slopes (picking up reliable changes in 20-60% of simulated time series if the overall change was large enough), to highly discontinuous changes (up to 100% accuracy).

Conclusion: The DaRCI may be particularly useful for identifying shifts in symptoms that appear relatively abrupt, which can help indicate when a patient is showing significant improvement or deterioration. Its ease of use makes it suitable for application in the clinical context, and is a promising method to explore different change durations in clinical populations.

Keywords: duration of change; within-person; reliable symptom change; repeated measurement; idiographic method

Background

Identifying if and when a patient's psychological symptoms have changed in a clinically relevant way is an integral part of many treatment settings, and studies of therapeutic interventions. Typically, methods to determine clinical change identify when a patient reaches a score over or under a certain threshold (e.g., within the range of a non-clinical population norm score [1]), or shows a change in scores that meets a cut-off (e.g., 50% reduction [2]) or combination of criteria (e.g., minimal score reduction and statistical significance [3]). The Reliable Change Index (RCI) is a widespread method that determines whether the variability in a person's measurements on a symptom questionnaire are more likely to be due to the instrument's precision (measurement error), or due to an actual clinical change [1,4–6]. Determining if a symptom change indicated by a questionnaire is reliable is important to establish whether a drop or increase in scores is not merely due to chance, and can aid decisions on whether to start, continue, end, or alter the intensity of treatment [7].

What most reliable change methods do not incorporate, however, is the time frame over which a change occurred. Some methods focus on pre- to post-treatment change — which can take weeks or months — whereas other methods aim to identify symptom shifts as they occur between or even *within* therapy sessions [8–10]. For instance, in research on depression and anxiety, “sudden gains” and “sudden losses” in symptoms are identified as changes between therapy sessions that combine a pre-defined minimum magnitude of change with the requirement that it takes place within a short period of time, most often a week [11–14]. Yet, symptom reductions of 50% or more over three to four weeks of treatment have also been described as “rapid” in studies of early response, and are certainly considered clinically relevant [15,16].

Differences in the duration of change should be taken into account when identifying a period of significant improvement, as the time it takes to change can be clinically meaningful in itself [17,18]. Imagine two individuals who show the same reliable reduction in scores (e.g., –15 points). For person A the cut-off is met relatively suddenly, from one week to the next, while person B shows slower, gradual improvement and meets the same reduction only after three weeks. These individuals both show a reliable improvement according to this criterion, but due to

the different timings, the process appears as a sudden change for one (person A) and a gradual improvement for the other (person B). Here, using a single criterion is effective at detecting a minimal reduction for both people, but it does nothing to capture the difference in “velocity” of the changes.

Now compare the changes shown by those two individuals to those of a third person. If person C showed a pronounced decline over multiple weeks (e.g., steps of -12, -9, -11 points), this pattern meets the minimal change criterion of -15 points after the second step (-21 compared to the first point), but unlike person A, the changes between two adjacent points (over one week) never meet the minimal change criterion. Yet, the overall change shown by person C is larger overall (-32) and may also be considered a clinically relevant and “rapid” improvement, despite the longer time frame (compared to person A; cf. [15,16]). Using a single cut-off that is unadjusted for the duration of a symptom change allows one to determine reliable change over one interval (person A), but runs the risk of missing smaller within-week changes that continue over multiple weeks and culminate into a reliable change over a longer period (person C). Thus, if different durations could be accounted for by a change criterion that requires that longer changes must also be larger as a whole, person C’s consistent improvement may be identified as a reliable change that is similar in its clinical relevance to the rapid one-week change shown by person A. In short, testing one criterion for changes with no regard for time, or considering only shifts of one specific duration (e.g., one week, as illustrated above) disregards clinically relevant changes that occur over consecutive time points.

As routinely measuring (former) patients’ psychological complaints and collecting time series of repeated assessments within individuals has become common practice in the context of therapy and relapse prevention [19–24], studies mapping repeated symptom assessments have shown that psychopathological change is often characterized by nonlinearity and abrupt changes [25–28]. Clinically, this is relevant as particularly sudden shifts may indicate that the patient may have experienced a ‘transition’ to a better or worsened state, which could be predictive of their treatment outcomes [11,13,14,28–30]. Sudden gains (symptom improvements), for instance, may occur when a therapy session or intervention has been especially effective [14,31–33], while sudden losses (deteriorations) indicate when a patient is less likely to benefit from treatment, and should be identified as

soon as possible to prevent the treatment from failing [10,34]. A pattern of steady early improvement over the first few treatment sessions has also been linked to better treatment outcomes [16,31,32,35,36], and conversely, early changes that did not conform to the expected response patterns have been linked to poorer outcomes [37]. Thus, change patterns of varying duration and magnitude have been described in the psychotherapy context [38–40], but flexible methods to detect shifts, as they unfold, and over different durations, are lacking.

To summarize, various methods exist to determine whether a relevant change in symptoms has occurred at the within-person level, but these typically make no particular assumption about the time it took for the symptoms to change. Furthermore, even those that examine sudden gains and losses in repeated assessment data do not provide solutions to identify abrupt changes that extend over multiple time points or therapy sessions. More data-intensive methods may focus on testing the significance of an overall symptom change with a regression model [6,41–43], or try to identify abrupt shifts with a change-point model [44,45], but these methods may be difficult to implement in real-time in the course of clinical practice, as they require more data than may be available in early stages of treatment, and also a fair level of statistical knowledge to be conducted (cf. [46]). Thus, adjusting a simple method like the RCI may help to identify both reliable symptom changes that are more sudden, as well as reliable slower gradual symptom changes over time.

In this brief report, I propose a method, based on the well-established Reliable Change Index [4], which allows researchers and clinicians to explore the presence of symptom changes of varying durations in individual patients' time series: the Duration-adjusted Reliable Change Index (DaRCI). A simulation study is conducted to test the DaRCI's ability to pick up periods of relevant change, and discontinuous change in particular, in the context of a larger overall symptom time series.

Method

Material

This study used the Dutch Symptom Checklist-90 (SCL-90) depression subscale [47,48] as a basis for our illustration of the DaRCI method. This questionnaire consists of sixteen items that ask to what extent one was bothered by particular depressive symptoms (e.g., “feeling blue”) on a five-point scale ranging from “not at all” to “very much”.

Reliable Change Index (RCI)

The RCI was developed as a method to ensure that any identified pre- to post-treatment change was a reliable change that could be distinguished from measurement error [1,4]. The RCI can be used to calculate a threshold at which the difference between a pre- and post-measurement for one person is, with a 95% two-tailed confidence level, “unlikely to occur without actual change” (Ref. [4], p. 14). It uses the standard error of measurement (SE_m) from a population norm of a given instrument to calculate the minimum score necessary to exceed a change that could be due to measurement inaccuracy (SE_{diff} , standard error of difference). Note that the RCI uses between-persons information for the standard error of measurement¹ regarding the spread of the distribution of test-retest reliability given no change, and uses this to test whether within-person changes are reliable. It is defined as:

$$RCI = SE_{diff} \times Z,$$

$$SE_{diff} = \sqrt{2(SE_m)^2}.$$

For this study, $SE_m = 4.37$; taken from the Dutch SCL-90 depression subscale, based on the psychiatric outpatient norm group of 5,621 patients (cf. SCL-90 manual; Ref. [47]). Thus, $SE_{diff} = \sqrt{2(4.37)^2} = 6.18$, and $RCI_{95} = 6.18 \times 1.96 = 12.113$. Where 1.96 is the Z-score for a 95% confidence level, and a score of ~ 13 is a reliable change between two observations on the SCL-90 depression

¹ Ideally, the same time frame is used for the calculation of the standard error of measurement (based on the test-retest reliability [6,41]) as for the investigated difference score. However, in practice, this time frame is usually not considered and the RCI is calculated with the available information and applied to pre-posttest scores.

subscale. The score is rounded up to ensure the confidence level is maintained, as the (difference) scores are always integers, and obtaining a score of 12.113 is not possible.

Duration-adjusted RCI (DaRCI)

The DaRCI is proposed as an adaptation of the RCI, with the aim to capture symptom changes of varying durations, particularly changes that appear as sudden or large in overall scope. The DaRCI requires setting a fixed time period between two points (e.g., one week) as a basis for extension when more points are added (i.e., when testing change over longer durations), and allows one to calculate thresholds for each additional increment of time (e.g., each added week). Like the regular RCI, the DaRCI tests the difference score between two points, but it accounts for instances where the two compared points are farther apart by proportionally increasing the change threshold.

To detect reliable change from a given starting point (t_{start}) to the last observation in a chosen period (t_n) while considering the additional time between these observations, the original RCI threshold (based on n observations = 2) is divided by 2, and multiplied by the number of observations in the range of interest. To calculate the DaRCI critical change threshold for a particular Z -value (confidence level) and number of observations (n):

$$\text{DaRCI} = \left(\frac{SE_{\text{diff}}}{2} \right) \times Z \times n.$$

Essentially, the RCI threshold is reduced to a range of uncertainty around a single point, and then proportionally extended for the number of observations (i.e., period of time) at hand, while maintaining the chosen confidence level (e.g., 95%). In doing so the DaRCI provides a way for researchers to detect symptom changes over various durations with the same degree of reliability, even if some shifts take longer. Applied to our illustrative sample and instrument, with the SE_{diff} for the SCL-90 depression subscale incorporated: $\text{DaRCI} = (6.18/2) \times Z \times n$. For 95% confidence, this can be rewritten as: $\text{DaRCI}_{95} = (6.18/2) \times 1.96 \times n = 6.0564n$. Where 6.0564 is the aforementioned range of uncertainty around one point. The DaRCI thresholds for different increments are presented in **Table 6.1**. Using this method, higher and lower confidence levels may also be calculated and explored,

and all DaRCI critical threshold scores are rounded up to maintain the cut-off $\geq Z$ requirement.

Alternatively, we can calculate the Z -score for a particular change over time (i.e., the difference between two assessments (Δy), divided by the number of observations (n) in the range of interest. This can be useful to compare the relative magnitude of multiple identified changes. In formula form:

$$\text{DaRCI}_z = \left(\frac{2 \frac{\Delta y}{n}}{SE_{\text{diff}}} \right) \times Z \times n.$$

The DaRCI does not prescribe over which time period the change must take place. Instead, the time interval between two observations serves as a basis for the other increments over which it is extended. Researchers must choose the duration of time over which detecting a change would be of interest (e.g., based on the clinical literature, pilot studies, or other conceptual grounding). For instance, if the change between t_{start} and t_2 occurs over one week, then the DaRCI will be proportionally extended for two, three, four weeks, etc. (as the limits of a given scale allow), if t_{start} to t_2 occurs over two weeks, reliable change can be calculated for four, six, eight weeks, etc. It is thus an important theoretical and clinical consideration what durations of “change” are relevant to capture, although the DaRCI lends itself precisely to exploring a number of different durations.

Table 6.1

The Duration-adjusted Reliable Change Index (DaRCI) over different durations and confidence levels for the SCL-90 depression subscale.

	<i>n</i> = 2	<i>n</i> = 3	<i>n</i> = 4	<i>n</i> = 5
	e.g., 1 week	e.g., 2 weeks	e.g., 3 weeks	e.g., 4 weeks
90% confidence $Z \geq 1.645$	10.17 $\approx$ 11	15.25 $\approx$ 16	20.33 $\approx$ 21	25.42 $\approx$ 26
95% confidence $Z \geq 1.96$	12.11 $\approx$ 13	18.17 $\approx$ 19	24.23 $\approx$ 25	30.28 $\approx$ 31
99% confidence $Z \geq 2.58$	15.94 $\approx$ 16	23.92 $\approx$ 24	31.89 $\approx$ 32	39.86 $\approx$ 40

Note. To calculate the cut-off scores, the following formula was used: $\text{DaRCI} = (6.18/2) \times Z \times n$. Where, n is the number of observations over which the change occurs. The scores are rounded up to maintain the required minimum Z -score. Change between two observations is equal to the original RCI.

Analysis

To test the accuracy of the DaRCI thresholds for change over two (Tn2), three (Tn3) and four (Tn4) time points, the frequency at which modeled shifts were correctly identified in simulated repeated symptom assessments was examined. A set of 10,000 time series with a length of 15 points was simulated, to reflect a typical duration of psychological treatment [49,50]. Each time series was drawn from a randomized normal distribution with a mean of 0 and variance concurrent with the SCL-90 depression subscale SE_m of 4.37. The time series were then fitted to different overall symptom reductions ($L = -10, -20, -30, -40$), and different degrees of discontinuity. Specifically, five shapes of increasingly abrupt change around the midpoint of the time series were modeled: (a) gradual change as a linear function (no internal shift); (b) a smooth curve formed by a sigmoid function with a slope of 1 at the origin (k), and the midpoint set to point 7.5 (the steepest decline occurs between observation 7 and 8); three mean shifts where change takes place over an increasingly short time period: (c) a step-function with the decrease occurring over four points ($n = 4$); (d) a step-function with the decrease occurring over three points ($n = 3$); (e) a step-function with the shift occurring from one point to the next ($n = 2$).

These change patterns were chosen to explore the interplay between the strength of the overall slope and different degrees of discontinuity. For instance, at an overall reduction of 10 points and a linear function, any identified ‘change’ would be due to a random fluctuation (as the (Da)RCI thresholds start at 13 points for $n = 2$). At the other extreme, at -40 points of overall change, we would expect that, even if the discontinuities are more gradual (like in the sigmoid function, or the step change over four points), the DaRCI thresholds will often detect the modeled shifts around the correct time period. Furthermore, three out of the five models are step-functions, which allow certainty about when a shift starts and ends, as this is explicitly modeled (in contrast to the sigmoid curve). Moreover, the durations of those shifts correspond to the different durations of the DaRCI thresholds we tested: over two, three and four points, and should thus provide a good test of the thresholds’ sensitivity.

In short, $DaRCI_{95}$ thresholds were calculated for change over two (Tn2), three (Tn3) and four observations (Tn4) and applied to the simulated time

series to identify periods over which the criteria for a reliable change were met. Particularly, we examined the sensitivity of the DaRCI thresholds for finding the modeled discontinuities in the middle of the dataset, and the extent to which they were specific to indicating the intended shift, rather than random fluctuations.

Results

Case illustration

To give a visual illustration of the ability of the DaRCI to identify the intended periods of relevant change, see **Figure 6.1**. The indicated changes for the simulated time series in **Figure 6.1** demonstrate clearly how the DaRCI was able to pick up on the periods of increased discontinuity, and thus, relevant change — especially when overall change is at least 30 points. At a lower overall reduction, of 20 points, only the more discontinuous changes were detected in this case example.

Another noteworthy point is that the DaRCI provides thresholds for reliable changes of different duration, but does not itself specify what the most best fitting duration of an identified shift is. Instead, if a score difference meets criteria for the DaRCI at multiple durations, the transitions will simply overlap. This can be seen, for instance, in the bottom row of plots (Step ($n=2$)) in **Figure 6.1**, where change is modeled between two points, but the DaRCI over four points ($n=4$) also picks up this shift as reliable. While this is correct in that the Tn4 threshold is met, visually it is apparent that the true shift occurs over a shorter duration. It is apparent that the DaRCI over four points cannot identify the exact location of the shift by looking at the multiple starting points on the x-axis. This is in contrast to the DaRCI over two points, which finds change only at one point (between observation 7 and 8). The reverse also happens, where the DaRCI over two points cannot identify the location of a shift in the Step-function over $n = 3$ and $n = 4$, and thus places multiple starting points in these longer change periods. However, the relative strength of these changes can be differentiated by their Z-scores, with higher scores indicating the best fit for that particular change. For instance, the aforementioned shift (**Figure 6.1**, -40, Step ($n=2$)) has a Z-score of -7.3 when the DaRCI is calculated for the change over two points (Tn2), and -5 at Tn3, and -3.7 for the Tn4 change duration.

Figure 6.1

Case demonstration of a simulated time series, with increasing levels of overall decline, and increasingly abrupt shifts around the middle of the time series.

Note. Increasing levels of overall score reductions (–10 to –40) are shown on the general X-axis (see headers), and increasing discontinuity around the midpoint on the general Y-axis (from a linear model without a shift, to an abrupt step-change between two points).

Periods where reliable changes were indicated by the DaRCI (with durations between one to three weeks) are shown with the colored areas (i.e., purple, green, orange), with the point markers indicating the ‘start’ of a reliable change.

The red vertical line indicates the simulated ‘midpoint’ of the sigmoid pattern, and the blue vertical lines indicate simulated start and end points for the step function.

Results of the simulation

In **Figure 6.2**, the results of the 10,000 simulations are visually represented as a heatmap, with higher values (darker purple) representing a higher frequency of DaRCI-indicated change start points for different durations (Tn2, Tn3 and Tn4).

Linear. Looking at the first row of subplots, for Linear change, we see a low degree of false positives across the different thresholds. There is a slightly higher degree of false positives for the DaRCI threshold for changes between two points (Tn2), indicating that the criterion picked up approximately 5% random fluctuations rather than true changes – this also applied to the DaRCI for change over two points in the other overall change models. The DaRCI criteria for changes over three or four points showed negligible rates of false positives, 0.5% at Tn3 and 0% at Tn4.

Sigmoid. Looking at the sigmoid change curves, where the exact starting point of change is less apparent, the DaRCI method started to pick up changes from an overall change of -30 : about 20% for all three different duration thresholds. When the overall amount of change was -40 , the DaRCI thresholds identified the correct period of increased discontinuity in about 35–58% of the simulations.

Step ($n=4$). For the third row of subplots in **Figure 6.2**, changes started to be noticeably detected in the -20 overall change model. Because the change was modeled to take place over four time points, the Tn4 threshold yields the highest specificity: 24% of changes were determined to start at time point 6, as modeled. At higher levels of overall change, this increased to 83% and 100%. This while in the overall change -20 models, changes over two (Tn2) and three points (Tn3) were still similar in accuracy: about 19% and 21% of simulations identified a change in the correct period, respectively. For Tn2, this improved to 36% at -30 , and 58% at -40 overall change, while Tn3 accurately captured 62%, and 91% at those levels of overall change.

Step ($n=3$). When change was modeled as a step function over three time points, the DaRCI over three change points picked this up accurately at 62% at -20 points overall change, and 98% and 100% at larger changes (-30 , -40). To compare, the DaRCI at Tn2 also showed about 37% accuracy at identifying the location of the mean shift at -20 , but the identified changes were divided over time point 7 and 8, as the one-week duration could not capture the entire change

Figure 6.2

Heatmap of DaRCI-detected starting points (t_{start}) of reliable changes in the simulated time series.

Note. Visualization of the frequency of indicated starting points of reliable change across the 1000 simulations. Varying levels of overall change (–10 to –40) are shown on the general X-axis (see headers), and increasing discontinuity on the general Y-axis (from no shift within the time series, to an abrupt shift between two points).

The red vertical lines indicate the midpoint of the sigmoid function, and the blue lines indicate the start and end of the step functions.

Tn2 = (Da)RCI calculated for change between two points; Tn3 = DaRCI calculated for change over 3 points; Tn4 = DaRCI calculated for change over 4 points.

Note that the location of t_{start} contributes to the value count, and the accuracy of the DaRCI must be inferred from whether the simulated change period falls within the range of the threshold's duration (Tn2, Tn3 and Tn4).

period that was modeled. Conversely, for Tn4, the changes were harder to identify because often the overall required change (Tn4 = 25) to meet that threshold was not met (only in about 23% of cases at -20 decline). These durations improved when the overall change increased to -30: with about 68% of simulations indicating changes over two time points, and 82% finding shifts over four points as well.

Step (n=2). In the bottom row of subplots, where a mean shift was modeled as occurring between two time points, the RCI was still able to pick up the correct location of the shift for about 36% of cases at Tn2 and for about 9% at Tn3 at the lowest level of overall score reduction (-10). For the -20 change model, the DaRCI over two points picked up the change point accurately in 90% of simulations (compared to 62% at Tn3, and 24% at Tn4). At -30, only the Tn4 model sometimes did not pick up the modeled shift (the other thresholds picked up ~100%), yet in 82% of simulations it indicated a shift started somewhere in the range of time point 5 to 7. Finally, at the highest level of overall change (-40), the change was correctly identified in 100% of cases for all DaRCI thresholds.

Discussion

This paper provided a first illustration of a newly proposed method to identify reliable changes in symptoms of varying duration. Based on the current simulation study of symptom time series, it appears that the DaRCI is a valuable extension of the standard RCI for detecting reliable shifts over multiple time points in the course of treatment. The simulations showed that the DaRCI thresholds over more than two observations (Tn3 and Tn4) were able to pick up the modeled points and periods of discontinuity with high accuracy, especially when the overall change was large (30 points or more). The longer durations also showed fewer false positive values than the DaRCI for change over two points (Tn2, which is equivalent to the standard RCI, but set to a chosen time interval). The DaRCI over three time points (Tn3) seemed to be particularly well-suited to identifying both the relatively abrupt changes modeled with the step function over two and three observations, and the more gradual changes in the sigmoid function and the slowest step function over four points. The DaRCI over four observations (Tn4) also performed well, even though it showed slightly more dispersion of the identified start points. This was not necessarily a problem of its performance, instead it indicates that the midpoint of a transition could be captured accurately, but the starting point of a change was less precisely estimated due to the larger range of the DaRCI at that duration. Similarly, the threshold for change over two points sometimes picked up reliable changes within the context of a larger overall shift (thus identifying a partial transition). To tease apart overlapping reliable changes, the period that shows relatively (to the time it took) the largest change, can be identified with the DaRCI by looking at which threshold an identified transition has the highest Z-score. Taken together, the results of this paper show that the DaRCI has the potential to explore symptom changes of different duration in the context of repeated outcome monitoring.

Apart from the accuracy of the DaRCI thresholds, the simulations also show that investigating reliable changes over different time periods with a purposely adapted index has the potential to uncover clinically relevant symptom changes that may be modest, step-by-step, but large overall. Employing only the standard RCI cut-off and testing difference scores without regard for time (any change that meets the criterion counts), or with a mere repetition of the same criterion for each increment (test the differences for point 1-2, point 2-3, etc., change is found

only when those adjacent points meet the cut-off), would overlook these kinds of continued changes. Clinically, it can already be of interest to study whether multiple change thresholds are met by one individual, or which kinds of changes occur within a given study population (e.g., response patterns in depression [24,38,51], or mood shifts in bipolar disorder [52,53]). The DaRCI may be particularly useful for identifying a discontinuous period within the course of a larger change process. However, whether the DaRCI will be able to detect such periods of relevant change relies on the chosen time period between observations, and exploring the time scale of clinical change processes is still an important topic of study in and of itself [54–57]. However, given the extant literature on the importance of particular change patterns for the outcome of treatment (e.g., Refs. [10,18,25,34,39,40]), the DaRCI method may provide a novel way to explore reliable changes of varying durations within the context of treatment.

One may note that the DaRCI thresholds becoming more demanding over longer periods, is statistically unexpected. Commonly, when more data is available, the additional power allows smaller changes to be detected (e.g., in a linear regression). While this is a fair observation, the aim is not to determine a statistically significant change [1,4,7]. Instead, the DaRCI may provide a simple method to add to a clinician's toolbox, which allows periods of reliable change to be uncovered within the course of a longer time series, even when little data is available. The DaRCI uses between-persons information (the SE_m from the instrument's norm group) to allow within-individual changes to be differentiated from measurement error. Certainly, some individuals may still show variations far beyond the expected range, seemingly showing reliable changes almost every other step; just like some individuals are more likely to vary very little over time, and never show changes that meet the threshold, even if they experience them as meaningful. Once more data is available, more refined methods may be useful in determining significant individual changepoints or overall change [44,46,58].

This method is not without some limitations. First is the challenge of setting a baseline duration for the RCI between two time points, from which the DaRCI calculated the extended thresholds. Researchers must decide for their study what the relevant periods of change may be, which may require pilot studies and in-depth clinical knowledge of the population and change process under study to come to an educated best guess. Another limitation is that although the DaRCI is very

flexible and can be adapted to any chosen symptom measurement instrument, the RCI relies on clinical norm scores for optimal performance [4]. In our simulations we used the SCL-90 norms as a reference as they are based on a large sample ($N = 5,621$). Other instruments may not have such norm scores available, which makes calculations of the standard error of measurement and consequently the (Da)RCI potentially less accurate [1,7]. It is also worth remarking upon the time period underpinning the SE_m : the reliability coefficient on which the SE_m is based, is drawn from test-retest reliability (although internal consistency can be used as well), which would likely be based on a repeated assessment after several months [7]. This is quite different from the time periods (weeks) discussed in this paper for the DaRCI. However, this problem may be minor, as the chosen SE_m is likely to have resulted in more conservative thresholds, as test-retest reliability would be higher if measurements occurred close together in time (and the SE_m smaller).

A strength of this method is that it has the potential for easy application in the clinical context. Once the thresholds have been calculated, clinicians can check whether new symptom assessments (e.g., as gathered with routine outcome monitoring) meet the shorter or longer duration thresholds of the DaRCI. Thus, using this method to examine within-person symptom time series for the presence changes of different durations, particularly while they are in treatment, could yield a novel view of change processes during therapy

Future research should aim to validate this method with clinical data, to provide insight into the kind of symptom changes that can be detected in real-world symptom assessments. It would further be of interest to expand the application of the DaRCI to other commonly used measurement instruments to test if it remains equally effective. Furthermore, comparing it to existing models of change within the course of treatment, such as “sudden gains” would be worthwhile, although the two methods may have slightly different objectives (i.e., identifying changes of various durations, versus change between therapy sessions).

To conclude, the DaRCI provides a simple adaptation of a well-established method for identifying reliable change [3,7,41], and may encourage researchers to consider exploring (discontinuous) symptom shifts of varying durations in the context of psychological treatment.

References

1. Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67(3), 300–307. <https://doi.org/10.1037/0022-006X.67.3.300>
2. Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, 1(2), 138–155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>
3. Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in individual change assessment. *Applied Psychological Measurement*, 40(8), 559–572. <https://doi.org/10.1177/0146621616664046>
4. Jacobson, N. S., & Truax, P. (1991). Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. In *Journal of Consulting and Clinical Psychology* (Vol. 59, Issue 1). <https://doi.org/10.1037/0022-006X.59.1.12>
5. Maassen, G. H. (2000). Kelley's formula as a basis for the assessment of reliable change. *Psychometrika*, 65(2), 187–197. <https://doi.org/10.1007/BF02294373>
6. Maassen, G. H., Bossema, E., & Brand, N. (2009). Reliable change and practice effects: Outcomes of various indices compared. *Journal of Clinical and Experimental Neuropsychology*, 31(3), 339–352. <https://doi.org/10.1080/13803390802169059>
7. Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment*, 82(1), 60–70. https://doi.org/10.1207/s15327752jpa8201_11
8. Keller, M. B. (2003). Past, present, and future directions for defining optimal treatment outcome in depression. *JAMA*, 289(23), 3152–3160. <https://doi.org/10.1001/jama.289.23.3152>
9. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
10. Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasch, C., Mocanu, S., Vocks, S., Schulte, D., & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research*, 23(1), 14–24. <https://doi.org/10.1080/10503307.2012.693837>
11. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
12. Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444–447. <https://doi.org/10.1037/0022-006X.70.2.444>
13. Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(1), 93–101. <https://doi.org/10.1037/a0026455>
14. Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review*, 76, 101827. <https://doi.org/10.1016/j.cpr.2020.101827>

15. Ilardi, S. S., & Craighead, W. E. (1994). The Role of Nonspecific Factors in Cognitive-Behavior Therapy for Depression. *Clinical Psychology: Science and Practice*, 1(2), 138-155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>
16. Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*, 58(9), 1157-1172. <https://doi.org/10.1002/jclp.10044>
17. Paul, R., Andlauer, T. F. M., Czamara, D., Hoehn, D., Lucae, S., Pütz, B., Lewis, C. M., Uher, R., Müller-Myhsok, B., Ising, M., & Sämann, P. G. (2019). Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. *Translational Psychiatry*, 9(1), 1-15. <https://doi.org/10.1038/s41398-019-0524-4>
18. Stulz, N., Lutz, W., Leach, C., Luccock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75(6), 864-874. <https://doi.org/10.1037/0022-006X.75.6.864>
19. aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*, 32(6), 510-523. <https://doi.org/10.1016/j.cpr.2012.05.007>
20. Barkham, M., Stiles, W. B., & Shapiro, D. A. (1993). The shape of change in psychotherapy: Longitudinal assessment of personal problems. *Journal of Consulting and Clinical Psychology*, 61(4), 667-677. <https://doi.org/10.1037/0022-006X.61.4.667>
21. Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., & Harrington, H. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. *JAMA Network Open*, 3(4), e203221. <https://doi.org/10.1001/jamanetworkopen.2020.3221>
22. Ebner-Priemer, U. W., Kubiak, T., & Pawlik, K. (2009). Ambulatory Assessment. *European Psychologist*, 14(2), 95-97. <https://doi.org/10.1027/1016-9040.14.2.95>
23. Schiepek, G., Aichhorn, W., Gruber, M., Strunk, G., Bachler, E., & Aas, B. (2016). Real-time monitoring of psychotherapeutic processes: Concept and compliance. *Frontiers in Psychology*, 7, 604. <https://doi.org/10.3389/fpsyg.2016.00604>
24. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2013). Nomothetic and idiographic symptom change trajectories in acute-phase cognitive therapy for recurrent depression. *Journal of Consulting and Clinical Psychology*, 81(4), 615-626. <https://doi.org/10.1037/a0032879>
25. Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, 27(6), 715-723. <https://doi.org/10.1016/j.cpr.2007.01.008>
26. Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of Counseling Psychology*, 63(4), 379-395. <https://doi.org/10.1037/cou0000150>
27. Schiepek, G. (2009). Complexity and nonlinear dynamics in psychotherapy. *European Review*, 17(02), 331. <https://doi.org/10.1017/S1062798709000763>
28. Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, 88(2), 119-127. <https://doi.org/10.1037/ccp0000469>
29. Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444-447. <https://doi.org/10.1037/0022-006X.70.2.444>

30. Aderka, I. M., & Shalom, J. G. (2021). A revised theory of sudden gains in psychological treatments. *Behaviour Research and Therapy*, 139, 103830. <https://doi.org/10.1016/j.brat.2021.103830>
31. Tadić, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V., & Hegerl, U. (2010). Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders*, 120(1–3), 86–93. <https://doi.org/10.1016/j.jad.2009.04.014>
32. Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, 118(1–3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>
33. Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden gains in cognitive-behavior therapy for treatment-resistant depression: Processes of change. *Journal of Consulting and Clinical Psychology*, 84(8), 726–736. <https://doi.org/10.1037/ccp0000101>
34. Thompson, M., Thompson, L., Gallagher-Thompson, D., & Alto, P. (1995). Linear and nonlinear changes in mood between psychotherapy sessions: Implications for treatment outcome and relapse risk. *Psychotherapy Research*, 5(4), 327–336. <https://doi.org/10.1080/10503309512331331436>
35. Finch, A. E., Lambert, M. J., & Schaale, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology & Psychotherapy*, 8(4), 231–242. <https://doi.org/10.1002/cpp.286>
36. Lambert, M. J. (2005). Early response in psychotherapy: Further evidence for the importance of common factors rather than “placebo effects”. *Journal of Clinical Psychology*, 61(7), 855–869. <https://doi.org/10.1002/jclp.20130>
37. Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology and Psychotherapy Clin. Psychol. Psychother.*, 9, 149–164. <https://doi.org/10.1002/cpp.333>
38. Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2015). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, 27(2), 478–488. <https://doi.org/10.1037/pas0000060>
39. Schiepek, G., Aichhorn, W., & Schöller, H. J. (2017). Monitoring change dynamics: A nonlinear approach to psychotherapy feedback. *Chaos and Complexity Letters*, 11(3), 355–375.
40. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2016). Defined symptom-change trajectories during acute-phase cognitive therapy for depression predict better longitudinal outcomes. *Behaviour Research and Therapy*, 87, 48–57. <https://doi.org/10.1016/j.brat.2016.08.008>
41. Ferrer, R., & Pardo, A. (2014). Clinically meaningful change: False positives in the estimation of individual change. *Psychological Assessment*, 26(2), 370–383. <https://doi.org/10.1037/a0035419>
42. Slofstra, C., Nauta, M. H., Bringmann, L. F., Klein, N. S., Albers, C. J., Batalas, N., Wichers, M., & Bockting, C. L. H. (2018). Individual negative affective trajectories can be detected during different depressive relapse prevention strategies. *Psychotherapy and Psychosomatics*, 87(4), 243–245. <https://doi.org/10.1159/000489044>
43. Maric, M., de Haan, E., Hogendoorn, S. M., Wolters, L. H., & Huizenga, H. M. (2015). Evaluating statistical and clinical significance of intervention effects in single-case experimental designs: An spss method to analyze univariate data. *Behavior Therapy*, 46(2), 230–241. <https://doi.org/10.1016/j.beth.2014.09.005>
44. Albers, C., & Bringmann, L. F. (2020). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *European Journal of Psychological Assessment*, 36(February), 492–499. <https://doi.org/10.1027/1015-5759/a000589>

45. Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods*, 49(3), 988–1005. <https://doi.org/10.3758/s13428-016-0754-9>
46. de Vries, R. M., & Morey, R. D. (2013). Bayesian hypothesis testing for single-subject designs. *Psychological Methods*, 18(2), 165–185. <https://doi.org/10.1037/a0031037>
47. Arrindell, W.A., & Ettema, J.H.M. (2003). *SCL-90: Manual to a multidimensional psychopathology indicator [SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator]* (Pearson, Ed.; 2nd ed.). Swets & Zeitlinger.
48. Derogatis, L. R. (1977). *SCL-90-R: administration, scoring and procedures manual-I for the R(vised) version*. John Hopkins University School Medicine.
49. Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, 9(3), 329–343. <https://doi.org/10.1093/clipsy.9.3.329>
50. Gloaguen, V., Cottraux, J., Cucherat, M., & Ivy-Marie Blackburn. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders*, 49(1), 59–72. [https://doi.org/10.1016/S0165-0327\(97\)00199-7](https://doi.org/10.1016/S0165-0327(97)00199-7)
51. Korf, J. (2014). Delayed mood transitions in major depressive disorder. *Medical Hypotheses*, 82(5), 581–588. <https://doi.org/10.1016/j.mehy.2014.02.015>
52. Bos, F. M., Schreuder, M. J., George, S. V., Doornbos, B., Bruggeman, R., van der Krieke, L., Haarman, B. C. M., Wichers, M., & Snippe, E. (2021). Anticipating manic and depressive shifts in patients with bipolar disorder using early warning signals. *Submitted*.
53. Kramlinger, K. G., & Post, R. M. (1996). Ultra-rapid and ultradian cycling in bipolar affective illness. *British Journal of Psychiatry*, 168(3), 314–323. <https://doi.org/10.1192/bjp.168.3.314>
54. Strunk, G., & Lichtwarck-Aschoff, A. (2019). Therapeutic chaos. *Journal for Person-Oriented Research*, 5(2), 81–100. <https://doi.org/10.17505/jpor.2019.08>
55. Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy*, 117, 40–53. <https://doi.org/10.1016/j.brat.2018.10.005>
56. Kazdin, A. E. (2001). Progression of therapy research and clinical application of treatment require better understanding of the change process. *Clinical Psychology: Science and Practice*, 73(4), 143–151. <https://doi.org/10.1093/clipsy/8.2.143>
57. Mahoney, M. J. (2004). Human change processes and constructive psychotherapy. In *Cognition and psychotherapy*, 2nd ed. (pp. 5–24). Springer Publishing Co.
58. de Vries, R. M., Hartogs, B. M. A., & Morey, R. D. (2014). A tutorial on computing bayes factors for single-subject designs. *Behavior Therapy*, 46(6), 809–823. <https://doi.org/10.1016/j.beth.2014.09.013>

7

Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression

Based on:

Helmich, M. A., Smit, A. C., Bringmann, L. F., Schreuder, M. J., Oldehinkel, A. J., Wichers, M., & Snippe, E. (Under review). Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression. *Manuscript under review.* <https://doi.org/10.31234/osf.io/vf86s>

Abstract

Background: The path to depressive symptom improvement during therapy is often complex, as many individuals experience periods of instability and discontinuous symptom change. If the process of remission follows dynamic systems principles, early warning signals (EWS) may precede such depressive symptom transitions.

Aims: We aimed to test whether EWS, in the form of rises in lag-1 autocorrelation and variance, occur in momentary affect time series preceding transitions towards lower levels of depressive symptoms during therapy. We also investigated EWS in patients without symptom transitions.

Method: In a sample of 41 depressed individuals who were starting psychological treatment, positive affect and negative affect (high and low arousal) were measured five times a day using ecological momentary assessments (EMA) for four months (521 observations per individual on average; 25,197 observations in total), and depressive symptoms were assessed weekly over six months. We used a moving window method and time-varying autoregressive generalized additive modeling to determine whether EWS occurred in these momentary affect measures.

Results: For the moving-window autocorrelation, 89% of individuals with transitions showed at least one EWS in one of the variables (versus 62.5% in the no-transition group), the proportion of EWS in the separate variables was consistently higher (~44% across affect measures) than for non-transitioning individuals (~27%). Rising variance was found for few individuals, both preceding transitions (~11%) and for individuals without transitions (~12%).

Conclusions: The process of symptom remission showed critical slowing down in at least part of our sample. Our findings indicate that EWS are not generic across all affect measures and may have limited value as a personalized prediction method.

Keywords: complexity; destabilization; sudden improvement; personalized models; idiographic change; replicated single-subject design

Background

Researchers have started drawing on dynamical systems theory for methods to improve their understanding of the processes of change in psychopathology [1–3]. In part, this renewed interest in dynamical systems is due the rise in studies revealing that changes in psychopathology are often nonlinear [4–10]. In the dynamical systems framework, people are viewed as complex systems of interacting components (e.g., behaviors, cognitions and emotions) in which the dynamics of those components can be studied to anticipate impending changes [11–13]. Particularly, the phenomenon of ‘critical slowing down’ may lead to early warning signals (EWS) in the temporal dynamics of the system variables before a critical transition from one dynamically stable state to another (e.g., from a depressed to non-depressed state). When a system critically slows, it destabilizes and gradually becomes less resilient to external shocks (e.g., daily events), which results in more extreme fluctuations (rising variance) and a longer time to return to the equilibrium position (rising autocorrelation). EWS have been described as generic indicators that a system may be losing resilience and nearing a tipping point where the likelihood of a critical transition is higher [14–17]. EWS are generic in the sense that they are expected to occur in a wide range of complex dynamical systems [16,18], and are beginning to be explored in the field of psychiatry [19–28].

An important reason that EWS have attracted such attention from psychopathology researchers is the promise that they may serve as a clinical tool to detect symptom shifts ahead of time. Descriptively, they fit with clinical observations: the process of change during psychotherapy is frequently characterized by destabilization and nonlinear patterns [29–34], such as symptom shifts between therapy sessions known as sudden gains and sudden losses [8,19,35–38], or rapid early response curves over the first few weeks of treatment [39–43]. These nonlinear patterns are often linked to better outcomes (e.g., in Refs. [8,29,33,37,39,40,44]), so using EWS to improve early detection of destabilization and potential shifts could provide therapists clues about their patients’ response potential, and allow them to personalize the treatment and timing of interventions accordingly. Thus, examining the promise of EWS is clinically important, as it may give patient and therapist valuable insight into whether the system is sensitive to therapeutic input (is destabilizing), and significant improvement could be imminent [2,34,45,46].

Moreover, studying the occurrence of EWS during therapy is theoretically relevant, as it takes the next step from describing the process of symptom remission from a complex systems angle, to actually testing whether dynamical systems phenomena such as critical slowing down apply to transitions in psychological systems [47]. EWS-like dynamics have shown promise in anticipating shifts in depressive symptoms in a few (group-level) studies [20,25,27,29,48–50], which correlated individuals' average autocorrelation or variance in mental states to their average symptom severity. However, to truly test the idea that EWS precede symptom transitions in psychopathology, it is imperative to investigate within-person *rises* in autocorrelation or variance prior to transitions [47]. To that end, Wichers et al. [21] first showed that EWS could be used as a personalized predictor of depressive relapse in a single patient, a result they replicated in another patient recently [22]. In the context of therapy, Olthof et al. [19] identified transitions in time series of daily self-ratings of problem intensity for a large sample of mood disorder patients during treatment. They found that EWS in the form of increased dynamic complexity predicted an increased probability of having a transition in symptom severity in the next four days. These results are very promising, as they show that destabilization and within-person defined transitions were positively associated. However, the study by Olthof et al. [19] used a multi-level model to predict transitions based on between-person differences in the level of destabilization, and thus does not provide a full within-person test of rising EWS prior to transitions. Therefore, this study aims to test whether impending transitions toward lower levels of depressive symptoms are consistently preceded by within-person rises in EWS.

Theoretically, we would expect EWS to occur in variables that are central to the system under consideration. A state of depression can be said to be composed of a persistent collection of individual depressive symptoms (e.g., anhedonia, sad mood, lack of energy), which in turn are expressed in shorter-lived feelings [51,52]. In the current study, we therefore look for EWS in the smallest building blocks that contribute to the overall course of a depressive disorder [51–54]: the moment-to-moment fluctuations in affect measured by ecological momentary assessments (EMA). Notably, the presence and strength of EWS may differ between variables [55–58]. Therefore, an important secondary goal in the investigation of EWS in the context of psychiatry is to examine which affective states most reliably anticipate transitions towards reduced symptoms of depression.

Gathering appropriate data is challenging, and only two single-subject studies of EWS prior to depressive relapse have been published [21,22]. To be able to capture and calculate EWS in affect observations before shifts in depressive symptoms, we need frequent momentary assessments over a period in which clinical change is likely. High resolution time series data [24,59–61] that captures moment-to-moment variations at a time scale that is short enough to cover the full range of fluctuations in the state of the system [12,62–65], and over a long enough period to capture the entire state change, is required to detect change in the system dynamics. Therefore, we collected intensive longitudinal data from 41 individuals who were starting treatment for depressive complaints as they entered the study, and were thus likely to show symptom improvement. We gathered four months of 5-times daily EMA affect observations to be able to detect EWS, and six months of weekly depressive symptom assessments for the identification of symptom shifts. As EWS are studied within-system (i.e., within each individual), these 41 individuals each represent a single-subject study and replicated test of our hypothesis.

The current study was uniquely designed to test whether EWS consistently occur prior to transitions toward depressive symptom improvement, and is the first to test this hypothesis empirically, at the within-person level, for multiple individuals. We aimed to establish whether destabilization in the form of EWS such as increasing trends in lag-1 autocorrelation and variance can be detected in momentary-measured affect before transitions towards improvement in depressive symptoms. We hypothesized that EWS would be most prevalent in variables congruent with the direction of the transition, that is, positive affect variables would be most likely to show changing dynamics leading up to a clinical improvement [21,22,25,66]. To provide the clearest picture of how well EWS work as personalized indicators of imminent change, we will also examine the rates of EWS in individuals without transitions. We hypothesized that individuals with clear symptom shifts would have relatively more and stronger EWS than individuals without transitions.

Method

Sample description

Individuals with current depressive symptoms who were about to start psychological treatment were recruited for this study. This population was chosen with the expectation that these patients would be likely to show symptom improvements during the study period. To be eligible for participation, individuals needed to be 18 years or older, have current depressive symptoms (a score ≥ 14 on the Inventory of Depressive Symptomatology—Self Report (IDS-SR) [67], and be bound for psychological treatment for depression within one month of the (intended) participation start date. Participants also needed to be capable of following the study procedures and operating a smartphone, have sufficient command of the Dutch language, and be willing and able to give informed consent. Exclusion criteria were chronicity of depressive complaints (persistence ≥ 2 years), current manic or psychotic symptoms, and a primary diagnosis for personality disorder.

The data collection took place between June 2017 and May 2020. The intended sample size was 45–50 participants, to arrive at a sufficient number of patients with a depressive symptom transition who could serve as replications of one another. The final sample consisted of 41 participants who completed the six-month study period (for an overview, see **Figure 7.1**). The participants were mostly female ($n = 35$, 85%), with a mean age of 40.1 years ($SD = 14.4$, min = 19, max = 70), at the time of starting the study. Written informed consent was obtained from all subjects. All procedures were approved on December 12, 2016 by the Medical Ethical Committee of the University Medical Center Groningen (reg. number: NL58848.04.16). Detailed procedures for this research project are described in Ref. [68].

Materials

Ecological momentary assessment (EMA): momentary mood

Over a period of four months, participants filled in a 27-item EMA questionnaire about their current feelings, activities and surroundings (available online, see <https://osf.io/a8572/>) five times a day at set times, with fixed intervals of three

Figure 7.1

Flowchart of the Transitions in Depression (TRANS-ID) Recovery participant inclusion steps

hours. Affect was measured with items like “I feel cheerful”, and “I feel down”, which were answered on a visual analogue scale ranging from “Not at all” to “Very much” (underlying scale ranged from 0 to 100).

With a compliance of 85%, this resulted in $M = 522$ ($SD = 42$, $\min = 423$, $\max = 591$) momentary mood measurements per individual, on which the EWS could be calculated (25,197 EMA observations in total). To that end, individual items were standardized, person-mean centered and averaged to create four variables based on the affect circumplex [69,70]: Positive Affect (PA) high arousal (“cheerful”, “energetic”), PA low arousal (“content”, “at ease”), Negative Affect (NA) high arousal (“restless”, “stressed”, “irritated”), and NA low arousal (“down”, “listless”, “tired”).

Depressive symptoms

Depressive symptoms were measured weekly with fourteen items of the Symptom Checklist-90 (SCL-90) Depression subscale [71–73]. Every weekend, over a period of six months, participants were asked to rate how much in the past week they were bothered by depressive symptoms on a five-point scale from “not at all”, to “extremely”. With a compliance of 98%, this gave $M = 29$ symptom observations per person ($SD = 1.4$, $\min = 23$, $\max = 30$), and 1,211 assessments in total. The full sixteen-item (Dutch) subscale has shown high internal consistency ($\alpha = .91$) and reliability ($\omega_k = .93$), as well as a strong relationship to the general factor psychological distress [71,74]. For this study, two questions on suicidal ideation were removed from the questionnaire to reduce mental burden for participants. The repeated measurements of the SCL-90 depression subscale were used to identify depressive symptom transitions (see below).

Psychological treatment

To optimize the chance of capturing transitions in symptoms during the study period, participation was started approximately three weeks before the start of a psychological treatment aimed at reducing depressive symptoms. Participants self-enrolled into a psychological treatment as part of care as usual, completely independent of the study.

Analysis

This analysis has been preregistered, for additional detail see <https://osf.io/xftuq/>. Briefly, we first identified transitions in depressive symptoms (based on weekly symptom data) and then tested (i) whether EWS (based on momentary affect assessments) preceded those transitions and (ii) how often EWS occurred in individuals without transitions. EWS were analyzed for each individual and each affective variable (PA high arousal, PA low arousal, NA high arousal, and NA low arousal) separately using two different methods, namely a moving-window approach (rises in autocorrelation and variance) and Time-Varying Autoregressive Generalized Additive Modeling (rises in autocorrelation). We elaborate on each analytical step below.

Transition identification

In line with descriptions of critical transitions, we aimed to identify symptom reductions that appeared relatively abrupt, were reliable, and remained stable for two weeks [15,17,21]. We did not know the time period over which a critical transition toward improvement would take place, but symptom reductions occurring over time periods of a week to a month have been described as ‘rapid’ in the clinical literature [75,76]. Therefore, we used an adapted version of the Reliable Change Index (RCI) [77], the Duration-adjusted RCI (DaRCI) [78], to examine whether symptoms improved reliably over one to maximum four weeks.

The RCI determines the threshold at which the difference between two points is larger than can be expected from measurement error [77], and is defined as: $RCI = SE_{diff} \times Z$, with $SE_{diff} = \sqrt{2(SE_m)^2}$. Drawing on psychiatric outpatient norm group scores for the SCL-90 depression subscale [71], we used a standard error of measurement (SE_m) of 4.37 for this study¹. With a Z-score of 1.96 to set a confidence level of 95%, this yielded $RCI_{95} = 6.18 \times 1.96 = 12.11 \approx 13$, indicating that a reliable change between two observations needed to be 13 points or larger (rounded up to preserve the confidence level).

Next, to calculate the thresholds for reliable change over two, three and four weeks, we used the DaRCI. Essentially, this method calculates the critical threshold of change from a starting point (t_{start}) to the last observation in that period (t_n) while accounting for the longer period between observations. This is done by reducing the RCI to a confidence range around one point, which is then proportionally extended over n points: $DaRCI = (SE_{diff} / 2) \times Z \times n$. This method allowed us to examine change that is reliable (maintains the confidence threshold) over various durations (for more detail, see Ref. [78]). Thus, transitions were identified as reductions of 13, 19, 25 or 31 points over one, two, three or four weeks, respectively [71]. To summarize, if the time between observations is larger, the size of the change must also be larger to be equally reliable.

To ensure that the identified transitions were not just temporary fluctuations in scores, we added a stability criterion based on the RCI_{95} that ensured that the mean of the two weeks prior to the first point of a transition ($t_{start-1}, t_{start-2}$), and the two weeks after a transition (t_{n+1}, t_{n+2}) was at least different by 8.6 points. This

¹Note that this SE_m is based on all sixteen depression subscale items of the SCL-90, while we only used fourteen items in this study. This meant the threshold we calculated is probably more conservative in identifying transitions.

threshold accounts for the additional power gained from applying it to a change between two means rather than two single assessments by dividing it by the square root of n , such that: $\text{RCI}_{95}/\sqrt{2} = 8.6$. For each individual, the earliest transition that met the DaRCI threshold and remained stable the two weeks after the transition was identified. This stability criterion was not preregistered, but it was considered prior to the analyses.

Finally, as this method for identifying transitions has not been tested previously, we examined the (Da)RCI at different confidence levels. We took 95% confidence as a default, and additionally explored transitions identified at 90% and 99% confidence.

Moving window method

In line with the two studies that examined EWS within-persons prior to recurrence of depressive symptoms [21,22], we used a metric-based moving window approach to examine rises in autocorrelation and variance in the momentary affect measures (cf. [61]). Inside a linearly detrended window of 70 observations (two weeks), with outliers winsorized (i.e., extreme values were set to the 5th and 95th percentile), we calculated the values of the lag-1 autocorrelation and the standard deviation of the affect variables within that window. The window was then moved forward one observation at a time, and the calculation of the autocorrelation and variance repeated, until the end of the time series (no transition group) or until the transition was reached. The Mann-Kendall test (M-K test, coefficient τ) was used to test for a monotonic trend in the entire EWS time series of window estimates of autocorrelation and the variance. To improve the specificity of our test, we maintained a minimally relevant effect size of the trend in the EWS time series of $\tau \geq 0.1$. Specifically, we hypothesized that EWS would be stronger for individuals with compared to individuals without transitions, and therefore, that the τ -coefficients for individuals with a transition would be distributed above zero, and more often have a significant positive trend of $\tau \geq 0.1$ in the EWS time series. In participants without a transition, we would expect the τ -coefficient be distributed around 0, and therefore to show a lower percentage of significant positive trends.

Sensitivity checks. We explored the effect of the chosen model settings on our findings by rerunning the analysis with the following parameters altered:

window sizes of 35, 105 and 140 observations (instead of 70), global detrending (instead of within-window), not removing outliers (instead of winsorizing). We also checked the effect of using only the M–K significance test at $\alpha < .05$ to detect positive trends in EWS, and explored the results when the dependency between window estimates (due to overlapping data) was corrected for, with the modified M–K test $\alpha < .05$, as proposed by Hamed and Rao [79]. In accordance with our preregistration, the single items “I feel cheerful” and “I feel down” were also analyzed. The results for those variables and the sensitivity checks are reported in the detailed **Supplementary Materials**.

Time-Varying Autoregressive Generalized Additive Modeling (TV-AR GAM)

To examine whether the autocorrelation-coefficient rises before a transition, we explored a recently developed method: TV-AR GAM [80]. This method accounts for non-stationarity in the data by simultaneously modeling the trend (rather than removing it in preprocessing), allowing both the intercept and autoregressive coefficient to vary over time. The TV-AR GAM uses thin-plate splines to fit smooth curves to the time series, and provides p -values for the intercept and autocorrelation parameter, and estimated 95% credible intervals for the fitted curve. A time-varying trend in the autocorrelation was indicated for curves with (a) a significant autocorrelation parameter, and (b) where a horizontal line could not be fitted within the credible intervals (i.e., the maximum value of the lower limit > the minimum value of the upper limit). To determine if there was truly an overall rise in the trend, three raters (authors M.A.H., L.F.B., and E.S) then visually inspected the diagnostic plots and the modeled smooth curves and decided by consensus on the direction of change – an increase, decrease, or a nonlinear trajectory (e.g., quadratic, cubic, or more complex nonlinear trajectories, apart from those that were very convincingly rising or declining).

Sensitivity checks. As the time series for the non-transitioning group had a higher maximum number of observations ($M = 616$) than the transition group, we performed a sensitivity analysis for the non-transition group with the time series shortened to the average length of time series in the transition-group: 334, 441, and 458 observations, for transitions identified at 90%, 95% and 99% confidence levels, respectively.

Results

Identified transitions

For nine out of 41 participants (22%), we identified reliable symptom reductions with the DaRCI at 95% confidence. The duration of these symptom transitions varied, with five transitions occurring over one-week, three transitions over two weeks, and one transition over three weeks. On average, transitions took place after 82 days ($SD = 29$, range = 51–165). With the DaRCI set to 90% confidence, fourteen individuals (34%) showed transitions, and at 99% confidence, six cases (15%) remained (for the analysis results for the transition groups at 90% and 99% confidence, see the **Supplementary Material, Tables S7.2 and S7.3**).

EWS with the moving window method

Autocorrelation. The results of the moving window method indicated that, of the nine individuals with a transition, 8 (89%) showed EWS in the autocorrelation of at least one affective state ($M = 1.78$ EWS per individual, $SD = 1.2$). By comparison, of the 32 individuals without transitions, 20 (62.5%) showed at least one EWS in autocorrelation ($M = 1.09$ EWS per individual, $SD = 1.17$).

When investigating EWS in the four variables separately (PA high and low arousal, NA high and low arousal), increases in autocorrelation over time occurred in a higher percentage of individuals with a transition than in individuals without a transition (see **Table 7.1**). For individuals with a transition, EWS occurred most often in NA high arousal (55.6%), and least often in PA low arousal (33.3%), and for the no-transition group EWS were most often indicated in NA high and low arousal (31.2% in both), and least often in PA high arousal (18.8%). Contrary to our expectations, PA high arousal did not show a higher number of true positive EWS than the other variables, instead showing the lowest sensitivity.

Variance. Of all nine individuals with a transition, 4 (44%) showed EWS in the variance of at least one affective state ($M = 0.44$ EWS per individual, $SD = 0.53$), as did 8 (25%) of the 32 individuals without transitions ($M = 0.47$ EWS per individual, $SD = 0.95$).

The moving window analysis showed lower proportions of EWS in the variance of the affect measures than the autocorrelation (see **Table 7.1**). For some variables, the percentage of EWS appeared to be higher in the transition group than in the non-transition groups (e.g., for PA low there were 3 (33%) EWS in the transition group and 2 (6.2%) in the non-transition group). However, averaging across the four variables revealed no discernible difference between the groups, with 11.1% true positive EWS in the transition group and 11.7% false positive EWS in the non-transition group. Again, PA high was not the best performing measure, showing 0% true positive EWS and 15.6% false positives in the variance.

Table 7.1

Results of the moving window and TV-AR GAM analyses: average τ -values and proportion of early warning signals in the autocorrelation and variance for individuals with and without transitions.

	Autocorrelation			Variance	
	TV-AR GAM	Moving Window method		Moving Window method	
	EWS % (N)	EWS % (N)	τ Mean (SD)	EWS % (N)	τ Mean (SD)
Transition (N = 9)					
PA high	0% (0)	33.3% (3)	-0.09 (0.37)	0% (0)	-0.35 (0.27)
PA low	11.1% (1)	44.4% (4)	0.13 (0.24)	33.3% (3)	-0.24 (0.42)
NA high	11.1% (1)	55.6% (5)	0.21 (0.36)	11.1% (1)	-0.35 (0.35)
NA low	0% (0)	44.4% (4)	-0.01 (0.39)	0% (0)	-0.36 (0.24)
Overall mean	5.6%	44.4%	0.06	11.1%	-0.33
No transition (N = 32)					
PA high	3.1% (1)	18.8% (6)	-0.08 (0.2)	15.6% (5)	-0.28 (0.42)
PA low	3.1% (1)	28.1% (9)	-0.07 (0.26)	6.2% (2)	-0.37 (0.32)
NA high	3.1% (1)	31.2% (10)	-0.06 (0.24)	9.4% (3)	-0.27 (0.36)
NA low	3.1% (1)	31.2% (10)	-0.05 (0.25)	15.6% (5)	-0.27 (0.35)
Overall mean	3.1%	27.3%	-0.07	11.7%	-0.30

Note. EWS = early warning signals; PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal; TV-AR GAM = time-varying autoregressive generalized additive model; SD = standard deviation; τ = coefficient of the Mann-Kendall trend test. The results for the TV-AR GAM represent cases where the modeled smooth curve was significantly increasing. The results for the Moving Window method are based on the analysis of transitions at 95% confidence, where individual time series were winsorized, locally detrended within windows of 70 observations, and the number of early warning signals based on cases with $\tau \geq .1$

Distribution of the τ -coefficients. We expected the Mann–Kendall τ -coefficients to be distributed above zero for individuals with transitions, and around zero in the group without transitions. However, the average τ -coefficients were lower than expected. As can be seen in **Table 7.1**, the average τ -coefficients for the autocorrelation of the different affect measures appeared to be distributed only marginally above zero in the transition group, and below zero for the no-transition group. Only the τ -coefficients of the autocorrelations of PA low and NA high had average values above the $\tau \geq 0.1$ threshold, and showed mostly slightly negative values in the group without transitions. For the variance, the τ -coefficients of all affect measures seemed to be more strongly negative on average, with similar values in both groups (for the distributions of the τ -values, see **Figures S7.1a** and **S7.1b** in the **Supplementary Materials**).

Sensitivity checks. For the autocorrelation, the finding that proportionally more EWS were found in the transition group was robust across all the altered model settings for the moving window analysis. The changes in the number of EWS for the autocorrelation found with these altered model settings were similar for people with a transition (difference in N EWS compared to the main analysis: Median = 0, first quartile (Q1) = -1, third quartile (Q3) = 0.75) and for people without a transition (Median = 0, Q1 = -1, Q3 = 1). Increasing the window size to 105 or 140 observations yielded an improvement of the accuracy for this sample, at 50% and 58% true positives and 33 and 31% false positives, respectively.

For the variance, the lack of a proportional difference in EWS between the transition and non-transition group was consistent across the sensitivity checks. For the altered analysis settings, similar changes in the number of indicated EWS were found for the transition group (difference in N EWS compared to the main analysis: Median = 0, Q1 = -0.75, Q3 = 0), and the people with no transition (Median = 0, Q1 = 0, Q3 = 1; see also **Tables S7.1a** and **S7.1b** in the **Supplementary Materials**). None of the altered model settings for the variance yielded notably improved accuracy.

EWS with the TV-AR GAM

The trajectories fitted by the TV-AR GAM indicated that, across the four affect variables, the proportion of significant increases in the autocorrelation was comparable between the transition group ($M = 5.6\%$) and the non-transition group ($M = 3.1\%$; see **Table 7.1**). Significant decreases and nonlinear trajectories were indicated more often than increases in both groups, with respective means of 8.3% and 27.8% in the transition group, and $M = 9.4\%$ and 25.8% in the non-transition group.

Sensitivity checks. Reducing the time series length for the non-transitioning individuals resulted in fewer cases with a significantly time-varying smooth curve (approximately 30% compared to 40% at full length), although the proportion of significant increases in autocorrelation remained similar (2.3%, compared to 3.1% at full length). With transitions identified at 90% or 99% confidence, the number of true-positive significant rises in autocorrelation was comparable to the results at 95% confidence (although at 99% confidence, no EWS remained in the transition group), as did the number of false positive EWS (see **Table S7.3** in the **Supplementary Materials**).

Exploratory findings

In addition to answering the main research question, we also aimed to investigate ways to potentially improve the balance of true and false positives of our results. We explored whether we could improve the sensitivity of the moving window analysis by separating participants according to a minimum cumulative number of EWS. Combining the results of the variance and autocorrelation for PA high, PA low, NA high and NA low showed that selecting individuals with a minimum of two EWS (out of a possible eight) improved the sensitivity, with 77.8% true positives compared to 44.4% in the main analysis, but also resulted in an increase in false positives from 27.3% in the main results, to 40.6% (for more detail, see **Figures S7.2** and **S7.3** in the **Supplementary Materials**).

Moreover, we explored different minimally relevant effect sizes by varying the threshold τ -value for the moving window analysis between 0 and 0.5. Our results did not reveal a clear cut-off value that improved the accuracy for all

variables (see **Figure S7.4** in the **Supplementary Material**). Overall, lowering or raising the τ -threshold led to a change the number of true positives that was approximately matched by the change in false positives.

Discussion

This is the first paper to provide an empirical investigation of whether EWS in the form of increasing trends in lag-1 autocorrelation and variance could be detected in individual affect time series in the period before a transition toward depressive symptom improvement. We examined the prevalence of EWS prior to the symptom transitions identified for nine individuals, as well as the rate of false positive EWS in the time series of 32 individuals without transitions. We found that rising autocorrelation was present in at least one of the affect measures for eight out of nine (89%) individuals with a transition, and for twenty out of the 32 (62.5%) individuals without transitions. When examining the affect measures separately in the moving window analysis, we found that EWS in the form of rising autocorrelation were present in proportionally more individuals with a transition (~44%) than without a transition (~27%). A rise in the variance was found in at least one variable for four out of nine (44%) individuals with a transition, and for eight of the 32 (25%) individuals without transitions. The proportion of EWS in the variance for the separate affect measures showed similar rates for individuals with a transition (~11%) and without a transition (~12%) and thus did not show the same differentiation between the groups. Furthermore, the TV-AR GAM analysis indicated at most one individual with a significant increase in autocorrelation for each of the four affect measures, in both the group with transitions (~5.6%) and in the no-transition group (~3.1%). In our extensive sensitivity checks, the findings for our analyses were robust. This strengthens our conclusion that the moving-window detected autocorrelation results were in the hypothesized direction, and the process of remission from depression showed signs of critical slowing down in part of the individuals in our sample.

Looking at the broad pattern of our results, our findings correspond reasonably well to earlier studies of EWS in depression [19,21,22]. Like the single-subject studies by Wichers et al. [21,22], our results showed that within-person

risers in autocorrelation occurred for some of the participants before a transitions in depressive symptoms – this also occurred relatively more often for people with a transition than for people with no transition. Unlike those studies by Wichers et al. [21,22], however, we did not find rising variance as an EWS. Furthermore, our finding that EWS occurred before symptom transitions during therapy is in line with the findings by Olthof et al. [19], who found that critical fluctuations had a higher probability of occurring in the four days before a sudden gain or loss. However, critical fluctuations are a measure of the spread of scores, like the variance, which did not function as an EWS in our study. The difference in findings may be explained by the fact that we looked at a different population than Wichers et al. [21,22] (individuals showing depressive symptom remission, rather than recurrence), and a much longer time period than Olthof et al. [19] for the rising fluctuations (over weeks or months, as opposed to days). Yet, taking in the broader evidence of all these studies, including the current study: significant destabilization in the form of rising autocorrelation could be identified prior to transitions in symptoms in the affect measures of depressed patients. In our sample, we found signs of critical slowing down, i.e., a rise in autocorrelation in at least one of the affect measures, for eight of the nine individual systems, which was in line with our hypothesis. This suggests that for those individuals the process of depressive remission showed the nonlinearity one would expect from a dynamic system, particularly in the context of therapy [2,30,33,34,50,81].

Although almost all individuals with transitions showed EWS in at least one affect measure, looking at the separate affect measures presents a less consistent picture of critical slowing down in our study. Each variable showed EWS for only a few people in the transition group, and, contrary to what we hypothesized, PA high arousal did not turn out to be the most sensitive affect measure to predict of symptom improvements. No single affect variable showed EWS for more than five people with transitions. Thus, EWS were not ‘generic’ indicators of impending symptom change when considering the variables separately. Theoretical works in ecology have found that variables can be ‘silent’ and lack EWS when they do not change in the same direction as the destabilization of the main system [55–58]. That is, perhaps the heterogeneity of which variables showed EWS was due to the fact that the destabilization occurred in different depressive symptoms for different people. By that reasoning, PA high arousal may have shown fewer EWS

if destabilization occurred more strongly in depressive symptoms that do not correlate highly with positive experiences (cheerfulness, feeling energetic)(cf. [20,66,82]). False negatives may also have occurred due to the nature of transitions in our study. Based on theory, critical slowing down and its corresponding EWS would be expected to occur primarily before critical transitions or shifts between states in systems that are dynamically stable around a single point and are slowly losing resilience [15,17]. It is possible that the transitions we detected did not follow this pattern, for example, because the states prior to and after the transition were not sufficiently stable or because the transition was not preceded by a gradual accumulation of instability [83]. The latter may happen if a transition was triggered by a large push (e.g., a psychologically impactful event) rather than gradual destabilization of the system [62,84–86]. Thus, it could be that only a subset of individuals in our sample met the theoretical requirements for detecting EWS, or that only some of the affect measurements picked up on the changing dynamics, which may have contributed to a lower sensitivity of EWS.

Furthermore, there were six to ten individuals without transitions who showed false positive EWS in our study. A potential explanation is that those individuals without transitions still experienced a destabilizing influence from entering therapy, and thus also showed changes in the temporal dynamics of their affect during the observation period, even if this did not precede a clear discontinuous symptom transition (or end in a change, at all). This interpretation is supported by the fact that, when we lowered our DaRCI confidence level to 90%, the number of true positive EWS increased (from 3–5, to 5–7 EWS per affect measure), while the false positives decreased (from 6–10, to 5–9 EWS per affect measure; see **Supplementary Table S7.2**). This suggests that some false positive EWS in the main analysis were likely due to individuals who did experience a relevant change, but it did not meet the transition criteria at the 95% confidence level. This is key to interpreting our findings, as our transition criterion may have been too conservative and selected only some of the relevant transitions, while categorizing others as belonging to the no-transition group, thereby potentially inflating the rate of false positives, and underestimating the number of true positives. Further explanations are derived from other fields, where critical slowing down has also been shown to occur before gradual transitions [18,87], as well as in periods with no significant system change [85,87]. However, most of these EWS studies have

been conducted in an ecological context, with very different data, and it is hard to tell which of the processes that influence the finding of EWS in other fields translate to our ability to detect EWS in psychological processes.

The accuracy of our moving-window detected autocorrelation analyses was not high enough to warrant any claims about clinical utility at this stage. Our exploratory analyses indicated that there was room to improve the method's performance, as we found improved accuracy for larger window sizes and when looking at the cumulative number of EWS within individuals (a minimum number of EWS threshold also worked well in two studies of EWS before depressive symptom recurrence, see Refs. [22,88]). Moreover, single predictors are rarely 100% accurate in isolation, and EWS may provide complementary information to early response indicators such as severity of depression [89–91]. Gaining a better understanding of which participants are likely to show detectable EWS could also help improve the sensitivity of this method.

The results for the variance indicated no proportional differences in EWS between individuals with and without transitions. In fact, most people in our study (with and without transitions) showed a negative trend in the variance over time. A clinical explanation for these downward trends in variance relates to our study population: given that our participants were entering treatment for depressive symptoms, they may have shown higher instability at the start of the observation period due to the destabilizing influence of entering therapy (and because high levels of variability are associated with higher levels of depression [92–94]). As they improved under the influence of therapy, their emotional responses may have become less volatile, and thus an overall downward trend in variance may have emerged in their time series. Another, practical reason for the declines in the variance is that participants may have gotten more habituated to filling out the EMA questionnaires, and thus used less and less of the available scale. Finally, previous EWS studies have also indicated that the variance may show decreases when less data is available and is a less robust as an indicator of impending change compared to autocorrelation [22,25,95].

Even though the autocorrelation worked reasonably well as an EWS with the moving window method, the more conservative TV-AR GAM did not show the same results. Possibly, this approach was not optimally suited for the purposes of testing for an overall rise in autocorrelation. Instead, the TV-AR GAM results highlighted

the high degree of nonlinearity in the autocorrelation of affect time series in both groups. While the M-K test may detect an overall significant trend over time even if the rise in autocorrelation or variance starts later in the time series, the GAM method is aimed at testing the overall shape of the data, and may instead identify similarly late rises as a curving, nonlinear pattern. We do not know over what time period a rise in EWS would be expected in our sample, but a method that allows significant rises to be detected over shorter periods (e.g., a few weeks prior to a transition), may provide an improved rate of EWS detection.

Another reason the TV-AR GAM may have resulted in lower numbers of EWS, is that this method requires a large amount of data to fit credible intervals that were narrow enough to meet our strict change criterion (the maximum value of the lower bound being larger than the lowest value of the upper bound). Furthermore, the TV-AR GAM is generally conservative and biased to detect zero change if little data is available (cf. Refs. [80,96,97]). While determining the direction of change through visual inspection, we noted that many individuals who showed linear change had confidence bounds that were too wide to identify an increase with certainty. The effect of (in)sufficient data on the TV-AR GAM findings was evident from the sensitivity check in the no-transition group: when the time series were shortened (e.g., from, on average, 616 to 411 observations in the main analysis), fewer significant smooth curves were found, and particularly the number of decreases were reduced. It thus seems that the available time series length may have played a limiting role in our ability to detect EWS with this method.

Because there were too few individuals with a transition to reliably test whether the proportional differences between the group with a transition and the group without a transition were significant, it is difficult to determine to what extent these findings generalize beyond our sample. However, this study was foremost a proof of principle to see if EWS could be found in individual time series of people receiving treatment for depressive symptoms. Our study was designed with the specific intent to detect EWS in individual time series, with several months intensive longitudinal within-person affect and symptom measurements (totaling 26,408 observations) gathered during a period where symptom improvement was likely. Given that our population differed from the previous studies of within-person EWS that looked at relapse in depression [21,22], and taking into account the new insights from recent studies [20,23,24], we preregistered and conducted

a very thorough analysis with a range of sensitivity checks. Because of this, we were able to discern the broad pattern of how EWS occurred in our dataset and can more confidently state that our findings were robust.

Conclusions

EWS in the form of a rise in the autocorrelation in NA high arousal, NA low arousal, PA high arousal, and PA low arousal preceded symptom transitions toward depressive symptom improvement in part of the participants, and occurred more often for individuals with transitions than for individuals without transitions. Almost all individuals (89%) with a transition had EWS in autocorrelation for at least one the affect measures, compared to about two-thirds of the participants without transitions. EWS based on the variance of affect did not systematically precede symptom transitions. It is too early to claim any clinical utility of EWS in the context of psychotherapy. However, this study provides much-anticipated empirical evidence that the process of symptom remission can follow dynamical systems principles, beyond a *post hoc* explanation of nonlinear patterns in psychopathology.

References

1. Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS One*, *11*(12), [e0167490]. <https://doi.org/10.1371/journal.pone.0167490>
2. Hayes, A. M., & Andrews, L. A. (2020). A complex systems approach to the study of change in psychotherapy. *BMC Medicine*, *18*(1), 1–13. <https://doi.org/10.1186/s12916-020-01662-2>
3. Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of Counseling Psychology*, *63*(4), 379–395. <https://doi.org/10.1037/cou0000150>
4. Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. <https://doi.org/10.1111/j.1467-8721.2009.01619.x>
5. Fisher, A. J., Newman, M. G., & Molenaar, P. C. M. (2011). A quantitative method for the analysis of nomothetic relationships between idiographic structures: Dynamic patterns create attractor states for sustained posttreatment change. *Journal of Consulting and Clinical Psychology*, *79*(4), 552–563. <https://doi.org/10.1037/a0024069>
6. Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10–15. <https://doi.org/10.1177/0963721416666518>
7. Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, *27*(6), 715–723. <https://doi.org/10.1016/j.cpr.2007.01.008>
8. Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, *88*(2), 119–127. <https://doi.org/10.1037/ccp0000469>
9. Delignières, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences*, *8*(4), 479–510.
10. Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L. J., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: Bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry*, *15*(1), 222. <https://doi.org/10.1186/s12888-015-0596-5>
11. Fried, E. I., & Robinaugh, D. J. (2020). Systems all the way down: Embracing complexity in mental health research. *BMC Medicine*, *18*(1), 205. <https://doi.org/10.1186/s12916-020-01668-w>
12. Jeronimus, B. F. (2019). Dynamic system perspectives on anxiety and depression. In E. S. Kunnen, N. M. P. de Ruiter, B. F. Jeronimus, & M. A. van der Gaag (Eds.), *Psychosocial Development in Adolescence: Insights from the Dynamic Systems Approach* (1st ed.). Routledge Psychology.
13. Boker, S. M., Staples, A. D., & Hu, Y. (2016). Dynamics of change and change in dynamics. *Journal for Person-Oriented Research*, *2*(1–2), 34–55. <https://doi.org/10.17505/jpor.2016.05>
14. Dakos, V., Carpenter, S. R., van Nes, E. H., & Scheffer, M. (2014). Resilience indicators: Prospects and limitations for early warnings of regime shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1659), 20130263–20130263. <https://doi.org/10.1098/rstb.2013.0263>
15. Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press. <https://press.princeton.edu/titles/8950.html>

16. Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I. A., Levin, S. A., & Van Nes, E. H. (2012). Anticipating critical transitions. *Science*, 338(6105), 344–348. <https://doi.org/10.1126/science.1225244>
17. Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, 234(4), 65–83.
18. Drake, J. M., & Griffen, B. D. (2010). Early warning signals of extinction in deteriorating environments. *Nature*, 467(7314), 456–459. <https://doi.org/10.1038/nature09389>
19. Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, 8(1), 25–35. <https://doi.org/10.1177/2167702619865969>
20. Schreuder, M. J., Hartman, C. A., George, S. V., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Early warning signals in psychopathology: What do they tell? *BMC Medicine*, 18(1), 1–12. <https://doi.org/10.1186/s12916-020-01742-3>
21. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>
22. Wichers, M., Smit, A. C., & Snippe, E. (2020). 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*, 6(1), 1–15. <https://doi.org/10.17505/jpor.2020.22042>
23. Helmich, M. A., Olthof, M., Oldehinkel, A. J., Wichers, M., Bringmann, L. F., & Smit, A. C. (2021). Early warning signals and critical transitions in psychopathology: Challenges and recommendations. *Current Opinion in Psychology*, 41, 51–58. <https://doi.org/10.1016/j.copsyc.2021.02.008>
24. Dablander, F., Pichler, A., Cika, A., & Bacilieri, A. (2020). Anticipating critical transitions in psychological systems using early warning signals: Theoretical and practical considerations theory of critical slowing down. *PsyArXiv*, 1–30. <https://doi.org/10.31234/osf.io/5wc28>
25. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
26. Kunkels, Y. K., Riese, H., Knapen, S. E., Lek, R. F. R. der, George, S. V., Roon, A. M. van, Schoevers, R. A., & Wichers, M. (2021). Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study. *Translational Psychiatry*, 11(1), 1–10. <https://doi.org/10.1038/s41398-021-01465-w>
27. Gijzel, S. M. W., Rector, J., van Meulen, F. B., van der Loeff, R. S., van de Leemput, I. A., Scheffer, M., Olde Rikkert, M. G. M., & Melis, R. J. F. (2020). Measurement of dynamical resilience indicators improves the prediction of recovery following hospitalization in older adults. *Journal of the American Medical Directors Association*, 21(4), 525–530.e4. <https://doi.org/10.1016/j.jamda.2019.10.011>
28. Bos, F. M., Schreuder, M. J., George, S. V., Doornbos, B., Bruggeman, R., van der Krieke, L., Haarman, B. C. M., Wichers, M., & Snippe, E. (2021). Anticipating manic and depressive shifts in patients with bipolar disorder using early warning signals. *Submitted*.
29. Olthof, M., Hasselman, F., Strunk, G., Aas, B., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Destabilization in self-ratings of the psychotherapeutic process is associated with better treatment outcome in patients with mood disorders. *Psychotherapy Research*, 30(4), 520–531. <https://doi.org/10.1080/10503307.2019.1633484>

30. Hayes, A. M., & Strauss, J. L. (1998). Dynamic systems theory as a paradigm for the study of change in psychotherapy: An application to cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 66(6), 939–947. <https://doi.org/10.1037/0022-006X.66.6.939>
31. Schiepek, G., Heinzel, S., Karch, S., Plöderl, M., & Strunk, G. (2016). Synergetics in psychology: Patterns and pattern transitions in human change processes. In G. Wunner & A. Pelster (Eds.), *Selforganization in Complex Systems: The Past, Present, and Future of Synergetics* (pp. 181–208). Springer International Publishing. http://link.springer.com/10.1007/978-3-319-27635-9_12
32. Hayes, A. M., Yasinski, C., Barnes, J. B., & Bockting, C. L. H. (2015). Network destabilization and transition in depression: New methods for studying the dynamics of therapeutic change. *Clinical Psychology Review*, 41, 27–39. <https://doi.org/10.1016/j.cpr.2015.06.007>
33. Heinzel, S., Tominschek, I., & Schiepek, G. (2014). Dynamic patterns in psychotherapy—Discontinuous changes and critical instabilities during the treatment of obsessive compulsive disorder. *Nonlinear Dynamics, Psychology, and Life Sciences*, 18(2), 155–176.
34. Strunk, G., & Lichtwarck-Aschoff, A. (2019). Therapeutic chaos. *Journal for Person-Oriented Research*, 5(2), 81–100. <https://doi.org/10.17505/jpor.2019.08>
35. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
36. Aderka, I. M., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(1), 93–101. <https://doi.org/10.1037/a0026455>
37. Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review*, 76, 101827. <https://doi.org/10.1016/j.cpr.2020.101827>
38. Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.-A., Jorasch, C., Mocanu, S., Vocks, S., Schulte, D., & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research*, 23(1), 14–24. <https://doi.org/10.1080/10503307.2012.693837>
39. Stulz, N., Lutz, W., Leach, C., Luccock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, 75(6), 864–874. <https://doi.org/10.1037/0022-006X.75.6.864>
40. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2016). Defined symptom-change trajectories during acute-phase cognitive therapy for depression predict better longitudinal outcomes. *Behaviour Research and Therapy*, 87, 48–57. <https://doi.org/10.1016/j.brat.2016.08.008>
41. Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology*, 70(9), 886–903. <https://doi.org/10.1002/jclp.22077>
42. Husen, K., Rafaeli, E., Rubel, J. A., Bar-Kalifa, E., & Lutz, W. (2016). Daily affect dynamics predict early response in CBT: Feasibility and predictive validity of EMA for outpatient psychotherapy. *Journal of Affective Disorders*, 206, 305–314. <https://doi.org/10.1016/j.jad.2016.08.025>
43. Kelley, M. E., Dunlop, B. W., Nemeroff, C. B., Lori, A., Carrillo-Roa, T., Binder, E. B., Kutner, M. H., Rivera, V. A., Craighead, W. E., & Mayberg, H. S. (2018). Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse. *Depression and Anxiety*, 35(10), 992–1000. <https://doi.org/10.1002/da.22832>
44. Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*, 58(9), 1157–1172. <https://doi.org/10.1002/jclp.10044>

45. Schiepek, G., Aichhorn, W., & Schöller, H. J. (2017). Monitoring change dynamics: A nonlinear approach to psychotherapy feedback. *Chaos and Complexity Letters*, 11(3), 355–375.
46. Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, 16, 49–74. <https://doi.org/10.1146/annurev-clinpsy-102419>
47. Bos, E. H., & De Jonge, P. (2014). ‘Critical slowing down in depression’ is a great idea that still needs empirical proof. *Proceedings of the National Academy of Sciences*, 111(10), E878–E878. <https://doi.org/10.1073/pnas.1323672111>
48. Gijzel, S. M. W., van de Leemput, I. A., Scheffer, M., Roppolo, M., Olde Rikkert, M. G. M., & Melis, R. J. F. (2017). Dynamical resilience indicators in time series of self-rated health correspond to frailty levels in older adults. *The Journals of Gerontology: Series A*, 72(7), 991–996. <https://doi.org/10.1093/gerona/glx065>
49. Kuranova, A., Booij, S. H., Menne-Lothmann, C., Decoster, J., Van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., Van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. *BMC Medicine*, 18(1), 1–11. <https://doi.org/10.1186/s12916-020-1500-9>
50. Schiepek, G. (2009). Complexity and nonlinear dynamics in psychotherapy. *European Review*, 17(02), 331. <https://doi.org/10.1017/S1062798709000763>
51. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(7), 1349–1360. <https://doi.org/10.1017/S0033291713001979>
52. Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, 7(4), 362–367. <https://doi.org/10.1177/1754073915590623>
53. Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, 1–7. <https://doi.org/10.1017/S0033291719002514>
54. Riese, H., & Wichers, M. (2019). Comment on: Eronen MI (2019). The levels problem in psychopathology. *Psychological Medicine*, 1–2. <https://doi.org/10.1017/S0033291719003623>
55. Patterson, A. C., Strang, A. G., & Abbott, K. C. (2021). When and where we can expect to see early warning signals in multispecies systems approaching tipping points: Insights from theory. *The American Naturalist*, 198(1), E12–E26. <https://doi.org/10.1086/714275>
56. Dakos, V. (2018). Identifying best-indicator species for abrupt transitions in multispecies communities. *Ecological Indicators*, 94, 494–502. <https://doi.org/10.1016/j.ecolind.2017.10.024>
57. Boerlijst, M. C., Oudman, T., & Roos, A. M. de. (2013). Catastrophic collapse can occur without early warning: Examples of silent catastrophes in structured ecological models. *PLOS ONE*, 8(4), e62033. <https://doi.org/10.1371/journal.pone.0062033>
58. Lever, J. J., Leemput, I. A. van de, Weinans, E., Quax, R., Dakos, V., Nes, E. H. van, Bascompte, J., & Scheffer, M. (2020). Foreseeing the future of mutualistic communities beyond collapse. *Ecology Letters*, 23(1), 2–15. <https://doi.org/10.1111/ele.13401>
59. Liu, R., Chen, P., Aihara, K., & Chen, L. (2015). Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. *Scientific Reports*, 5(1), 17501. <https://doi.org/10.1038/srep17501>
60. Cabrieto, J., Adolf, J., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2018). Detecting long-lived autodependency changes in a multivariate system via change point detection and regime switching models. *Scientific Reports*, 8(15637), 1–15. <https://doi.org/10.1038/s41598-018-33819-8>

61. Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H., & Scheffer, M. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One*, 7(7). <https://doi.org/10.1371/journal.pone.0041010>
62. Van Der Bolt, B., van Nes, E. H., & Scheffer, M. (2021). No warning for slow transitions. *Journal of the Royal Society, Interface*, 18(176), 20200935. <https://doi.org/10.1098/rsif.2020.0935>
63. Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322. <https://doi.org/10.1177/1754073915590619>
64. Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99(6), 1042–1060. <https://doi.org/10.1037/a0020962>
65. Haslbeck, J. M. B., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*. <https://doi.org/10.1080/00273171.2021.1896353>
66. Wichers, M., Schreuder, M. J., Goekoop, R., & Groen, R. N. (2019). Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology. *Psychological Medicine*, 49(3), 380–387. <https://doi.org/10.1017/S0033291718002064>
67. Rush, A. J., Carmody, T., & Reimnitz, P.-E. (2000). The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research*, 9(2), 45–59. <https://doi.org/10.1002/mpr.79>
68. Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Recovery: Study protocol for a repeated intensive longitudinal n = 1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *PsyArXiv, February*. <https://doi.org/10.31234/osf.io/fertg>
69. Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77(3), 600–619. <https://doi.org/10.1037/0022-3514.77.3.600>
70. Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734. <https://doi.org/10.1017/S0954579405050340>
71. Arrindell, W. A., & Ettema, J. H. M. (2003). *SCL-90: Manual to a multidimensional psychopathology indicator [SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator]* (Pearson, Ed.; 2nd ed.). Swets & Zeitlinger.
72. Arrindell, W. A., & Ettema, J. H. M. (1981). Dimensional structure, reliability and validity of the Dutch version of the Symptom Checklist (SCL-90): Data based on a phobic and a 'normal' population. [Dutch]. *Nederlands Tijdschrift Voor de Psychologie En Haar Grensgebieden*, 36(2), 108.
73. Derogatis, L. R. (1977). *SCL-90-R: administration, scoring and procedures manual-I for the R(vised) version*. John Hopkins University School Medicine.
74. Smits, I. A. M., Timmerman, M. E., Barelds, D. P. H., & Meijer, R. R. (2015). The Dutch Symptom Checklist-90-Revised: Is the use of the subscales justified? *European Journal of Psychological Assessment*, 31(4), 263–271. <https://doi.org/10.1027/1015-5759/a000233>
75. Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, 1(2), 138–155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>

76. Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2015). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, 27(2), 478–488. <https://doi.org/10.1037/pas0000060>
77. Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
78. Helmich, M. A. (2020). The Duration-adjusted Reliable Change Index (DaRCI): Defining clinically relevant symptom changes of varying durations. *PsyArXiv Preprint*. <https://doi.org/10.31234/osf.io/q7ch9>
79. Hamed, K. H., & Rao, A. R. (1998). A modified Mann–Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204(1–4), 182–196. [https://doi.org/10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X)
80. Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425. <https://doi.org/10.1037/met0000085>
81. Shapiro, Y. (2015). Dynamical Systems Therapy (DST): Theory and practical applications. *Psychoanalytic Dialogues*, 25(1), 83–107. <https://doi.org/10.1080/10481885.2015.991245>
82. Barrett, L. F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74(4), 967–984. <https://doi.org/10.1037//0022-3514.74.4.967>
83. Boettiger, C., Ross, N., & Hastings, A. (2013). Early warning signals: The charted and uncharted territories. *Theoretical Ecology*, 6(3), 255–264. <https://doi.org/10.1007/s12080-013-0192-6>
84. Hastings, A., & Wysham, D. B. (2010). Regime shifts in ecological systems can occur with no warning. *Ecology Letters*, 13(4), 464–472. <https://doi.org/10.1111/j.1461-0248.2010.01439.x>
85. Nazarimehr, F., Jafari, S., Perc, M., & Sprott, J. C. (2020). Critical slowing down indicators. *EPL*, 132(1), 18001. <https://doi.org/10.1209/0295-5075/132/18001>
86. van Nes, E. H., Arani, B. M. S., Staal, A., van der Bolt, B., Flores, B. M., Bathiany, S., & Scheffer, M. (2016). What do you mean, ‘tipping point’? *Trends in Ecology & Evolution*, 31(12), 902–904. <https://doi.org/10.1016/j.tree.2016.09.011>
87. Kéfi, S., Dakos, V., Scheffer, M., Van Nes, E. H., & Rietkerk, M. (2013). Early warning signals also precede non-catastrophic transitions. *Oikos*, 122(5), 641–648. <https://doi.org/10.1111/j.1600-0706.2012.20838.x>
88. Smit, A. C., Helmich, M. A., Bringmann, L. F., Oldehinkel, A. J., Wichers, M., & Snippe, E. (2021). Do early warning signals precede transitions toward higher levels of depressive symptoms in individual patients? *Submitted*.
89. Vries, Y. A. de, Roest, A. M., Bos, E. H., Burgerhof, J. G. M., Loo, H. M. van, & Jonge, P. de. (2019). Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: Individual patient data meta-analysis. *The British Journal of Psychiatry*, 214(1), 4–10. <https://doi.org/10.1192/bjp.2018.122>
90. Geschwind, N., Nicolson, N. A., Peeters, F., van Os, J., Barge-Schaapveld, D., & Wichers, M. (2011). Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. *European Neuropsychopharmacology*, 21(3), 241–247. <https://doi.org/10.1016/j.euroneuro.2010.11.004>
91. Allen, N. B., Nelson, B. W., Brent, D., & Auerbach, R. P. (2019). Short-term prediction of suicidal thoughts and behaviors in adolescents: Can recent developments in technology and computational science

- provide a breakthrough? *Journal of Affective Disorders*, 250, 163–169. <https://doi.org/10.1016/j.jad.2019.03.044>
92. Thompson, R. J., Mata, J., Jaeggli, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, 121(4), 819–829. <https://doi.org/10.1037/a0027978>
93. Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, 6(3), 383–391. <https://doi.org/10.1037/1528-3542.6.3.383>
94. Bos, E. H., de Jonge, P., & Cox, R. F. A. (2018). Affective variability in depression: Revisiting the inertia-instability paradox. *British Journal of Psychology*. <https://doi.org/10.1111/bjop.12372>
95. Dakos, V., Van Nes, E. H., D’Odorico, P., & Scheffer, M. (2012). Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, 93(2), 264–271. <https://doi.org/10.1890/11-0889.1>
96. Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, 56(1), 120–149. <https://doi.org/10.1080/00273171.2020.1743630>
97. Wood, S. N. (2017). Generalized additive models: An introduction with R. In *Generalized Additive Models: An Introduction with R, Second Edition* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781315370279>

Supplementary Material

o. Brief introduction

In this Supplementary Material, we show various additional plots and tables to provide further insight into the effects of different model settings (sensitivity checks) on our results, using different confidence levels for the transition identification, and the findings of our further pre-registered exploratory work.

With the low number of transition cases, it is prudent to interpret the proportion of EWS in the groups and for individual variables with caution, and we recommend looking mainly at overall patterns.

For the comparison of the analyses at different confidence levels, it is also worth noting that there is relevant overlap between the groups at different thresholds, which may affect the number of false positive and false negative EWS. For those who change transition status, the analysis is conducted as if they have no transition – over their entire time series.

7

1. Moving Window method

1.1 Main analysis

The main moving window analysis was conducted with transitions identified at 95% confidence level (CL), a minimally relevant effect size of $\tau \geq .1$, a window size of 70 observations, outliers winsorized and within-window detrending.

Figure S7.1a

Density plot of the τ -values found in the moving window analysis of the autocorrelation.

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal. Note that group sizes were unequal: $N = 9$ individuals with a transition, and $N = 32$ without a transition, at a 95% confidence threshold. For this analysis, individual time series were winsorized, detrended within windows of 70 observations. Number of early warning signals indicated in the upper right corner of each plot for both groups, meet the criterion of $\tau \geq .1$. The dotted vertical line indicates 0, and the orange vertical line indicates the $\tau \geq .1$ minimally relevant effect size threshold.

Distribution of τ for the autocorrelation

In **Figure S7.1a**, we can see that the distributions of the τ showed different spreads and peaks in the two groups — with the modes in the transition group appearing relatively more on the right (higher, more positive values of the τ), and the modes of the no-transition group appearing generally closer to zero or below. Furthermore, it is noteworthy that many people showed negative values of τ , indicating significant decreases in autocorrelation over time (both for individuals with and without transitions).

Figure S7.1b

Density plot of the τ -values found in the moving window analysis of the variance.

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal. Note that group sizes were unequal: $N = 9$ individuals with a transition, and $N = 32$ without a transition, at a 95% confidence threshold. For this analysis, individual time series were winsorized, detrended within windows of 70 observations. Number of early warning signals indicated in the upper right corner of each plot for both groups, meet the criterion of $\tau \geq .1$. The dotted vertical line indicates 0, and the orange vertical line indicates the $\tau \geq .1$ minimally relevant effect size threshold.

Distribution of τ for the variance

In **Figure S7.1b**, we can see that the differences between groups on the six variables is negligible, with the distributions largely overlapping. However, it is noteworthy that in both groups many people showed negative values of τ : significant decreases in variance over time (prior to a transition, or during the entire study period). For the people with transitions in particular, this is contrary to what we expected from the theory of early warning signals.

Table S7.1a

Summary of the main findings and sensitivity checks: number and proportion of early warning signals in the autocorrelation indicated for the various altered model settings.

	Reference results	M-K trend $p < .05$	M-K trend HR $p < .05$	Global detrend	Not winsorized	Window size 35	Window size 105	Window size 140	Shortened time series
Transition (N = 9)									
PA high	33.3% (3)	33.3% (3)	11.1% (1)	22.2% (2)	33.3% (3)	22.2% (2)	33.3% (3)	33.3% (3)	-
PA low	44.4% (4)	55.6% (5)	22.2% (2)	44.4% (4)	33.3% (3)	44.4% (4)	55.6% (5)	77.8% (7)	-
NA high	55.6% (5)	66.7% (6)	33.3% (3)	77.8% (7)	55.6% (5)	44.4% (4)	66.7% (6)	77.8% (7)	-
NA low	44.4% (4)	44.4% (4)	33.3% (3)	22.2% (2)	44.4% (4)	33.3% (3)	44.4% (4)	44.4% (4)	-
	44.4%	50%	25%	41.7%	41.7%	36.1%	50%	58.3%	-
cheerful	33.3% (3)	44.4% (4)	11.1% (1)	33.3% (3)	44.4% (4)	22.2% (2)	44.4% (4)	55.6% (5)	-
down	66.7% (6)	66.7% (6)	22.2% (2)	44.4% (4)	66.7% (6)	33.3% (3)	44.4% (4)	44.4% (4)	-
No transition (N = 32)									
PA high	18.8% (6)	28.1% (9)	6.2% (2)	31.2% (10)	18.8% (6)	18.8% (6)	25% (8)	25% (8)	25% (8)
PA low	28.1% (9)	31.2% (10)	15.6% (5)	25% (8)	25% (8)	25% (8)	31.2% (10)	31.2% (10)	37.5% (12)
NA high	31.2% (10)	37.5% (12)	15.6% (5)	28.1% (9)	28.1% (9)	21.9% (7)	37.5% (12)	37.5% (12)	31.2% (10)
NA low	31.2% (10)	31.2% (10)	12.5% (4)	34.4% (11)	31.2% (10)	21.9% (7)	37.5% (12)	31.2% (10)	31.2% (10)
	27.3%	32%	12.5%	29.7%	25.8%	21.9%	32.8%	31.2%	31.2%
cheerful	12.5% (4)	15.6% (5)	6.2% (2)	21.9% (7)	12.5% (4)	9.4% (3)	15.6% (5)	12.5% (4)	25% (8)
down	21.9% (7)	25% (8)	6.2% (2)	31.2% (10)	18.8% (6)	9.4% (3)	18.8% (6)	25% (8)	18.8% (6)

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal; M-K = Mann-Kendall rank correlation; HR = Hamed-Rao correction. Reference results are based on the analysis of transitions identified at 95% confidence level, where individual time series were winsorized, locally detrended within windows of 70 observations. Number of early warning signals based on cases with a $\tau \geq .1$

1.2 Sensitivity checks

Sensitivity checks were run to check the results of the reference moving window analysis (i.e., window size $n = 70$, outliers winsorized and within-window detrend, minimally relevant effect size of $\tau \geq .1$, transition identified at 95% confidence) against:

- Using significance of the Mann–Kendall test only (no minimally relevant effect size)
- Hamed-Rao window-dependency correction for the Mann–Kendall p -value.
- Global detrend
- No winsorizing (outliers unchanged)
- Alternate window sizes: 35 observations, 105 observations, 140 observations (1, 3, and 4 weeks, respectively)
- Different transition confidence levels: 90% and 99%

Altered model settings for the autocorrelation

Table S7.1a reveals that the number of individuals with early warning signals in the autocorrelation tended to be higher in the group without transitions, but *proportionally* (in terms of percentage) the group with transitions consistently showed more EWS across the different sensitivity checks.

Of all the checks, the Hamed-Rao correction for the M–K p -value yielded the most conservative results: true positives reduced to 25% across the four circumplex variables (i.e., PA high/low — NA high/low), and the false positives to 12.5%. Conversely, increasing the window size to 140 observations seemed to improve the rate of true positives (from 44.4% to 58.3% in the circumplex variables), along with a small increase in false positives (from 27.3% to 31.2% in the circumplex variables). However, most model setting changes led to a slight increase or decrease (ca. 2–5%) in both the true and false positive rates.

Shortening the time series in the non-transition group also appeared to result in an increase in false positive EWS for some variables, particularly for cheerful and the PA variables. For individual variables, this proportional difference—more true positives than false positives—was almost always maintained in the various analyses (some exceptions occurred in the global detrend analysis in particular), though the percentual differences were not always large.

Table S7.1b

Summary of the main findings and sensitivity checks: number and proportion of early warning signals in the variance indicated for the various altered model settings.

	Reference results	M-K trend $p < .05$	M-K trend HR $p < .05$	Global detrend	Not winsorized	Window size 35	Window size 105	Window size 140	Shortened time series
Transition (N = 9)									
PA high	0% (0)	11.1% (1)	0% (0)	0% (0)	11.1% (1)	22.2% (2)	0% (0)	11.1% (1)	-
PA low	33.3% (3)	33.3% (3)	22.2% (2)	33.3% (3)	33.3% (3)	22.2% (2)	33.3% (3)	22.2% (2)	-
NA high	11.1% (1)	11.1% (1)	11.1% (1)	11.1% (1)	22.2% (2)	0% (0)	22.2% (2)	11.1% (1)	-
NA low	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	-
	11.1%	13.9%	8.3%	11.1%	16.7%	11.1%	13.9%	11.1%	-
cheerful	33.3% (3)	33.3% (3)	0% (0)	33.3% (3)	33.3% (3)	22.2% (2)	22.2% (2)	22.2% (2)	-
down	22.2% (2)	33.3% (3)	0% (0)	11.1% (1)	11.1% (1)	33.3% (3)	22.2% (2)	22.2% (2)	-
No transition (N = 32)									
PA high	15.6% (5)	18.8% (6)	12.5% (4)	18.8% (6)	15.6% (5)	15.6% (5)	18.8% (6)	18.8% (6)	21.9% (7)
PA low	6.2% (2)	9.4% (3)	6.2% (2)	6.2% (2)	9.4% (3)	6.2% (2)	6.2% (2)	9.4% (3)	12.5% (4)
NA high	9.4% (3)	12.5% (4)	9.4% (3)	12.5% (4)	15.6% (5)	9.4% (3)	12.5% (4)	15.6% (5)	12.5% (4)
NA low	15.6% (5)	18.8% (6)	15.6% (5)	15.6% (5)	15.6% (5)	15.6% (5)	15.6% (5)	15.6% (5)	18.8% (6)
	11.7%	14.8%	10.9%	13.3%	14.1%	11.7%	13.3%	14.8%	16.4%
cheerful	12.5% (4)	12.5% (4)	9.4% (3)	15.6% (5)	12.5% (4)	12.5% (4)	15.6% (5)	21.9% (7)	12.5% (4)
down	9.4% (3)	12.5% (4)	3.1% (1)	9.4% (3)	6.2% (2)	6.2% (2)	6.2% (2)	6.2% (2)	21.9% (7)

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal; M-K = Mann-Kendall rank correlation; HR = Hamed-Rao correction. Reference results are based on the analysis of transitions identified at 95% confidence level, where individual time series were winsorized, locally detrended within windows of 70 observations. Number of early warning signals based on cases with a $\tau \geq .1$

Altered model settings for the variance

Across the sensitivity checks shown in **Table S7.1b**, the variance showed EWS in only a few variables for the individuals with transitions (range = 0 – 33.3%), while those without transitions showed at least one EWS in all variables (range = 3.1 – 21.9%). The lack of a proportional difference in EWS between individuals with and without a transition was consistent across the sensitivity checks. Averaged across the affect circumplex variables, true positives ranged from 8.3–16.7% in the various analyses, and false positives ranged from 10.9–16.4%. None of the altered model settings led to a relevant change in the results, all showing only a slight increase or decrease (ca. 2–5%) in both the true and false positive rates.

Like for the autocorrelation, the Hamed–Rao adjusted M–K p -value resulted in the most conservative outcomes, especially for the single items. Conversely, the shortened time series in the non-transition group led to the strongest increase in false positives in the circumplex variables, however, this was still only a ~5% increase compared to the main analysis. Notably, NA low arousal never showed EWS in the transition group in any of the analyses, while PA low arousal showed EWS in 2 or 3 individuals in the transition group (22.2–33.3%), with a similar 2 to 4 false positives in the no-transition group (6.2–12.5%).

Moving window results at different transition thresholds

Table S7.2 details the proportion of EWS found in the moving window analyses for the autocorrelation and variance when transitions were defined at a lower confidence level (90%), and at a higher confidence level (99%), compared to the main analysis (95%). Across the analyses for the autocorrelation, the group with transitions (consisting of 14 individuals at 90% CL, and 6 individuals at 99% CL), the proportion of identified EWS was consistently higher than in the group of individuals without a transition – with the exception of NA high and NA low arousal at 90% CL, and PA low arousal at 99% CL, where the proportion of EWS was similar for both groups. For the variance, the lack of proportional difference in EWS between groups also extended to these alternate analyses.

Table S7.2

Comparison of the number and proportion of early warning signals in the autocorrelation and variance found in the moving window analysis when different confidence levels were used to identify transitions.

	Variable	Autocorrelation			Variance		
		90% CL	95% CL	99% CL	90% CL	95% CL	99% CL
Transition	PA high	50% (7)	33.3% (3)	50% (3)	7.1% (1)	0% (0)	16.7% (1)
	PA low	50% (7)	44.4% (4)	33.3% (2)	21.4% (3)	33.3% (3)	16.7% (1)
	NA high	35.7% (5)	55.6% (5)	66.7% (4)	14.3% (2)	11.1% (1)	16.7% (1)
	NA low	35.7% (5)	44.4% (4)	50% (3)	7.1% (1)	0% (0)	16.7% (1)
No transition	cheerful	46.2%	44.4%	50%	13.5%	11.1%	16.7%
	down	42.9% (6)	33.3% (3)	33.3% (2)	28.6% (4)	33.3% (3)	16.7% (1)
		57.1% (8)	66.7% (6)	50% (3)	35.7% (5)	22.2% (2)	16.7% (1)
	PA high	18.5% (5)	18.8% (6)	17.1% (6)	18.5% (5)	15.6% (5)	14.3% (5)
	PA low	33.3% (9)	28.1% (9)	31.4% (11)	7.4% (2)	6.2% (2)	8.6% (3)
	NA high	33.3% (9)	31.2% (10)	31.4% (11)	11.1% (3)	9.4% (3)	14.3% (5)
	NA low	33.3% (9)	31.2% (10)	28.6% (10)	14.8% (4)	15.6% (5)	17.1% (6)
		22.2%	27.3%	27.1%	13%	11.7%	13.6%
	cheerful	11.1% (3)	12.5% (4)	14.3% (5)	14.8% (4)	12.5% (4)	17.1% (6)
	down	18.5% (5)	21.9% (7)	20% (7)	11.1% (3)	9.4% (3)	8.6% (3)

Note. EWS = Early warning signals; CL = confidence level; PA = Positive Affect, NA = Negative Affect; high = high arousal; low = low arousal. Total $N = 41$. At 90% CL: 13 transitions (one person had insufficient data before the transition to be analysed). At 95% CL: 9 transitions. At 99% CL: 6 transitions. Main analysis at 95% CL **in bold**. For the other CLs, all other analysis settings were equal: individual time series were winsorized, locally detrended within windows of 70 observations. Number of early warning signals based on cases with a $\tau \geq .1$

Figure S7.2

Total number of EWS per person in autocorrelation and variance combined, indicated by the moving window analysis

Note. $N = 9$ individuals with a transition, and $N = 32$ without a transition.

1.3 Further exploration

1.3.1 Combining EWS

The moving window results were further explored to see if the number of EWS per individual could be a simple way to improve the specificity of the results. We would expect that individuals with transitions show EWS in multiple variables in both autocorrelation and variance (4 variables $\times$ 2 EWS), whereas individuals without transitions would show no or fewer EWS.

Differentiating groups by number of EWS per person

Based on **Figure S7.2**, we see that the individuals with a transition all ($N = 9$) showed at least one EWS—out of a possible 8—in one of the four variables (PA high/low – NA high/low), in either the autocorrelation or variance results of the moving window analysis. In the group without transitions, most people ($N = 19$) had 0 or only 1 EWS, but there were also 13 individuals with at least two EWS.

Thus, if combining EWS and setting the threshold to a minimum of two EWS per person, this results in 77.8% true positives, and 40.6% false positives. Raising the threshold by one step ($N \text{ EWS} \geq 3$) lowers the sensitivity again, capturing 33.3% true positives, and 18.8% false positives.

Figure S7.3

Number of EWS per person in autocorrelation and variance separately, indicated by the moving window analysis.

Note. $N = 9$ individuals with a transition, and $N = 32$ without a transition.

Number of EWS per person in autocorrelation and variance separately

On the left side of **Figure S7.3**, we see that all but one (thus, 88.89%) of the individuals with a transition showed at least one EWS for the autocorrelation (out of a possible 4), compared to 62.5% in the group without transitions. Raising the minimum number of EWS in autocorrelation by one ($N \text{ EWS} \geq 2$) leads to 55.5% true positives from the transition group, and 28.1% false positives. For minimally 3 EWS in autocorrelation, the transition group shows 22.2% true positives, while the no-transition group shows 12.5% false positives.

For the variance, on the right side of **Figure S7.3**, 4 out of 9 (44.4%) individuals with a transition had one variable (out of a possible 4) with EWS, compared to 8 (25%) in the group without transitions. The remaining people with transitions had no EWS in the variance at all (i.e., 55.5% false negatives), while 12.5% of non-transitioning individuals was still found with a threshold $N \text{ EWS} \geq 2$.

1.3.2 MRES: τ exploration

As indicated in the preregistration, we examined different minimally relevant effect sizes (MRES) than the chosen τ of 0.1, to compare the balance of true positives (EWS in individuals with a transition) to false positive EWS (EWS in individuals with no transition) at different increments of τ in the moving window results. To that end, we calculated the balanced accuracy (BA; the average of the proportions of correct classifications) for the autocorrelation and variance, as follows:

1. Sensitivity or True positive rate (TPR):

$$N_{\text{EWS in the transition group}} / N_{\text{transition group}}$$
2. Specificity or True negative rate (TNR):

$$1 - (N_{\text{EWS in the non-transition group}} / N_{\text{non-transition group}})$$
3. Balanced accuracy (BA):

$$(TPR + TNR) / 2$$

Figure S7.4

Balanced Accuracy for the autocorrelation and variance results of the moving window analysis at different minimally relevant effect size thresholds (of τ)

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal. $N = 9$ individuals with a transition, and $N = 32$ without a transition. Balanced accuracy (BA): (True Positive Rate + True Negative Rate) / 2

Balanced accuracy for different increments of τ

Figure S7.4 shows the balanced accuracy for the EWS found in autocorrelation and variance with the moving window method. For the BA, a value of 50% means the rate of true positives to true negatives is equal, no better than chance; values below 50% indicate that there are more false positives or false negatives; values above 50% indicate a higher rates of true positives or negatives. We can see that across the various increments of τ , some variables remain stable in the trade-off between sensitivity and specificity indicated by the BA percentages (e.g., PA high and NA low in autocorrelation, and NA high and NA low in variance), whereas others have a higher BA at some thresholds of τ . For instance, the single item “down” seems to find a higher BA at a τ between 0 and 0.2, for both the autocorrelation and variance. Note also that for the autocorrelation, NA high appears to work relatively well across all increments. However, overall, this figure shows the BA values for our sample are modest at best, and particularly the variance performs close to chance. Given these low BA values and the inconsistent ‘optimal’ threshold for different variables, it is not possible to make any strong recommendations about appropriate τ thresholds for future research.

2. TV-AR GAM method

2.1 Main analysis and sensitivity check

The main TV-AR GAM analysis results show the different time-varying patterns that were found for the transition group, and the group without a transition (columns in bold in the table below). Additional sensitivity checks were conducted to examine the effect of the longer time series in the no-transition group, and the analysis was repeated with the transitions identified at 90% and 99% confidence level.

TV-AR-GAM results for different CL and shortened time series

Table S7.3 details the proportion of EWS found in the autocorrelation, as modeled by the TV-AR-GAM method, when transitions were defined at a lower confidence level (90%), and at a higher confidence level (99%), compared to the main analysis (95%).

Across these analyses at different confidence levels, the number of cases with a significant increase in autocorrelation is similar in the transition group and non-transition group and remarkably low. At 90%, 95% and 99%, the transition group had 6%, 5.6% and 0% EWS across the four affect circumplex variables, respectively, while the individuals without transitions showed 3.1%, 3.1% and 3.3% significant increases in autocorrelation across those four variables.

Another pattern across analyses, variables and the two groups, was that half the individuals or more (range = 48% to 87%) did not show a significant change over time at all.

Among the significant smooth curves, the proportion of cases with nonlinear changes (i.e., cubic, quadratic, or more complex curves) stands out. For instance, summarized across the four affect circumplex variables, 27.8% of the significant changes were nonlinear in the 95% CL transition group, and 25.8% in the non-transition group. The percentages at the other confidence levels were 10.7% and 25.3% at 90% CL, and 22.2% and 29% at 99% CL for the transition and no-transition groups, respectively.

Of further note are the results of the sensitivity check of the effect of time series length for individuals without transitions. For all confidence levels, the shortened time series led to fewer significant smooth curves, particularly to fewer cases with a decrease. Likely, the loss in significant smooth curves is because the TV-AR GAM had less data available in the shortened time series, and could not fit the 95% credible intervals as narrowly as it could for the full-length time series. Wider credible intervals would result in fewer cases where change could be identified with certainty.

Table S7.3

Comparison of the TV-AR GAM results: direction of the modeled smooth curves and proportion of EWS at different confidence levels and with the shortened time series in the no-transition group.

Variable	90% CL			Reference: 95% CL			99% CL		
	(Transition $N = 14$, No transition $N = 27$)			(Transition $N = 9$, No transition $N = 32$)			(Transition $N = 6$, No transition $N = 35$)		
	Prop. EWS	↗	↘	Prop. EWS	↗	↘	Prop. EWS	↗	↘
Transition									
PA high	7.1%	1	1	0%	0	2	0%	0	1
PA low	7.1%	1	0	11.1%	1	0	0%	0	0
NA high	7.1%	1	1	11.1%	1	1	0%	0	0
NA low	0%	0	0	0%	0	0	0%	0	0
cheerful	7.1%	1	1	0%	0	1	0%	0	1
down	7.1%	1	0	11.1%	1	0	0%	0	0
No transition									
PA high	0%	0	2	3.1%	1	3	2.9%	1	3
PA low	3.7%	1	3	3.1%	1	4	5.7%	2	4
NA high	3.7%	1	1	3.1%	1	2	2.9%	1	2
NA low	3.7%	1	3	3.1%	1	3	2.9%	1	3
cheerful	0%	0	5	0%	0	5	0%	0	5
down	7.4%	2	1	6.2%	2	1	5.7%	2	1

90% CL			Reference: 95% CL			99% CL									
(Transition $N=14$, No transition $N=27$)			(Transition $N=9$, No transition $N=32$)			(Transition $N=6$, No transition $N=35$)									
Variable	Prop. EWS	↗ ↘ ~ —	Prop. EWS	↗ ↘ ~ —	Prop. EWS	↗ ↘ ~ —	Prop. EWS	↗ ↘ ~ —							
Sensitivity check:															
No-transition ts set to average transition ts length															
PA high	0%	0	1	6	20	0	0	7	25	0%	0	0	6	29	
PA low	0%	0	1	9	17	0%	0	2	10	0%	0	2	11	22	
NA high	3.7%	1	1	5	20	3.1%	1	1	8	22	5.7%	2	2	8	23
NA low	3.7%	1	2	9	15	6.2%	2	1	10	19	5.7%	2	2	10	21
cheerful	0%	0	2	6	19	3.1%	1	1	8	22	2.9%	1	2	6	26
down	7.4%	2	0	6	19	3.1%	1	0	5	26	0%	0	1	7	27

Note. ↗ = Increasing trend. ↘ = Decreasing trend. ~ = Nonlinear trend. — = No trend. PA = Positive Affect; NA = Negative Affect; ts = time series; Prop. EWS = Proportion of EWS, which is calculated only over the number of cases with a significant increase in autocorrelation. Average time series length in the transition group = 334, 441 and 458 at 90%, 95% and 99% confidence, and in the no transition group 616, 616 and 615, respectively. Main analysis at 95% CL in bold. For the other CLs, all other analysis settings were equal.

8

Personalized detection of impending symptom transitions in depression using early warning signals during and shortly after antidepressant discontinuation

Based on:

Smit, A. C., **Helmich, M. A.**, Bringmann, L. F., Oldehinkel, A. J., Wichers, M., & Snippe, E. (Submitted). Personalized detection of impending symptom transitions in depression using early warning signals during and shortly after antidepressant discontinuation. *Manuscript submitted*.

Abstract

Background: Research suggests that transitions towards higher levels of depressive symptoms may be preceded by critical slowing down (CSD), which could form the basis for clinical tools predicting recurrence in individual patients. Surprisingly, nearly all empirical support for this *within-person* phenomenon has been based on *between-person* comparisons. The current study aimed to investigate if indicators of CSD called early warning signals (EWS) consistently precede recurrence in depression at the within-person level, to test if personalized risk markers based on CSD are feasible.

Method: A replicated single-subject design was used to test if EWS (increases in window autocorrelation and variance) systematically preceded transitions in depression in 37 adults who were in remission at baseline and (gradually) discontinued their antidepressant medication during the study period. Affect was measured five times daily using ecological momentary assessment over a period of four months, yielding 19,395 completed questionnaires (median 545 per participant).

Results: Significant EWS were more common in participants with a transition, in 97.3% of the preregistered model settings. Though the results suggested a good specificity (around 83.8% depending on the model settings used), no EWS were found in a substantial proportion of the participants with a transition, leading to a low sensitivity (around 32.9% depending on the model settings used).

Conclusions: Overall, the results provided an essential within-person addition to existing between-person evidence for the phenomenon of CSD *on average*. However, at the current stage EWS did not precede transitions consistently enough to be applied as a clinical tool for monitoring the risk of recurrence in *individual patients*.

Keywords: complex dynamical systems; personalized prediction models; symptom transitions; depressive symptom recurrence; replicated single-subject design

Introduction

Depression is one of the main contributors to the worldwide burden of disease, not in small part due to its recurrent nature [1]. Being able to predict if an individual will experience a recurrence of symptoms in the near future would be a major step towards improved prevention of depression in clinical practice. The vast majority of studies assessing the risk of depressive recurrence have used only cross-sectional between-person risk markers (for a review, see Ref. [2]). However, such cross-sectional methods only provide a snapshot of differences between individuals at a specific time, whilst being unable to shed light on how an individual's psychological state develops and changes over time. This has been pointed out as an important mismatch between the predominantly between-person studies, and the predominantly within-person way in which clinical practice is conducted [3–5]. It has indeed been demonstrated that group-level results can lead to severely incorrect inferences about within-person processes, which can in turn lead to poor recommendations for clinical practice [6,7]. It is therefore not surprising that in recent years there has been an increasing interest in studying changes in risk of recurrence over time at the within-person level.

Along with this increased interest in within-person research paradigms, there has been an increase in theories for conceptualizing depression as a dynamical system [3,8–10]. In this framework, each individual is viewed as a dynamical system of interacting components (e.g., emotions, cognitions, symptoms, or combinations between these) that vary and influence each other over time and may together form a dynamic stable state. Furthermore, the dynamics of these components can convey information about the resilience of the system and possibly also about upcoming transitions to alternate states.

A promising within-person risk marker for transitions within dynamical systems is Critical Slowing Down (CSD) [11,12]. CSD occurs when a system gradually becomes slower in recovering from perturbations, which signals a process of destabilization that leads to an increasing risk that even small perturbations may eventually result in large, potentially detrimental, transitions (e.g., the transition from remission to a state in which depressive symptoms are elevated). CSD has indeed been shown to precede such 'critical transitions' in a wide range of simulated dynamical systems, including simulations aimed at understanding depression

[8,13]. Importantly, CSD is also empirically measurable, as it leads to increases in the autocorrelation and variance of time-series of observations of a wide range of dynamical systems, and these so-called Early Warning Signals (EWS) have been shown to anticipate upcoming transitions in a wide range of empirical time-series, such as ecology, climate change, geology, neurology, microbiology, chemistry, physics, and engineering [14–17]. As it has become apparent that depression is a highly heterogeneous concept [18–20], such broadly applicable EWS may provide a consistent and reliable warning for upcoming transitions towards higher levels of depression, even if the precise mechanisms and stressors leading up to the recurrence of depression vary from patient to patient.

Though it has been suggested before that EWS may be useful when the goal is to anticipate acute severity transitions in chronic diseases such as depression [17], the body of literature that tests this idea empirically is not without limitations [16]. Some evidence has been found that autocorrelation and variance in momentary affect data may be higher prior to increases in depressive symptoms [21–25]. However, due to the between-persons design used in these studies, we do not know whether there truly was a *rise* in autocorrelation and variance before depressive symptoms increased [26]. Only if we find that EWS systematically precede transitions towards higher levels of depressive symptoms within single individuals, we will know that these transitions are indeed preceded by the process of destabilization that is predicted by dynamical systems theory [26]. Hence, only by investigating EWS within-persons will we be able to determine whether preventative treatments should be focused on individuals who destabilize over time, for example when they go through vulnerable periods.

To examine whether CSD takes place within individuals, it is necessary to intensively study the period before transitions towards higher levels of depressive symptoms in individual patients. Such a radically new design to study this within-person process was introduced by Wichers et al. [27], who measured mood prospectively using Ecological Momentary Assessment (EMA) [28–31] multiple times a day over the course of several months. By starting when depression was still in remission and continuing until after recurrence, they captured the entire period that led up to recurrence. Promisingly, the results from this N=1 study [27], and from studies that applied different analyses to the data of that same individual [32,33], were in line with the theory that the recurrence of depressive symptoms

is preceded by a period of within-person destabilization indicated by EWS, such as rises in autocorrelation and variance. A confirmatory study using a new dataset found the same result in a second individual who experienced a transition towards higher levels of depression [34].

Given that EWS have been studied on the within-person level in only two individuals with a transition towards higher levels of depression, it is clear that these results need to be repeated in many more individuals. To date, this has not been done, as finding many patients willing to fill out a questionnaire multiple times a day for a period of several months during a potentially vulnerable period is a challenge, and keeping participants motivated to provide high-quality data requires a major investment of time and resources. The current study analyzed 37 individuals that were in remission of depression at baseline, who (gradually) discontinued their antidepressant medication during the study. Participants were monitored during this period, as patients are more vulnerable to experience transitions towards higher levels of depressive symptoms when they discontinue their medication [35]. In each of these individuals, mood was measured five times a day during a four-month period. Repeating such an intensive longitudinal design in this many individuals enabled us to study whether we could systematically replicate the finding that CSD occurs before transitions towards higher levels of depressive symptoms.

In line with theoretical literature on EWS and the empirical evidence found so far in depression, we hypothesized that (a) autocorrelation and variance of EMA-measured affect show a significant increase in the period before symptoms recur, and (b) that this rise is absent or significantly smaller in participants that remain in remission. Based on previous findings and theoretical considerations, we expected to find the strongest EWS before transitions towards higher levels of depressive symptoms in the closest proxy of depression [9,21,36]. Therefore, we focused the analysis primarily on negative affect with low arousal (NA low) in the current data, though other variables based on the affect circumplex were also tested [37,38]. The presence of EWS in the majority of patients with recurring symptoms, and absence in the majority of others, could be a major breakthrough in the monitoring and management of recurrence risk in depression for individual patients.

Methods

Participants

Data from the TRANS-ID Tapering project was used for the current study. For an in-depth description of the study protocol and more information on participants see <https://osf.io/zbwkp/> or Refs. [39,40].

The sample consisted of adults (age ≥ 18) who were on antidepressant medication at baseline because of a past depressive episode, and made a shared decision with their general practitioner or psychiatrist to discontinue their antidepressant medication. The study was advertised in local and national newspapers, online (e.g., via social media), and through a Dutch pharmacy when patients ordered antidepressant tapering strips [41]. Participants were included between June 2017 and October 2018. Exclusion criteria were (a) not being in remission at baseline, (b) psychotic or manic episode at baseline (a and b were assessed using the mini-SCAN [42]), (c) reported diagnosis of personality disorder, (d) starting another antidepressant, or (e) not being able to work with a smartphone. The flow chart (**Figure 8.1**) summarizes recruitment and inclusion.

All participants provided written informed consent. The study procedures were reviewed and approved by the Medical Ethical Committee of the University Medical Center Groningen. All study procedures complied with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Ecological momentary assessment (EMA)

The EMA measurements consisted of five smartphone-based measurements a day, over the course of four months (~600 prompts per participant). Measurements took about two minutes to complete, and were set at fixed intervals of three hours regarding the sleep-wake habits of the participant. The full EMA-questionnaire is available on OSF (<https://osf.io/zbwkp/>; [39]).

The EMA-questionnaires included ten momentary affect items based on the affect circumplex [37,38], which were answered on a visual analog scale from ‘not at all’, to ‘very much’. Items were standardized within-person, and averaged

Figure 8.1*Flow chart summarizing recruitment and inclusion.*

at each time point to form the four mental states used in the analysis. Negative affect low arousal (NA low) included: down, listless, and tired; negative affect high arousal (NA high) included: restless, stressed, and irritated; positive affect low arousal (PA low) included: content, and at ease; and positive affect high arousal (PA high) included: cheerful, and energetic. A total mood score was also obtained, by averaging all standardized items, reversing scores on positively stated items.

The compliance was good (median 88.5%), yielding a median of 545 observations per individual (range = 344-609), and 19,395 completed questionnaires in total.

Transitions towards higher levels of depression

A combination of quantitative and qualitative criteria was used to determine which participants experienced a transition towards higher levels of depression that was (a) statistically reliable, (b) persistent, and (c) clinically relevant. The reliable change index [43] was used to test if depressive symptoms, measured using the Symptom Checklist-90 (SCL-90) depression subscale [44], showed a *statistically reliable* increase in depressive symptoms compared to the average of the SCL-90 depression scores obtained during the first two weeks of the study (i.e., when participants were still in remission). This change was considered *persistent* if the SCL-90 score remained elevated for at least three consecutive weeks, and/or treatment was started or increased, and/or tapering was interrupted. To assess *clinical relevance*, three independent raters inspected qualitative data obtained in evaluation interviews, three to four-weekly phone calls during the research period, and comment sections that were included in all EMA and SCL-90 questionnaires, and determined if participants had indeed experienced a substantial increase in depressive symptoms that had caused them discomfort in daily life. For a more in-depth description of criteria used, see Ref. [40].

In previous studies [27,34], the rise in EWS started long before the transition, hence a transition shortly after the four-month EMA period may still show EWS in the EMA data. Therefore, transitions were investigated for a six-month period, also including the two months directly after the EMA data collection had ended.

Analysis

At the within-person level, the analysis closely resembled the analysis of the previous N=1 studies [27,34]. A more detailed version of the analysis plan, including more information on preprocessing and model choices, was published on OSF before the data was analyzed (accessible online at <https://osf.io/x58nc/>).

First, the lag-1 autocorrelation and variance in NA low were estimated in a window of 28 days (max. 140 observations per window). Within the window, data was first detrended using robust linear regression, and then Winsorized (i.e., extreme values were capped at the 5th and 95th percentile). Missing data were not imputed, and during the calculation of the autocorrelation observations were not

regressed on the previous observation if it was missing, or if it was the previous day. The calculations were repeated, each time moving the window forward one observation, until the final observation was reached.

After these time series of window autocorrelation and variance were obtained, the Mann–Kendall (M–K) trend test was used to estimate the trend in these window time-series of autocorrelation and variance. All available windows were used for participants without a transition; in participants with a transition the windows including observations made after the start of the transition were omitted from the analysis. To improve reliability, the test was only performed for individuals for whom at least 140 window estimates could be included in the analysis, leaving 19 participants with a transition, and 18 participants without a transition. To improve the specificity of the test, a minimum relevant effect size of $\tau > 0.1$ was used, and in the remainder of the manuscript an increase in EWS will only be called ‘significant’ if both $p < 0.05$, and $\tau > 0.1$. We hypothesized such significant increases would be found in a higher percentage of participants that experienced a transition (i.e., sensitivity) than in participants that remained in remission (i.e., $1 - \text{Specificity}$).

The distribution of within-person estimates of τ -values summarizing the trend in autocorrelation and variance in each individual were used to gain insight in group-level differences in EWS. We hypothesized that the τ -values would be distributed (a) above zero for participants with a transition, indicating an increase in autocorrelation and variance before transitions, and (b) around zero for participants that remained in remission, indicating no trend in autocorrelation and variance. We tested if the average τ -value was significantly higher (i.e., one-sided) for participants with a transition using Welch’s t-test. Additionally, we tested how many days before the transition the EWS became significant and remained significant until the transition, and if EWS were more common in participants with a sudden transition compared to a gradual transition (defined in Ref. [40]).

Though the method above was expected to yield the most defensible and powerful results, the robustness of these results was tested by repeating the analysis using a range of 17 different preprocessing and model settings that were specified in the preregistration (<https://osf.io/x58nc/>, see also **Supplement 2**). First, the analysis was repeated using different variables: NA high, PA low, PA high, total mood, and single variables ‘down’, ‘listless’, and ‘tired’. Then robustness against model settings was tested using (a) window sizes of 105 and 70, (b) including

also participants for whom EWS could be estimated in 70 to 140 windows, (c) using exactly 140 windows for each participant, (d) detrending globally or not at all, (e) removing outliers or leaving them in, and (f) omitting participants that experienced a transition more than two weeks after the last EMA observation.

Results

Demonstration cases: Individual analysis

For each individual patient with a transition, we hypothesized that autocorrelation and variance would increase significantly before the transition. To demonstrate this principle, **Figure 8.2** shows window estimates of autocorrelation and variance for one participant with a transition. For this participant, visual inspection seems to indicate an increase in both EWS in this case, which was confirmed by the M-K trend test (autocorrelation: $\tau = 0.41$, $p < 0.001$; variance: $\tau = 0.68$, $p < 0.001$). As the results were both significant, and the effect size was larger than the prespecified minimum relevant effect size of 0.1, we conclude that this patient demonstrated a significant rise in autocorrelation and variance before the transition. **Figure 8.2** shows the window estimates for a participant without a transition. The autocorrelation did not increase in this case ($\tau = -0.04$, $p = 0.26$), and the variance even decreased significantly over time ($\tau = -0.65$, $p < 0.001$).

Main analyses: NA low arousal

A total of 19 of the 33 participants with a transition had sufficient data to estimate EWS in at least 140 windows before the transition, and 18 of the 19 participants without a transition had sufficient data for at least 140 window estimates. Out of the 19 participants with a transition, 7 (36.8%) showed a significant increase in the lag-1 autocorrelation of NA low and 5 (26.3%) a significant increase in the variance of NA low before the transition (see **Supplement 1** for overlap between these groups). As hypothesized, of the 18 participants without a transition, the number of significant EWS was smaller, as only 4 (22.2%) and 2 (11.1%) showed a significant increase in autocorrelation and variance respectively.

Figure 8.2

Window estimates of autocorrelation and variance for a participant with and without a transition in depressive symptoms.

Note. Dots represent window estimates in autocorrelation (left), and variance (right), calculated using all available data in the 28 days prior to the indicated dot. This means no window estimates are available during the first 28 days, and for the participant with a transition (top), and estimates before the transition only include data from before the start of the transition. Two visually clear cases were selected for demonstration purposes.

As hypothesized, the average M-K trend in autocorrelation and variance was higher for participants with a transition (mean M-K trend in: autocorrelation = -0.019 ; variance = -0.189) than for participants without a transition (mean M-K trend in: autocorrelation = -0.175 ; variance = -0.468). This difference did not reach significance for autocorrelation ($t(34.955) = 1.3674$, $p = .090$, Cohen's $d = 0.45$), but was significant for variance ($t(33.672) = 1.8262$, $p = .038$, Cohen's $d = 0.60$). The density plot in **Figure 8.3** shows the distribution of the M-K trend in autocorrelation and variance for participants with a transition and without a transition. Notice that the average M-K trend was negative in both EWS and both

Figure 8.3

Between person distribution of the M-K trend for the autocorrelation and variance.

Note. Orange vertical lines indicate the minimum relevant effect size of $\tau > 0.1$ used in this study.

groups, indicating that on average the autocorrelation and variance tended to decrease over time. **Figure 8.3** also shows that there was a substantial overlap between groups, illustrating why the difference in average trend did not lead to a large difference in the number of participants showing EWS in these groups.

For participants who showed a significant increase in autocorrelation or variance, on average, the M-K trend became significant 56.7 and 37.8 days, respectively, before the start of the transition in depressive symptoms (range: autocorrelation = 3–111 days; variance = 12–69 days) and remained significant until the transition. This shows that true positive increases in autocorrelation and variance typically start long before a transition in depressive symptoms.

Though it was hypothesized that EWS would be more likely to appear in participants with a sudden transition in depressive symptoms compared to participants with a gradual transition, this was not confirmed by the data. Participants with a sudden transition ($N = 9$) showed a lower percentage of significant rises in autocorrelation (22.2%) and variance (11.1%) than participants with a gradual transition ($N = 10$; 50.0% and 40.0% respectively).

Figure 8.4

Proportion of true and false positives in different variables for different EWS.

Note. ACor = lag-1 autocorrelation; Var = variance; NA = negative affect; PA = positive affect; EWS = early warning signals.

Secondary analyses: NA high, NA low, PA high, total mood and separate items

We repeated the tests for increases in autocorrelation and variance using the variables NA high, PA low, and PA high, yielding a total of 8 EWS tested for each individual. In all variables, proportionally more increases in autocorrelation and variance were found in the group of participants with a transition, compared to participants without a transition (see **Figure 8.4**). In 7 of the 8 EWS, the number of true positives (i.e., pre-transition EWS) was more than double the number of false positives (i.e., EWS in the non-transition group), see **Figure 8.4**. A full overview of all significant EWS for each individual can be found in **Supplement 1**.

Out of the 8 combinations of variable and EWS, the average number of EWS found in participants with a transition was more than three times larger than for participants without a transition (2.79 per participant with a transition, versus 0.89 per participant without a transition). Though participants with a transition had more EWS on average, which EWS were significant differed from person to person (see **Table S8.1 in Supplement 1**). Based on this, it seems likely

that the predictive performance of EWS can be improved substantially by using a combination of variables. Even a simple distinction between participants with 3 or more versus fewer than 3 significant rises across the total of 8 tests in these 4 variables would have resulted in detecting 10/19 (52.6%) of transitions, at the cost of only 2/18 (11.1%) of false positives. A decision rule based on 5 or more significant rises would have resulted in detecting 6/19 (31.6%) true positives, at the cost of no false positives (see **Figure 8.4**). Interestingly, simply combining all items into a “total mood” variable did not have the same beneficial effect (see **Figure 8.4**).

Sensitivity analyses

To test whether the group difference in M–K trends in autocorrelation and variance were maintained across different model settings, we varied: window size, minimum and maximum number of windows used in the analysis, and how to deal with trends and outliers. **Figure 8.5** shows a summary of the Cohen’s *d* and *p*-value of the t-tests testing if the average M–K trend in autocorrelation and variance was larger

Figure 8.5
Distribution of results across the 17 different model settings used in the sensitivity check.

for participants with versus without a transition, for all 17 analyses with different model settings. The M–K trend in autocorrelation was not significantly larger for participants with a transition compared to participants without a transition in most model settings, confirming the results of the main analysis that the effect size for this EWS was too low to be detected with the current sample size. In contrast, the M–K trend in variance was significantly larger in participants with a transition in 11 out of 17 model settings. Also, the effect size tended to be larger for variance than for autocorrelation. The direction of the effect was highly robust, as the effect had the hypothesized direction in nearly all of the preregistered settings (16/17 = 94.1% for autocorrelation, and 17/17 = 100% for variance). This robustness also translated to the individual level: the percentage of true positives (mean = 32.9%, $SD = 6.3\%$) was larger than the percentage of false positives (mean = 16.2%, $SD = 9.9\%$) in the same settings. The one setting that showed a reversed effect for autocorrelation, was when the analysis was limited to 140 windows instead of all available windows. Model settings and results for each of the tested model settings separately can be found in **Supplement 2, Table S8.2**.

Discussion

At the individual level, results indicated that significant increases in both the autocorrelation and variance of the separate affect measures were more common in participants with a transition than in participants without a transition. We did not only find this for NA low, on which our primary hypothesis was focused, but also for NA high, PA low, and PA high. At the group level, we found evidence that the average trend in variance of these affect measures tended to be significantly higher for participants with a transition than for participants that remained in remission. Though the effect had the same direction for autocorrelation, the effect size was somewhat smaller for this EWS, and the difference between participants with and without a transition did not reach significance. Our findings were highly robust: in 94.1% and 100% of the preregistered model settings rises in autocorrelation and variance respectively were more commonly found in the participants with a transition than in participants that remained in remission, and the trend in EWS was, on average, higher in participants with a transition. Though the results suggested a good specificity (around 83.8% depending on the model settings used), no EWS

were found in a substantial proportion of the participants with a transition towards higher levels of depression, leading to a low sensitivity (around 32.9% depending on the model settings used). Overall, the results of the current study show that, despite the evidence for the phenomenon of CSD *on average*, at the current stage EWS did not precede transitions consistently enough to be applied as a clinical tool for monitoring the risk of recurrence in *individual patients*.

Unlike previous studies using between-persons designs to study destabilization in affect dynamics [21–25], results from the current study cannot be explained by stable between-person differences [26]. Therefore, the current study is an essential addition to existing literature as it provides evidence that a process of within-person destabilization over time, which can be detected using EWS, occurs more often and is stronger on average before transitions towards higher levels of depression.

As the two earlier studies that investigated EWS at the within-person level over time only included a single patient each [27,34], the current study was the first to replicate these studies in a larger sample and the first to obtain estimates of the sensitivity and specificity of using EWS to predict transitions. Though it has been suggested that EWS based on CSD may be a fruitful basis for clinical interventions based on group-level studies [16,17,21], the current results suggest such application may not be realistic in the near future. EWS were only found in a limited percentage of individuals with transitions, underlining the importance of postponing conclusions about within-persons processes of change until within-persons studies are performed [3–7,26].

The absence of EWS in the majority of participants with a transition, may indicate that in the majority of patients, the recurrence of depressive symptoms is *not* a critical transition in which small perturbations lead to large and relatively sudden increases in depressive symptoms. However, from theoretical work on EWS, it is known that in some cases, systems can go through a critical transition without showing CSD, e.g., when a large perturbation causes a transition even though the system was still quite stable, effectively skipping the process of destabilization that would cause CSD (see Ref. [45] for additional examples). In addition, even if CSD is present, it may not always be detectable using EWS, e.g., because it may appear in variables that could not be obtained (see Ref. [46] for additional examples). The absence of EWS in some individuals with a transition towards higher levels of

depression should therefore not be interpreted as evidence that individuals do not function like a dynamical system.

A more practical explanation for why the difference in EWS at the group-level did not lead to a high sensitivity, would be the unexpected result that the majority of participants showed a downward trend in both EWS. This downward bias may have been caused by participants answering more precisely and using less of the available scale when they become more accustomed to filling out the questionnaire. Though a formal investigation of this effect is needed to gain more insights in the causes for this average downward trend, it is likely that this bias contributed substantially to the low number of significant EWS that were found in both groups, leading to a low sensitivity and a high specificity. Correcting for this average decrease in EWS would increase the sensitivity, though this may come at the cost of a lower specificity.

Even though the sensitivity of separate EWS in separate variables was quite low, the results indicated that participants who experienced a transition showed more than three times as many EWS as participants who remained in remission. With this observation, the current study provides preliminary evidence that the performance of EWS can be improved by combining multiple variables and EWS. In the current sample, the sensitivity could be improved substantially: from around 34.9% (depending on the variable and EWS used) for one variable and EWS, to 52.6% when interpreting three or more of these separate metrics as an EWS, without negatively impacting the specificity, which remained at 88.9%. Optimizing the way in which multiple EWS and variables can be combined to form an ‘ensemble EWS’ that performs well at the individual level would be an interesting topic for future investigation.

Though the specificity in the current study was quite good, some participants without a symptom transition still showed EWS. These false positives may be explained by the fact that some individuals discontinuing antidepressant medication experience a period in which they are less resilient, even though this may not always lead to symptom transitions. Recently the term ‘dynamic indicators of resilience’ has been coined to refer to the same metrics used as EWS, to better reflect the fact that these metrics aim to detect destabilization which does not always lead to transitions [47]. To summarize, we need more insight in (a) the possibility of improving the accuracy of EWS, and (b) the reasons behind false

negatives and false positives in this specific type of transition, before EWS can be used for clinical purposes.

A limitation of the current study was its limited between-persons sample size. As Voelkle et al. [48] pointed out: in a realistic world, researchers are forced to make the trade-off between maximizing the number of participants, and maximizing the number of observations per participant. Furthermore, this study was designed as a replicated single-subject design with the aim to systematically replicate the finding that CSD precedes recurrence of depressive symptoms in multiple individuals. Though we obtained the required intensive longitudinal data (median number of observations per individual = 545, range = 344–609) in a sample that was unprecedentedly large for such an intensive within-person design ($N = 37$), the between-person sample size was still on the small side for statistical testing at the between-person level. The combination between the medium between-person sample size, and the small to medium effect size (Cohen's d around 0.36, depending on the model settings used) meant that the difference in the trend in autocorrelation did not reach significance. Despite this medium between-person sample size, the difference between the groups in the trend in variance reached significance, as the effect size was medium to large (Cohen's d around 0.62, depending on the model settings used).

This study is a strong addition to the existing literature, as it has an unprecedentedly large sample of individual patients, who were followed intensively with EMA during and after antidepressant discontinuation, allowing us to capture what happens just before depressive symptom transitions in multiple individuals. In line with previous between-persons and $N = 1$ studies [21–25,27,34], our findings provide important empirical evidence that the process of within-person destabilization in affect over time is stronger on average and occurs more often in participants with transitions than in individuals that remain in remission. Despite this, we found that the autocorrelation and variance of positive and negative affect variables did not show a significant increase preceding transitions in most individuals. Though preliminary evidence suggests that the percentage of EWS found in participants with a transition may be improved substantially by monitoring multiple EWS in multiple variables simultaneously, at the current stage the accuracy of EWS in predicting depressive recurrence seems insufficient for clinical applications.

References

1. Eaton, W. W., Shao, H., Nestadt, G., Lee, B. H., Bienvenu, O. J., & Zandi, P. (2008). Population-based study of first onset and chronicity in major depressive disorder. *Archives of General Psychiatry*, 65(5), 9.
2. Buckman, J. E. J., Underwood, A., Clarke, K., Saunders, R., Hollon, S. D., Fearon, P., & Pilling, S. (2018). Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clinical Psychology Review*, 64, 13–38. <https://doi.org/10.1016/j.cpr.2018.07.005>
3. Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder a review. *JAMA Psychiatry*, 74(5), 528–534. <https://doi.org/10.1001/jamapsychiatry.2017.0001>
4. Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy*, 117, 40–53. <https://doi.org/10.1016/j.brat.2018.10.005>
5. Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, 16, 49–74. <https://doi.org/10.1146/annurev-clinpsy-102419>
6. Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
7. Hamaker, E. L. (2012). Why researchers should think ‘within-person’: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 43–61). Guilford.
8. Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS One*, 11(12), [e0167490]. <https://doi.org/10.1371/journal.pone.0167490>
9. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(7), 1349–1360. <https://doi.org/10.1017/S0033291713001979>
10. Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, 7(4), 362–367. <https://doi.org/10.1177/1754073915590623>
11. Strogatz, S. H. (2015). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering* (2nd ed.). CRC Press.
12. Wissel, C. (1984). A universal law of the characteristic return time near thresholds. *Oecologia*, 65(1), 101–107.
13. van Nes, E. H., & Scheffer, M. (2007). Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *The American Naturalist*, 169(6), 738–747. <https://doi.org/10.1086/516845>
14. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53–59. <https://doi.org/10.1038/nature08227>
15. Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I. A., Levin, S. A., & Van Nes, E. H. (2012). Anticipating critical transitions. *Science*, 338(6105), 344–348. <https://doi.org/10.1126/science.1225244>
16. Trefois, C., Antony, P. M., Goncalves, J., Skupin, A., & Balling, R. (2015). Critical transitions in chronic disease: Transferring concepts from ecology to systems medicine. *Current Opinion in Biotechnology*, 34, 48–55. <https://doi.org/10.1016/j.copbio.2014.11.020>

17. Olde Rikkert, M. G. M., Dakos, V., Buchman, T. G., Boer, R. de, Glass, L., Cramer, A. O. J., Levin, S., van Nes, E., Sugihara, G., Ferrari, M. D., Tolner, E. A., van de Leemput, I., Lagro, J., Melis, R., & Scheffer, M. (2016). Slowing down of recovery as generic risk marker for acute severity transitions in chronic diseases. *Critical Care Medicine*, *44*(3), 601–606. <https://doi.org/10.1097/CCM.0000000000001564>
18. Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(72). <https://doi.org/10.1186/s12916-015-0325-4>
19. Fried, E. (2017). Moving forward: How depression heterogeneity hinders progress in treatment and research. *Expert Review of Neurotherapeutics*, *17*(5), 423–425. <https://doi.org/10.1080/14737175.2017.1307737>
20. Monroe, S. M., & Anderson, S. F. (2015). Depression: The shroud of heterogeneity. *Current Directions in Psychological Science*, *24*(3), 227–231. <https://doi.org/10.1177/0963721414568342>
21. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 87–92. <https://doi.org/10.1073/pnas.1312114110>
22. Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, *12*(2), 283–289. <https://doi.org/10.2337/db10-1371>
23. Schreuder, M. J., Hartman, C. A., George, S. V., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Early warning signals in psychopathology: What do they tell? *BMC Medicine*, *18*(1), 1–12. <https://doi.org/10.1186/s12916-020-01742-3>
24. Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., Thiery, E., Delespaul, P. H., & van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, *124*(1–2), 191–195. <https://doi.org/10.1016/j.jad.2009.11.010>
25. Panaite, V., Rottenberg, J., & Bylsma, L. M. (2020). Daily affective dynamics predict depression symptom trajectories among adults with major and minor depression. *Affective Science*, *1*(3), 186–198. <https://doi.org/10.1007/s42761-020-00014-w>
26. Bos, E. H., & De Jonge, P. (2014). 'Critical slowing down in depression' is a great idea that still needs empirical proof. *Proceedings of the National Academy of Sciences*, *111*(10), E878–E878. <https://doi.org/10.1073/pnas.1323672111>
27. Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
28. Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 35–54). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_3
29. Myin-Germeyns, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine*, *39*(9), 1533–1547. <https://doi.org/10.1017/S0033291708004947>
30. Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment*, *21*(4), 457–462. <https://doi.org/10.1037/a0017653>

31. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>
32. Cabrieto, J., Tuerlinckx, F., Kuppens, P., Hunyadi, B., & Ceulemans, E. (2018). Testing for the presence of correlation changes in a multivariate time series: A permutation based approach. *Scientific Reports*, 8(769), 1–20. <https://doi.org/10.1038/s41598-017-19067-2>
33. Albers, C., & Bringmann, L. F. (2020). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *European Journal of Psychological Assessment*, 36(February), 492–499. <https://doi.org/10.1027/1015-5759/a000589>
34. Wichers, M., Smit, A. C., & Snippe, E. (2020). 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*, 6(1), 1–15. <https://doi.org/10.17505/jpor.2020.22042>
35. Borges, S., Chen, Y.-F., Laughren, T. P., Temple, R., Patel, H. D., David, P. A., Mathis, M., Unger, E., Yang, P., & Khin, N. A. (2014). Review of maintenance trials for major depressive disorder: A 25-year perspective from the us Food and Drug Administration. *The Journal of Clinical Psychiatry*, 75(03), 205–214. <https://doi.org/10.4088/JCP.13r08722>
36. Wichers, M., Schreuder, M. J., Goekoop, R., & Groen, R. N. (2019). Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology. *Psychological Medicine*, 49(3), 380–387. <https://doi.org/10.1017/S003329718002064>
37. Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734. <https://doi.org/10.1017/S0954579405050340>
38. Yik, M. S. M., Russell, J. A., & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77(3), 600–619. <https://doi.org/10.1037/0022-3514.77.3.600>
39. Smit, A. C., Helmich, M. A., Kunkels, Y. K., Riese, H., Snippe, E., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Tapering: Study protocol for a repeated intensive longitudinal n=1 study design to search for personalized early warning signals of critical transitions towards depression. *OSF*. <https://osf.io/zbwkp/>
40. Smit, A. C., Snippe, E., Hoenders, H. J. R., & Wichers, M. (2020). Transitions In Depression: If, how, and when depressive symptoms increase during and after tapering of antidepressant medication. *Under Review*.
41. Groot, P. C., & van Os, J. (2018). Antidepressant tapering strips to help people come off medication more safely. *Psychosis*, 10(2), 142–145. <https://doi.org/10.1080/17522439.2018.1469163>
42. Nienhuis, F. J., Van De Willige, G., Rijnders, C. A. T., De Jonge, P., & Wiersma, D. (2010). Validity of a short clinical interview for psychiatric diagnosis: The mini-SCAN. *British Journal of Psychiatry*, 196(1), 64–68. <https://doi.org/10.1192/bjp.bp.109.066563>
43. Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19. <https://doi.org/10.1037/0022-006X.59.1.12>
44. Arrindell, W. A., & Ettema, J. H. M. (2003). *SCL-90: Manual to a multidimensional psychopathology indicator [SCL-90: Handleiding bij een multidimensionale psychopathologie-indicator]* (2nd ed.). Swets & Zeitlinger.
45. Boettiger, C., Ross, N., & Hastings, A. (2013). Early warning signals: The charted and uncharted territories. *Theoretical Ecology*, 6(3), 255–264. <https://doi.org/10.1007/s12080-013-0192-6>

46. Boerlijst, M. C., Oudman, T., & Roos, A. M. de. (2013). Catastrophic collapse can occur without early warning: Examples of silent catastrophes in structured ecological models. *PLOS ONE*, 8(4), e62033. <https://doi.org/10.1371/journal.pone.0062033>
47. Scheffer, M., Elizabeth Bolhuis, J., Borsboom, D., Buchman, T. G., Gijzel, S. M. W., Goulson, D., Kammenga, J. E., Kemp, B., van de Leemput, I. A., Levin, S., Martin, C. M., Melis, R. J. F., van Nes, E. H., Michael Romero, L., & Olde Rikkert, M. G. M. (2018). Quantifying resilience of humans and other animals. *Proceedings of the National Academy of Sciences of the United States of America*, 115(47), 11883–11890. <https://doi.org/10.1073/pnas.1810630115>
48. Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, 49(3), 193–213. <https://doi.org/10.1080/00273171.2014.889593>

Supplementary Materials

Supplement 1

Table S8.1

Overview of all significant combinations of variable and EWS per participant.

Transition										No transition									
NA low		NA high		PA low		PA high		Sum		NA low		NA high		PA low		PA high		Sum	
1	0	1	0	1	1	1	0	5		0	1	0	0	0	0	0	0	1	
0	1	0	0	0	1	0	1	3		0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	0	1		1	0	0	0	1	0	0	0	2	
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	
1	1	0	1	1	1	0	0	5		0	0	0	0	0	0	0	0	0	
1	0	1	1	1	1	1	0	6		0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	
0	0	1	0	1	0	1	0	3		1	1	0	0	1	0	0	0	3	
0	0	0	0	0	0	0	0	0		0	1	0	0	0	0	0	0	1	
0	0	0	0	1	0	0	0	1		0	0	0	0	0	0	0	0	0	
1	0	1	1	0	1	1	0	5		0	0	1	1	0	0	0	0	1	
1	1	1	0	1	1	1	1	7		0	0	0	0	0	0	0	0	0	
1	1	1	1	1	1	1	1	8		0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	1	0	1	
0	0	0	0	0	0	0	1	1		0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0		1	0	1	0	0	0	0	0	0	
0	1	0	0	0	0	0	0	1		1	1	0	1	1	0	1	0	2	
0	1	0	1	0	0	0	1	3		1	0	0	0	0	0	0	0	4	
0	0	1	1	1	0	1	0	4		0	0	0	0	1	0	0	0	1	
0.37	0.26	0.37	0.32	0.42	0.37	0.42	0.26	2.79		0.22	0.11	0.11	0.11	0.17	0.06	0.06	0.06	0.89	

Note. A cell shows “1” if the trend in that variable and EWS if the trend was significant, and $\tau > 0.1$. For participants with a transition (left), the bottom row shows the proportion of true positives in the sample (Sensitivity); for participants without a transition (right), the bottom row shows the proportion of false positives in the sample (1-Specificity). The last column shows the number of significant EWS per participant, and the cell at the bottom right shows the average number of EWS per participant.

Supplement 2

Based on pilot data [27,34], simulations, and theory, the analysis we expected to perform best was preregistered as “main analysis”. To (a) test the robustness of these main analyses, and (b) explore the data to better understand the results, a total of 16 sensitivity checks were also preregistered. These alternative analyses were replications of the main analysis, except each time one variable or setting was changed, to investigate how the analysis would react to this change.

Below, we first describe the settings of the main analysis. For each alternative analysis, we then describe what was different from the main analysis in that iteration, all other settings were identical to the main analysis. **Table S8.2** below shows an overview of the results for each of these analyses separately. More information on model choices can be found in the preregistration (<https://osf.io/x58nc/>).

1. Main analysis: Used variable NA low, window size 4 weeks, participants were included if at least 140 windows could be obtained for that participant, the analysis used all windows available for that participant, detrending occurred in each window using robust linear regression, outliers were Winsorized capping extreme values to the 5th and 95th percentile, and participants with a transition in the two months after the EMA data collection had ended were included in the analysis.
2. Used variable NA high
3. Used variable PA low
4. Used variable PA high
5. Used variable total mood
6. Used variable “I feel down”
7. Used variable “I feel listless”
8. Used variable “I feel tired”
9. Used window size 3 weeks
10. Used window size 2 weeks
11. Participants were included if at least 70 windows could be obtained for that participant
12. The analysis used exactly 140 windows for each participant
13. Detrending occurred globally, using Gaussian local linear regression with a bandwidth of 1 week (this means 2 weeks will determine ~68.2% of the weight)
14. Analysis was performed without detrending
15. Extreme values lower than the 5th percentile or higher than the 95th percentile were removed
16. Extreme values were left in the analysis
17. Participants with a transition more than two weeks after the EMA data collection had ended were excluded from the analysis

Table S8.2
Overview of results using all preregistered model settings.

Settings	Autocorrelation					Variance						
	TP	FP	T τ	NT τ	p-value	Cohen's d	TP	FP	T τ	NT τ	p-value	Cohen's d
1	36.8	22.2	-0.02	-0.17	0.090	0.45	26.3	11.1	-0.19	-0.47	0.038	0.60
2	36.8	11.1	-0.06	-0.17	0.201	0.28	31.6	11.1	-0.20	-0.48	0.027	0.66
3	42.1	16.7	-0.03	-0.14	0.190	0.29	36.8	5.6	-0.22	-0.46	0.054	0.54
4	42.1	5.6	0.06	-0.26	0.006	0.88	26.3	5.6	-0.21	-0.48	0.045	0.57
5	36.8	33.3	-0.04	-0.17	0.168	0.32	26.3	5.6	-0.26	-0.53	0.023	0.68
6	42.1	27.8	-0.06	-0.14	0.254	0.22	47.4	11.1	-0.03	-0.40	0.009	0.82
7	42.1	27.8	-0.04	-0.11	0.258	0.22	31.6	16.7	-0.13	-0.34	0.099	0.43
8	36.8	38.9	-0.03	-0.08	0.360	0.12	31.6	16.7	-0.19	-0.33	0.205	0.27
9	33.3	15.8	-0.06	-0.21	0.079	0.45	28.6	5.3	-0.15	-0.48	0.005	0.85
10	31.8	5.3	-0.05	-0.19	0.034	0.58	31.8	5.3	-0.13	-0.45	0.002	0.95
11	36.4	21.1	-0.05	-0.19	0.107	0.39	31.8	10.5	-0.13	-0.48	0.009	0.76
12	26.3	44.4	-0.13	-0.02	0.732	0.21	42.1	16.7	-0.12	-0.40	0.058	0.53
13	31.6	16.7	0.00	-0.18	0.071	0.49	26.3	11.1	-0.17	-0.50	0.020	0.70
14	31.6	22.2	-0.13	-0.15	0.418	0.07	21.1	16.7	-0.28	-0.44	0.124	0.39
15	26.3	11.1	-0.04	-0.14	0.142	0.36	26.3	11.1	-0.19	-0.48	0.037	0.60
16	31.6	27.8	-0.03	-0.13	0.202	0.28	26.3	11.1	-0.19	-0.42	0.067	0.50
17	36.8	22.2	-0.02	-0.17	0.090	0.45	26.3	11.1	-0.19	-0.47	0.038	0.60

Note. Settings are explained in the text. 'TP' and 'FP' refer to the percentage of true and false positives. 'T τ ' and 'NT τ ' refer to the average M-K trend in autocorrelation and variance for participants with a transition and participants with no transition, respectively. The *p*-value and Cohen's *d* summarize a Welch's *t*-test testing if the average M-K trend was larger (i.e., one-sided) for participants with a transition.

9

Summary & General discussion

In this thesis, I aimed to investigate the way depression evolves over time, with a focus on the dynamics that occur when individuals experience symptom improvements. The studies in this thesis used intensive longitudinal data gathered with ecological momentary assessments (EMA) to test hypotheses based on dynamical systems theory, with the objective to reveal the temporal patterns in symptoms and emotions that can be informative about a patient's likelihood to recover. In particular, I have studied idiographic change over the course of therapy and have tested the role of affect dynamics, indicators of instability like early warning signals (EWS), and overall change patterns as predictors of symptom improvement and treatment response. Taken together, the studies in this thesis provide important insights that allow us to reflect on the theoretical and clinical promise of complex dynamical systems approaches to the study of change in depression. In this final chapter, I will summarize the main findings, discuss what general conclusions we can draw from this body of work, and provide a discussion of the practical and conceptual challenges researchers may face when pursuing a similar line of study.

Summary of the main findings

In **Chapter 2**, we investigated the occurrence and clinical relevance of discontinuous changes in daily measurements of self-rated problem severity in a group of 329 patients with mood disorders receiving treatment. We modelled the shape of overall individual change as linear, loglinear or one-step change trajectories, identified sudden gains, and examined how these patterns related to treatment response. Most patients showed complex, nonlinear change patterns (not fitted by our simplified trajectories), which were also significantly more common in treatment responders than in those who did not improve. For individuals whose repeated scores were best fitted with one of our pre-defined trajectories, the one-step pattern dominated, and often showed a sudden gain within that pattern. Sudden gains were related to treatment response status in our sample. We concluded that clinically relevant improvements in depression often follow a nonlinear pattern, within which abrupt shifts can occur, and that this finding allows for the possibility that the remission process functions according to dynamical systems principles.

Next, we investigated the predictive value of destabilization in momentary affect prior to symptom improvement. In **Chapter 3**, we studied whether signs of destabilization in momentary affect dynamics play a role in predicting the rate of response in depressive symptoms, and remission status over eighteen months in a sample of 46 depressed patients. Particularly, we used instability and variability of negative affect measured the week before psychological treatment as predictors of the rate of treatment response and of whether people remitted during treatment, versus after treatment or did not remit. Higher mean levels of negative affect predicted a stronger reduction in symptoms over five months. However, instability or variability of negative affect were not predictive indicators, and were also unrelated to remitting during treatment. We concluded that assessing instability and variability in negative affect prior to treatment may not give a good indication of patients' response or remission potential, whereas the mean of negative affect may have some clinical utility as an indicator of the likelihood of response to treatment and could be explored further. A more theoretically appropriate study of increasing destabilization – and EWS specifically – before relevant symptom improvements requires analysing longer, within-person time series over the course of therapy to capture affect dynamics as they change.

This consideration and others were reviewed in **Chapter 4**: we outlined three main conceptual and practical challenges that researchers face when intending to conduct a truly empirical, within-person test of whether EWS precede symptom transitions in psychopathology. First, we recommended defining *a priori* what relevant symptom transitions —potential state shifts— look like in the study sample of interest, and how these can be reliably identified. Second, we discussed measurement timing: in a study of EWS, the time scale at which transitions take place must be accounted for in the length of the observation period, the frequency of assessment, and in choosing a period in which symptom change is likely to occur—all the while balancing these choices with participant burden. Finally, the practical and theoretical choice of constructs for the detection of EWS involves selecting variables that will change as symptoms change, can be measured at high-resolution, and will provide sufficient variability for the data-demanding analyses. We concluded that balancing the theoretical basis of one's methodological choices with practical considerations gleaned from pilot studies can help build empirically effective research designs that allow a true test of EWS-phenomena in psychopathology.

In line with these recommendations, we designed a study to capture depression as a changing process, of which the procedures are described in **Chapter 5**. Particularly, we aimed to intensively monitor the symptoms and momentary experiences (affect, cognitions, and context) of patients with depressive symptoms in a period during which they received psychological treatment, and were therefore more likely to experience improvement of their complaints. In this study, 41 participants completed a year of intensive longitudinal data collection. The first four months consisted of ambulatory assessments: continuous actigraphy, twice daily heart rate recordings, and five times daily ecological momentary assessments. Weekly symptom assessments were started at the same time and continued for six months, followed by six monthly assessments in the final half year of the study period. This design allowed us to gather intensive longitudinal data to examine whether EWS could be found consistently before symptom transitions at the within-person level.

In **Chapter 6**, I developed a methodology that can help to identify symptom transitions of different durations reliably. We need to be able to detect such relevant shifts in symptoms in multiple individuals without knowing *a priori* what time frames would be relevant in the context of depressive symptom improvement. To do so, I developed a threshold method, named the Duration-adjusted Reliable Change Index (DaRCI), with the aim to enable the detection of reliable change over multiple time points. This index is adapted from a well-established method (the RCI), which uses the standard error of measurement of a given instrument to determine what the range of uncertainty is that must be surpassed for a change between two points to be reliable. In this chapter, I demonstrate the ability of the DaRCI to detect reliable changes of varying duration (e.g., from one observation to the next, or from one observation to the observation two or three time points later) in simulated symptom time series data. The DaRCI was most effective at picking up highly discontinuous changes (mean shifts over two to four points) but also detected gradual changes when the overall change was steep.

With these rich time series of 41 individuals with depressive complaints collected and a method to identify transitions, we were able to study whether EWS, in the form of rising lag-1 autocorrelation and variance, could be detected in momentary mood prior to large symptom improvements in individual patients. As described in **Chapter 7**, we found rising autocorrelations in a larger proportion

of individuals with a transition, compared to those with no transition in symptoms. Rising variance occurred in approximately similar rates in participants with and without transitions, and hence did not function well as an EWS. These results remained robust across the various sensitivity checks we conducted. Our findings suggest that EWS are not generic indicators of impending symptom shifts — they were found in less than half of the participants with a transition — but that the process of symptom remission did show signs of critical slowing down in part of the depressed individuals, in accordance with principles of dynamical systems.

In **Chapter 8**, we aimed to answer the same question regarding EWS in a sample of 56 previously depressed individuals who were tapering their antidepressant medication and were consequently at risk of experiencing depressive symptom recurrence. Thus, we investigated whether lag-1 autocorrelation and variance rose before transitions toward increased depressive symptoms. We found that the number of significant EWS was higher in individuals with transitions for both the autocorrelation and variance, compared to individuals without transitions. Moreover, the average trends in variance were also significantly higher for individuals in the transition group than for individuals in the no-transition group. Again, these findings were robust across the sensitivity checks. Additionally, in this study, combining EWS from multiple variables and for autocorrelation and variance together improved the number of true positives (correct identification of individuals with transitions) substantially, while not raising the number of false positives. Thus, while EWS did not occur for all individuals with a transition, these results provided important empirical confirmation that some symptom changes appear to behave in line with what would be expected in a complex dynamical system.

Reflections on the main findings

What's in a mood? On dynamic changes in depression

In this thesis, I focused on scrutinizing moment-to-moment experiences for clues about people's response potential. The reasoning behind this being that if a diagnosis of depression is determined by — among other things — sad mood that is present for at least two weeks (a core symptom of depression), it seems intuitive that how one's mood or affect changes and varies over time may contain relevant information about the overarching syndrome, and the course of the disorder as a whole [1–5].

The chapters in this thesis have contributed to a conceptual debate in the existing literature: are the dynamics in mood measurements trait-like, or process-like? Or, as Boker [6] summarizes it: are we interested in the dynamics of change, or in the change of dynamics? The “dynamics of change” concern the trait-like variability that individuals show around their own stable affective mean (or attractor), which essentially reflects how strongly their emotions are regulated from moment to moment [7–11]. Someone with a strong attractor would be said to have high emotion regulation, which shows in momentary affect as low variability (less extreme responses to stimuli), and low inertia (less resistance to change) [12–15]. This rests on the assumption that the period over which the dynamics are calculated does not involve meaningful change and that the process is stationary (which may hold true, if the measurement period is brief) [6,9,16]. The “change of dynamics”, in contrast, refers to how people change over a longer time period and is not based on the assumption of stationarity. Instead, this point of view focuses on the ways in which people meaningfully vary and change as time passes – a more “process”-like approach. This naturally fits with the dynamical systems approach to psychology, which also views individuals as defined by their complex, layered, and time-varying nature [2,17–25].

The findings of the studies in this thesis suggest that mood is seldom stable for long, and that the dynamics in momentary affect are process-like, at least in the context of psychological treatment and discontinuation of antidepressants. The intensive longitudinal datasets that we used highlighted the high degree of changeability and nonlinearity that individuals show in their everyday affective

experiences, particularly when examined over longer periods. For instance, our analyses in **Chapter 2** revealed that the overall improvement patterns formed over 100 days of patients' day-to-day ratings of problem intensity were varied and complex, and not readily captured by the simplified trajectories we modelled. Furthermore, **Chapter 7** and **8** illustrated that the dynamics of affect ratings varied over the four months of data collection for many individuals. Likewise, several other studies have shown that the dynamics of affect and related constructs are multistable and non-stationary when viewed over longer periods: this includes daily mood and anxiety measured over 50 days [26], mood in bipolar patients monitored over one to two and a half years [27], and self-esteem assessed over 512 days [28].

A logical way to continue the study of within-person affect dynamics in depression is to focus on their process-like nature and conduct studies in periods where clinical change can be expected (e.g., in the context of therapy). Although we can calculate affect dynamics like the variability, lag-1 autocorrelation, and mean square of successive differences (MSSD) over a short EMA period in which a person may appear relatively stable [14,29,30], it seems likely that these dynamics may be different a few weeks later. Rather than assuming stationarity or detrending the time series, one can deliberately model the non-stationarity in the data to gain a better understanding of the change patterns patients show as they move from a depressive episode to remission, or vice versa [6,31–33]. For instance, using analysis techniques that allow one to model a time-varying intercept or autocorrelation, change points, or other nonlinear shapes, or even simply including time as a predictor in the model can be a way to further explore intensive longitudinal data for predictors of response potential [34–39]. These kinds of analyses can be quite demanding in terms of the data needed to reach a reasonable estimation, which means data collection will be a challenge [40]. Nevertheless, it appears that there is a multitude of relevant information to be found by studying the dynamics of depression over extended periods, and continuing this line of inquiry is highly relevant.

One might argue that capturing affect dynamics over brief periods when no (or little) clinical change would be expected (e.g., in healthy individuals, or waitlist control groups) could still provide a dynamics-based marker of long-term potential to change. That is, current levels of destabilization in affect dynamics

could perhaps signal the likelihood that a person's mental state will be meaningfully different by a later point. A snapshot of one's mood dynamics, measured over a short enough time period (e.g., a few days), can exhibit approximate local stationarity such that this assumption may just be met for the statistical analyses [6,31]. In line with this, researchers have used affect measurements gathered over a few days to calculate dynamics measures like inertia, instability and variability to prospectively predict increases in general psychopathology symptoms a year later [41], depressive symptoms over six months [42], as well as the onset of depressive disorders [43], and bipolar disorder [44]. However, whether this approach applies to predicting the process of remission in depression is not yet clear. In **Chapter 3** of this thesis, we examined whether within-person levels of variance and instability in negative affect measured over six days directly before antidepressant treatment could be informative of response rate over five months and long-term remission status. In our study, only mean-levels predicted faster response curves during treatment, while the affect dynamics were not informative of later change. Similar studies have shown mixed but mostly comparable results (cf. [44–47]). Thus, while meaningful individual differences may still be gleaned from affect dynamics when they are measured at very short time scales, their predictive performance is not consistent for predicting depressive remission. Considering the empirical evidence in this thesis, and the well-known fact that people are always changing, I would argue that a better choice for further study is to explore what changing dynamics look like over longer time scales, in periods of meaningful overall change. Such studies are still scarce at the time of writing, but they are what is required to gain a better understanding of the intrinsic temporal fluctuations in people's affect and symptoms and to improve our knowledge of depressive remission and psychopathology.

Is it too early for early warning signals?

One of the driving research questions behind the works in this thesis (and the Transitions in Depression (TRANS-ID) project as a whole) was the notion that, if EWS were found consistently, they could provide a new method for detecting sensitive periods and destabilization for individual patients. Were EWS to be found with a high enough sensitivity (most true changes are flagged), and a high enough

specificity (changes that are flagged, are most often true), looking for EWS in real-time could then become clinically appealing: asking patients to monitor their mood during treatment or in periods of increased risk of symptom recurrence, and testing their data for EWS, could be a way to improve the timing of treatment interventions to encourage positive symptom change or prevent deterioration [48,49].

Taken together, the results of **Chapter 7** and **Chapter 8** showed that EWS preceded transitions towards both higher and lower levels of depressive symptoms in part of our samples, which was in line with our hypotheses. However, these studies also showed that the generalizability of EWS was limited, as they only occurred for some individuals, and not for all who experienced a relevant symptom change, nor were they found consistently across different affect measures. In other words, the sensitivity and specificity of our findings were too modest to conclude that EWS can deliver on the promise to serve as a clinical tool for accurate personalized predictions of impending symptom change (cf. [50,51]). Thus, it is too early for EWS to be used in clinical practice at this stage.

In order for EWS to reach an acceptable predictive performance to be useful in a clinical setting, we need to learn which analysis parameters can consistently improve the sensitivity and specificity of detected EWS. Our studies suggest that combining multiple EWS and multiple affect measures may result in a heightened rate of true positives, as may using large window sizes (see also Refs. [52,53]). Since studies with similar approaches are so rare and require researchers to make a lot of analysis choices, the cycle of exploration and replication is key to making progress in the study of EWS in psychopathology. Before application in clinical practice can even be considered, an informed set of optimal model choices must be obtained. We cannot be certain how our parameter settings translate to new datasets, but our studies, with their extensive sensitivity checks, may at least provide starting points for further exploration.

Another reason that it is too early for clinical applications of EWS is that the burden of the monitoring design is not justified by the predictive accuracy of EWS (as found in our studies). In a real-world setting, patients would need to rate various aspects of their momentary mood multiple times a day for several months to obtain the required data resolution for the analyses – and preferably they would start monitoring in a period that is relatively stable, or unlikely to show large symptom shifts (as described in **Chapter 4**). These are challenging

demands for patients who are potentially in a vulnerable state (e.g., currently struggling with mental disorder), or about to enter a period of uncertainty (e.g., starting the tapering of psychoactive medication). Even though the TRANS-ID studies in **Chapter 7** and **8** showed that intensive longitudinal data collection was feasible (with high compliance rates of ~87%) and even experienced as positive by participants in these patient groups, there was likely an element of self-selection, in that only highly motivated individuals entered or remained in the study (cf. [54,55]). Many also dropped out or declined to participate: the repeated questions about how they were feeling were too confronting for some, others could not fit the five EMA-measurements a day into their schedules or found it burdensome in an already mentally taxing period. At this stage, the time and effort required from patient and clinician for such intensive longitudinal data collection do not outweigh the potential benefit that EWS promised, especially not until the predictive performance of EWS is improved.

Future directions: Define, refine, measure

A first potential direction for further study is to investigate what predictive information and meaningful changes are contained in the larger overall pattern of affect and symptoms [18,31]. Given that EWS may not be the quickest step forward clinically, it is worthwhile looking beyond EWS and affect dynamics for ways to predict individual improvement from depression. Perhaps a more promising way to predict how individuals change over time lies in identifying whether and which symptoms change early in the course of therapy [56], and how this relates to treatment outcomes [57–59]. Measuring symptoms and affect regularly (weekly, or daily) and detecting changing trends therein may be a less burdensome way to predict response potential than the design needed to study EWS [60,61]. Alternatively, the specific shape and duration of changes that relate most strongly to depressive remission could be examined in more detail. Certain overall symptom change patterns and sudden gains and losses have already been linked to better treatment outcomes and increased response potential [58,59,62–65]. However, the timeframe for relevant symptom changes in the course of treatment is not consistently defined: “rapid” symptom change can refer to improvements that take anywhere between one to four weeks [66–68]. In fact, **Chapter 6** describes

a new method to define reliable symptom changes of various durations, since such a definition was simply lacking. This implies that there is room to study the various gradual, sudden and overall changes people experience in the context of treatment. To pinpoint exactly what kind of changes are most commonly linked to depressive improvement, a more thorough investigation on the frequency and form of these patterns is needed.

A second direction for future research is the refinement of models for use on within-person intensive longitudinal data, and using the dynamics in these long time series to predict changes in depression (see Refs. [22,31,32] for more elaborate methodological and conceptual discussions). Various techniques are being developed to explore the dynamics of psychological time series, including multilevel models to tease apart intra-individual trait change and inter-individual state variability [3,69–72], time-varying autoregressive models that take into account trends in the overall pattern by deliberately including them in the model [34,35,73], and complexity based methods like recurrence quantification analysis and spectral analysis that are designed to study nonlinearity and individual systems [74–80]. Adopting appropriate statistical methods to analyse time series data is necessary [81], and it is important to learn in more detail what these models can and cannot do for us [82]. Although the psychological time series we collected for **Chapter 7** and **8** are of exceptional length (~550 within-person measurements), they are still considered short in comparison to the length of time series in other fields, and we must be cautious to avoid underpowered models [40,53,83]. Furthermore, psychological data is notoriously noisy, which may complicate the application and interpretation of some more complex models [83,84]. Simulation studies can be a helpful way to manage expectations of what can realistically be detected with certain techniques. In short, an important aim is to continue refining our understanding of newly adopted statistical models so that we can make optimal use of intensive longitudinal within-person time series [72,85].

A third course for future study is the conceptual question of whether we can really expect to accurately capture EWS and relevant affect dynamics in psychological data. An important requirement for being able to properly measure the fluctuations in a variable over time is that the time between observations is shorter than the time it takes for the system to return to its equilibrium position [17]. We may be limited in our ability to measure the required

psychological constructs (affect, symptoms) at sufficiently high resolution if the rate at which they change, recover and change again, is faster than the EMA-measurements we are performing [2]. Any analyses we then do would be biased by the insufficient resolution of the measurements [83]. A potential solution to this problem may be to explore the relevant time scales of affect and symptoms with continuous time models, which use differential equations to estimate the rate of change of a variable [86–89]. Another option for predicting changes in the system (i.e., in an individual's state of wellbeing) may be to look for clues in the changing dynamics of non-psychological time series data, such as actigraphy and heart rate measurements. These kinds of data can be measured continuously over long periods, with low burden to the participant, resulting in much denser, higher-resolution data than EMA self-ratings [77,90–92].

Finally, a deeper examination of the temporal relationship between symptoms and affect may be worthwhile. Wichers [1,23] postulated the notion that emotions may constitute the smallest building blocks of depression: one's moment-to-moment emotions can form an enduring negative mood, which in turn can be identified as a depressive symptom, and together with other psychological complaints can form a cluster of symptoms that we recognize as depression. This makes intuitive sense and this conceptualization has been readily adopted by other researchers in the field to demonstrate associations of current psychopathology with positive and negative affect levels and dynamics in the course of daily life [44,93–97]. Yet those studies link symptoms and affect with a more trait-like approach, in which the variables are assumed to reflect a stable cross-sectional relationship. A formal study of how symptoms and momentary affect change together over time has not been conducted. In line with studying depression as a process, it would be relevant to see how closely a week's worth of EMA-measured momentary mood correlates with a weekly symptom assessment.

To summarize, underpinning any study of affect dynamics and their predictive value is the notion that we are able to measure momentary feelings well enough with EMA-designs. However, the fundamental properties of affect dynamics themselves deserve a more in-depth investigation, so that future empirical work can be more thoroughly grounded in a theoretical understanding that grows as the length of datasets grows.

References

1. Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, 7(4), 362–367. <https://doi.org/10.1177/1754073915590623>
2. Jeronimus, B. F. (2019). Dynamic system perspectives on anxiety and depression. In E. S. Kunnen, N. M. P. de Ruiter, B. F. Jeronimus, & M. A. van der Gaag (Eds.), *Psychosocial Development in Adolescence: Insights from the Dynamic Systems Approach* (1st ed.). Routledge Psychology.
3. Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, 83(4), 825–836. <https://doi.org/10.1037/ccp0000026>
4. Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, 16, 49–74. <https://doi.org/10.1146/annurev-clinpsy-102419>
5. Fried, E. I., & Robinaugh, D. J. (2020). Systems all the way down: Embracing complexity in mental health research. *BMC Medicine*, 18(1), 205. <https://doi.org/10.1186/s12916-020-01668-w>
6. Boker, S. M., Staples, A. D., & Hu, Y. (2016). Dynamics of change and change in dynamics. *Journal for Person-Oriented Research*, 2(1–2), 34–55. <https://doi.org/10.17505/jpor.2016.05>
7. Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry*, 11(3), 129–141. https://doi.org/10.1207/S15327965PLI1103_01
8. Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, 5(2), 208–225. <https://doi.org/10.1037/1528-3542.5.2.208>
9. Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99(6), 1042–1060. <https://doi.org/10.1037/a0020962>
10. Brans, K., Koval, P., Ku, P. V., Yan, L., Lim, L., & Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion*. <https://doi.org/10.1037/a0032400>
11. Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, 7(4), 355–361. <https://doi.org/10.1177/1754073915590617>
12. Kuppens, P., Allen, N. B., & Sheeber, L. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, 21(7), 984–991. <https://doi.org/10.1177/0956797610372634>
13. Hollenstein, T. (2015). This time, it's real: Affective flexibility, time scales, feedback loops, and the regulation of emotion. *Emotion Review*, 7(4), 308–315. <https://doi.org/10.1177/1754073915590621>
14. Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, 13(6), 1132–1141. <https://doi.org/10.1037/a0033579>
15. Bos, E. H., de Jonge, P., & Cox, R. F. A. (2018). Affective variability in depression: Revisiting the inertia-instability paradox. *British Journal of Psychology*. <https://doi.org/10.1111/bjop.12372>
16. Hamaker, E. L. (2012). Why researchers should think 'within-person': A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 43–61). Guilford.
17. Thelen, E., & Smith, L. B. (1994). A dynamic systems approach to development and action. MIT Press.
18. Van Der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review*, 99(3), 395–417. <https://doi.org/10.1037/0033-295X.99.3.395>

19. Hayes, A. M., & Strauss, J. L. (1998). Dynamic systems theory as a paradigm for the study of change in psychotherapy: An application to cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 66(6), 939–947. <https://doi.org/10.1037/0022-006X.66.6.939>
20. Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of Counseling Psychology*, 63(4), 379–395. <https://doi.org/10.1037/cou0000150>
21. Granic, I. (2005). Timing is everything: Developmental psychopathology from a dynamic systems perspective. *Developmental Review*, 25(3–4), 386–407. <https://doi.org/10.1016/j.dr.2005.10.005>
22. Lichtwarck-Aschoff, A., van Geert, P., Bosma, H., & Kunnen, S. (2008). Time and identity: A framework for research and theory formation. *Developmental Review*, 28(3), 370–400. <https://doi.org/10.1016/j.dr.2008.04.001>
23. Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, 44(7), 1349–1360. <https://doi.org/10.1017/S0033291713001979>
24. Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, 1–7. <https://doi.org/10.1017/S0033291719002514>
25. Vallacher, R. R., Van Geert, P., & Nowak, A. (2015). The intrinsic dynamics of psychological process. *Current Directions in Psychological Science*, 24(1), 58–64. <https://doi.org/10.1177/0963721414551571>
26. Bornas, X., Balle, M., Morillas-Romero, A., & Aguayo-Siquier, B. (2015). Allometric control of daily mood and anxiety fluctuations. *Motivation and Emotion*, 39(4), 571–579. <https://doi.org/10.1007/s11031-015-9471-4>
27. Gottschalk, A., Bauer, M. S., & Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry*, 52(11), 947–959. <https://doi.org/10.1001/archpsyc.1995.03950230061009>
28. Delignières, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences*, 8(4), 479–510.
29. Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, 13(4), 354–375. <https://doi.org/10.1037/a0014173>
30. Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, 121(4), 819–829. <https://doi.org/10.1037/a0027978>
31. Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322. <https://doi.org/10.1177/1754073915590619>
32. Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, 26(1), 10–15. <https://doi.org/10.1177/0963721416666518>
33. Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder a review. *JAMA Psychiatry*, 74(5), 528–534. <https://doi.org/10.1001/jamapsychiatry.2017.0001>
34. Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425. <https://doi.org/10.1037/met0000085>

35. Albers, C., & Bringmann, L. F. (2020). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *European Journal of Psychological Assessment*, 36(February), 492–499. <https://doi.org/10.1027/1015-5759/a000589>
36. Cabrieto, J., Adolf, J., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2018). Detecting long-lived autodependency changes in a multivariate system via change point detection and regime switching models. *Scientific Reports*, 8(15637), 1–15. <https://doi.org/10.1038/s41598-018-33819-8>
37. Du, H., & Wang, L. (2018). Reliabilities of intraindividual variability indicators with autocorrelated longitudinal data: Implications for longitudinal study designs. *Multivariate Behavioral Research*, 53(4), 502–520. <https://doi.org/10.1080/00273171.2018.1457939>
38. Wang, L., & McArdle, J. J. (2008). A simulation study comparison of bayesian estimation with conventional methods for estimating unknown change points. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(1), 52–74. <https://doi.org/10.1080/10705510701758265>
39. Wang, L. (Peggy), & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, 20(1), 63–83. <https://doi.org/10.1037/met0000030>
40. Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 495–515. <https://doi.org/10.1080/10705511.2017.1392862>
41. Kuranova, A., Booi, S. H., Menne-Lothmann, C., Decoster, J., Van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., Van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. *BMC Medicine*, 18(1), 1–11. <https://doi.org/10.1186/s12916-020-1500-9>
42. Panaite, V., Rottenberg, J., & Bylsma, L. M. (2020). Daily affective dynamics predict depression symptom trajectories among adults with major and minor depression. *Affective Science*, 1(3), 186–198. <https://doi.org/10.1007/s42761-020-00014-w>
43. Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, 12(2), 283–289. <https://doi.org/10.2337/db10-1371>
44. Sperry, S. H., Walsh, M. A., & Kwapil, T. R. (2020). Emotion dynamics concurrently and prospectively predict mood psychopathology. *Journal of Affective Disorders*, 261, 67–75. <https://doi.org/10.1016/j.jad.2019.09.076>
45. Bosley, H. G., Soyster, P. D., & Fisher, A. J. (2019). Affect dynamics as predictors of symptom severity and treatment response in mood and anxiety disorders: Evidence for specificity. *Journal for Person-Oriented Research*, 5(2), 101–113. <https://doi.org/10.17505/jpor.2019.09>
46. Husen, K., Rafaeli, E., Rubel, J. A., Bar-Kalifa, E., & Lutz, W. (2016). Daily affect dynamics predict early response in CBT: Feasibility and predictive validity of EMA for outpatient psychotherapy. *Journal of Affective Disorders*, 206, 305–314. <https://doi.org/10.1016/j.jad.2016.08.025>
47. Höhn, P., Menne-Lothmann, C., Peeters, F., Nicolson, N. A., Jacobs, N., Derom, C., Thiery, E., van Os, J., & Wichers, M. (2013). Moment-to-moment transfer of positive emotions in daily life predicts future course of depression in both general population and patient samples. *PLoS One*, 8(9), e75655. <https://doi.org/10.1371/journal.pone.0075655>
48. Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. <https://doi.org/10.1159/000441458>

49. Wichers, M., Smit, A. C., & Snippe, E. (2020). 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*, 6(1), 1–15. <https://doi.org/10.17505/jpor.2020.22042>
50. Bos, F. M., Schreuder, M. J., George, S. V., Doornbos, B., Bruggeman, R., van der Krieke, L., Haarman, B. C. M., Wichers, M., & Snippe, E. (2021). Anticipating manic and depressive shifts in patients with bipolar disorder using early warning signals. *PsyArXiv*. <https://doi.org/10.31234/osf.io/v5b7t>
51. Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, 8(1), 25–35. <https://doi.org/10.1177/2167702619865969>
52. Lenton, T. M., Livina, V. N., Dakos, V., Van Nes, E. H., & Scheffer, M. (2012). Early warning of climate tipping points from critical slowing down: Comparing methods to improve robustness. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1662), 1185–1204. <https://doi.org/10.1098/rsta.2011.0304>
53. Dablander, F., Pichler, A., Cika, A., & Bacilieri, A. (2020). Anticipating critical transitions in psychological systems using early warning signals: Theoretical and practical considerations theory of critical slowing down. *PsyArXiv*, 1–30. <https://doi.org/10.31234/osf.io/5wc28>
54. Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M., & van der Krieke, L. (2019). Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. *Psychiatric Services*, 70(11), 983–991. <https://doi.org/10.1176/appi.ps.201900050>
55. Bos, F. M., Snippe, E., Bruggeman, R., Doornbos, B., Wichers, M., & van der Krieke, L. (2020). 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*, 8(1), 38. <https://doi.org/10.1186/s40345-020-00201-5>
56. Snippe, E., Doornbos, B., Schoevers, R. A., Wardenaar, K. J., & Wichers, M. (2021). Individual and common patterns in the order of symptom improvement during outpatient treatment for major depression. *Journal of Affective Disorders*, 290, 81–88. <https://doi.org/10.1016/j.jad.2021.04.097>
57. Haas, E., Hill, R. D., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *Journal of Clinical Psychology*, 58(9), 1157–1172. <https://doi.org/10.1002/jclp.10044>
58. Rubel, J., Lutz, W., Kopta, S. M., Köck, K., Minami, T., Zimmermann, D., & Saunders, S. M. (2015). Defining early positive response to psychotherapy: An empirical comparison between clinically significant change criteria and growth mixture modeling. *Psychological Assessment*, 27(2), 478–488. <https://doi.org/10.1037/pas0000060>
59. Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, 118(1–3), 60–68. <https://doi.org/10.1016/j.jad.2009.01.019>
60. Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Recovery: Study protocol for a repeated intensive longitudinal n = 1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *PsyArXiv, February*. <https://doi.org/10.31234/osf.io/fertq>
61. Helmich, M. A., Olthof, M., Oldehinkel, A. J., Wichers, M., Bringmann, L. F., & Smit, A. C. (2021). Early warning signals and critical transitions in psychopathology: Challenges and recommendations. *Current Opinion in Psychology*, 41, 51–58. <https://doi.org/10.1016/j.copsyc.2021.02.008>
62. Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2016). Defined symptom-change trajectories during acute-phase cognitive therapy for depression predict better longitudinal outcomes. *Behaviour Research and Therapy*, 87, 48–57. <https://doi.org/10.1016/j.brat.2016.08.008>

63. Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review*, 76, 101827. <https://doi.org/10.1016/j.cpr.2020.101827>
64. Thomas, C., & Persons, J. B. (2013). Sudden gains can occur in psychotherapy even when the pattern of change is gradual. *Clinical Psychology: Science and Practice*, 20(2), 127–142. <https://doi.org/10.1111/cpsp.12029>
65. Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy*, 43(6), 703–714. <https://doi.org/10.1016/j.brat.2004.06.002>
66. Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(6), 894–904. <https://doi.org/10.1037/0022-006X.67.6.894>
67. Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, 1(2), 138–155. <https://doi.org/10.1111/j.1468-2850.1994.tb00016.x>
68. Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology*, 70(9), 886–903. <https://doi.org/10.1002/jclp.22077>
69. Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait-state model. *Journal of Research in Personality*, 41(2), 295–315. <https://doi.org/10.1016/j.jrp.2006.04.003>
70. Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, 40(2), 207–233. https://doi.org/10.1207/s15327906mbr4002_3
71. McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60(1), 577–605. <https://doi.org/10.1146/annurev.psych.60.110707.163612>
72. Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, 53(6), 820–841. <https://doi.org/10.1080/00273171.2018.1446819>
73. Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, 56(1), 120–149. <https://doi.org/10.1080/00273171.2020.1743630>
74. Heino, M. T. J., Knittle, K., Noone, C., Hasselman, F., & Hankonen, N. (2021). Studying behaviour change mechanisms under complexity. *Behavioral Sciences (Basel, Switzerland)*, 11(5), 77. <https://doi.org/10.3390/bs11050077>
75. Hasselman, F., & Bosman, A. M. T. (2020). Studying complex adaptive systems with internal states: A recurrence network approach to the analysis of multivariate time-series data representing self-reports of human experience. *Frontiers in Applied Mathematics and Statistics*, 6(April), 1–14. <https://doi.org/10.3389/fams.2020.00009>
76. Olthof, M., Hasselman, F., & Lichtwarck-Aschoff, A. (2020). Complexity in psychological self-ratings: Implications for research and practice. *BMC Medicine*, 18(317). <https://doi.org/10.1186/s12916-020-01727-2>
77. Kunkels, Y. K., Riese, H., Knapen, S. E., Lek, R. F. R. der, George, S. V., Roon, A. M. van, Schoevers, R. A., & Wichers, M. (2021). Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study. *Translational Psychiatry*, 11(1), 1–10. <https://doi.org/10.1038/s41398-021-01465-w>

78. Fisher, A. J., Newman, M. G., & Molenaar, P. C. M. (2011). A quantitative method for the analysis of nomothetic relationships between idiographic structures: Dynamic patterns create attractor states for sustained posttreatment change. *Journal of Consulting and Clinical Psychology*, 79(4), 552–563. <https://doi.org/10.1037/a0024069>
79. Woyshville, M. J., Lackamp, J. M., Eisengart, J. A., & Gilliland, J. A. M. (1999). On the meaning and measurement of affective instability: Clues from chaos theory. *Biological Psychiatry*, 45(3), 261–269. [https://doi.org/10.1016/S0006-3223\(98\)00152-8](https://doi.org/10.1016/S0006-3223(98)00152-8)
80. Strunk, C., & Lichtwarck-Aschoff, A. (2019). Therapeutic chaos. *Journal for Person-Oriented Research*, 5(2), 81–100. <https://doi.org/10.17505/jpor.2019.08>
81. Walls, T. A., & Schafer, J. L. (2012). Models for intensive longitudinal data. In *Models for Intensive Longitudinal Data*. <https://doi.org/10.1093/acprof:oso/9780195173444.001.0001>
82. Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
83. Haslbeck, J. M. B., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*. <https://doi.org/10.1080/00273171.2021.1896353>
84. Speelman, C., Olivetti, E., Kessler, B., Stamey, J., Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in $n = 1$ psychological autoregressive modeling. *Frontiers in Psychology*, 1, 1038. <https://doi.org/10.3389/fpsyg.2015.01038>
85. Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, 26(1), 10–15. <https://doi.org/10.1177/0963721416666518>
86. Voelkle, M. C., & Oud, J. H. L. (2015). Relating latent change score and continuous time models. *Structural Equation Modeling*, 22(3), 366–381. <https://doi.org/10.1080/10705511.2014.935918>
87. Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein–Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, 74(3), 395–418. <https://doi.org/10.1007/s11336-008-9106-8>
88. Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous time approach to intensive longitudinal data: What, Why and How? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (Issue January). Springer. <https://doi.org/10.1007/978-3-319-77219-6>
89. Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, 53(6), 782–805. <https://doi.org/10.1080/00273171.2018.1496813>
90. Kunkels, Y. K., Smit, A. C., George, S. V., Snippe, E., van Roon, A. M., Wichers, M., & Riese, H. (2020). Risk Ahead: Behavioral early-warning signals of increases in depressive symptoms during anti-depressant tapering. *Open Science Framework*, preregistration.
91. Bayani, A., Hadaeghi, F., Jafari, S., & Murray, G. (2017). 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*, 34(2), 235–245. <https://doi.org/10.1080/07420528.2016.1272608>
92. Valenza, G., Nardelli, M., Lanata, A., Gentili, C., Bertschy, G., Kosel, M., & Scilingo, E. P. (2016). Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics. *IEEE Journal of Biomedical and Health Informatics*, 20(4), 1034–1043. <https://doi.org/10.1109/JBHI.2016.2554546>

93. Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, 141(4), 901-930. <https://doi.org/10.1037/a0038822>
94. Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, 6(3), 383-391. <https://doi.org/10.1037/1528-3542.6.3.383>
95. Schoevers, R. A., van Borkulo, C. D., Lamers, F., Servaas, M. N., Bastiaansen, J. A., Beekman, A. T. F., van Hemert, A. M., Smit, J. H., Penninx, B. W. J. H., & Riese, H. (2020). Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. *Psychological Medicine*, 1-10. <https://doi.org/10.1017/S0033291720000689>
96. Funkhouser, C. J., Kaiser, A. J. E., Alqueza, K. L., Carrillo, V. L., Hoffman, L. M. K., Nabb, C. B., Auerbach, R. P., & Shankman, S. A. (2021). Depression risk factors and affect dynamics: An experience sampling study. *Journal of Psychiatric Research*, 135(January), 68-75. <https://doi.org/10.1016/j.jpsychires.2021.01.007>
97. Oren-Yagoda, R., Björngvinsson, T., & Aderka, I. M. (2017). The relationship between positive affect and negative affect during treatment for major depressive disorder. *Psychotherapy Research*, 1-11. <https://doi.org/10.1080/10503307.2017.1292066>

Nederlandse samenvatting

Dankwoord / a word of thanks

My PhD in mood measurements

Curriculum Vitae

List of publications

SHARE

Nederlandse samenvatting

Achtergrond

De meeste mensen voelen zich wel eens een dagje of een paar dagen somber en ellendig, maar voor mensen met depressie zijn gevoelens van leegte, somberheid en gebrek aan plezier een voortdurende staat van zijn. Therapie en antidepressiva kunnen helpen, maar depressieve klachten keren ook vaak terug en niet iedereen heeft baat bij psychologische of medicamenteuze behandeling. Helaas weten we nog weinig over welke personen meer kans hebben om op te knappen tijdens een behandeling en welke factoren een voorspoedig herstelproces voorspellen.

Depressie is een stemmingsstoornis en de ervaring daarvan wordt wel eens beschreven als een onaflatende donkere wolk die alle gevoelens dempt zodat vooral somberheid en een gevoel van leegte overblijven. Toch toont onderzoek aan dat er ook voor mensen met depressieve klachten wel veranderingen zijn in hun stemming van moment tot moment. En juist in die kortstondige wisselingen zou mogelijk informatie kunnen zitten die van belang kan zijn voor het voorspellen van het verdere beloop van iemands depressieve klachten.

Dit idee ontleen we aan de complex dynamische systeemtheorie, van waaruit we mensen kunnen beschrijven als een ‘psychologisch systeem’ waarin gedrag, gevoelens en gedachten met elkaar samenhangen, elkaar beïnvloeden en over verschillende tijdsschalen veranderen (zoals bijvoorbeeld een irritatie kortstondig kan zijn en de sombere gevoelens van een depressie tenminste twee weken aanhouden). In een complex dynamisch systeem kunnen we, door het over de tijd te meten en monitoren, in speciale gevallen een verandering op één niveau (bijvoorbeeld symptoomniveau) voorspellen door te kijken naar veranderingen in de dynamiek van een ander niveau (bijvoorbeeld veranderlijke emoties). Kortom, door depressie als een veranderlijk, dynamisch proces te benaderen, kunnen we onderzoeken of de fluctuaties in emoties binnen de dag voorspellende waarde hebben voor de manier waarop iemands symptomen verbeteren over een tijdsspanne van weken of zelfs maanden.

In dit proefschrift onderzoek ik de manier waarop depressie zich ontwikkelt over de tijd bij individuele personen, met een specifieke focus op de emotiedynamiek die voorafgaat aan het opknappen van klachten tijdens psychologische behandeling.

Hiervoor gebruik ik intensieve longitudinale datasets, verzameld met *ecological momentary assessment* (EMA) technieken, waarbij mensen door de dag heen hun gevoelens ‘in het moment’ registreren. Bijvoorbeeld: “Op dit moment voel ik mij somber”, of “ik voel me energiek”, met antwoordschalen die doorgaans lopen van “helemaal niet” tot “heel erg”. Door mensen steeds opnieuw te vragen hoe ze zich voelen, kunnen we ons een beeld vormen van de wisselingen in stemming over de dag.

Met dergelijke data kunnen we dan kijken naar de mate van emotionele instabiliteit binnen personen en in hoeverre dit samenhangt met de manier waarom iemand opknapt. Ook kunnen we onderzoeken of een belangrijk principe uit de dynamisch systeemtheorie consistent voorkomt bij het opknapproces, namelijk vroege waarschuwingssignalen (*early warning signals*). Dit zijn indicatoren van toenemende instabiliteit in het systeem, die aangeven dat de kans dat een plotselinge transitie zal plaatsvinden steeds groter wordt. In de context van depressie kan een dergelijke grote verandering voor iemand bijvoorbeeld een flinke vermindering van klachten betekenen, of juist een onverwachte terugval na een periode van remissie. Als transities in depressieve klachten ook worden voorafgegaan door zulke kenmerkende vroege waarschuwingssignalen dan zou het monitoren van de stemming van mensen voor de klinische praktijk een waardevolle toevoeging kunnen zijn, bijvoorbeeld voor terugvalpreventie, of om te zien of een behandeling effect heeft. Het doel van dit proefschrift was daarom om dynamiek in gevoelens en de veranderingen in symptomen te bestuderen vanuit een complex dynamisch systeem-perspectief, en de praktische en conceptuele implicaties daarvan te belichten.

Belangrijkste bevindingen per hoofdstuk

Allereerst onderzochten we in **hoofdstuk 2** in hoeverre discontinue veranderingen in dagelijkse metingen van ‘ernst van problemen’ voorkwamen in een groep van 329 patiënten met stemmingsstoornissen die in behandeling waren. Om dit te onderzoeken bekeken we de vorm van verandering en of deze meer een lineair, log-lineair of éénstapspatroon liet zien, en of er ook zogeheten ‘*sudden gains*’ (letterlijk: plotselinge winsten) in plaatsvonden. Vervolgens onderzochten we in hoeverre deze patronen samenhangen met behandelrespons. De meeste mensen,

vooral zij die opknapten tijdens de behandeling ('responders'), lieten complexe, non-lineaire veranderingen zien, die niet pasten binnen onze simpele modellen. Binnen de groep waarbij onze gedefinieerde patronen wel het beste pasten, kwam het éénstapspatroon het vaakst voor en ging dit vaak samen met een *sudden gain*. Degenen met een *sudden gain* hadden vaker een positieve uitkomst van de behandeling dan degenen die geen dergelijke sprong vooruit hadden. Uit deze resultaten concludeerden we dat klinisch relevante verbeteringen in depressieve klachten vaak een non-lineair beloop hebben waarbinnen er abrupte sprongen kunnen plaatsvinden. Dit kan er mogelijk op wijzen dat het proces van herstel van depressie zich gedraagt volgens de principes van de dynamische systeemtheorie.

Vervolgens onderzochten we de voorspellende waarde van destabilisatie in momentane (d.w.z., gemeten 'in het moment') gevoelens voorafgaand aan symptoomverbeteringen. In **hoofdstuk 3** onderzochten we of de mate van destabilisatie in de dynamiek van EMA-gemeten negatieve gevoelens samenhang met een snellere behandelrespons over vijf maanden en met het bereiken van remissie over achttien maanden in een groep van 46 depressieve patiënten. Specifiek onderzochten we of instabiliteit en variabiliteit in negatieve gevoelens, gemeten in de week voor de start van de behandeling, voorspellend waren voor 1) de snelheid van de behandelresponse en 2) of patiënten remissie bereikten tijdens de behandeling, of pas na de behandeling, of helemaal niet. Onze resultaten toonden aan dat een hoger gemiddeld niveau van negatieve gevoelens een sterkere reductie in klachten over vijf maanden voorspelde. Daarentegen waren de mate van variabiliteit en instabiliteit voorspellend voor noch een snellere behandelrespons, noch het bereiken van remissie tijdens de behandelperiode. Hieruit maakten we op dat instabiliteit en variabiliteit gemeten vlak voor de start van behandeling waarschijnlijk geen nuttige indicatoren zijn voor de herstelpotentie van een patiënt, terwijl het gemiddelde niveau van negatieve gevoelens in het dagelijks leven wel enige klinische relevantie zou kunnen hebben en verder onderzocht zou kunnen worden. Een theoretisch gepastere manier om de rol van toenemende destabilisatie — en vroege waarschuwingssignalen in het bijzonder — voorafgaand aan relevante symptoomverbeteringen te onderzoeken vereist een studie waarin langere tijdsreeksen binnen personen worden verzameld tijdens een behandeltraject, zodat *verandering* in emotiedynamiek kan worden gemeten vlak voor symptoomreductie plaatsvindt.

Deze en andere overwegingen werden samengebracht in de beschouwing van **hoofdstuk 4**. In dit hoofdstuk sneden we drie belangrijke conceptuele en praktische uitdagingen aan waar onderzoekers mee te maken krijgen als ze een studie willen uitvoeren voor een empirische, binnen-persoonstest van het fenomeen van vroege waarschuwingssignalen voorafgaand aan symptoomtransities in psychopathologie. Ten eerste raadden we aan om van tevoren (*a priori*) te bedenken hoe relevante transities in symptomen eruitzien en of ze betrouwbaar kunnen worden vastgesteld in de populatie die men wil onderzoeken. Ten tweede bespraken we het belang van de timing van metingen: als je vroege waarschuwingssignalen wilt bestuderen moet de tijdsschaal waarop transities plaatsvinden bepalend zijn voor de lengte van de observatieperiode, de frequentie van de metingen moet hoog genoeg zijn om de relevante dynamiek te vangen en de studie moet plaatsvinden in een periode waarin verandering verwacht kan worden. De methodologische keuzes binnen een dergelijk intensief design moeten echter ook praktisch haalbaar zijn, en dienen in verhouding te zijn met een mate van belasting die voor deelnemers acceptabel is. Tot slot gingen we in op de theoretische en praktische overwegingen voor welk constructen het beste zullen werken om vroege waarschuwingssignalen in te detecteren: men zal een variabele willen kiezen die verandert als symptomen veranderen, die met een hoge resolutie gemeten kan worden (bijvoorbeeld meerdere keren per dag) en waarin voldoende variantie kan worden gemeten voor de analyses. We sloten onze beschouwing af met het advies om de theoretische basis van de methodologische keuzes aan te vullen met de praktische conclusies die men uit pilot-studies kan trekken, om op deze wijze een gedegen, empirisch effectief onderzoeksdesign te bereiken waarmee een ware test van vroege waarschuwingssignalen in psychopathologie mogelijk is.

In lijn met deze aanbevelingen ontwierpen we een studie (als onderdeel van het Transities in Depressie onderzoeksproject; TRANS-ID) waarin we depressie als veranderend proces hebben onderzocht; de onderzoeksprocedures van het TRANS-ID Herstel project staan beschreven in **hoofdstuk 5**. In dit onderzoek hebben we zowel symptomen als momentane ervaringen (gevoelens, gedachten en context) van personen met depressieve klachten intensief gemonitord in een periode waarin zij psychologische behandeling ontvingen en dus een verhoogde kans hadden om (plotselinge) verbetering van hun klachten te vertonen. Voor

deze studie hebben 41 deelnemers elk een jaar lang intensieve longitudinale data verzameld. In de eerste vier maanden bestond de studie met name uit ambulante metingen: continue actigrafie, hartslagmetingen tweemaal daags en vijf keer per dag een EMA-dagboekmeting. Wekelijkse symptoomvragenlijsten begonnen gelijktijdig en liepen zes maanden door, om vervolgens nog een keer per maand te worden afgenomen gedurende de laatste zes maanden van de studieperiode. Met dit design konden we intensieve longitudinale data verzamelen waarmee het mogelijk was om te onderzoeken of vroege waarschuwingssignalen consistent voorkwamen voorafgaand aan symptoomtransities binnen individuen.

In **hoofdstuk 6** beschreef ik een methodologie die ik heb ontwikkeld om op een betrouwbare manier veranderingen in symptomen vast te stellen die over meerdere weken kunnen plaatsvinden. We zochten naar een manier om voor meerdere personen betrouwbare symptoomveranderingen te kunnen identificeren, zonder dat we *a priori* wisten welke tijdsbestek relevant zou zijn voor het opknappen van depressieve klachten. Dit zou namelijk heel goed anders kunnen zijn voor verschillende personen. Daarom ontwikkelde ik de zogenaamde Duration-adjusted Reliable Change Index (DaRCI), waarmee drempelwaardes bepaald kunnen worden voor betrouwbare veranderingen die plaatsvinden over meerdere tijdspunten. Deze index is een aangepaste versie van een veelgebruikte methode (de RCI), waarbij de standaardmeetfout van een vragenlijst wordt gebruikt om een onzekerheidsbereik te bepalen dat overschreden moet worden voordat een verandering tussen twee punten als betrouwbaar kan worden beschouwd (dus: groter dan de verwachte meetfout). In dit hoofdstuk heb ik met gesimuleerde symptoom-tijdsreeksdata laten zien dat de DaRCI in staat was om betrouwbare veranderingen van verscheidene tijdsduren te detecteren — bijvoorbeeld van de ene observatie naar de volgende, maar ook van de ene observatie naar de observatie twee of drie tijdspunten later. De DaRCI was zeer effectief in het oppikken van sterk discontinue veranderingen (een abrupte verschuiving in het gemiddelde over twee tot vier tijdspunten), maar kon ook graduele veranderingen ontdekken als de algehele verandering groot genoeg was.

Met het afronden van de dataverzameling van 41 gedetailleerde datasets van individuen met depressieve klachten en een methode om transities te bepalen, konden we vervolgens de stap nemen om te onderzoeken of vroege waarschuwingssignalen, in de vorm van lag-1 autocorrelatie en variantie,

gedetecteerd konden worden in momentaan gemeten affect, voorafgaand aan een grote symptoomverbetering in individuele patiënten. In **hoofdstuk 7** vonden we dat een toename in autocorrelatie in specifieke gevoelens iets vaker voorkwam bij individuen met een transitie dan bij individuen zonder transitie. Een toename in de variantie in specifieke gevoelens kwam slechts zelden en ongeveer even vaak voor, ongeacht of iemand een transitie had of niet, en werkten dus niet voldoende als vroeg waarschuwingssignaal. Deze resultaten waren robuust en hielden stand in de verscheidene sensitiviteitschecks die we hadden uitgevoerd. Onze bevindingen wijzen erop dat vroege waarschuwingssignalen in specifieke gevoelens geen generieke indicatoren zijn van aankomende symptoomveranderingen — de sensitiviteit (het percentage positieven onder mensen met een transitie) per variabele was hoogstens 50% — maar dat het proces van remissie in depressie wel voor een deel van de depressieve individuen tekenen liet zien van de destabilisatie die je zou verwachten op basis van dynamische systeemprincipes.

In **hoofdstuk 8** hebben we dezelfde vraag over vroege waarschuwingssignalen onderzocht, binnen een groep voormalig depressieve individuen die hun antidepressivum afbouwden en daardoor een verhoogd risico hadden om een terugkeer van depressieve klachten te ervaren. We onderzochten of de lag-1 autocorrelatie en variantie een toename lieten zien voorafgaand aan plotselinge toenames in depressieve klachten. Onze resultaten lieten zien dat zowel de autocorrelatie als variantie vaker steeg onder individuen met transities dan onder individuen zonder transities in klachten. De gemiddelde trends in de variantie waren ook significant hoger voor degenen met een transitie dan voor de groep zonder transities. De verschillende sensitiviteitschecks toonden aan dat deze bevindingen robuust waren. In deze studie vonden we ook dat toenames in autocorrelatie of variantie in tenminste drie van de variabelen veel vaker voorkwam onder mensen met een transitie vergeleken met mensen zonder een transitie. Dit minimum van drie vroege waarschuwingssignalen leidde tot een verhoogde sensitiviteit en ging niet ten koste van een verhoogd aantal vals-positieven. Alhoewel vroege waarschuwingssignalen dus niet werden gevonden voor alle individuen met een transitie, zijn onze resultaten wel een belangrijke empirische bevestiging dat sommige symptoomveranderingen richting depressie een patroon volgen dat overeenkomt met wat men zou verwachten van een complex dynamisch systeem.

Conclusie

Samenvattend geven de studies in dit proefschrift op meerdere manieren inzicht in het herstelproces van depressieve klachten bij individuele personen. Door mensen intensief te bemeten over langere periodes hebben we kunnen laten zien dat de manier waarop depressieve stemming en symptomen veranderen over de tijd veelal complex, non-lineair en lastig te voorspellen is. Hoewel we ons met de TRANS-ID datasets in een unieke positie bevonden om vroege waarschuwingssignalen te vinden, bleek het toch te vroeg om te stellen dat deze binnenkort enige klinische toepassing zullen genieten. Over het geheel gezien tonen onze bevindingen aan dat een complex dynamische systeembenadering gepast lijkt te zijn om veranderingsprocessen in depressie te beschrijven, maar dat het tevens erg lastig is om in psychologische data de beoogde dynamiek en patronen zuiver te meten en detecteren. Daarom concludeer ik in de discussie van dit proefschrift dat we toe moeten werken naar een beter begrip van de relevante tijdsschalen van emoties en symptomen, zodat we de daadwerkelijke processen van verandering ook echt in kaart brengen. Bovendien zijn we de manier waarop mensen opknappen over de tijd nog maar net met intensieve tijdsreeksdata gaan onderzoeken en is er dus nog veel ruimte voor simpele doch fundamentele beschrijvingen van de verschillende soorten fluctuaties, onverwachte veranderingen, en het algehele beloop dat individuen laten zien tijdens hun eigen herstelproces.

Dankwoord / a word of thanks

Om dit boekje mooi af te sluiten is het tijd voor een emotionele speech, maar dan op papier: in dit dankwoord wil ik de velen die me hebben aangemoedigd, aangehoord en bij tijden afgeleid met leuke dingen, een blijvend ereplekje geven.

Allereerst, mijn super supervisors, te beginnen met **Marieke**. Ik weet het moment waarop je mij belde om mij een plek aan te bieden als PhD-student in het TRANS-ID team nog precies; ik zat in Londen in een bus op weg naar mijn werk. Ik aarzelde niet en een oogwenk later belandde ik in Groningen. Wat een waardevolle transitie is dat gebleken. Onder jouw begeleiding heb ik ontzettend veel geleerd; inhoudelijk over psychopathologie en complex systems, maar ook dat je als wetenschapper je eigenzinnigheid en onbevangen enthousiasme niet hoeft op te geven. En het allerbelangrijkste: de waarde van een goed team om je heen. Door jouw steun en supervisie had ik een basis waar vanuit ik ook in jouw afwezigheid kon doorgroeien. Zonder jou was ik niet de trotse, zelfverzekerde wetenschapper die ik nu ben; ik zal je daar altijd dankbaar voor zijn.

Evelien, als dagelijks begeleider zocht ik jou vaak op voor steun en advies als ik weer eens vastliep — bij schokkende codefouten (zeldzaam), gebrek aan overzicht (vaak), en ook momenten dat ik er even doorheen zat (iets vaker dan soms). Je stond altijd klaar om te luisteren en met mij samen dingen weer behapbaar en overzichtelijk te maken. Als geen ander kan jij met één vraag tot het hart van de zaak komen, een kwaliteit die ik mij ook meester probeer te maken. Met jou als begeleider heb ik ervaren hoe fantastisch het is om te werken met mensen die, net als ik, giechelen om flauwe grapjes, aan hangriness lijden, maar ook onverbloemd blij kunnen worden van het doen van onderzoek.

Laura, in een jaargesprek dat plaatsvond ver voordat je officieel bij mijn begeleidingsteam werd betrokken had ik al aangegeven hoe leuk ik het zou vinden als ik met jou meer samen kon werken, omdat ik dacht dat wij wel eens compatibele ideeën zouden kunnen hebben. Dat je mij hebt geadopteerd halverwege mijn traject was voor mij dus echt een feestje. Ik heb—zoals gehoopt—veel plezier gehad van onze samenwerking en het delen van ideeën middels de vele leerzame discussies die we hebben gehad, ook binnen LaBLab. Bovendien heb ik behalve van

jouw methodologische expertise ook ontzettend veel geleerd van jouw krachtige houding over de wetenschap, eerlijkheid en openheid; daarin ben je voor mij echt een voorbeeld.

En dan de laatste in deze line-up van fantastische begeleiders, **Tineke**. Toen jij ten tonele kwam, bracht je met jouw ervaring en praktische nuchterheid al gauw de nodige rust terug nadat ik me al een tijdje wat stuurloos had gevoeld. Jouw bemoedigende woorden en oprechte vertrouwen dat ik er goed voorstond, waren heel welkom en hebben ervoor gezorgd dat ik de laatste fase van mijn PhD ook met veel plezier heb kunnen afronden. Ik ben je ontzettend erkentelijk dat je bereid was om zo laat in mijn promotietraject als begeleider aan te treden, en ik ervaar het als een zeer positieve wending om onder jouw kalme, toezien oog de puntjes op de i te hebben kunnen zetten.

Vervolgens moet ik natuurlijk mijn geweldige **YAMFARM**-vrienden bedanken, te beginnen met de kamergenoten waarmee ik de afgelopen vijf jaar het acroniem **YAM** (en een volkstuin) heb gedeeld. Eigenlijk begon mijn PhD met het grote ‘Yoram is 46 en verft zijn baard’-verhaal, en daarmee was de toon van rare streken en ‘geloof ze nou niet op hun woord!’-trend wel gezet. Ik sloot mij daar graag bij aan als medeplichtige, en mede dankzij jullie bodemloze geduld tegenover mijn oeverloze gebabbel, kijk ik met veel plezier terug op onze tijd samen op K1.15a. **Yoram**, ik zie jou een beetje als een kwajongen, want naast je ijver op werk had je ook altijd nog tijd voor off topic ongein op kantoor — en voor opinieblogs schrijven, je koeken-stash aanvullen, en een ‘illegale’ plek verwerven in Arts en auto. Je laat je niet altijd makkelijk lezen, maar over de jaren ontwikkelde zich toch een soort kamer-buizenradar, waardoor we ook over de uitdagende dingen van het leven fijn hebben kunnen praten. **Arnout**, ik vind het geweldig hoe wij elkaar met zoveel gemak kunnen aanschieten voor lange gedachtewisselingen over van alles en nog wat: statistiek en dynamische systemen, de beste manier om iets te fermenteren, of over hoe je het beste deelnemers kan overtuigen dat ze zoveel mogelijk moeten invullen; het houdt nog altijd niet op. Ik waardeer jouw bereidheid om te luisteren naar mijn rants en steevast iets zinnigs te antwoorden op de vraag ‘wat ga je vanavond eten?’ De vriendschap die we na deze vijf jaar hebben opgebouwd had ik wellicht niet voorspeld, maar ik ben heel dankbaar dat je van begin tot eind naast me staat.

En dan de eerste van **FARM!** Lieve **Fionneke**, wij hebben ons op een gegeven moment in elkaar herkend en zijn simpelweg niet opgehouden met praten en delen. Nu gooien we er zo een muntje in (liefst met een glas wijn erbij) en kunnen we uren losgaan over wetenschap, feminisme, flauwe memes, twitterdraadjes, klagen over prikkels, enzovoorts. Ik kan bij jou altijd rekenen op je eerlijkheid en enthousiasme, waardoor ik zelf ook altijd zonder schroom aan kan geven hoe ik ervoor sta, ook al is dat soms een beetje hypomaan of juist #overprikkelbaars. Het was al een genot en eer om onderdeel te zijn van jouw promotie, en het is een hartverwarmend vooruitzicht om straks de rollen om te draaien en met jou mijn eigen verdediging te vieren. Ik ben dankbaar voor onze vriendschap, en stel voor dat we vooral doorgaan met lekker genieten.

Next up, dubbel dokter **Anouk**, ik word altijd vrolijk van jouw energieke non-sense houding en gekke verhalen, en als ik zie hoe fanatiek jij kan genieten van het spelen van verstopperijtje in een donker huis of het leren jongleren, steek ik daar zelf ook altijd weer wat van op. **Robin**, jij was eigenlijk de katalysator voor mijn komst naar Groningen omdat jij de vacature voor mijn PhD positie hebt doorgestuurd – bonuspuntjes voor jou dus. Al tien jaar(!) bewandelen we al ongeveer hetzelfde academische pad, en dat is toch een gezellige wandeling geweest, hopelijk verliezen we elkaar niet uit het oog nu we elk een eigen weg op gaan. **Marieke**, ik kijk op tegen jouw vermogen om enerzijds een werkelijk power house van wetenschappelijke kennis en kunnen te zijn, en anderzijds ook een heel normaal primair-reagerend lief mens te zijn, en kan ontzettend genieten van jouw droge constatering. Ik hoop op nog vele fietstochtjes en (niet-)schrijfweken in de toekomst met heel **YAMFARM**, jullie vriendschap is zonder meer het dierbaarste dat ik uit Groningen meeneem.

Uiteraard is mijn dank ook groot naar de andere leden van het TRANS-ID team. **Harriëtte**, jouw inzet voor het project — zowel het regelen van de apparatuur als later het overnemen van een leidende rol — hebben ervoor gezorgd dat we met zijn allen onze ‘knalpapers’ hebben kunnen schrijven in een lastige tijd. Dit, en je aanstekelijke lach, zal ik niet snel vergeten. **Sandip**, when you joined the team, I was curious to see what an astrophysicist would make of all our psychological data, but I really loved the ideas you brought to the table. Thanks for all the fun discussions (and memes), and send me those foodie photos from London! **Ando**,

bij het bouwen, en later tweaken, van de TRANS-ID website heb ik met veel plezier samen met jou zitten nederen over het design en heb ik veel geleerd over workflows in github – je was misschien minder zichtbaar, maar toch een onmisbaar onderdeel van het team.

De helden van de Herstel dataverzameling, **Elly, Margriet, Abel, René, en Anneke**. Zonder jullie was de TRANS-ID Herstel dataset niet wat ze nu is, en had ik mijn handen niet vrij gehad om aan de hoofdstukken in dit boekje te werken. Naast jullie fantastische inzet heb ik ook met elk van jullie kunnen genieten van een fijne, gezellige samenwerking.

Ook bedank ik graag de leden van **RoQua** voor het programmeren van de vragenlijst- en dagboekprotocollen en het beheren van de data, en **Inge** voor het invoeren van de vragen en al het detailwerk dat daarbij kwam kijken. Dank ook aan alle **secretariaten, behandelaars en instellingen** die zich hebben ingezet voor het werven voor TRANS-ID, hetzij via flyers, via intakegesprekken of andere paden.

Tot slot, mijn grote dank aan **de deelnemers** van TRANS-ID Herstel. Tijdens een pittige periode waarin depressieve klachten en behandeling een grote plaats innamen in jullie leven waren jullie toch bereid om ook nog plaats te maken voor ons onderzoek, dat blijft voor mij bijzonder. Zonder jullie was het niet mogelijk geweest!

Dan, een ode aan de vele gezellige collega's van het ICPE. Waaronder ook **Daan, Olga, Anya, Vera, Michelle, Jojanneke, Hanneke, Larisa** en **Sanne**; dank jullie allen voor de gezelligheid bij (internationale) congressen, project-brainstorms en de vele borrels en -mibo's. **Lino**, ik zal je nooit niet verdenken bij (bord)spelletjes en hoop dat we jou ook bij toekomstige spelletjesavonden tegen alle waarschijnlijkheid in zullen ontmaskeren. **Maurits**, dankzij jou heb ik het aangedurfd om een prachtige wielrenfiets te kopen, én door jou ben ik altijd bang dat mijn *p*-waarden verdacht zijn, maar ben ik ook fervent open science voorstander geworden en daardoor altijd goed voorbereid op de argusogen van eventuele critici. Ik wil ook **Esther** en **Margo** heel erg bedanken voor al het geweldige geregeld en georganiseerd dat ze voor mij en anderen hebben gedaan over de jaren – dat, en de gezellige praatjes waar jullie meestal wel voor te porren waren.

Another group that deserves a big shout out are the LaBlabers, **Timon, Marie, Jannis, Anna, and Sebas**. Thank you for always being so welcoming and for showing interest in my work and wellbeing, even if I was kind of an ‘honorary member’. I really enjoyed our discussions and loved learning from you, and, of course, ziplining at the team onion - thanks LaBlabers!

Members of the reading committee, **prof. Hayes, prof. Jacobs, and prof. Oude Voshaar**, I am very grateful for your willingness to read and assess my dissertation, and for your participation in the defense itself!

My thanks also to my **co-authors**, and in particular to **Günther** and **Frenk**, who trusted me with their datasets and supported me in writing the first two chapters of this thesis. **Merlijn**, als frequent co-auteur verdien je een special thanks: samen hebben we toch een aantal heel erg coole papers geschreven, waarbij we juist doordat we over dezelfde onderwerpen net iets anders nadenken tot een beter eindresultaat hebben kunnen komen.

To my dear UCU friends (in no particular order, I love you all equally <3): **Prashanthi, Veerle, Kim Lien, Bibi, Joey, Judith, Minh Tue, Mischa, Emiel, Matthijs**. In one way shape or form, you have all inspired me over the years and made my life just so much more fun. Sometimes that meant remembering that I actually love to spend time with friends (even if I am also a bit of a hermit at times), sometimes that meant encouraging me to fight for what I deserve or taking the rest I need, and sometimes that meant traveling to see you all and discover beautiful new places along the way. You guys are awesome, thanks for being my friends.

Dan nog mijn lieve ouders, **Jan** en **Martine**. Jullie hebben me eigenlijk altijd vooral mijn eigen pad laten kiezen, en me gesteund en terug in evenwicht gebracht als ik dat nodig had. Toen ik verhuisde naar Groningen kon ik het niet zo goed overzien om een huis te zoeken (papa to the rescue om een bezichtiging te regelen), en daarna had ik natuurlijk wéér eens verhuishulp nodig, maar ook daarvoor stonden jullie klaar (voor Noorwegen zal ik het zelf regelen, beloofd). Ik hoop dat jullie trots zijn en met mij samen zullen genieten van deze kroon op mijn wetenschappelijke opvoeding, want dat is toch ook een beetje te danken aan jullie support als opvoeders vanaf de eerste dag.

Mijn grote broer, **Thom**, ik vind het fijn dat we na vele jaren elkaar het bloed onder de nagels vandaan halen toch goed met elkaar op kunnen schieten. We zien elkaar niet zo vaak, maar dat heeft ons niet weerhouden om ineens dezelfde sport te gaan beoefenen en allebei te genieten van natuur, spicy eten en gamen. Ik vind het ontzettend tof dat je erbij bent op mijn grote dag, en hoop dat je met me komt klimmen in Noorwegen.

Dear **Leonardo**, there's a reason my data show a burst of positive feelings around the time I met you – you left quite an impression. Being with you means lots of time being silly, warmed, loved, and sharing in activities we both enjoy. Having your support (read: coffee in the morning) and quiet (read: annoying drumming) presence with me while I went through the last phase of my PhD was a source of comfort that made it all seem more doable. I'm grateful for your willingness to put up with me, make me laugh, care for me, and work towards building a life together. Thank you, boy, ik heb je lief. To many more adventures <3

The course of my PhD in 7000+ mood measurements

Over four years of my PhD, I collected more than 7300 measurements of my own feelings, cognitions, and activities by completing five questionnaires a day ('dagboektijd!' was at 10:00, 13:00, 16:00, 19:00, and 22:00). What started as a way to test our study protocol and better relate to our participants' experiences, has resulted in my own idiographic intensive longitudinal dataset, which I felt deserved a mention in this dissertation. Visualisations of the smoothed curves of 20 variables (legend below) and a scatter plot of the 10 primary affect items (**bold** in English) are included on the following pages.

content	■	Ik voel me tevreden
at ease	■	Ik voel me op mijn gemak
look forward	■	Ik heb zin in de rest van de dag/morgen
self-thoughts	■	Mijn gedachten over mezelf zijn nu
cheerful	■	Ik voel me opgewekt
concentrate	■	Ik kan me goed concentreren
energetic	■	Ik voel me energiek
physical activity 3h	■	De afgelopen drie uur was ik lichamenlijk actief
tired	■	Ik voel me moe
spoken with others 3h	■	De afgelopen drie uur heb ik gepraat met een ander(en)
desire company	■	Ik heb zin om met mensen te zijn
overstimulated	■	Ik voel me overprikkeld
physical discomfort	■	Ik ervaar lichamenlijk ongemak
chaotic	■	Ik voel me chaotisch
down	■	Ik voel me somber
listless	■	Ik voel me lusteloos
restless	■	Ik voel me onrustig
irritated	■	Ik voel me geïrriteerd
stressed	■	Ik voel me gestrest
worrying	■	Ik ben aan het piekeren

Note. The chosen annotated life events are illustrative, not exhaustive

Curriculum Vitae

Marieke Helmich was born on February 13, 1992 in Haaren, The Netherlands. In 2010, she started her Bachelor in Liberal Arts & Sciences at University College Utrecht. As part of the bachelor, she spent the fall semester of 2012 at Nanyang Technological University in Singapore — from where she eagerly traveled around, and picked up a passion for cooking along the way. She graduated *cum laude* in 2013, with a major in Psychology, and a minor in Philosophy as well as in Methods & Statistics.

In September of the same year, she started the selective Research Master programme in Cognitive and Clinical Neuroscience with a specialisation in Psychopathology at Maastricht University. Most of the second year of the programme was spent as a research internship, for which Marieke moved to London, United Kingdom, at the end of 2014. She completed her master's thesis —a study of the relation between daily life social behaviour and emotion recognition in patients with psychosis— while in London in the summer of 2015, and remained living there until she found a PhD with the Transitions in Depression (TRANS-ID) research team at the University Medical Center Groningen in November 2016.

Under the expert supervision of prof. dr. Marieke Wichers, dr. Evelien Snippe, and later also dr. Laura Bringmann and prof. dr. Tineke Oldehinkel, she managed the TRANS-ID Recovery data collection —recruiting depressed individuals seeking treatment to monitor their mood and behaviour intensively over four months— and developed strong theoretical and methodological knowledge on the topic of dynamical systems applications (and early warning signals in particular) in the context of depressive symptom change during treatment.

From January 2022 onwards Marieke will join the Complexity in treatment Outcome, Psychopathology and Epidemiology (COPE) research group at the University of Oslo, Norway, as a postdoctoral research fellow. She looks forward to continuing her work on how individuals change over time in the context of psychopathology and psychological treatment.

List of publications

Peer-reviewed publications

Helmich, M. A., Olthof, M., Oldehinkel, A. J., Wichers, M., Bringmann, L. F., & Smit, A. C. (2021). Early warning signals and critical transitions in psychopathology: challenges and recommendations. *Current Opinion in Psychology*, 41, 51–58. <https://doi.org/10.1016/j.copsyc.2021.02.008>

Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, 88(2), 119–127. <https://doi.org/10.1037/ccp0000469>

Bringmann, L.F., **Helmich, M.A.**, Eronen, M.I., & Völkle, M.C. (In press) *Complex systems approaches to psychopathology*. In Krueger, R.F. & Blaney, P.H. (Eds.), *Oxford Textbook of Psychopathology*. Oxford University Press.

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., **Helmich, M. A.**, Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, 8(1), 25–35. <https://doi.org/10.1177/2167702619865969>

Submitted publications

Helmich, M. A., Wichers, M., Peeters, F., & Snippe, E. (Under review). Daily dynamics of negative affect: indicators of rate of response to treatment and remission from depression? *Manuscript under review*. <https://doi.org/10.31234/osf.io/hv65n>

Helmich, M. A., Smit, A. C., Bringmann, L. F., Schreuder, M. J., Oldehinkel, A. J., Wichers, M., & Snippe, E. (Under review). Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression. *Manuscript under review*. <https://doi.org/10.31234/osf.io/vf86s>

Smit, A. C., **Helmich, M. A.**, Bringmann, L. F., Oldehinkel, A. J., Wichers, M., & Snippe, E. (Submitted). Personalized detection of impending symptom transitions in depression using early warning signals during and shortly after antidepressant discontinuation. *Manuscript submitted*.

In preparation

Hol-Steegstra, A., **Helmich, M. A.**, Riese, H., & Snippe, E. (In preparation). Is anxiety associated with active behaviour in patients diagnosed with Major Depressive Disorder? *Manuscript in preparation*

Helmich, M. A. (In preparation). The Duration-adjusted Reliable Change Index (DaRCI): defining clinically relevant symptom changes of varying durations. *PsyArXiv Preprint*. <https://doi.org/10.31234/osf.io/q7ch9>

Open materials

Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Recovery: study protocol for a repeated intensive longitudinal n = 1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *Open Science Framework*. <https://doi.org/10.31234/osf.io/fertq>

Smit, A. C., **Helmich, M. A.**, Kunkels, Y. K., Riese, H., Snippe, E., & Wichers, M. (2020). Transitions in Depression (TRANS-ID) Tapering: study protocol for a repeated intensive longitudinal n = 1 study design to search for personalized early warning signals of critical transitions towards depression. *Open Science Framework*. <https://osf.io/zbwkp/>

Research institute SHARE

This thesis is published within the Research Institute SHARE (Science in Healthy Ageing and HealthcaRE) of the University Medical Center Groningen and University of Groningen. Further information regarding the institute and its research can be obtained from our website: <http://www.share.umcg.nl/>

More recent theses can be found in the list below ((co-)supervisors in parentheses).

2021

Wijnen, A.

Rehabilitation policies following total hip arthroplasty; across borders
(*prof. S. K. Bulstra, dr. M. Stevens, dr. D. Lazovic*)

Narmandakh, A.

The night before the day after; connections between sleep problems and anxiety across the lifespan
(*prof. A. J. Oldehinkel, dr. A. M. Roest*)

Birkenhäger-Gillesse, E. G.

"More at Home with Dementia"; effects of psychosocial interventions in the community and in nursing homes
(*prof. S. U. Zuidema, prof. W. P. Achterberg, dr. S. I. M. Janus*)

Sheferaw, E. D.

Improving respectful maternity care in Ethiopia
(*prof. J. Stekelenburg, prof. T. H. van den Akker, dr. Y. M. Kim*)

Douma, K. W.

Clinical muscle strength measurements: reference values and reliability
(*prof. C. P. van der Schans, prof. P. U. Dijkstra, dr. W. P. Krijnen*)

Bootsma, J. M.

Effects of task difficulty and old age on motor learning and its neural mechanisms
(prof. T. Hortobagyi, dr. S. R. Caljouw)

Fraszczyk, E.

DNA methylation in type 2 diabetes and metabolic health
(prof. H. Snieder, dr. M. Luijten, dr. A. M. W. Spijkerman, dr. J. V. van Vliet-Ostaptchouk)

Ayuningtyas, D. A.

Will you quit smoking for me? The role of spouses in the smoking behaviour and intention to quit of Indonesian male smokers
(prof. M. Hagedoorn, dr. M. A. Tuinman)

Lackova Rebicova, M. L.

Adverse childhood experiences among adolescents with emotional and behavioural problems
(prof. S. A. Reijneveld, prof. A. Madarasova-Geckova, dr. J. P. van Dijk, dr. Z. Dankulincova)

van der Veen, D. C.

Anxiety in late-life depression
(prof. R. C. Oude Voshaar, prof. R. A. Schoevers, dr. H. Comijs, dr. W. H. van Zeist)

van Amerongen, C. C. A.

Contact allergy; occurrence and susceptibility
(dr. M. L. A. Schuttelaar, prof. P. J. Coenraads)

Hernandez, L. G.

Health economic models for relapsing-remitting multiple sclerosis
(prof. M. J. Postma, dr. H. Le)

For earlier theses visit the SHARE website.