Early-warning signals derived from physiological and behavioural measures

Is it worth a transition in clinical practice?

Y. K. Kunkels

The research in this thesis was performed within 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).

Graphical design: Jenita Moolhuijsen

Printed by: NBD Biblion

university of groningen

Early-warning signals derived from actigraphy and electrocardiogram time series data:

Is it worth a transition in clinical practice?

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 June 21st at 14:30 hours

by

Yoram K. Kunkels

born on 6 March 1986
in Haarlem

Supervisors

Dr. H. Riese

Prof. dr. M.C. Wichers

Co-supervisor

Dr. ir. A.M. van Roon

Assessment Committee

Prof. dr. D. Borsboom

Prof. dr. I. Germeys

Prof. dr. R. C. Oude Voshaar

Contents

1.

General Introduction

Chapter 1: General Introduction

Major Depressive Disorder (MDD) is a mental disorder wherein diagnosed patients experience, amongst others, depressed mood, decreased pleasure, sleep problems (hypersomnia or insomnia), psychomotor agitation or retardation, and fatigue (American Psychiatric Association, 2013). It is a common mental disorder and one of the most important factors of disability worldwide, with approximately 264 million persons suffering from its consequences (WHO, 2009).

Findings from earlier research suggest that MDD involves encountering various complex endogenous and exogenous triggers during a person's lifespan through a perpetual process of continuous changes in vulnerability (Beauchaine et al., 2011; Wichers et al., 2010). As various triggers and their timing are distinctly person-specific, significant changes, or transitions, in depressive symptom severity can be challenging to foresee. This is partly due to the commonly used nomothetic research designs – focussed on group-level studies and generalising knowledge – being the most commonly used perspective for studying depression so far. In order to translate nomothetic research findings on MDD into useful assets for clinical practice, a more person-centred (or idiographic) approach – focussed on individual-level studies and person-specific knowledge – is required (Molenaar, 2004; Zuidersma et al., 2020).

An idiographic focus can help to unravel the complex individual differences in fluctuations in symptoms typically encountered in patients in clinical practice. Given the inability of methods currently used in psychological research to process such individual differences, a different approach is required to gain insight into the development of depression vulnerability and to be able to support personalised interventions. This new appreciation for idiographic research highlights the need for novel approaches to better grasp the developmental trajectories of disorders in individual patients. Discovering personalised mechanisms could potentially contribute to improved treatments for MDD and associated disorders. In this general introduction, I will elaborate on why approaching depression as a complex dynamical system could contribute to the field and which data acquisition methods are needed to obtain the intensive data to longitudinal monitor changes in depressive symptoms.

Illustration of complex dynamic systems

Complex dynamic system theory has shown its merit in understanding switches between alternate stable states, seeing application in, for example, biological, financial, and climate science contexts (Scheffer et al., 2009). The underlying conceptual principles are relatively simply explained though via a practical demonstration of a *cusp* model, a type of catastrophe model (Van der Maas, 2004; Wagenmakers et al., 2005). In short, the cusp model is a model with one behavioural axis and two control axes, which can show the various potential stable and unstable states. For example, try holding a credit card between two fingers of one hand (or see figure 1). When some pressure is applied, the credit card will bend, ending up in one of two final states; being bent to the left, or being bent to the right, while any state between these two is unstable. A transition, for example, the credit card going from being bent to the right side to being bent to the left side, can be forced by applying pressure from the side. Notable here is that it takes quite some pressure to force the credit card into the other stable state, that the movement of the credit card from right to left does not progress linearly, and that this transition can occur very suddenly. This simple example of a system with two stable states can be used as a starting point for conceptualising mental disorders, such as depression, as complex dynamic systems. For instance, instead of the credit card now imaging a person with a history of experiencing depressive symptoms, but currently being symptom-free. This person's current state, not experiencing depressive symptoms, can be seen as one of the two stable states the credit card could be in. Now, when external or internal stressors start putting this person under pressure, the risk of transition towards the other stable state, experiencing depressive symptoms, will increase. However, as with the credit card, the transition from the not depressed state into the depressed state will not happen linearly and can occur quite suddenly and unexpectedly, thereby frustrating the timely prevention of mental disorder onset.

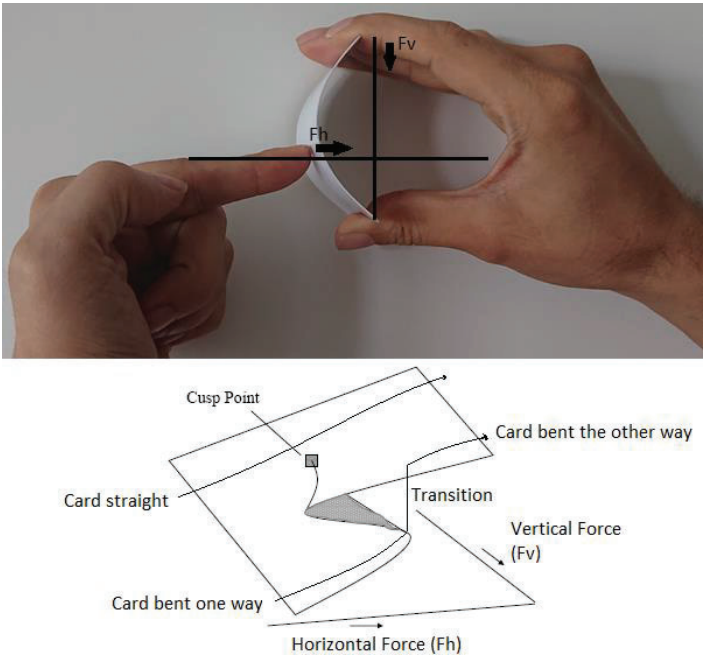


Figure 1: Example showing forces applied to a credit card (upper panel) and how they relate to a transition model (here a 'Cusp' catastrophe model).

Major Depressive Disorder conceptualized as complex dynamic systems

These conceptualisations can help to illustrate why previous methods struggled to identify transitions in mood disorders, such as MDD. For example, methods depending on a linear progression of a transition between mood states, such as being depressed or not, will fall short in detecting the non-linear progression often observed by mental health care professionals in patients suffering from MDD. The risk for transitioning towards, or away from, being depressed is highly dependent on the waxing and waning of various internal and external processes, which can be hard, if sometimes not impossible, to track reliably and consistently. This struggled to identify transitions in mood disorders is further complicated by the well-known large individual differences in psychopathological mechanisms.

The conceptualisation of mental disorders, such as depression, as complex dynamic systems was already suggested over twenty years ago (Hayes & Strauss, 1998; Van der Maas & Molenaar, 1992). However, most research into this construct is conducted in other fields. The

sudden manner in which lake water can transition from clear to turbid water quality, but also various pathologies such as epileptic seizures and asthma attacks, can be accurately modelled using measures derived from complex dynamic systems theory (Scheffer et al., 2009). In short, a complex dynamical system approach can potential predict the onset or remission of symptoms without fully understanding the mechanisms causing these changes. Both sudden rises in depressive symptoms and sudden disappearance of symptoms, are frequently seen in patients and suggest that the fragility of the system builds up gradually and eventually results in a sudden critical transition. This is a typical characteristic of complex dynamical systems. Scheffers and colleagues (2009, 2012) argue that critical transitions are preceded by a characteristic critical slowing down (Wissel, 1984). When a system approaches a critical transition (or a tipping point), it returns more slowly to its initial stable state under small perturbations. The return time to the stable state can index whether a critical change is near or not. Based on these phenomena statistical indices have been proposed to operationalise this, not simply for the eye detectable, slowing down (Nazarimehr et al., 2020). These indices are named Early Warning Signals (EWS). Although theoretically valuable and potentially clinical relevant, studies have not yet empirically and prospectively investigated whether EWS anticipate critical transitions in the severity of symptoms of mental disorders, such as MD. For this, innovative monitoring tools to gather intensive time series data of individual patients and analytical (pre-)processing tool are needed. In the following section, I will first elaborate on the main aim of this thesis, that is, predict transitions in depression with EWS and subsequently the time series data acquisition tool needed for this.

Early-warning signals

Conceptualising depression as a complex dynamical system allows us to apply promising statistical techniques to predict upcoming symptom transitions, which proposed a simple set of indices, called generic *early-warning signals* (EWS). EWS could help detect whether a transition from one state to an alternative one is afoot. For example, it is well-known that patients who are in stable remission and wish to taper their antidepressants use are extra vulnerable to experience a significant change in their depressive symptoms (Shelton, 2001). Here EWS indices may signal an increase in system instability, which is expected to occur before a transition (Dakos, Van Nes, et al., 2012; Scheffer et al., 2009). Common time series derived EWS indices include increasing *variance*, *autocorrelation*, and *kurtosis* (Biggs et al., 2009; Dakos, Carpenter, et al., 2012; Scheffer et al., 2009). Variance is a measure, which

indicates dispersion of data points around a certain value, say 100. Here, low variance would mean most data points would lie relatively close to 100, while with high variance the data points would lie varied from relatively close to relatively far away from 100. Autocorrelation is the correlation of something with a delayed version of itself, the delay often being specified as lag-*n*. For example, autocorrelation at lag-1 is the correlation of a measure with a delayed version of itself with a delay period of 1. Kurtosis is a measure of the "peakiness" of a probability distribution of a certain variable. More positive kurtosis corresponds to a more peaked distribution with less weight in the tails. More negative kurtosis on the other hand corresponds to a relatively flat probability distribution with more weight in the tails.

Predicting transitions in depressive symptoms

As outlined above, EWS are deemed *generic* because they are not dependent on the exact underlying model. In a pioneering case-study in an individual diagnosed with MDD tapering his antidepressant use, it was shown that increases in EWS, namely variance and autocorrelation, derived from daily fluctuations in experienced affect can be observed in anticipation of a transition in depression (Wichers & de Groot, 2016). This was the first study showing rising EWS within a single person anticipating a clinically and statistically significant transition in depressive symptoms. It highlighted that EWS measures and analyses tailored to such *single-subject designs* are helpful to successfully study the development of depression in individuals. As such, this study has laid the groundwork for subsequent investigations herein, such as the TRANS-ID (TRANSitions In Depression; www.transid.nl) study.

The TRANS-ID study was designed as a repeated single-subject study (TRANS-ID Tapering (see: https://osf.io/zbwkp/); TRANS-ID-Recovery (see: https://osf.io/85ngu)). Time series data were gathered during four months from over 100 participants with a background of MDD to investigate whether EWS in a broad range of data types would be detected preceding transitions in depression. The collected data were multi-faceted as it includes various data types, namely daily diary data (or experience sampling method, ESM), physical activity data, and electrocardiogram (ECG, or heart rate) time series data. ESM is a scientific method to optimize accuracy and validity of self-report of affective experience in normal daily life. Participants fill out a short questionnaire, multiple times a day. Earlier research showed that ESM monitoring, also for longer periods of time (weeks, months) is also feasible for affected psychiatric patients (Bos et al., 2019), although sometimes experiences as burdensome (van Genugten et al., 2020) as it requires conscious effort. Recent innovations in ambulatory

monitoring have made easy, non-invasive monitoring of physical activity and ECG activity with low participant burden possible. Although wearable monitors cannot measure emotions or mood symptoms directly, they do offer objective measures, while offering potentially low burden to participants. Various data modalities can help negate each other's disadvantages. Therefore, it makes sense to combine various data modalities and to monitor both ESM and physiological measures. Part of the studies presented in my thesis can be seen as complementary to the TRANS-ID ESM studies, as I investigate objective and physiological time series data in relation to (transitions in) depression. I used data collected by the participants via wrist-worn actigraphy monitors and inter beat-interval (IBI or heart rate) data collected with wireless electrocardiogram (ECG) monitors. Typically, actigraphy is assessed at the second level and IBI time series do constitute the most fine-grained type of data assessed at microsecond level. To the best of my knowledge, these types of data have not been used before to test the dynamical systems hypothesis.

In sum, ambulatory monitoring technology advances quickly, enabling and facilitating the long-term monitoring of actigraphy and ECG patterns. This offers an opportunity to combine the newest insights and ideas from the field of complex dynamical system theory with the newest technological possibilities and a novel personalized vision on psychiatry. The psychiatric field would move forward if we can indeed obtain personalized information on the likelihood for critical transitions in depressive symptoms using time series data from ambulatory monitoring of physical actigraphy and ECG. However, before continuing on actigraphy and ECG time series derived EWS to predict transitions in depression, a short overview is given on the link between depression and actigraphy and both ambulatory and non-ambulatory ECG data.

Conceptual links between actigraphy and ECG time series data and depression

In the following paragraph, the two main types of time series data linked to depression used in this thesis; physical activity data and ECG data, will be introduced.

The link between sleep and circadian rhythm disturbances and depression

Living organisms show cyclic rhythmicity in many of their physiological and behavioural processes, for instance, sleep and wake cycles. The suprachiasmatic nucleus (SCN), located in the anterior hypothalamus, functions as a master circadian (from the Latin *circa diem*;

meaning around one day) clock sustaining a near 24-hour cycle (Stephan & Zucker, 1972; Stetson & Watson-Whitmyre, 1976). Factors such as light (Van Someren et al., 1999) and (absence of) feeding (Damiola et al., 2000) play a role in maintaining the circadian rhythm, by entraining it to external *Zeitgebers*. There is an established link between experiencing depressive symptoms and specific physical activity patterns, such as psychomotor retardation, wherein patients experience a slowing down of and reduction in their physical movements. This psychomotor retardation is a key feature of MDD and it is one of the main symptoms used to diagnose MDD (American Psychiatric Association, 2013; Buyukdura, McClintock, & Croarkin, 2011). Given this knowledge we know that physical activity patterns change when patients start to experience more depressive symptoms, and thus expect physical activity to be an important component of depression. Patients experiencing mood disorders often report sleep issues, such as having problems falling asleep or waking too early (Tsuno et al., 2005). Moreover, earlier studies have suggested that sleep problems are related to negative clinical outcomes. For instance, persons experiencing insomnia are known to have a significantly increased risk for developing major depression (Breslau et al., 1996; Mallon et al., 2000). Moreover, insomnia was found to occur before the onset or reappearance of depressive symptoms (Perils et al., 1997; Riemann & Voderholzer, 2003). In addition to these sleep-related issues, problems with circadian rhythm have been linked to depression. For example, while healthy individuals commonly experience lower mood in the evening than in the morning, the reverse is true for individuals suffering from depression, who often report lower mood in the morning than in the evening (Gordijn et al., 1994; Tölle & Goetze, 1987). Such behavioural patterns are also known to change and normalise during treatment and recovery (Winkler et al., 2014). Based on such earlier works, disturbances in sleep and circadian rhythm are ought to play a considerable role in the development and maintenance of mood disorders (Germain & Kupfer, 2008). These disturbances are found to be so central to the characterization of depression that they are included as diagnostic criteria for depression; highlighting fatigue, sleep difficulties (insomnia or hypersomnia), and psychomotor retardation (American Psychiatric Association, 2013; Frankland et al., 2015; Parker et al., 2002). A non-invasive method for collecting data on sleep and circadian rhythm is actigraphy. Actigraphy uses wrist-worn accelerometers to measure data from participants in their normal daily life environment, and with minimal burden to the participant (Fuller et al., 2017).

The link between heart rate and depressive symptoms

Cardiac imbalances in the (para-)sympathetic nervous system are also known to be associated to the development and maintenance of depression. Such imbalances are reflected in heart rate or Inter-Beat Intervals (IBI), or heart rate variability (HRV) measures (Choi & Gutierrez-Osuna, 2010; Malik et al., 1996). A number of prior nomothetic studies have provided group-level knowledge about the relation between cardiac measures and depressive symptoms. For example, decreased HRV indicators were found to be associated with depressive symptoms (Vaccarino et al., 2008). As HRV decreases, heart rate will show less and less variation, making the heart rate less adaptive to triggers from the internal and external context or more rigid (Dekker et al., 2000). HRV rigidity is related to various unwanted health effects, such as myocardial infarctions (Kleiger et al., 1987) and increased mortality rates (Schouten et al., 1993; Dekker et al., 2000). Whether the heart rate can react flexibly or rigidly may influence how individuals can cope effectively with mental, physical, and environmental stressors (Colzato et al., 2018; Hel et al., 2021). Heart rate can change due to depressive symptoms, for example deregulated cardiac autonomic nervous function, including elevated heart rate (HR) and, decreased heart rate variability, have been associated with depressive symptoms (Carney, et al., 2005). As such, electrocardiogram (ECG) derived measures are expected to be an important component of depression. This makes ECG derived EWS measures potential markers to foresee changes in depressive symptoms. Yet, while earlier nomothetic studies do offer valuable knowledge, they inform us less about the development of depressive symptoms on an individual level.

Theoretically, IBI time series can be seen as potential biomarkers for depression. However, the hardware needed to collect such IBI time series data in ambulatory settings, combined with necessities such as wireless data collection, and online data uploading capabilities has not been available until recently. Conversely, statistical analyses commonly applied to assess and even predict mood transitions often require long-term datasets with high-resolution data, which ECG recordings do offer. Therefore, a thorough assessment and validation of wireless monitors suitable for long-term ambulatory ECG assessments is warranted. Therefore, I conducted the TRANS-ID Validation study on ECG monitors.

Gaps in the current knowledge

Currently, group-based scientific outcomes are those most reported in the scientific literature on mood disorders, such as MDD. However, it is unlikely that such studies can fully inform us on an individual person's risk (Zuidersma et al 2020). That is because with results from group-based studies researchers cannot check whether EWS really increase within persons before a change in depressive symptoms. The proposed EWS originate from ecology and related sciences, and have shown potential in their respective fields in which they were found able to foresee large changes in, for example, climate. However, an important gap in our current knowledge is whether it is even possible to use such a method within a new scientific field with its' own (data) peculiarities. As such, we hope by investigating EWS in psychiatric settings to narrow this knowledge gap down.

Another gap in our knowledge is whether the optimal hardware and software options are available to monitor, pre-process, and analyse the required data. Hence, such devices need to be validated before they can be used in clinical research while software to handle specific data and tasks might have to be developed. At the start of the research described in this thesis, long-term, high-resolution data sets with enough observations to perform the required longitudinal analyses were lacking. However, with the completion of the TRANS-ID study data collection (which collected physical activity and ECG data from over 100 participants for over four months) such high-resolution data sets finally came available, and the aforementioned gaps can be narrowed down further. In this thesis, I will focus on calculating EWS based on actigraphy and ECG time-series data from several studies (see Table 1). Studies were typically designed as between-subject studies (TRANS-ID validation study (Kunkels et al., 2021a), Bipolar study (Kunkels et al., 2021b), as well as repeated single-subject studies (TRANS-ID, (see: https://osf.io/zbwkp/ and https://osf.io/85ngu).

Aims and outline of the Thesis

In the presented studies, we aim to investigate whether actigraphy and ECG derived EWS can predict transitions in (depressive) mood symptoms, and whether available hardware and software are sufficient for this task. The first part of this thesis (chapters 2 until 5) focuses on investigating whether actigraphy-based EWS, and derived related complexity measures, can be predictive of upcoming transitions in mood symptoms. Moreover, software developed to automate various pre-processing steps of processing actigraphy data is presented. In the second part of this thesis (chapter 6 and 7), we will investigate whether IBI-

derived EWS and derived complexity measures, can be predictive of upcoming transitions in (depressive) mood symptoms.

Chapter 2: ACTman: Automated pre-processing and analysis of actigraphy data

In chapter 2, the ACTman software suite for automatically pre-processing and analysing bulk quantities of physical activity data is presented. This software suite will subsequently be used for pre-processing and analysing the collected actigraphy data sets in the subsequent studies. It also allows for the calculation of various relevant actigraphy and EWS quantifiers within a customizable moving window.

Chapter 3: Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study

In chapter 3 in a repeated single-subject design, we investigate whether EWS and spectral periodicity measures can facilitate the identifications of upcoming mood transitions in patients suffering from bipolar depression. Bipolar depression is a mental disorder wherein diagnosed patients experience distinct periods of depression, interwoven with periods of abnormally increased mood and energy (American Psychiatric Association, 2013). This sample offered unique characteristics, such as a higher change of observing changes in mood, than in unipolar depression. The participants measured their physical activity using actigraphy for over 180 days, allowing us to investigate whether we could identify upcoming transitions based on their activity.

Chapter 4: Risk Ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during anti-depressant discontinuation

In chapter 4, we investigated the effectiveness of EWS in detecting depressive mood transitions, but now in a sample of TRANS-ID Tapering participants who were tapering their anti-depressant medication. TRANS-ID data were collected specifically to test for within individual changes. Given that there were a number of participants who experienced a transition during tapering, but others did not, we were able to compare both groups to see

whether EWS perform differently in both groups. Additionally, sensitivity and specificity characteristics were checked to assess EWS performance.

Chapter 5: Complexity makes the difference: Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed

In chapter 5, actigraphy data collected for 30 days during the Mood and Movement in Daily Life (MOOVD) study were used. Group differences between depressed and non-depressed individuals on actigraphy based mean levels, circadian rhythm and complexity measures were tested. Novel complexity markers based on recurrence plots are presented and their effectiveness in detecting mood transitions are investigated.

Chapter 6: Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring

In chapter 6, two novel wireless ambulatory ECG monitors (the *Cortrium C3* and the *Ithlete finger sensor*) were investigated to assess their feasibility, validity, and reproducibility characteristics. These monitors were tested against a wired ECG reference monitor (the *VU-AMS*) to see whether obtained data was valid under various protocolled conditions. Moreover, the ability of participants to measure their ECG data in ambulatory settings, that is in real-life, was investigated.

Chapter 7: Using complexity of cardiac dynamics as a predictor of recurrence of depression

In chapter 7, we investigated whether complexity and variability indicators of cardiac dynamics decreased in the period before a transition in depressive symptoms in a sample of the TRANS-ID Tapering participants. We studied decreases in these indicators as they are known to be substantially lower in individuals suffering from depression.

Chapter 8: General summary, discussion and conclusion

In chapter 8, the main findings are summarised and are reflected upon in context of relevant literature. The potential impact of the current findings for clinical practice will be described, and recommendations for future studies are given.

Table 1: An overview of the studies and main data types presented in this thesis (more details and references given in the main text).

Study	Chapter	Main data type	Number of participants	Data collection context
ACTman	2	Actigraphy	1	Ambulatory
Bipolar study	3	Actigraphy	8	Ambulatory
TRANS-ID Tapering study	4	Actigraphy	16	Ambulatory
MOOVD study	5	Actigraphy	54	Ambulatory
TRANS-ID Validation study	6	IBI	51	Ambulatory & Laboratory
TRANS-ID Tapering study	7	IBI	14	Ambulatory

Note:IBI = inter-beat interval, TRANS-ID = TRANSitions In Depression*, MOOVD = Mood and Movement in Daily Life.*

References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)* (5th ed.). https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596

Beauchaine, T. P., Neuhaus, E., Zalewski, M., Crowell, S. E., & Potapova, N. (2011). The effects of allostatic load on neural systems subserving motivation, mood regulation, and social affiliation. *Development and Psychopathology*, *23*, 975–999. https://doi.org/10.1017/S0954579411000459

Biggs, R., Carpenter, S. R., & Brock, W. A. (2009). Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(3), 826–831. https://doi.org/10.1073/pnas.0811729106

Bos, F., 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

Breslau, N., Roth, T., Rosenthal, L., & Andreski, P. (1996). Sleep disturbance and psychiatric disorders: A longitudinal epidemiological study of young adults. *Biological Psychiatry*, *39*(6), 411–418. https://doi.org/10.1016/0006-3223(95)00188-3

Buyukdura, J., McClintock, S., & Croarkin, P. (2011). Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(2), 395-409. doi:10.1016/j.pnpbp.2010.10.019

Carney, R., Freedland, K., & Veith, R. (2005). Depression, the Autonomic Nervous System, and Coronary Heart Disease. Psychosomatic Medicine, 67, S29-S33. doi: 10.1097/01.psy.0000162254.61556.d5

Colzato, L., Jongkees, B., de Wit, M., van der Molen, M., & Steenbergen, L. (2018). Variable heart rate and a flexible mind: Higher resting-state heart rate variability predicts better task-switching. *Cognitive, Affective, & Behavioral Neuroscience, 18*(4), 730-738. https://doi.org/10.3758/s13415-018-0600-x

Choi, J., & Gutierrez-Osuna, R. (2010). Estimating mental stress using a wearable cardio-respiratory sensor. *Proceedings of IEEE Sensors*, 150–154. https://doi.org/10.1109/ICSENS.2010.5690677

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), e41010. https://doi.org/10.1371/journal.pone.0041010

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

Damiola, F., Le Minli, N., Preitner, N., Kornmann, B., Fleury-Olela, F., & Schibler, U. (2000). Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes and Development*, *14*(23), 2950–2961. https://doi.org/10.1101/gad.183500

Dekker, J. M., Crow, R. S., Folsom, A. R., Hannan, P. J., Liao, D., Swenne, C. A., & Schouten, E. G. (2000). Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes: The ARIC study. *Circulation*, *102*(11), 1239–1244. https://doi.org/10.1161/01.CIR.102.11.1239

Frankland, A., Cerrillo, E., Hadzi-Pavlovic, D., Roberts, G., Wright, A., Loo, C. K., Breakspear, M., & Mitchell, P. B. (2015). Comparing the phenomenology of depressive episodes in bipolar I and II disorder and major depressive disorder within bipolar disorder pedigrees. *Journal of Clinical Psychiatry*, *76*(1), 32–39. https://doi.org/10.4088/JCP.14m09293

Fuller, K. L., Juliff, L., Gore, C. J., Peiffer, J. J., & Halson, S. L. (2017). Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes. *Journal of Science and Medicine in Sport*, *20*, 756–760. https://doi.org/10.1016/j.jsams.2016.11.021

van Genugten, C., Schuurmans, J., Lamers, F., Riese, H., Penninx, B., & Schoevers, R. et al. (2020). Experienced Burden of and Adherence to Smartphone-Based Ecological Momentary Assessment in Persons with Affective Disorders. *Journal Of Clinical Medicine, 9*(2), 322. https://doi.org/10.3390/jcm9020322

Germain, A., & Kupfer, D. J. (2008). Circadian rhythm disturbances in depression. In *Human Psychopharmacology* (Vol. 23, Issue 7, pp. 571–585). NIH Public Access. https://doi.org/10.1002/hup.964

Gordijn, M. C. M., Beersma, D. G. M., Bouhuys, A. L., Reinink, E., & Van den Hoofdakker, R. H. (1994). A longitudinal study of diurnal mood variation in depression; characteristics and significance. *Journal of Affective Disorders*, *31*(4), 261–273. https://doi.org/10.1016/0165-0327(94)90102-3

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

Held, J., Vîslă, A., Wolfer, C., Messerli-Bürgy, N., & Flückiger, C. (2021). Heart rate variability change during a stressful cognitive task in individuals with anxiety and control participants. *BMC Psychology, 9*(1). https://doi.org/10.1186/s40359-021-00551-4

Kleiger, R. E., Miller, J. P., Bigger, J. T., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology*, *59*(4), 256–262. https://doi.org/10.1016/0002-9149(87)90795-8

Kunkels, Y., Roon, A., 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). https://doi.org/10.1111/psyp.13898

Kunkels, Y., Riese, H., Knapen, S., Riemersma - van der Lek, R., George, S., & van Roon, A. et al. (2021). Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study. *Translational Psychiatry, 11*(1). https://doi.org/10.1038/s41398-021-01465-w

Leistedt, S. J. J., Linkowski, P., Lanquart, J. P., Mietus, J. E., Davis, R. B., Goldberger, A. L., & Costa, M. D. (2011). Decreased neuroautonomic complexity in men during an acute major depressive episode: Analysis of heart rate dynamics. *Translational Psychiatry*. https://doi.org/10.1038/tp.2011.23

Macfarlane, B. (2007). Defining and Rewarding Academic Citizenship: The implications for university promotions policy. *Journal of Higher Education Policy and Management*, *29*(3), 261–273. https://doi.org/10.1080/13600800701457863

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Mallon, L., Broman, J. E., & Hetta, J. (2000). Relationship between insomnia, depression, and mortality: A 12-year follow-up of older adults in the community. *International Psychogeriatrics*, *12*(3), 295–306. https://doi.org/10.1017/S1041610200006414

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

Nazarimehr, F., Jafari, S., Perc, M., & Sprott, J. (2020). Critical slowing down indicators. *Europhysics Letters, 132*(1), 18001. https://doi.org/10.1209/0295-5075/132/18001

Nosek, B. A., & Errington, T. M. (2017). Making sense of replications. *ELife*, *6*. https://doi.org/10.7554/eLife.23383

Parker, G., Gladstone, G., & Hadzi-Pavlovic, D. (2002). Measuring psychomotor agitation by use of an actimeter. *Journal of Affective Disorders*, *72*(1), 91–94. https://doi.org/10.1016/S0165-0327(01)00426-8

Perils, M. L., Giles, D. E., Buysse, D. J., Tu, X., & Kupfer, D. J. (1997). Self-reported sleep disturbance as a prodromal symptom in recurrent depression. *Journal of Affective Disorders*, *42*(2–3), 209–212. https://doi.org/10.1016/S0165-0327(96)01411-5

Riemann, D., & Voderholzer, U. (2003). Primary insomnia: A risk factor to develop depression? *Journal of Affective Disorders*, *76*(1–3), 255–259. https://doi.org/10.1016/S0165-0327(02)00072-1

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. 10.1038/nature08227

Scheffer, M., Carpenter, S., Lenton, T., Bascompte, J., Brock, W., & Dakos, V. et al. (2012). Anticipating

Critical Transitions. Science, 338(6105), 344-348. https://doi.org/10.1126/science.1225244

Schouten, E., Dekker, J., Kok, F., Cessie, S., Van Houwelingen, H., Pool, J., & Vandnbroucke, J. (1993). Risk ratio and rate ratio estimation in case-cohort designs: Hypertension and cardiovascular mortality. Statistics In Medicine, 12(18), 1733-1745. doi: 10.1002/sim.4780121808

Shelton, R. (2001). Steps Following Attainment of Remission. *The Primary Care Companion To The Journal Of Clinical Psychiatry, 03*(04), 168-174. https://doi.org/10.4088/pcc.v03n0404

Stephan, F. K., & Zucker, I. (1972). Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions. *Proceedings of the National Academy of Sciences of the United States of America*, *69*(6), 1583–1586. https://doi.org/10.1073/pnas.69.6.1583

Stetson, M. H., & Watson-Whitmyre, M. (1976). Nucleus suprachiasmaticus: The biological clock in the hamster? *Science*, *191*(4223), 197–199. https://doi.org/10.1126/science.942799

Tölle, R., & Goetze, U. (1987). On the daily rhythm of depression symptomatology. *Psychopathology*, *20*(5–6), 237–249. https://doi.org/10.1159/000284507

Tsuno, N., Besset, A., & Ritchie, K. (2005). Sleep and depression. *Journal of Clinical Psychiatry*, *66*(10), 1254–1269. https://doi.org/10.4088/JCP.v66n1008

Vaccarino, V., Lampert, R., Bremner, J. D., Lee, F., Su, S., Maisano, C., Murrah, N. V., Jones, L., Jawed, F., Afzal, N., Ashraf, A., & Goldberg, J. (2008). Depressive symptoms and heart rate variability: Evidence for a shared genetic substrate in a study of twins. *Psychosomatic Medicine*, *70*(6), 628–636. https://doi.org/10.1097/PSY.0b013e31817bcc9e

Van Der Maas, H. L. J. (2004). *Liever een simpel model* (p. 15). Vossiuspers UvA. https://pure.uva.nl/ws/files/49527391/46201_maas.pdf

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

Van Someren, E. J. W., Swaab, D. F., Colenda, C. C., Cohen, W., McCall, W. V., & Rosenquist, P. B. (1999). Bright light therapy: Improved sensitivity to its effects on rest- activity rhythms in Alzheimer patients by application of nonparametric methods. *Chronobiology International*, *16*(4), 505–518. https://doi.org/10.3109/07420529908998724

Wagenmakers, E. J., Molenaar, P. C. M., Grasman, R. P. P. P., Hartelman, P. A. I., & van der Maas, H. L. J. (2005). Transformation invariant stochastic catastrophe theory. *Physica D, 211*, 263-276

WHO. (2009). Global Health Risks: Mortality and burden of disease attributable to selected major risks. In *Bulletin of the World Health Organization* (Vol. 87). https://doi.org/10.2471/BLT.09.070565

Wichers, M, Geschwind, N., Van Os, J., & Peeters, F. (2010). Scars in depression: is a conceptual shift necessary to solve the puzzle? *Psychological Medicine*, 359–365. https://doi.org/10.1017/S0033291709990420

Wichers, M., & de Groot, P. C. (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

Wissel, C. (1984). A universal law of the characteristic return time near thresholds. Oecologia, 65(1), 101-107. https://doi.org/10.1007/bf00384470

Winkler, D., Pjrek, E., Lanzenberger, R., Baldinger, P., Eitel, D., Kasper, S., & Frey, R. (2014). Actigraphy in patients with treatment-resistant depression undergoing electroconvulsive therapy. *Journal of Psychiatric Research*, *57*(1), 96–100. https://doi.org/10.1016/j.jpsychires.2014.06.006

Zuidersma, M., Riese, H., Snippe, E., Booij, S., Wichers, M., & Bos, E. (2020). Single-subject research in psychiatry: Facts and fictions. *Frontiers In Psychiatry*, 11. https://doi.org/10.3389/fpsyt.2020.539777

ACTman: Automated preprocessing and analysis of actigraphy data

This chapter has been published as:

Kunkels, Y. K., Knapen, S. E., Zuidersma, M., Wichers, M., Riese, H. Emerencia, A. C. (2020). ACTman: Automated preprocessing and analysis of actigraphy data. Journal of Science and Medicine in Sport, 23(5), 481-486. https://doi.org/10.1016/j.jsams.2019.11.009.

2.

Chapter 2: ACTman: Automated preprocessing and analysis of actigraphy data

Yoram K. Kunkels[1],*, Stefan E. Knapen[2,4] , Marij Zuidersma[1], Marieke Wichers[1], Harriëtte Riese[1], Ando C. Emerencia[3]

[1] University of Groningen, University Medical Center Groningen (UMCG), Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotional Regulation (ICPE), The Netherlands

[2] University of Groningen, University Medical Center Groningen, Department of Psychiatry, Research School of Behavioral and Cognitive Neurosciences(BCN), ICPE, The Netherlands

[3] University of Groningen, Department of Developmental Psychology, The Netherlands

[4] Department of Neurology, Leiden University Medical Centre (LUMC), The Netherlands

This chapter has been published as; Kunkels, Y. K., Knapen, S. E., Zuidersma, M., Wichers, M., Riese, H., & Emerencia, A. C. (2020). ACTman: Automated preprocessing and analysis of actigraphy data. *Journal of Science and Medicine in Sport, 23*(5), 481-486. https://doi.org/10.1016/j.jsams.2019.11.009.

Abstract

Objectives: To introduce a novel software-library called Actigraphy Manager (ACTman) which automates labour-intensive actigraphy data preprocessing and analyses steps while improving transparency, reproducibility, and scalability over software suites traditionally used in actigraphy research practice.

Methods: Use cases are described for performing a common actigraphy task in ACTman and alternative actigraphy software. Important inefficiencies in actigraphy workflow are identified and their consequences are described. We explain how these hinder the feasibility of conducting studies with large groups of athletes and/or longer data collection periods. Thereafter, the information flow through the ACTman software is described and we explain how it alleviates aforementioned inefficiencies. Furthermore, transparency, reproducibility, and scalability issues of commonly used actigraphy software packages are discussed and compared with the ACTman package.

Results: It is shown that from an end-user perspective ACTman offers a compact workflow as it automates many preprocessing and analysis steps that otherwise have to be performed manually. When considering transparency, reproducibility, and scalability the design of the ACTman software is found to outperform proprietary and open-source actigraphy software suites. As such, ACTman alleviates important bottlenecks within actigraphy research practice.

Conclusions: ACTman facilitates the current transition towards larger datasets containing data of multiple athletes by automating labour-intensive preprocessing and analyses steps within actigraphy research. Furthermore, ACTman offers many features which enhance user-convenience and analysis customization, such as moving window functionality and period selection options. ACTman is open-source and thus fully verifiable, in contrast with many proprietary software packages which remain a black box for researchers.

Practical implications:

- The feasibility of large scale actigraphy studies is often constrained by the associated analysis software.

- Open-source actigraphy software makes actigraphy analysis more transparent and verifiable.
- Automated batch processing of actigraphy data allows researchers to study more and longer datasets from multiple athletes in less time, while reducing human error.

Introduction

Over the years, actigraphy has gained popularity as it allows researchers to study athletes and regular participants in their own living environment, with minimal disturbance.[1] Actigraphy is considered to assess more ecological valid measures of physical activity, sleep duration, and circadian rhythm compared to subjective measures as it minimizes the risk of recall and social desirability bias.[2,3] Although polysomnography is the golden standard in sleep research,[4] its invasive nature and high costs make it an unfeasible method when studying large samples of multiple athletes in ecological valid daily life settings for longer periods of time (i.e. weeks, months). Actigraphy on the other hand, is non-invasive, relatively cheap, and easy to use.[3–5] In actigraphy studies, participants often continuously wear wireless, lightweight, accelerometers that measure (tri-axial) movement in small intervals (and thus high resolution). Such accelerometers have been validated against polysomnography (PSG) and are an accepted alternative for long-term sleep assessment during daily life.[3,4,6] The usefulness of actigraphy in studying circadian rhythms is supported by the finding that actigraphy-defined sleep-wake cycles accurately predicted sleep-wake cycles defined by simultaneously measured polysomnography.[7]

However, while accelerometer hardware is important, the software required for subsequent analyses plays an often overlooked role in determining whether the analysis of large data sets is feasible and whether the results are of sufficient quality. There is a limited number of software packages which cater towards the actigraphy researcher. The most common is the default software included with some actigraphy devices. Examples hereof are the MotionWare software, included with the MotionWatch 8 accelerometer, and the Actiware software, included with the Philips Actical and Actiwatch 2 accelerometers.[1,8] Next to the proprietary software, there are a limited number of open-source alternatives, for example: *nparACT*,[9] *acc*,[10] and *GGIR*[11] for R statistical software.[12]

There are, however, some limitations to existing software suites regarding transparency, reproducibility, and scalability. Transparency can be defined as the availability and complete observability of the source code of the software for all stakeholders. Moreover, transparency is found to be a crucial software design choice that influences stakeholder behavior.[13,14] Making the source code transparent thus makes it verifiable and auditable. Relevant and easily checkable indicators of transparency are whether the software is open-source; and whether the software has a public code repository where the full source code is available and observable.

Reproducibility is defined as being able to obtain identical numeric results at a later time point while using the same data,[15] and is regarded to be one of the defining characteristics of science.[16,17] Indeed, reproducibility and transparency are closely related as a lack of transparency hinders identification of the root causes for low reproducibility. Furthermore, a lack of reproducibility can be caused by human error. For example, consider two actigraphy researchers named Alice and Bob. Alice originally conducted a study collecting actigraphy data of one athlete for 365 days using actigraphy. Alice then uses a software suite to run the analyses which requires her to manually input the athlete's bed and wake up times from a pen-and-paper sleep-log. As Alice is very focused and precise she inputs the information from the sleep-log into the software without error and after running the analyses receives a set of results. Bob, on the other hand, read about Alice's results and decides to try to reproduce them. While Bob is also very dedicated, he unintentionally mixed up some times from the sleep-log when inputting these manually into the actigraphy software. Hence, as he tries to replicate Alice's results Bob will find deviating results meaning that he could not fully reproduce the results reported by Alice. However, this lack of reproducibility is caused by human error, not deficiencies in theory or experimental design. Given this example the importance of reducing human error to improve reproducibility becomes clear. Moreover, it was found that humans are especially susceptible for making errors during repetitive and monotonous tasks, while automation is regarded as a method to reduce such human error and thus improve reproducibility.[18,19] A relevant indicator of reproducibility is how much monotonous manual processes are automated to prevent human error.

Scalability is defined as the ability to process an increasing amount of work and elements, while facilitating effortless enlargement.[20,21] Given such a definition, a scalable process will keep performing at an acceptable level under load while the performance of a less scalable process will deteriorate, possibly causing longer turnaround times and/or

increased error rates. Moreover, less scalable processes will require large amounts of work to add improvements, while a scalable process is designed with ease of implementation in mind. Relevant indicators of scalability are, for example, the ease of adaptability of the process to interact with and support new types of hardware, and the degree to which performance under increased workload is bounded by manual processing steps.

To address these issues, we developed the Actigraphy Manager (ACTman) software-library (software repository available online[22]) for the statistical programming language R.[12] ACTman automates preprocessing and analysis of actigraphy data, thereby automatically performing a multitude of labour-intensive tasks usually done by hand. ACTman aims to improve transparency, reproducibility, and scalability over existing actigraphy software packages by offering a solution that is open-source, automates monotonous manual processes, and is easily expandable. Moreover, the data preprocessing steps in ACTman add functionalities which can benefit actigraphy researchers, for example: (1) the possibility for selecting a subset of the data from a specific time period, (2) removing tails of zero activity, (3) binning the data into 60 s epochs, (4) reformatting dates and times to a standardized format, (5) plotting 48-h or 24-h actograms, (6) offering moving window functionality, that is, being able to analyze a distinct period, for example 14 days, in a larger dataset and then moving this period forward in user-defined steps while iterating the analysis, and (7) being able to take plain marker button data and automatically convert this into a machine-readable sleep-log which can then be used in subsequent sleep analyses.

ACTman automatically preprocesses input data into a uniform file format for subsequent analyses. ACTman currently supports two actigraphs, namely the MotionWatch 8 (CamNTech), and the Actiwatch 2 (Philips Respironics), while offering a scriptable framework which can be readily extended to include other actigraphs. After the preprocessing steps ACTman applies validated sleep and circadian algorithms in order to calculate commonly used output variables. The major and novel contribution of ACTman is that it offers researchers both a comprehensive actigraphy software solution for automating necessary preprocessing steps, as well as a tool for calculating validated circadian rhythm and sleep analysis output measures. Moreover, replacing manual preprocessing by automated preprocessing reduces chances for human error.

Materials and methods

The main goal of the ACTman data preprocessing part is transforming device-specific data into generic data files (also see Fig. 1). These generic data files are then used for subsequent sleep or circadian rhythm analyses.

During preprocessing, ACTman first indexes all the actigraphy data files and sleep-logs in the specified working directory. Thereafter, the actigraphy data files are read and transformed into a data frame, out of which only relevant data columns (date, time, and activity) are extracted. Reading in the actigraphy data files requires a device-specific approach as the native layout of these actigraphy files differs per device. ACTman then checks if the data is binned in epochs of 60 s; a common epoch length required for various actigraphy analyses. If 30 s epochs are found instead, ACTman will automatically bin these into 60 s epochs. Then, the dates and times are combined and reformatted into a standardized format (e.g. 2019-01-31 12:30:00). Hereafter, a check for missing data is performed and the user is informed when the dataset contains a considerable amount of missings. If missing data is found and the user has specified to omit missing values, then these are removed in a row wise manner wherein incomplete cases are also removed. The user can also specify missings to be imputed using predictive mean matching.[23] In the next step it is checked if there is activity in the tail of the dataset, as researchers can sometimes forget to immediately stop the actigraph after receiving it from the athlete. If such chunks of trailing zeroes are found, they are removed from the dataset. Hereafter a new folder is created within the working directory wherein the processed actigraphy files are saved for use in subsequent analyses.

Regarding circadian rhythm analysis, the literature describes both parametric as well as non-parametric methods.[5] In the ACTman R-package, we chose to implement non-parametric methods as they do not require any a priori assumptions about the waveform of the activity data. Seven commonly used, well-established non-parametric variables are implemented in ACTman. These are inter-daily stability (IS; the stability of the rhythm from day to day), intra-daily variability (IV; the fragmentation of the rhythm), total activity of the ten most active (M10) and five least active hours (L5), onset times for M10, and onset times for L5. Their calculation is explained in detail elsewhere.[5,24]

Sleep measurements in actigraphy studies are often accompanied by a sleep-log, either in digital or paper-and-pencil form, in which times for going to bed and waking up are logged. ACTman automatically reads data from digital sleep-logs and combines these with

the data from the actigraph for the sleep analysis. Sleep identification is performed by calculating per epoch whether the subject is asleep or awake. To attain this, activity scores are created based on the total activity counts of the specified epoch and its direct neighbours, through the use of weighted factors specific to the epoch length.[25]

The performance of ACTman in comparison with alternative software suites will be illustrated and evaluated by presenting use case descriptions for a common actigraphy task, namely performing sleep analysis. By describing software using use case descriptions, we can systematically clarify interactions between the software and its users while also highlighting workflow.[26] Furthermore, ACTman will be evaluated regarding transparency, reproducibility, and scalability using the aforementioned indicators, while being compared to existing proprietary software. By doing so, we illustrate important inefficiencies in actigraphy processing and analyses. Short description regarding software runtimes and analysis output is available in the supplementary materials.

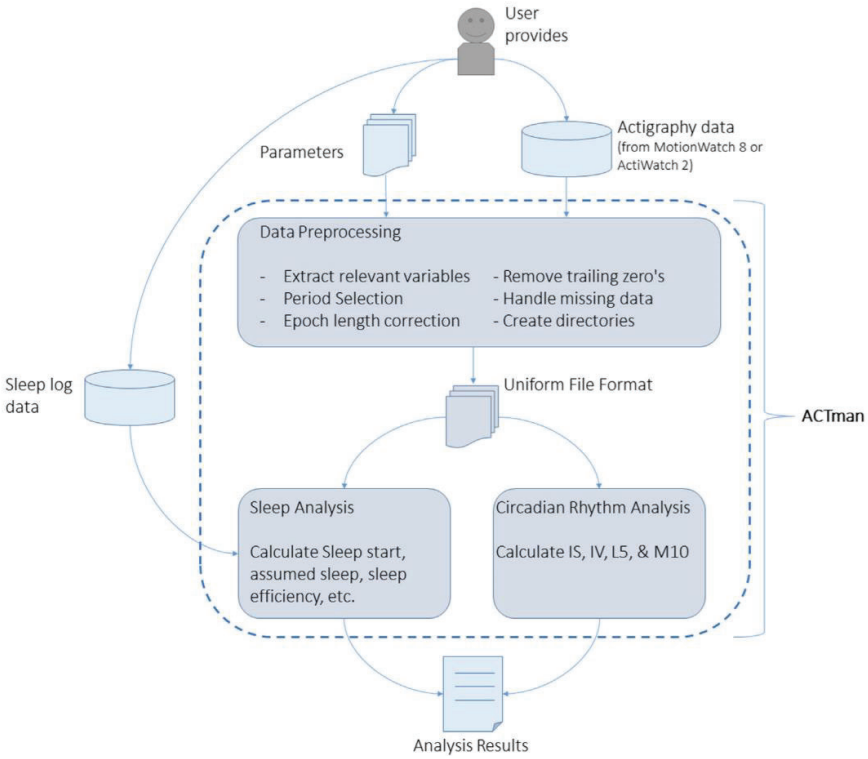


Fig. 1. Overview of ACTman information flow. The dotted line delineates the ACTman software. First, the data from either the MotionWatch 8 or the Actiwatch 2 is entered into the data-preprocessing module to obtain a uniform file format for use in subsequent analyses. Thereafter, the data in the uniform file format is used as input for the circadian rhythm analysis or the sleep analysis. If sleep analysis is required, the data from the sleep-log is read in and used in the subsequent analyses.

Results

Tables 1 and 2 show use case descriptions for performing sleep analysis in two different actigraphy software packages, respectively MotionWare and ACTman. The dataset contained multiple days, which thereby provided options for both single- and multirepetition analyses. A comparison of both tables reveals that the basic flow of the MotionWare software includes more manual steps requiring user input, which further increases as the analysis gets repeated over multiple days. Moreover, as the MotionWare software can only perform sleep analysis one day at a time, user workload will increase strongly when multiple nights of sleep are analyzed. In such a case, the user will need to repeat steps 11 till 16 until all required nights are analyzed. Such a requirement for manual involvement forms a bottleneck in facilitating actigraphy research in larger datasets as preprocessing and analysis performance is then ultimately bounded by the user performing manual tasks. Conversely, the ACTman

Table 1
Use case description of performing sleep analysis with the MotionWare software on a single dataset with multiple days.

Perform sleep analysis (in MotionWare)
Brief description: The actor preprocesses an actigraphy dataset of multiple days and performs a sleep analysis to obtain reportable results
Actors: Researcher, athlete and/or coach, student
Preconditions: The to be analyzed actigraphy files and corresponding sleep logs are available on the actor's computer
Basic flow of events:
1 The actor starts the system
2 The system authenticates the actor and starts a session
3 The actor locates and opens the actigraphy data file using the system
4 The system returns a selectable icon of the actigraphy data file
5 The actor clicks the *Sleep Analysis* button
6 The system returns a plot of every day available in the actigraphy data
7 The actor chooses *Tools > Sleep Summary Table* to open the sleep log window
8 The actor manually inputs the bed- and wakeup times of all days to be included in the analysis from the sleep diary into the system's sleep log window
9 The actor saves the changes made in the sleep log window
10 The systems returns to the plot of every day available in the actigraphy data
11 The actor selects the required period by double-clicking and holding the left mouse button while dragging and scrolling the selection tool downwards over the plotted actigraphy data
12 The system returns a summary view of the selected period
13 The actor chooses the *Sleep* option from the *Analysis Function* menu
14 The actor manually drags two markers, indicating bed- and wake up time, over the plot of the actigraphy data to the times corresponding to those in the sleep log
15 The system calculates and returns sleep analysis results
16 The actor transfers the sleep analysis results by copy-pasting the results into a text editor, or writing them down
17 The actor leaves the system
Extensions:
3a. The system cannot authenticate the actor
• The system informs and halts the actor
8a. Bed- and wakeup times from sleep diary are not available.
• Sleep analysis cannot be correctly performed and current night has to be skipped
16a. The actor requires the analysis of a longer period than the one day processed and analyzed up until now
• The actor repeats steps 11 till 16 until the whole required period is preprocessed and analyzed.
Post-conditions: Results are outputted on-screen and have to be transferred to another software program for saving and formatting
Special requirements: The actor has acquired the required licenses and registration to use the software

Table 2
Use case description of performing sleep analysis with the ACTman software on a single dataset with multiple days.

Perform Sleep analysis (in ACTman)
Brief description: The actor preprocesses an actigraphy dataset of multiple days and performs a sleep analysis to obtain reportable results
Actors: Researcher, athlete and/or coach, student
Preconditions: The to be analyzed actigraphy files and corresponding sleep logs are available on the actor's computer
Basic flow of events:
1 The actor starts the system
2 The system starts a session
3 The actor specifies the path where the actigraphy files are located
4 The actor runs the ACTman script
5 The system preprocesses the data into a uniform file format
6 The system reads in the sleep diary
7 The system performs the sleep analysis
8 The system returns the sleep analysis results in a dedicated spreadsheet file
9 The actor leaves the system
Extensions:
4a. The actor requires the sleep analysis of a longer period than one day.
• When the corresponding user-defined parameters are provided, the system automatically performs the required analysis over multiple days.
6a. Bed- and wakeup times from sleep diary are not available.
• The system searches for a marker button file instead and, when available, uses that to generate the sleep log, otherwise the system informs the actor
Post-conditions: A preprocessed dataset is created and used to perform the sleep analysis. Results are outputted in a dedicated results file and on-screen
Special requirements: none

package does not require additional manual steps for every extra night of sleep, as it automates such preprocessing steps. Regarding reproducibility, we identified that automation of monotonous manual processes decreases human error and facilitates better reproducibility rates. The aforementioned use case descriptions have furthermore shown that the ACTman workflow does automate processes in actigraphy research that previously required large amounts of monotonous manual labour. Especially when compared to steps 11 till 16 in the MotionWare use case description, we can see that the ACTman package substantially alleviates the workload for the actigraphy researcher. As such, by using ACTman, actigraphy researchers can decrease chances for human error which facilitates better reproducibility rates.

For transparency it is important that code is available and observable, hence we evaluated transparency by reporting whether the compared software packages are open-source and whether there is a repository where stakeholders can find the full source code. As making a software package open-source instead of proprietary is often a dedicated design choice, both the ACTman package as well as alternatives such as *nparACT*[9] are necessarily transparent in this sense. For example, both have easy to find repositories where the source code is available for stakeholders.[22,27] Conversely, proprietary actigraphy software is by definition not as transparent, thereby obscuring the exact workings of the software from stakeholders. Such obscurity can have substantial ramifications as it could take a long time before bugs in calculations are caught, if ever. The open-source actigraphy software packages, on the other hand, are fully verifiable, even offering users to make their own improvements and expansions of the software, where needed.

When considering scalability, the ability of a software package to accommodate an increasing amount of work and to support new types of hardware were identified as important characteristics. Here, software suites which require user input in many of the processing steps are necessarily limited in the amount of actigraphy files they can process within a set time period. Hence, scalability of these software suites is ultimately bounded by how quick a researcher can manually process actigraphy files without error; thus constituting a human-bound scaling limit. As such, the ACTman package, which automate these manual steps, is more scalable as its performance under increased workload is not bounded by manual processing steps. Another feature of scalability in actigraphy software is being able to support multiple types of actigraphy hardware. Here, proprietary software suites are at a disadvantage as they only offer dedicated support for one device, or in the best case, for multiple related

devices from the same manufacturer. In this case, open-source packages could theoretically offer better scalability then proprietary ones as they are not per se constrained to one type of actigraph device or manufacturer. However, many opensource actigraphy packages do cater only to one specific device (e.g., the *"ActivityIndex"* R-package[28]) or require actigraphy data which is already preprocessed into a specific data format (e.g., the *"nparACT"* R-package[9]), thereby limiting their scalability. ACTman on the other hand, preprocesses actigraphy files into a uniform format which is then used for subsequent analyses. This allows for easy support of other actigraphy devices as adding support for a new device would only require adding a module to read the device output files and transform it to the uniform file format. This not only future-proofs the ACTman package, but also allows for heterogeneous test environments of different actigraph devices, in which the processing of data is guaranteed to be the same.

Discussion

In this paper we introduced ACTman; a R-library which automates various data preprocessing steps and performs circadian rhythm and sleep analysis. ACTman is thereby able to quickly process data from two actigraphs, the MotionWatch 8 and the Actiwatch 2, into a generic file format which are then analyzed. The ACTman package thereby tackles existing issues in actigraphy research regarding transparency, reproducibility, and scalability. By offering transparent source code and a public repository, ACTman guarantees stakeholders that the code is both auditable and verifiable. This is in contrast with many proprietary software packages which obscure their source code and thus prevent stakeholders from identifying possible issues. Furthermore, as ACTman is open-source it allows stakeholders to use and modify the ACTman source code for their own specific goals and contexts. Hence, the ACTman code could thus serve as a stepping stone for other developers to create novel actigraphy applications.

Additionally, we identified an important bottleneck in actigraphy research; namely that with common actigraphy software, quality and speed of both preprocessing and analysis is ultimately constrained by user performance in conducting specific manual tasks. This bottleneck was exemplified by presenting a use case description of actigraphy software which requires a large amount of manual involvement, and the ACTman R-package which automates most of such manual tasks. ACTman thereby also tackles an important threat to

producing reproducible results, namely human error, which is especially common during repetitive and monotonous tasks, such as those normally required when processing and analyzing actigraphy data. Automating such repetitive and monotonous steps also improves scalability of actigraphy research, which can support the current trend in actigraphy research towards longer datasets including large numbers of athletes measured over extended time periods. Moreover, as ACTman transforms data into a generic data format prior to analyses, it is receptive for extensions to include various other actigraphs. Moreover, this design offers researchers heterogeneous test environments for different actigraphs, wherein data processing steps are guaranteed to be the same. In addition to these main points, ACTman offers actigraphy researchers a suite of additional features, including: comprehensive moving window functionality with a large degree of user customization, the ability to convert marker button presses to a sleep-log, and actigraphy plotting options.

There are, however, some limitations of the current work. For example, proprietary software packages often include extensive graphical user interfaces (GUI), which are easy to use for novel users, whereas open-source software sometimes lack refined user interfaces. This is also the case for ACTman, especially when compared to software such as MotionWare. However, as the R programming language is commonly used amongst sport researchers and data analysts, and offers extensive community tech support, this is not expected to deter its use. Moreover, R-users routinely use R through integrated development environments, such as R studio, which add functional user interfaces. A second limitation is that ACTman currently only offers calculation of non-parametric circadian rhythm variables, whereas parametric methods such as acrophase, MESOR (Midline Statistic of Mean), and cosinor analysis could bring complementary information. However, we prioritized non-parametric methods as they do not require any assumptions on the waveform underlying the rest-activity circadian rhythm. In summary, the current paper presented the ACTman R-library, which aims at facilitating large scale actigraphy research with the MotionWatch 8 (CamNTech) and the Actiwatch 2 (Philips Respironics) actigraphs. ACTman realizes this by taking away an important hurdle in actigraphy data processing, namely its current labour intensity, while addressing transparency, reproducibility and scalability issues. ACTman thereby offers sport and actigraphy researchers a software solution that improves on many common inefficiencies in current actigraphy software libraries.

Appendix A. Supplementary data Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.jsams.2019.11.009

References

1. Fuller, K., Juliff, L, Gore, C. et al. (2017). Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes. *Journal of Science and Medicine in Sport* ; *20*(8):756–760. http://dx.doi.org/10.1016/j.jsams.2016.11.021.

2. Evenson, K., Catellier, D., Gill, K. et al. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences*; *26*(14):1557–1565. http://dx.doi.org/10.1080/02640410802334196.

3. Kaplan, K., Talbot, L., Gruber, J. et al. (2012). Evaluating sleep in bipolar disorder: comparison between actigraphy, polysomnography, and sleep diary. *Bipolar Disorder; 14*(8):870–879. http://dx.doi.org/10.1111/bdi.12021.

4. Littner, M., Hirshkowitz, M., Kramer, M. et al. (2003). Practice parameters for using polysomnography to evaluate insomnia: an update. *Sleep*; *26*(6):754–760. http://dx.doi.org/10.1093/sleep/26.6.754.

5. van Someren, E., Swaab, D., Colenda, C. et al. (1999). Bright light therapy: improved sensitivity to its effects on rest-activity rhythms in Alzheimer patients by application of nonparametric methods. *Chronobiology International*; *16*(4):505–518. http://dx.doi.org/10.3109/07420529908998724.

6. Ancoli-Israel, S., Cole, R., Alessi, C. et al. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*; *26*(3):342–392. http://dx.doi.org/10.1093/ sleep/26.3.342.

7. Pollak, C., Tryon, W., Nagaraja, H. et al. (2001). How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep*; *24*(8):957–965. http://dx.doi.org/10.1093/sleep/24.8.957.

8. Rosenkranz, R., Weber, C., Rosenkranz, S. (2010). Validity of the actical accelerometer step count function in children. *Journal of Science and Medicine in Sport*; *13*:e97. http://dx.doi.org/10.1016/j.jsams.2010.10.667.

9. Blume, C., Santhi, N., Schabus, M. (2016). 'nparACT' package for R: a free software tool for the non-parametric analysis of actigraphy data. *MethodsX; 3*:430–435. http://dx.doi.org/10.1016/j.mex.2016.05.006.

10. Song, J., Cox, M. (2016). R package version 1.3.3. Retrieved from: acc: Exploring Accelerometer Data, https://CRAN.R-project.org/package=acc.

11. van Hees, V. R package version 1.5-9. Retrieved from: GGIR: Raw Accelerometer Data Analysis, 2017 https://CRAN.R-project.org/package=GGIR.

12. R Core Team, Retrieved from: R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2017. https://www.R-project.org/ .

13. Brunswicker, S., Jensen, B., Song, Z. et al. (2018). Transparency as design choice of open data contests. *Journal of the Association for Information Science and Technology; 69*(10):1205–1222. http://dx.doi.org/10.1002/asi.24033.

14. West, J., O'mahony, S. (2008). The Role of Participation Architecture in Growing Sponsored Open Source Communities. *Ind Innovation*; *15*(2):145–168. http://dx.doi.org/10.1080/13662710801970142.

15. Epskamp, S. (2019). Reproducibility and Replicability in a Fast-Paced Methodological World. *Advances in Methods and Practices in Psychological Science*; *2*(2):145–155. http://dx.doi.org/10.1177/2515245919847421.

16. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*; *349*(6251). http://dx.doi.org/10.1126/science.aac4716.

17. Pauliuk, S., Majeau-Bettez, G. et al. (2015). Lifting industrial ecology modeling to a new level of quality and transparency: a call for more transparent publications and a collaborative open source software framework. *Journal of Industrial Ecology*; *19*(6):937–949. http://dx.doi.org/10.1111/jiec.12316.

18. Jessop-Fabre, M., Sonnenschein, N. (2019). Improving reproducibility in synthetic biology. *Frontiers in Bioengineering and Biotechnology* ; *7*. http://dx.doi.org/10.3389/fbioe.2019.00018.

19. Yeow, J., Ng, P., Tan, K. et al. (2014). Effects of stress, repetition, fatigue and work environment on human error in manufacturing industries. *Journal of Applied Sciences*; *14*(24):3464–3471. http://dx.doi.org/10.3923/jas.2014.3464.3471.

20. Bondi, A. Characteristics of scalability and their impact on performance, *Proceedings of the Second International Workshop on Software and Performance — WOSP 2000*, ACM Press, 2000, p. 195–203. http ://dx.doi.org/10.1145/350391.350432.

21. Gupta, A., Christie, R., Manjula, P., R. (2017). Scalability in internet of things: features, techniques and research challenges. *International journal of computational intelligence research*; *13*(7):1617–1627.

22. Kunkels, Y., K., Knapen, S., E., Emerencia, A. C., Retrieved from Compsy/ACTman, 2018 https://github.com/compsy/ACTman.

23. Buuren, S., Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* ; *45*(3). http://dx.doi.org/10.18637/jss.v045.i03.

24. Witting, W., Kwa, I., Eikelenboom, P. et al. (1990). Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. *Biological Psychiatry* 1990; *27*(6):563–572. http://dx.doi.org/10.1016/0006-3223(90)90523-5.

25. Camntech, Retrieved from: https://www.camntech.com/bulletins (user login required) Information Bulletin No. 3 (information bulletin no. 3 sleep algorithms.docx), 2013.

26. Zhu, Q., Nakata, T., Mine, M. et al. (2004). System-on-chip validation using UML and CWL, *Proceedings of the 2nd IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, IEEE, 2004, p. 92–97. http://dx.doi.org/10.1145/1016720.1016745.

27. Blume, C., Santhi, N., Schabus, M., Retrieved from cran/nparACT, 2016 https://github.com/cran/nparACT.

28. Bai, J., Di, C., Xiao, L. et al. (2017). An activity index for raw accelerometry data and its application in older adults. *Innovation Aging*; *1*(Suppl. 1). http://dx.doi.org/10.1093/geroni/igx004.4497, 1239-1239.

3.

Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: an actigraphy study

This chapter has been published as:

Kunkels, Y. K., Riese, H., Knapen, S. E., Riemersma - van der Lek, R. F., George, S. V. van Roon, A. M., 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). https://doi.org/10.1038/s41398-021-01465-w

Chapter 3: Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study

Yoram K. Kunkels, MSc[1*], Harriëtte Riese, PhD[1], Stefan E. Knapen, PhD[1, 3], Rixt F. Riemersma - van der Lek, PhD[1], Sandip V. George, PhD[1], Arie M. van Roon, PhD[2], Robert A. Schoevers, PhD[1], Marieke Wichers, PhD[1]

[1] University of Groningen, University Medical Center Groningen, Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), The Netherlands.

[2] Department of Vascular Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[3] Department of Neurology, Leiden University Medical Centre (LUMC), Leiden, The Netherlands. *

This chapter has been published as; Kunkels, Y. K., Riese, H., Knapen, S. E., Riemersma - van der Lek, R. F., George, S. V., van Roon, A. M., 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). https://doi.org/10.1038/s41398-021-01465-w

Abstract

Early-warning signals (EWS) have been successfully employed to predict transitions in research fields such as biology, ecology, and psychiatry. The predictive properties of EWS might aid in foreseeing transitions in mood episodes (i.e. recurrent episodes of mania and depression) in bipolar disorder (BD) patients. We analyzed actigraphy data assessed during normal daily life to investigate the feasibility of using EWS to predict mood transitions in bipolar patients. Actigraphy data of 15 patients diagnosed with BD Type I collected continuously for 180 days were used. Our final sample included eight patients that experienced a mood episode, three manic episodes and five depressed episodes. Actigraphy data derived generic EWS (variance and kurtosis) and context-driven indices (autocorrelation at lag-720) were used to determine if these were associated to upcoming bipolar episodes. Spectral analysis was used to predict changes in the periodicity of the sleep/wake cycle. The study procedures were pre-registered. Results indicated that in seven out of eight patients at least one of the EWS did show a significant change-up till four weeks before episode onset. For the generic EWS the direction of change was always in the expected direction, whereas for the context-driven indices the observed effect was often in the direction opposite of what was expected. The actigraphy data derived EWS and spectral analysis showed promise for the prediction of upcoming transitions in mood episodes in bipolar patients. Further studies into false positive rates are suggested to improve effectiveness for EWS to identify upcoming bipolar episode onsets.

Introduction

Patients diagnosed with bipolar disorder (BD) suffer from recurrent episodes of depression and mania, interchanged with stable or euthymic periods[1] . Rapid transitions in mood, behavior, psychomotor agitation, and sleep may occur [2 – 4] with a debilitating impact on patients and their families. One of the key goals of treatment is to maintain euthymic state and prevent relapse. Ideally, treatment is tailored to counter upcoming symptom transitions but both patients and clinician are often late in signalling that a new episode is developing. A promising approach to foresee upcoming transitions comes from complex systems literature, which suggests that a set of generic early-warning signals (EWS) could identify whether resilience to change is declining. Such decreasing resilience can suggest that a transition from one state to another is forthcoming, for example, in global financial markets, biological phenomena, and ecological systems [5] .

Examples of commonly used EWS include rising variance, autocorrelation, and kurtosis[5–7] . Variance indicates how much values of interest deviate from the mean and each other, autocorrelation indicates how similar a variable is to a delayed copy of itself, and kurtosis informs on the shape of the probability distribution. It was shown that these three EWS substantially increased right before a transition [5] . As EWS are generic they do not depend on specific contexts or topics, and are expected to perform similarly within different complex systems. If the typical BD characteristics of transitions towards either manic or depressive episodes also behave as a complex system, detection of EWS may provide a new approach to foresee these transitions. In this conceptualization, we expect the transitions towards a bipolar episode are indicated by the proximity of so-called tipping points. At these tipping points critical slowing down is expected to occur; meaning that the dominant eigenvalue corresponding to recovery rate will go to zero [5] . Such events can —under the right circumstances —be identified by declining resilience indicators, such as the aforementioned EWS.

Practically, we would thus expect to find increased variance, autocorrelation and kurtosis, before the onset of a bipolar episode. Moreover, we speculate that mean physical activity levels could help differentiate between whether the episode change predicted by EWS is either manic or depressive in nature. We expect to find higher levels of physical activity before patients enter a manic episode and lower activity levels before a depressed episode.

Besides EWS, other more content-specific indices might also facilitate prediction of upcoming episodes, such as psychomotor-oriented indices. A growing number of studies are unravelling the deregulation of the circadian sleep/wake cycle by analyzing psychomotor agitation and sleep disturbances -symptoms typically seen in Bipolar patients [2,8] . For example, Bipolar patients showed increased sleep duration, less daytime sleep, and larger contrast between day and night activity during their euthymic period compared to their manic/mixed episodes[8,9] . Moreover, compared to controls, Bipolar patients showed a more fragmented sleep/wake cycle as indicated by more variability within days and less stability over multiple days in their actigraphy data[10]. A related finding is that variability in the sleep/wake cycle can be indicative for the upcoming onset of a depressive episode among bipolar patients during their euthymic periods[11]. A analysis method well suited for studying such changes in the sleep/wake cycle is spectral analysis; a technique wherein the variation in the time domain of time-series data is decomposed in their respective frequencies[12]. Consider, for example, actigraphy data of a healthy individual. Healthy humans have a near 24-h sleep/wake cycle, so their spectrum will show the largest peak close to the 24-h frequency even under atypical circumstances such prolonged isolation from natural external Zeitgebers[13]. Hence, deviations from the 24-h frequency may indicate disturbances in the sleep/wake cycle, as for example a 48- h sleep/wake cycle was detected prior to a transition from a depressed state into a manic state in bipolar patients[14] . Disturbances in spectral periodicity are expected to be indicative of increased risk for transition to a manic or depressive episode as they signal dysregulation of the sleep/wake cycle; a feature observed in bipolar patients.

Given the impact of disturbances in activity and the sleep/wake cycle on mood episodes in bipolar patients, actigraphy is especially suited to investigate such upcoming transitions. However, while the aforementioned studies on this topic offer useful hypothesis generating information, only short-term actigraphy time-series data or questionnaires were used. Moreover, none of the aforementioned studies included live transitions in mood episodes while being monitored with actigraphy. As such, prior conclusions can be considered to be limited regarding the effects of sleep/wake disturbances on mood transitions. To study whether changes in EWS predict upcoming manic or depressive episodes in bipolar patients, a study design in which patients are monitored for multiple weeks is required. Actigraphy can be employed to continuously assesses time-series data of physical activity from which onsets of episodes could be predicted. Such continuous activity measurements

are better suited than short-term activity measurements to fully capture some of the hallmark symptoms of BD, namely the disturbances in activity and sleep/wake cycles over time. In actigraphy studies, patients wear lightweight, wireless, wrist-worn accelerometers which measure (triaxial) movement. Actigraphy is relatively simple and has been successfully validated against polysomnography for predicting sleep/wake cycles[15]. As such, its ease of use and its objective nature make actigraphy well suited for monitoring patients with bipolar disorder.

To investigate how transitions towards the onset of either a depressive or manic episode in Bipolar patients can be predicted, we will examine whether declining resilience —as indicated by the EWS —can help anticipate such transitions. Moreover, we will investigate whether disturbances in the periodicity of the sleep/wake cycle can also aid in predicting such transitions. Actigraphy data collected for six months by Bipolar patients will be used. We hypothesise that in the period before the onset of either a manic or depressive episode, actigraphy patterns of Bipolar patients will show: (1) rising variance, (2) rising kurtosis, (3) rising autocorrelation in actigraphy activity patterns, and (4) spectral indices indicating disturbances in the typical 24-h wake/sleep cycle. Moreover, we hypothesize that the period before an episode patients will show: (5) mean activity levels congruent with the type of episode (i.e., finding higher mean levels of activity before a manic episode, and lower mean levels of activity before a depressive episode). Our study was pre-registered, meaning that we disclosed our hypotheses and analysis plans before conducting the study and before looking at the data for meaningful patterns, thereby optimizing transparency and replicability. Moreover, we endeavoured to make our materials and procedures as open as possible by storing these on a publicly accessible repository (see: https://osf.io/63d8w/).

Materials and methods

Sample

An existing dataset of patients diagnosed with BD type I was used for the current study. Details of the current data collection are described elsewhere[16]. Patients were mainly enrolled from the outpatient clinic of the University Center Psychiatry (UCP) within the University Medical Center Groningen (UMCG), and secondarily from the Dutch patient society (“Plusminus”). The inclusion criteria were: (1) diagnosed with bipolar disorder type I, (2) having suffered from at least 1 episode in the past 2 years, and (3) being motivated to

participate in a long-term study. The exclusion criteria were: (1) suffering from somatic diseases which could interfere with the actigraphy measurements, and (2) suffering from somatic sleep disorders, such as for example sleep apnea. Patients participated in 180 days of mood monitoring once a day, weekly symptom monitoring, and continuous activity monitoring. In total, 15 patients provided informed consent for their participation in the study. Of these 15 patients, one dropped out during the study due to personal reasons, whilst the data of one patient was found to be largely incomplete and had to be excluded as well. Of the 13 remaining patients, 11 did experience a transition towards a mood episode (4 experienced a manic episode and 7 experienced a depressive episode). The two patients who did not show any transition towards a mood episode during the study period were excluded. Lastly, of the remaining 11 patients three experienced so many mood symptoms that their designated “euthymic ” episode was not evidently euthymic anymore. These patients were thus excluded from the analyses. Eight patients were included in the final sample.

Symptom indices

Depressive and manic episodes were de fined with validated questionnaires. Every week patients filled out the Inventory of Depressive Symptomatology - Self Rating (IDS-SR[17]), and the Altman Self Rating Scale for Manic symptoms (ASRM[18]). In order to be designated as being in a manic episode, patients had to score higher than 5 points on the ASRM for two consecutive weeks, assuring that there is at least one full week of manic symptoms present. The criteria of scoring at least 5 points on the ASRM for two weeks was included because ASRM ratings can reflect the mental state on the day of completing the ASRM more than the mental states of several days before[19]. Depressive episodes, on the other hand, required patients to maintain a score higher than 25 points on the IDS-SR for at least three consecutive weeks, assuring that there are at least two full consecutive weeks of symptoms. A transition was de fined as starting to fulfil the aforementioned criteria for depressive and manic episodes. For our analyses, we studied the first transition of each patient.

Activity

Actigraphy time-series data were collected with a wrist-worn MotionWatch 8 (MW8; CamNTech) actigraph. Patients were instructed to wear the MW8 continuously, only removing the device under rare conditions, such as sauna visits where a combination of high humidity and temperature could induce technical difficulties. The MW8 was initialized to use one-minute epoch lengths, no data compression, and no light detection. An electronic sleep diary was filled-out every morning.

Analyses

Pre-processing and analysis of the data was performed in the statistical programming language R[20]. For preprocessing and analyzing the actigraphy data, the R package ACTman[21] was used. Generic EWS such as autocorrelation and kurtosis, are sensitive to detrending [7] . Therefore, we removed linear trends from the data by calculating the least squares regression line to estimate the growth rate, and then subtract differences from the least squares fit line from the data. For the spectral analysis, we calculated the spectral periodogram with a fast Fourier transformation without smoothing in R statistical software[20].

EWS indices were calculated from 1 min actigraphy time-series averages for each participant independently. Variance and kurtosis were calculated over the minute-level actigraphy data in a moving window. We used a window size of 7 days which means that every window includes at least one Saturday and one Sunday, thereby equalizing any possible effects from weekend days. The size of the steps at which the moving window was moved over the data was set to one day; the algorithm first calculated the EWS for the first 7-day window, then moved the window 1 day ahead, and repeated these steps until the last full window was calculated. Autocorrelation indicates how much a variable correlates with itself at a later lagged instance of itself. Based on conceptual reasoning, we chose to investigate autocorrelation at lag 720 min (acf-720). Autocorrelations at lag-720 are informative about how activity is correlated to the amount of activity 12 h (=720 min) earlier: for example, the correlation between the amount of activity at 12:00 to the amount of activity at 00:00 midnight. The autocorrelation of activity separated by 12 h is expected to be negative, whereas autocorrelation of activity separated by 24 h is expected to be positive. However, if the normal sleep/ wake cycle gets deregulated and the contrast between sleep and waking hours diminishes, we expect acf-720 to approximate zero or even positive values; either due

to an individual becoming more restless during sleeping hours (an expected symptom of a manic episode), or by getting less active during waking hours (an expected symptom of a depressive episode).

We investigated whether strong increases in EWS occur up till four weeks before the onset of an episode. This four-week period was chosen to allow for a plausible extent of time for increases in EWS to develop form. To test whether the increase in EWS is significant the Mann-Kendall test[22,23], a commonly used non-parametric test for detecting significant monotonic trends in time-series data, was used. In order to estimate disturbances in spectral periodicity we examined the ratio between the fundamental and second harmonic frequencies. Here, the fundamental frequency represents the lowest frequency of a periodic waveform, whereas the harmonic frequencies are frequencies that operate at (whole-number) multiples of the fundamental frequency. The ratio between these two estimates the likeliness that a disturbance in periodicity is afoot as it indicates the strength of the fundamental 24-h frequency versus the strength of an alternative frequency (the 12-h frequency in this case). The ratio between the fundamental and second harmonic frequencies was calculated by dividing the power spectral density value of the second harmonic by the same the fundamental frequency. Harmonic frequencies were calculated by one divided by n times the fundamental period. Lastly, for the mean activity level analyses we considered 7-day periods around the onset of the episode, operationalized as starting one week before episode onset. These mean levels were then compared with mean activity levels 7-day periods from the euthymic period.

TABLE 1

OUTCOMES OF THE MANN-KENDALL TREND TESTS FOR EWS ASSESSMENT

ID	Episode type	Early warning signal	z-scores	N	p-values	Direction
1	D	Variance	273	28	0.001**	increase
1	D	Kurtosis	236	28	0.066	increase
1	D	Autocorr_lag720	1.663	28	0.096	increase
2	M	Variance	214	28	0.336	increase

2	M	Kurtosis	203	28	0.597	increase
2	M	Autocorr_lag720	-2.736	28	0.006**	decrease
3	D	Variance	212	28	0.377	increase
3	D	Kurtosis	139	28	0.05	increase
3	D	Autocorr_lag720	-5.719	28	< 0.001**	decrease
4	M	Variance	240	28	0.045*	increase
4	M	Kurtosis	95	28	0.001**	increase
4	M	Autocorr_lag720	2.894	28	0.004**	increase
8	M	Variance	109	28	0.001**	increase
8	M	Kurtosis	243	28	0.034*	increase
8	M	Autocorr_lag720	-2.341	28	0.019*	decrease
9	D	Variance	194	28	0.860	increase
9	D	Kurtosis	238	28	0.055	increase
9	D	Autocorr_lag720	-3.935	28	< 0.001**	decrease
11	D	Variance	107	21	0.929	increase
11	D	Kurtosis	119	21	0.420	increase
11	D	Autocorr_lag720	-0.799	21	0.424	decrease
15	D	Variance	297	28	0.001**	increase
15	D	Kurtosis	111	28	0.002**	increase
15	D	Autocorr_lag720	-1.490	28	0.136	decrease

Note: **: $p <= 0.01$; * $p <= 0.05$; episode type indicates whether a depressive episode (D) or a manic episode (M) occurred; ID's are numbered according to their original ID designations to facilitate comparison between studies.

Results

Early-warning signals

Results are presented in Table 1, and plots of the EWS for all patients can be found on an online repository (see: https://osf.io/63d8w/). An example of an EWS which significantly increased before an episode is given in Fig. 1. In one patient (ID 11) the transition occurred so early during data collection that only three weeks of data are available before the transition. Mann-Kendall trend tests showed significant trend increases in EWS up till four weeks before the onset of either a manic or depressive episode in seven patients. In two patients, their episode onset was preceded by significant changes in all three EWS. In five patients, their episode onset was preceded by significant changes in acf-720 only. In one patient, the episode onset was preceded by significant increases in variance only. Lastly, in one patient, the episode onset was preceded by significant increases in both variance and kurtosis. However, although the direction of the change was always in the expected direction for variance and kurtosis (i.e. increasing), for acf-720 it significantly decreased in four patients preceding episode onset, while in one patient it significantly increased. Finding effects in the opposite direction for acf-720 might suggest that it operates more like a general measure of instability, wherein any change from the normal rhythm is indicative of an upcoming episode, not necessarily only increases. Post-hoc analyses using Fisher's test for combined pvalues were performed for each EWS to investigate whether overall significant trends were present. By combining the p-values of each EWS, outcomes are less dependent on the individual patient and should thus be more generalizable. Fisher's test for combined p-values is performed by taking the p-values for one indicator and from this calculate the chi-values. Lastly, taking one minus the calculated chi-squared cumulative density distribution using a transformation to N(0,1) will yield the required combined group p-values. Group p-values obtained via this method were (1) Variance, group pvalue < 0.001, (2) Kurtosis, group p-value < 0.001, (3) Acf720, group p-value < 0.001. This suggests that the individual quantifiers were significant overall, and that the three quantifiers are reliable indicators of upcoming transitions.

Disturbances in periodicity

In one patient (no. 11) the onset of an episode was preceded by the hypothesized rhythm transition and the increased frequency ratio associated with it. This change in

periodicity was observed nine days before the start of the episode. During that day the ratio between the fundamental frequency and its second harmonic reached a value of 0.98, which indicates an upcoming change from a 24-h rhythm to an atypical 12-h rhythm, as presented in Figs. 2 and 3. None of the other patients showed the hypothesized change.

Post-hoc exploratory analysis of the spectral results showed that one other patient (no. 3) also showed similar disturbances in periodicity, albeit from a 24 h rhythm towards an atypical 4-h rhythm. This change in periodicity preceded episode onset by 28 days. Moreover, next to the aforementioned rhythm transition of patient 11, this patient showed another transition wherein the end of the euthymic period was preceded by a change in periodicity. Here, the periodicity change preceded the end of the euthymic period by 17 days, while the ratio between the fundamental frequency and the second harmonic reached a value of 0.96. Here the ratio between the fundamental frequency and the second harmonic already approximated one, five days before the end of the episode. From these results it may be concluded that the used spectral periodicity indices act more as a general indicator of instability in mood than a specific indicator for one particular type of transition.

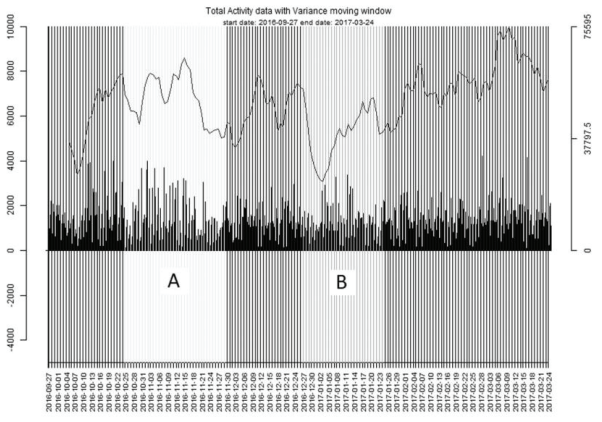


Fig. 1 Variance marker of patient 1, calculated from a moving window of the activity data (see method section for details). The left set of light-gray vertical lines indicated with an "A" represent the euthymic period. The right set of light-gray vertical lines indicates with a "B" represent the episode period. The y-axis shows the variance of the activity while the x-axis represents time in days.

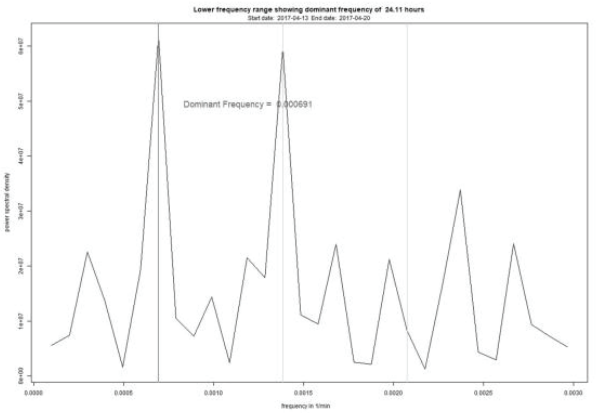


Fig. 2 Frequency spectrum of patient 11 showing the dominant frequency (dark gray) and its two subsequent harmonic frequencies (light gray), one day before transitioning towards an atypical 12-h rhythm. The y-axis shows the power spectrum density in U^2 while the x-axis shows the frequency in 1/min.

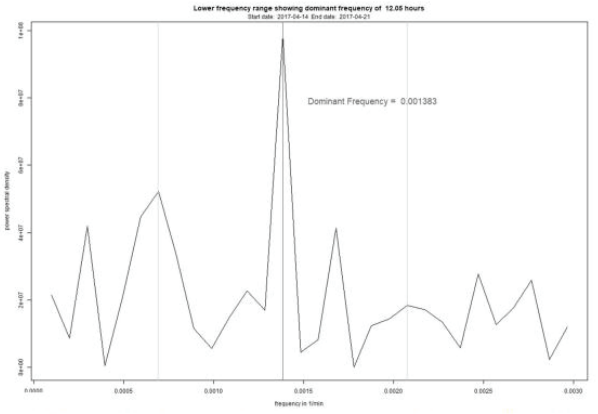


Fig. 3 Frequency spectrum of patient 11 showing the new dominant frequency (dark gray) after transitioning from a typical 24-h rhythm to an atypical 12-h rhythm. The y-axis shows the power spectrum density in U^2 while the x-axis shows the frequency in 1/min.

Mean activity levels

The results of mean level changes in anticipation of the de fined transition episodes are presented in Table 2. All patients who developed a depressive episode showed higher activity levels during the first seven days before their episode when compared to the first seven days of their euthymic period. One of the three patients who developed a manic episode showed less activity during the first seven days before entering a manic episode when compared to the first seven days of their euthymic period. In six of the eight patients mean activity levels did not develop in the expected direction and thus, did not contribute to the differentiation whether an upcoming episode is manic or depressive in nature. Post-hoc exploratory analyses

were performed to ascertain that the aforementioned results were not caused by artefacts due to mean levels of activity during episodes that were in the expected direction. After correcting for this, five of the eight patients showed an effect in the expected direction; three showed less activity during their depressed episode than in their euthymic period, and two showed more activity during their manic episode than in their euthymic period (see Table 3).

TABLE 2

MEAN ACTIVITY DURING SEVEN DAY PERIODS BEFORE EPISODES AND DURING EUTHYMIC PHASES

ID	Episode type	Mean euthymic	Mean episode	Interpretation
1	D	96.111	139.135	More activity before episode
2	M	153.808	124.585	Less activity before episode
3	D	100.444	109.135	More activity before episode
4	M	51.167	61.670	More activity before episode
8	M	91.874	92.871	More activity before episode
9	D	171.252	211.028	More activity before episode
11	D	186.168	192.013	More activity before episode
15	D	110.661	166.197	More activity before episode

Note: episode type indicates whether a depressive episode (D) or a manic episode (M) emerged; mean actigraphy values are given in MotionWatch count units

TABLE 3

MEAN ACTIVITY DURING EPISODES AND EUTHYMIC PHASES

ID	Episode type	Mean euthymic	Mean episode	Interpretation
1	D	108.403	87.236	Less activity during episode
2	M	145.258	144.386	Less activity during episode
3	D	94.089	96.359	More activity during episode
4	M	49.156	54.576	More activity during episode
8	M	94.559	112.914	More activity during episode
9	D	196.669	196.399	Less activity during episode
11	D	186.464	168.594	Less activity during episode
15	D	121.871	137.651	More activity during episode

Note: episode type indicates whether a depressive episode (D) or a manic episode (M) emerged; mean actigraphy values are given in MotionWatch count units

Discussion

In the current study, we applied generic Early-Warning Signals (EWS) and spectral periodicity analysis calculated from actigraphy time-series data to investigate whether we could predict upcoming mood transitions in patients suffering from bipolar disorder (BD). We tested whether three EWS (i.e. variance, kurtosis, and autocorrelation at lag-720), showed significant changes up till four weeks before the onset of a manic or depressive episode. We found that in seven out of eight patients a significant change in at least one of these three EWS could be identified up till four weeks before the onset of an episode. For the variance and kurtosis EWS, the effect was in the expected direction, thereby con firming our first two hypotheses. For the acf-720 EWS, the effect was in the expected direction in one patient, but in the opposite direction in four patients, thereby rejecting our third hypothesis. Such a result suggests that acf-720 seems to act as a more general EWS which can signal both increases and decreases, which implies that shifts in either direction can precede episode. Acf-72 was able to detect episode onsets in three patients that the variance and kurtosis EWS did not pick up. Moreover, when considering more large scale trends, both increases and decreases could be observed. Yet, our finding that autocorrelation effects were often in the opposite direction of what was expected do cast doubt on whether the observed transitions in bipolar patients are best described and predicted with zero-eigenvalue tipping points. For the fourth hypothesis we expected to find disturbances in the typical 24-h sleep/wake cycle before the onset of an episode. However, the hypothesized pattern was only observed in one out of eight patients. We thus rejected our fourth hypothesis. Our fifth hypothesis concerned testing whether mean activity levels are congruent with the episode type before the start of the episode. We have rejected this hypothesis as only two out of eight patients showed the expected effect. Nonetheless, post-hoc analyses showed mean activity levels congruent with the episode type when data during the episode was analyzed instead of data from before the episode.

The EWS results did support the hypothesis from complex system theory that actigraphy derived EWS, did precede transitions such as onsets of bipolar disorder episode onset. Commonly used EWS such as variance and kurtosis seem to operate complementary to more contextdriven EWS, such as autocorrelation at lag-720. These findings suggest that a

combination of effective, personalized EWS could be potential useful in clinical practice. Such a clinical tool could be used to monitor a patient ' s risk for developing a clinical relevant manic or depressive episode, and warn clinicians in time to temper or even prevent upcoming BD episodes. As this clinical tool used actigraphy data, it offers patients a low burden method to monitor BD episode risk. Moreover, more research is required to assess the effectiveness of these, and other EWS in larger samples to investigate whether more effective combinations of EWS can be found. Here combining an actigraphy-based method with subjective time-series data, derived with the experience sampling method, may be the next step forward to enhance BD episode risk assessment.

Regarding the spectral indices, we found that the investigated spectral periodicity indices appeared to be somewhat sensitive to multiple types of transitions (amongst others, towards the start of the episode or towards the end of the euthymic period). However, results were not as clear-cut as found in earlier studies, wherein for example, distinct 48-h sleep/wake cycles were found in bipolar patients who transitioned from depressed to manic episodes14. Here, the spectral periodicity indices thus behaved more like general mood instability indicators than as an exclusive indicator of episode onsets†.

Despite the innovative character of our study, there are a number of issues that need to be addressed. First, precollected data were used, whereas only performing the analysis while the patient is still monitoring herself would allow for real-time detection of changes in BD episodes. However, there is a current lack of actigraphy hardware that can send information in real-time for long-term continuous monitoring and calculation of indices. Therefore, extensive cooperation between researchers, clinicians, patients, and actigraphy hardware manufacturers is needed to develop the infrastructure required for such real-time monitoring of bipolar patients. Second, another issue is the relatively low number of patients in our sample. Yet, this is somewhat offset by the relatively large number of data points collected by each patient (1440 observations each day, for approx. 180 consecutive days). While the large number of observations does offer confidence in the robustness of the within patient findings, a replication study with a larger number of patients for the same time period could provide improved generalizability of the obtained results. Third, the current study did not investigate the false positive rate, the number of times the EWS would have falsely suggested that a transition is afoot, while in fact none is. As such false negatives can reduce the effectiveness of EWS as a clinical tool, this point should be investigated further in future studies. Fourth, although both variance and autocorrelation are promising resilience indicators for upcoming

critical transitions, variance was found to be not as robust as autocorrelation23. That is, when environmentally triggered changes that affect the equilibrium of a system itself was found to be able to decrease rather than to increase before an upcoming transition. This lower robustness for variance will be more profound if the system 's own rate of change is relatively slow when compared to the frequency of the environmentally triggered changes. Fifth, as our data is time-correlated the obtained trends in the indicators could be due to chance. Bootstrapping is typically a viable strategy to prevent such chance findings 7 ,24. However, as the sleep/wake cycle of physical activity introduces a strong daily periodicity in our data, commonly used bootstrapping methods would not be suitable25. Given the intricacies of selecting and performing a bootstrap strategy suitable for the current data, such additional analyses would be beyond the scope of this work. Yet we do recognize that a suitable bootstrap analysis in for example a future study, could provide additional evidence for the hypotheses investigated in the current study.

Development and application of EWS in the field of psychiatry is still quite novel. Future studies could aim to elucidate basic EWS properties in psychiatric samples by aiming to answer fundamental questions relevant for the search of EWS such as: (1) "At what time scale do changes occur in this psychiatric sample? ", (2) "What exact marker (actigraphy, heart rate, mood, etc.) would be best to search for EWS in? ", (3) "How can we help increase the number of n = 1 studies with large enough samples to establish EWS sensitivity and specificity? ", (4) "Which combination of EWS would outperform most single EWS indices? ", or (5) "How is EWS performance affected by external factors, such as life events or medication use? ". The answers to these kinds of questions may be helpful to unravel if and how dynamical systems theory fulfils its promise for psychiatric research and implementation in clinical practice.

In summary, we found that both EWS and spectral periodicity indices could facilitate the prediction of upcoming mood episodes in bipolar patients. With the tested EWS we were able to identify upcoming BD episode onsets. Yet, before this method can be used in clinical practice further studies are required to investigate how the tested EWS perform in the absence of transitions; thereby investigating their false positive rates. Contextdriven actigraphy based EWS such as autocorrelation at lag-720, require further conceptual study in order to leverage their predictive capabilities, especially regarding their timing, as currently only a period of four weeks was considered. Furthermore, we investigated whether disturbances in periodicity preceded episode onset. Such periodicity disturbances were found to show performance that

is more akin to general instability indices than as singular indices for episode onset. While such findings were unexpected, and might be less helpful from a clinical perspective (e.g., when the end instead of the start of the episode is predicted), they do offer useful theoretical knowledge to assess their effectiveness and limitations in EWS interpretation. The current study thus provides exploratory information on the opportunities and pitfalls of analyzing EWS from actigraphy data. As such, the pioneering work presented in this study can be used as a stepping-stone for future studies examining the possibility to predict mood transitions by using actigraphy data, both in patients suffering from bipolar disorders as well as in other mental disorders.

1. López-Muñoz, F., Vieta, E., Rubio, G., García-García, P. & Alamo, C. Bipolar disorder as an emerging pathology in the scientific literature: a bibliometric approach. *J. Affect Disord. 92*, 161 –170 (2006).

2. Harvey, A. Sleep and circadian functioning: critical mechanisms in the mood disorders? *Annu Rev. Clin. Psychol. 7*, 297 –319 (2011).

3. Jackson, A., Cavanagh, J. & Scott, J. A systematic review of manic and depressive prodromes. *J. Affect Disord. 74*, 209 –217 (2003).

4. Sierra, P., Livianos, L., Arques, S., Castelló, J. & Rojo, L. Prodromal symptoms to relapse in bipolar disorder. *Aust. NZ J. Psychiat. 41*, 385 –391 (2007).

5. Scheffer, M. et al. Early-warning signals for critical transitions. *Nature 461* , 53 –59 (2009).

6. Biggs, R., Carpenter, S. & Brock, W. Turning back from the brink: Detecting an impending regime shift in time to avert it. Proc. Natl Acad.Sci. USA 106 , 826 –831 (2009). 7. Dakos, V. et al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS ONE 7*, e41010 (2012).

8. Alloy, L., Ng, T., Titone, M. & Boland, E. Circadian rhythm dysregulation in bipolar spectrum disorders. *Curr Psychiatry Rep. 2017*;19. https://doi.org/ 10.1007/s11920-017-0772-z

9. Salvatore, P. et al. Circadian activity rhythm abnormalities in ill and recovered bipolar I disorder patients. *Bipolar Disord. 10*, 256 –265 (2008).

10. Jones, S., Hare, D. & Evershed, K. Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder. *Bipolar Disord. 7*, 176 –186 (2005).

11. Ng, T., Chung, K., Ng, T., Lee, C. & Chan, M. Correlates and prognostic relevance of sleep irregularity in inter-episode bipolar disorder. *Compr. Psychiatry 69* , 155 –162 (2016).

12. Box, G., Jenkins, G. & Geinsel, G. *Time series analysis forecasting and control*. New Jersey: Wiley; 2008.

13. Basner, M. et al. Mars 520-d mission simulation reveals protracted crew hypokinesis and alterations of sleep duration and timing. *PNAS 110* , 2635 –2640 (2013).

14. Wehr, T. et al. 48-hour sleep-wake cycles in manic-depressive illness. *Arch. Gen. Psychiatry 39*, 559 (1982).

15. Pollak, C., Tryon, W., Nagaraja, H. & Dzwonczyk, R. How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep 24*, 957 –965 (2001).

16. Knapen, S. E. Rhythm & Blues: Chronobiology in the pathophysiology and treatment of mood disorders (Doctoral dissertation, Rijksuniversiteit Groningen, Groningen, the

Netherlands). 2019. Retrieved from: https://www.rug.nl/research/portal/en/publications/rhythm--blues(94b2329b-7ff7-44ee-a095-b330d92735a6).html.

17. Rush, A., Gullion, C., Basco, M., Jarrett, R. & Trivedi, M. The inventory of depressive symptomatology (IDS): psychometric properties. *Psychol. Med. 26* , 477 –486 (1996).

18. Altman, E., Hedeker, D., Peterson, J. & Davis, J. The Altman self-rating mania scale. *Biol. Psychiatry 42*, 948 –955 (1997).

19. aan het Rot, M., Hogenelst, K. & Schoevers, R. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. *Clin. Psychol. Rev. 32*, 510–523 (2012).

20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. https://www.Rproject.org/.

21. Kunkels, Y. K. et al. ACTman: Automated preprocessing and analysis of actigraphy data. *J. Sci. Med. Sport*. https://doi.org/10.1016/j.jsams.2019.11.009 (2019).

22. Mann, H. Nonparametric Tests Against Trend. *Econometrica 13*, 245 (1945).

23. Dakos, V., van Nes, E., D'Odorico, P. & Scheffer, M. Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology 93*, 264–271 (2012).

24. Dakos, V. et al. (2008). Slowing down as an early warning signal for abrupt climate change. IOP Conference Series: *Earth and Environmental Science, 6*, p.062012.

25. Politis, D. N. (2001). Resampling time series with seasonal components. Frontiers in Data Mining and Bioinformatics: Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics, Orange County, California, June 13–17, pp. 619–621. (Proceedings).

Risk ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during anti-depressant discontinuation

This chapter is published as:
Kunkels, Y. K., Smit, A. C., Minaeva, O., Snippe, E., George, S. V., van Roon, A. M., Wichers, M., Riese, H. (2023). Risk Ahead: Actigraphy-based early-warn-ing signals of increases in depressive symptoms during anti-depressant discontinuation. Clinical Psychological Science, p.216770262211481. doi:10.1177/21677026221148101.

4.

Chapter 4: Risk Ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during anti-depressant discontinuation

Yoram K. Kunkels[1*], Arnout C. Smit[1,2], Olga Minaeva[1], Evelien Snippe[1], Sandip V. George[1,3], Arie M. van Roon[4], Marieke Wichers[1], Harriëtte Riese[1]

[1]University of Groningen, University Medical Center Groningen, Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), The Netherlands. [2]Clinical Psychology, Faculty of Behavioral and Movement Sciences, VU Amsterdam, Amsterdam, the Netherlands. [3] University College London, Gower Street, London WC1E 6BT, United Kingdom. [4]Department of Vascular Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

***Corresponding author.** University of Groningen, University Medical Center Groningen, Groningen, Department of Psychiatry, ICPE, PO Box 30.001 (CC 72), 9700 RB Groningen, The Netherlands; Tel: +31 50 6319005; Email: y.k.kunkels@umcg.nl

This chapter is published as; Kunkels, Y. K., Smit, A. C., Minaeva, O., Snippe, E., George, S. V., van Roon, A. M., Wichers, M., & Riese, H. (2023). Risk ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during antidepressant discontinuation. *Clinical Psychological Science*, 216770262211481.

https://doi.org/10.1177/21677026221148101

Abstract

Antidepressant discontinuation increases the risk of experiencing depressive symptoms. In a repeated single-subject design, we tested if transitions in depression were preceded by increases in actigraphy-based critical slowing down based Early Warning Signals (CSD-based EWS; variance, kurtosis, autocorrelation), circadian rhythm-based indicators, and decreases in mean activity levels.

Four months of data from 16 individuals with a transition and 9 without a transition in depression were analyzed using a moving window method.

As expected, more participants with a transition showed at least one EWS (50% true positives; 22.2% false positives). Surprisingly, increases in circadian rhythm variables and decreases in activity levels were more common in participants without a transition (respectively, 25.0% and 37.5% true positives; 44.4% and 44.4% false positives).

None of the tested risk indicators could confidently predict upcoming transitions in depression, but some evidence was found that CSD-based EWS were more common in participants with a transition.

Keywords

actigraphy, antidepressant discontinuation, early warning signals, personalized psychiatry, repeated single-subject design

Introduction

Major depressive disorder (MDD) is a severely debilitating condition affecting approximately 264 million individuals worldwide (WHO, 2019). For this condition, the prescription of antidepressant medication is widespread and increasing in Europe and the United States (Gusmão et al., 2013; Pratt, Brody, & Gu, 2017). Discontinuation of antidepressant use is often desired by patients for different reasons but raises the risk of increases in depressive symptoms (Geddes et al., 2009; Glue et al., 2010; Sim et al., 2016). While there are a number of established risk factors from (group-based) epidemiological research for depression relapses and increases in depressive symptoms, such as comorbid psychopathology and negative cognitive styles (Burcusa & Iacono, 2007; Buckman et al., 2018), these effects might not always apply on an individual level (Molenaar, 2004; Hamaker, 2012; Zuidersma et al., 2020). As such, there seems to be a clear lack of individual-based research to reveal within-person risk quantifiers of increases in depressive symptoms following antidepressant discontinuation. Moreover, established risk factors do not yet convey much information about the timing of potential upcoming transitions in depressive symptoms in individual patients. This hinders clinical practice, wherein clinicians are yet unable to accurately monitor patient progress and potential risks of depressive symptoms returning during and shortly after medication discontinuation attempts.

Complex dynamical systems theory could aid in such personalized transition detection based on an individual's collected time-series data. This theory presumes a set of *critical slowing down*-based *early warning signals* (EWS) for a broad subset of dynamical systems, which can identify if a transition into another (mood) state is approaching in some dynamical systems (Scheffer et al., 2009). Near such a tipping point, we expect to find *critical slowing down*, wherein the return rate to equilibrium after minor disturbances goes to zero (Scheffer et al., 2009; Wissel 1984; Strogatz, 2018). EWS preceding critical transitions are assumed to

be present in various systems, from ecological systems wherein climate can change to global financial systems wherein markets can deteriorate and collapse (Scheffer et al., 2009). EWS are calculated as relatively straightforward statistical indices, such as variance, kurtosis, or autocorrelation at lag 1 (acf-1), which are expected to increase before the transition occurs and to peak somewhere around the transition moment (Scheffer et al., 2009; Biggs, Carpenter & Brock, 2009; Dakos et al., 2012a). These increases occur because when a system recovers more slowly from perturbations, its state spends more time away from the equilibrium.

In psychiatry, several studies report evidence that EWS may precede transitions in depression (van de Leemput et al., 2014; Wichers & Groot, 2016; Wichers, Smit & Snippe, 2020; Helmich et al., submitted). In these studies, EWS were examined in time-series data collected through the *experience sampling method* (ESM, also known as ecological momentary assessment (Csikszentmihalyi & Larson, 1987). Here, individuals fill out short questionnaires multiple times a day on mobile devices, such as smartphones. While ESM time-series data can offer insight into how momentary affect develops over time, it can be relatively limited in the number of daily assessments. This is mainly due to studies having to balance the number of presented questionnaires against the potential burden for individuals having to fill-out multiple questionnaires each day (van Genugten et al., 2020). This is why there is a need to investigate if EWS could be applied to certain types of time-series data, which are less burdensome to collect, such as *actigraphy* or *accelerometer* data (Kunkels et al., 2021). Moreover, by doing so, we could investigate whether the predictive capabilities of actigraphy-based EWS can improve over those of ESM-based EWS. Such actigraphy data on physical activity is anticipated to include relevant information for identifying transitions in depressive symptoms, as the normal physical activity pattern is expected to change when nearing such a transition. Actigraphy data are collected by having individuals continuously wear lightweight accelerometers, which can provide data in intervals from, for example, 60

second periods, down to 1 second periods. Intensive longitudinal data are assessed (e.g., 1440 measurements a day when using 60 sec. periods) without individuals having to put in any conscious effort (Kunkels et al., 2021).

Another advantage of using critical slowing down based EWS on actigraphy time-series data is the established conceptual link between depressive symptoms and physical activity, as psychomotor retardation is a key feature of MDD (Buyukdura, McClintock & Croarkin, 2011). While ESM is also based on a conceptual link between ESM and depression, the link between physical activity and depression might have a different pathway and thus could provide new information. Slowing down in MDD patients can be observed in gross psychomotor movements, including diminished hand and leg movements (Sobin, Mayer & Endicott, 1998). Also, it was found that patients who clinically improved showed significantly higher movement intensities after four weeks, while patients who did not improve did not show increased movement intensities (Todder, Caliskan & Baune, 2009). As such, we expect to detect decreases in the mean level of physical activity, as measured by actigraphy, before an increase in depressive symptoms.

When investigating actigraphy time-series data, the interdaily stability (IS) and intradaily variability (IV) are commonly used non-parametric methods. Both are well-established circadian rhythm variables and provide information about the stability and fragmentation of the circadian rhythm (van Someren 1999; Witting et al., 1990). IS indicates the association between the circadian rhythm and external Zeitgebers (stability). That is, IS indicates how stable the circadian rhythm is from day to day, and the higher IS, the more stable the rhythm. IV marks the intensity and frequency of changes in rest and activity (fragmentation). IV thus indicates how much the circadian rhythm is fragmented within a day, wherein higher IV indicates a more fragmented and unstable circadian rhythm. Given their role in general actigraphy research, we hypothesize these quantifiers will also be highly

informative in our sample. In contrast with the critical slowing down-based EWS, quantifiers such as IS and IV can be considered *circadian rhythm variables*. Hence, preceding a transition towards increases in depressive symptoms, we would expect increased IS and decreased IV, as it could indicate rigidity changes in the system, causing the system to have problems coping with external stressors. We also propose a third circadian rhythm variable, acf-1440 (autocorrelation at lag-1440), a circadian variant of acf-1, which provides information about the autocorrelation of the actigraphy data over 1440 minutes (24 hours). As acf-1440 corresponds to roughly one circadian cycle, it is expected to carry information about the individual's day-to-day activity levels and thus may be more informative than acf-1 in this case as it captures a longer and more *circadian* component of the time series than acf-1. This notion is further supported by earlier studies investigating alternative autocorrelation lags, such as acf-720 (Kunkels et al., 2021). We expect acf-1440 to show an increase near transitions, as measurements were taken 24-h apart are expected to become more alike.

While there are a number of studies that investigated depression using actigraphy (Todder, Caliskan & Baune, 2009; Lemke, Puhl & Broderick, 1999; Minaeva et al., 2020; Raoux et al., 1994; Difrancesco et al., 2019), most only used relatively short assessment periods of a number of days up to a few weeks. Such short-term periods are too short to fully capture the (gradual) discontinuation of antidepressant (tapering) periods, involving several weeks or sometimes even months. Most notably, none of the actigraphy studies in MDD patients focused on within-person changes in context-driven actigraphy-based indicators for upcoming increases in depressive symptoms following antidepressant discontinuation. To investigate potential quantifiers of upcoming transitions in MDD, a study design is required, which allows us to investigate per participant whether changes in EWS preceded upcoming transitions.

Therefore, in 25 single-subject time series, all obtained within the same TRANS-ID study (for the complete study protocol, see: https://osf.io/zbwkp), we will investigate whether EWS and context-driven risk quantifiers calculated over actigraphy data precede increases in depressive symptoms in individuals in remission who were discontinuing their antidepressant medication. We expect to detect in each of the single subject time-series that prior to transitions in depressive symptoms: (1) increasing critical slowing down based EWS (variance, kurtosis, and acf-1), (2) increasing IS and acf-1440, and decreasing IV, and (3) decreasing mean levels of physical activity. In Table 1, an overview of critical slowing down-based EWS (variance, kurtosis, acf-1), circadian rhythm variables (IS, IV, acf-1440) and mean level, and expected direction of effect are given. The reported variables will be studied at the individual level, which allows for examining of individual differences as well as whether the presence of EWS can be replicated across individuals. The analysis of this study, including its hypotheses, was pre-registered on the Open Science Framework (see: https://osf.io/dfmw3) and the complete study protocol is available online (https://osf.io/zbwkp).

Table 1

Overview of quantifiers and expected direction of change to identify an upcoming transition in depressive symptoms.

Quantifiers	Transition	No transition	Reference
H1: Variance	Increase	No significant change	Scheffer et al., 2009; Biggs, Carpenter & Brock, 2009; Dakos et al., 2012a
H1: Kurtosis	Increase	No significant change	Scheffer et al., 2009; Biggs, Carpenter & Brock, 2009; Dakos et al., 2012a
H1: acf-1	Increase	No significant change	Scheffer et al., 2009; Biggs, Carpenter & Brock, 2009; Dakos et al., 2012a
H2: IS	Increase	No significant change	van Someren 1999; Witting et al., 1990*
H2: IV	Decrease	No significant change	van Someren 1999; Witting et al., 1990*
H2: acf-1440	Increase	No significant change	None**
H3: Mean activity	Decrease	No significant change	Buyukdura, McClintock & Croarkin, 2011; Sobin, Mayer & Endicott, 1998; Todder, Caliskan & Baune, 2009

Note: H1, hypothesis 1; H2, hypothesis 2, H3, hypothesis 3; acf-, autocorrelation at lag-.

* While the cited references offer the theoretical basis for IS and IV, the expected directions of these quantifiers in this study were introduced in the current study.

** Autocorrelation at lag-1440 was introduced in the current study as a circadian (24 hours) variant of autocorrelation at lag-1.

Methods

Sample: An overview of individual inclusion and exclusion is shown in Figure 1. Individuals were recruited through a pharmacy and online means. In total, 69 individuals were included in the study who fulfilled the criteria of a past diagnosis of major depressive disorder (MDD) according to DSM-IV criteria. These formerly depressed individuals made a shared decision with their mental health professional to taper their antidepressant medication and did not meet the criteria for MDD at baseline. Of these individuals, 13 dropped out, and for five individuals, it was not possible to clearly define whether a transition towards higher levels of depression had occurred and were thus excluded. One participant was excluded due to technical issues causing actigraphy data loss. Of the remaining 50 individuals, 31 did experience a transition in depressive symptoms, while 19 did not. Of those with a transition, we excluded: two individuals due to incomplete actigraphy data files, three individuals because their actigraphy time-series data had more than seven consecutive days of missing data, five individuals because the transition occurred before or after the measurement period, and five individuals because there were less than 30 days of data before the transition, which would be not enough data to properly conduct the analyses. This period was doubled when compared to the 14 days first described in our pre-registration as we have also doubled the window size from 7 to 14 days for our analyses. This increase is expected to maintain reliable calculation of the quantifiers, while also maintaining the weekly periodicity of actigraphy data. As such, the final sample included 16 individuals with a transition. Regarding the 19 individuals without a transition, one had to be excluded due to incomplete actigraphy data files, while nine had to be excluded because there were more than seven consecutive days of actigraphy data missing. The final sample included nine individuals without a transition. As such, the available final sample differed in size from that reported in the pre-registration (see: https://osf.io/dfmw3). Further details on the used study protocol are available online (see: https://osf.io/zbwkp/). The study was approved by the Medical Ethical Committee of the

University Medical Center Groningen (UMCG, METc2016.443). All patients were informed that they could stop their participation at any time and were asked to read and provide written informed consent prior to participation.

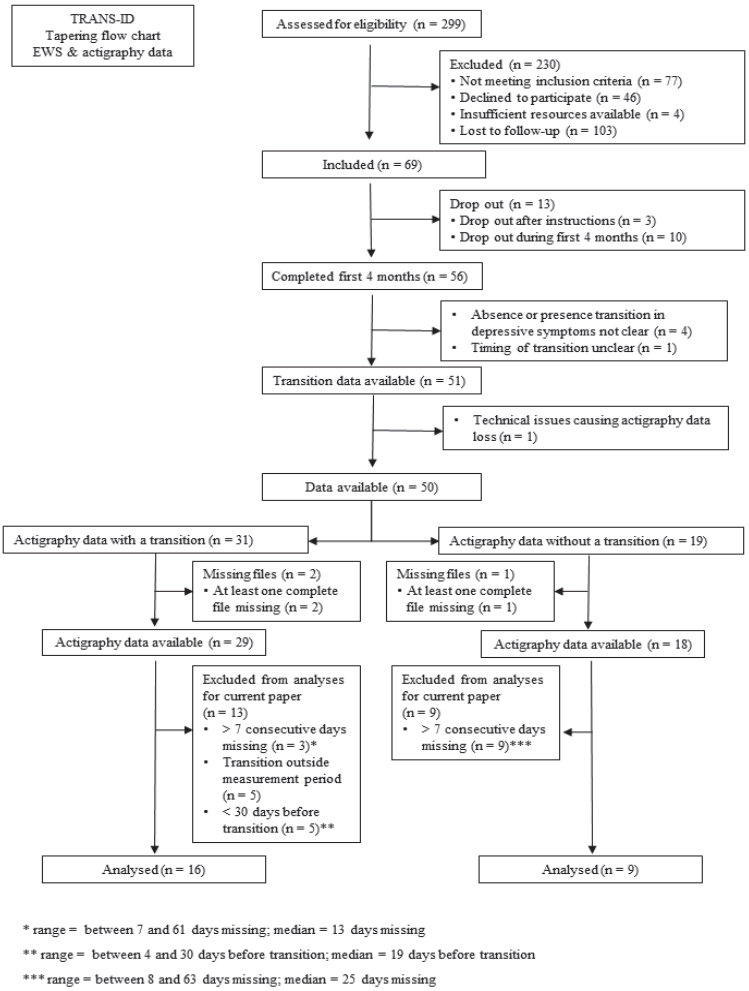


Figure 1: Flowchart TRANS-ID Antidepressant discontinuation and Actigraphy study

Actigraphy assessment: Physical activity time-series data were collected with a wrist-worn MotionWatch 8 (MW8, *CamNTech*) accelerometer. The MW8 was initialized to assess 60 seconds epoch lengths, while light detection and data compression were disabled. Individuals were instructed to continuously wear the MW8, only removing the device under rare conditions, such as sauna visits. Moreover, individuals were instructed to press the MW8 event marker button to register the times at which the individual got out of bed and when the individual went to sleep. As battery and memory capacity was valid for data assessment for two months, MW8 actigraphs were replaced halfway through the four-month monitoring period. Participants received the first MW8 during a personal interview at the start of the monitoring period and the replacement MW8 through registered mail. These MW8 devices were initialized before sending them through postal services, and participants only had to switch the old MW8 for the new MW8 while returning the old MW8 via postal services.

Actigraphy data pre-processing*:* Activity count data from the MW8 accelerometers were extracted with the native Motionware software (version 1.2.28). Due to replacing the MW8 devices halfway through monitoring, the two actigraphy files assessed from each participant were merged prior to analyses. Although the native Motionware software offers the merging functionality, such straightforward merging led to the mismatched merging of files in the current application. Hence, three raters were employed to systematically merge these files. Files were visually checked on their activity and sleep patterns to provide optimal matches of files on the minute level. Missing data between files were imputed with zeroes to merge the two files into one continuous file, as subsequent analysis software could not handle missing data. The imputation of data was only done on small-scale instances when it was required to merge multiple data files from one participant. In Table 2, a column is added to show the information on how much data was missing at maximum in relation to the size of the moving

window. From these percentages, we infer that it is unlikely that the applied imputation strategies could have substantially tainted the analysis outcomes.

Transitions in depression: Transitions towards higher levels of depressive symptoms were defined using weekly SCL-90 depressive symptom data, the evaluation interview, and other qualitative data (Smit et al., In Press). Patients had to fulfill a criterion of a reliable change (Jacobson & Truax, 1991) in depressive symptoms on the SCL-90 depression subscale, a criterion on the persistence of this depressive symptom increase, and a criterion on the clinically meaningful change as experienced by patients (qualitative consensus rating based on interviews and open-ended questions).

Actigraphy outcome variables: From the actigraphy data, the following outcome variables were calculated: (1) EWS (variance, kurtosis, and acf-1), (2) circadian rhythm variables (IS, IV, and acf-1440), and (3) mean physical activity. Comprehensive overviews of IS and IV calculation are described in more detail elsewhere (van Someren 1999; Witting et al., 1990). Missing data were handled by the following exclusion criteria. That is when more data were missing than the moving window analyses could process (e.g., when complete files were missing) or when more than seven consecutive days were missing (details given in Figure 1).

Statistical analyses: To detect significant changes in the EWS, modified Mann-Kendall (MK) tests (Hamed & Ramachandra Rao, 1998) were used as this method is better suited to deal with autocorrelation between consecutive windows than the normal MK-test and is, therefore, less vulnerable to false positives. The Mann-Kendall trend test is a commonly used test in the literature on EWS (Dakos et al., 2012b). An advantage of the Mann-Kendall trend test is that it is better suited to detect nonlinear monotonic trends than a Pearson correlation. Additionally, the Kendall correlation coefficient requires much fewer points than Pearson or Spearman correlation coefficients for detecting the same trend in data (Bonett & Wright,

2000). As we expected some of the indicators to overlap, we employed the *effective number of tests* (*Meff*) method (Cheverud, 2001) to correct for multiple testing, which takes into account the cross-correlations between quantifiers, considering individual participants. For individuals with a transition, only data obtained before the transition were analyzed; for individuals without a transition, the full research period was analyzed.

Traditionally, a statistical test with a p-value as an outcome parameter is used to detect a change in an EWS, and this significant change ($p<p_\alpha$) is expected to predict, or detect, a transition. We used the *Meff* method to adjust the standard p_α of 0.05 to a lower lever (approximately 0.02) to correct for multiple testing and correlations between EWS. We used a single-sided test (change in the pre-defined direction, Table 1, p_α applied to the upper side of the probability distribution only).

Calculating EWS: The ACTman software package (Kunkels et al., 2019, see; https://github.com/compsy/ACTman/) for R statistical software, version 4.0.4 (R Core Team, 2019) was used to preprocess the actigraphy data. The employed moving window method spanned 14 days and was moved over the actigraphy data in one-day steps. In the pre-registration, a window size of seven days was described. However, such short window sizes might lead to more variable estimates which could also provide more unreliable EWS and risk indicator estimations. In order to investigate whether there are substantial differences in study outcomes when choosing between a 14-day moving window and a 7-day moving window, we did also perform the analyses in a 7-day moving window, the results of which are available in the supplementary materials, Table S6. No substantial difference was found herein.

Results

Descriptives

Of the sixteen participants who did experience a transition, the mean age was 51.3 years old (range: 27 – 67), and 87.5% were female. Of the nine participants who did not experience a transition, the mean age was 44.8 (range: 25 – 61), and 77.8% were female.

Critical slowing down-based EWS

The results on the critical slowing down-based EWS quantifiers are given in Table 2. Significant increases in variance preceded transitions in three individuals (18.8% true positives; 11.1% false positives). Significant increases in kurtosis preceded transitions in four individuals (25.0% true positives; 11.1% false positives). Significant increases in acf-1 preceded transitions in three individuals (18.8% true positives; 11.1% false positives), one of whom also showed an increase in variance. As such, in eight out of sixteen individuals (50.0%) with a transition, at least one critical slowing down-based EWS preceded the transition (i.e., true positives). Regarding the nine individuals who did not experience a transition, two out of nine individuals (22.2%) falsely showed significant increases in at least one EWS (i.e., false positives). Here, kurtosis and acf-1 both showed a false positive in the same individual, while variance showed a false positive in one other individual.

Table 2

Performance of EWS, calculated over participants' actigraphy data, in identifying upcoming transitions in depressive symptoms.

	Critical slowing down based	EWS		Context EWS			Mean		Max % imputed of moving window
Transition group (N= 16)									
ID	Variance	Kurtosis	acf-1	IS	IV	acf-1440	Mean	**TOTAL**	
1036	1	0	1	0	0	0	0	2	5.22%
1045	0	0	0	0	0	0	0	0	0.00%
1046	1	0	0	1	0	0	0	2	0.00%
1052	0	1	0	0	0	0	0	1	0.00%
1074	0	0	0	0	0	0	1	1	0.00%
1075	0	1	0	0	0	0	0	1	0.00%
1076	0	0	0	0	0	0	0	0	0.00%
1077	0	0	1	0	0	0	0	1	0.00%
1108	0	1	0	0	0	0	1	1	0.00%
1133	0	0	0	0	0	1	0	1	72.17%
1173	0	0	1	0	0	0	1	2	0.00%
1181	0	0	0	0	0	0	1	1	0.00%
1193	0	0	0	0	1	0	0	1	10.62%

1224	1	1	0	0	0	0	0	2	0.00%
1264	0	0	0	0	1	0	1	1	0.00%
1293	0	0	0	0	0	0	1	1	0.00%
TOTAL	3	4	3	1	2	1	6		
Percentage:	18.75%	25.00%	18.75%	6.25%	12.50%	6.25%	37.50%	87.50%	

Non-Transition group (N = 9)

1041	0	0	0	0	0	0	1	1	24.22%
1059	0	0	0	0	0	0	0	0	0.00%
1067	0	0	0	0	0	1	1	2	0.00%
1110	0	1	1	0	0	0	1	3	0.00%
1178	1	0	0	0	1	0	0	2	0.00%
1180	0	0	0	0	1	0	0	1	0.00%
1255	0	0	0	1	0	0	1	2	3.27%
1280	0	0	0	0	0	0	0	0	0.00%
1295	0	0	0	0	0	0	0	0	32.84%
TOTAL	1	1	1	1	2	1	4		
Percentage:	11.11%	11.11%	11.11%	11.11%	22.22%	11.1%	44.44%	66.7%	

Note: The upper panel shows the transition group wherein green cells indicate the true positives. The lower panel shows the non-transition group wherein red cells indicate the false positives. Yellow cells indicate zeroes. acf-1, autocorrelation at-lag-1; acf-1440, autocorrelation at-lag-1440; IS, interdaily stability; IV, intradaily variability. *These two percentages are calculated by counting how many participants in each group had at least one significant EWS, and dividing it by total group size.

Circadian rhythm variables

Results in the circadian rhythm variables are given in Table 2. Significant increases in IS were found to precede a transition in one individual. (6.3% true positives; 11.1% false positives). Significant increases in IV were found to precede a transition in two other individuals. (12.5% true positives; 22.2% false positives). Significant increases in Acf-1440 were found to precede a transition in another individual. (6.3% true positives; 11.1% false positives). In sixteen individuals with a transition, four participants showed increases in circadian rhythm variables in the period prior to the transition (25.0%). Regarding the nine individuals who did not experience a transition, four out of nine individuals (44.4%) incorrectly showed significant increases in at least one EWS (i.e., false positives, indicating that false positives were more common, among the investigated circadian rhythm variables than true positives).

Mean levels

When investigating whether decreases in the mean levels of physical activity precede transitions in depression, we did find six such decreases in the sixteen individuals (37.5%) with a transition (true positives, see Table 2). However, we also found four decreases in mean activity in the nine individuals (44.4%) without a transition (false positives). This indicates that false positives were more common, among the investigated mean levels than true positives.

Post-hoc analyses

In our preregistered analysis plan, we described tests on the expected direction of effects. However, as we observed during analysis that many EWS showed substantial changes in the unexpected direction, and we decided to also perform post-hoc analyses using two-sided tests. In other words, it seemed that the performance of some EWS may be improved by excluding the predicted direction of the effect, albeit at a higher false positive rate. ROC curves were calculated post-hoc and plotted to investigate the true positive rate and false positive rate characteristics of the EWS and circadian rhythm variables. ROC curves show the true positive rate (sensitivity) against the false positive rate (Egan, 1975; Fawcett, 2006). A description of the ROC curves is given in S1, the ROC curves are shown in Figure S2, and the results are given in Table S3. Here, one can observe that the two-sided tests mostly outperformed the one-sided tests. Regarding individual EWS, the two-sided plot for kurtosis and IV runs close to the top-left corner. This can be interpreted as that these EWS performed relatively well in preceding transitions while not suffering as much from giving false positives. A description hereof is given in S4, and the corresponding results are given in Table S5. Here, the number of true positives for the 2-sides test was that 15 out of 16 unique transitions were preceded by a significant change in at least one EWS (93.75%; instead of 14 out of 16 (87.5%) described for the 1-sided tests above). Unfortunately, the false positive rate also increased substantially when testing two-sided. Instead of 66.7% of false positives in the one-sided test, the false positive rate increased to 100% when testing two-sided. Additionally, we exploratively combined multiple risk quantifiers into aggregate measures that increased the accuracy of the method for predicting the same data used to fit the model (see S7 and S8). Here we found that combinations of EWS were able to outperform single EWS under specific circumstances, however further studies are needed to confirm such findings in larger samples.

Discussion

A repeated single-subject design was used to test if circadian rhythm variables could predict an increase of depressive symptoms to a clinically relevant level in individuals who discontinued their antidepressant medication. At the individual level, we found that in eight out of sixteen participants (50.0% true positives), an upcoming transition in depressive symptoms was preceded by at least one CSD-based EWS (variance, kurtosis, and acf-1), compared to 22.2% false positives in participants without a transition, which was in line with our first hypothesis. However, the performance of individual critical slowing down-based EWS was lower, ranging from 18.7% to 25.0% true positive rates versus an 11.1% false positive rate. We also found that the circadian rhythm variables (IS, IV, and acf-1440) did not signal upcoming transitions, as false positives were more common than true positives. Regarding mean activity levels, we also found that false positives were more common than true positives. Therefore, we conclude that no evidence was found for our second and third hypotheses. In the current study, we did not investigate the potential effects of life events on mean activity levels and circadian rhythm. However, as such events may provide an alternative explanation for some of the detected changes, we suggest future studies to investigate this in more detail. From these results, we found some support for the first hypothesis that increases in at least one critical slowing down-based EWS (variance, kurtosis, and acf-1) precedes transitions in depressive symptoms. With respect to this finding, we can conclude that increases in at least one critical slowing down-based EWS were found to be more prevalent in participants who experienced a transition than in participants who did not experience such a transition. As such, single EWS does not yet seem to be able to differentiate between participants with and without transitions. No evidence was found for the second hypothesis that circadian rhythm variables precede transitions. Furthermore, no strong support was found for the third hypothesis.

Earlier research examined whether circadian rhythm variables could differentiate healthy individuals from depressed individuals or whether depressive episodes or the timing of the episodes could be detected based on actigraphy-based measures (Minaeva et al., 2020; Zanella-Calzada et al., 2019). One of the unique features of the current study is that we examined if within-person changes in these actigraphy-derived quantifiers occurred just before individuals transitioned towards higher levels of depressive symptoms. Whereas previous studies have shown a number of circadian rhythm variables to be associated with depressed mood states (Esaki et al., 2021), the current study shows that the investigated circadian rhythm variables (IS, IV, and acf-1440) do not function as early indicators of an upcoming recurrence of depression.

When considering what the best predictor was for transitions in depressive symptoms, our post-hoc investigations into combinations of EWS provide some information. That is, our analyses yielded evidence that combining multiple EWS may improve the prediction of transitions, for example, the combination of acf-1440, acf-1, and kurtosis, whose point was found to be on the line of optimal solutions for equal costs for true and false positives. Due to the limited sample size, no cross-validation was possible, and thus, we cannot rule out this finding as a chance finding. However, these exploratory results may be a stepping stone for future research into EWS combinations on actigraphy time-series data from a larger sample. Using that data to find EWS combinations would then be validated through cross-validation. Only after that studies with a more confirmatory character can be used to investigate these EWS combinations for predictions of transitions in depression.

Given advances in actigraph technology above ESM and the current availability of cloud data storage and analysis, as signaled by mainstream adaptation of commercial actigraphs such as *Fitbit*, future research could consider developing a software tool for automatically calculating actigraphy-derived transition quantifiers that can be presented to

end-users (patients or clinicians) in an intuitive way. Such a tool may have the potential to more adequately estimate upcoming transitions than tools used currently. However, whether such future actigraphy quantifiers should be based on critical slowing down, context-driven, circadian, or alternative theories or measures is still very much open to debate. Moreover, as we will discuss in the next paragraph, there are several methodological challenges that need attention as well.

Given the mixed findings at the individual level, we post-hoc investigated the effects of the a priori formulated expected direction of the EWS post-hoc (see S4 and S5). In general, testing two-sided instead of one-sided did increase the true positive rate somewhat, but it increased the false positive rate equally or even more. It is interesting to note that in acf-1440, the true positive rate increased substantially more than the false positive rate, which may indicate that our initial hypothesis that acf-1440 would become more rigid and therefore decrease over time may have been wrong. Future studies investigating whether the expected directions of CSD-based EWS and circadian rhythm variables hold robustly in actigraphy data could be worthwhile, especially for acf-1440. Additionally, such research could also investigate whether, instead of becoming more rigid, the system might become more irregular instead (Servaas et al., 2021).

While the used TRANS-ID dataset offers a unique and rich high-resolution longitudinal dataset with multiple datatypes, the current study found that trying to predict transitions through the investigated actigraphy-based quantifiers on a more idiographic, individual-centered basis is not feasible. Perhaps firstly, the identification of more homogeneous activity subgroups is required, along with the study of the ranges and possible cut-off values for proposed transition quantifiers. While such studies would require sample sizes that are too large to easily study in academic settings due to financial and other constraints, commercial parties, such as *Fitbit*, *Garmin, Apple*, or *Huawei,* do process such

large quantities of (near) real-time data. Hence, perhaps future studies could aim at improving industry-academia cooperation in developing potential transition detection and prediction measures and methods.

This study had a number of limitations that should be taken into account. First, the study was designed for repeated single-subject analyses meaning that power calculations were based on the number of data points within one participant needed for statistical analyses instead of the number of participants in a group. This meant that the results were descriptive, and differences in EWS between participants with and without a transition could not be tested statistically. Therefore, the results from the between-person analyses should be interpreted carefully, and confirmation using a larger sample is needed. Second, the core analyses in this paper were applied after data collection was completed. The retrospective EWS analysis strategy involved that we first had to determine the transitions and subsequently could test if EWS could predict these transitions. To become clinically relevant, this should be the other way around to be able to provide ample warning time before a transition occurs. Real-time methods such as statistical process control (SPC) have recently been proposed (Smit et al., 2022), tested in simulated ESM data (Schat et al., 2021), and shown to have value in foreseeing recurrence of depression using empirical ESM time-series data (Smit, Snippe, Wichers, 2019; Smit & Snippe, 2022, Snippe, et al., under review). When considering to apply EWS in real-time, researchers will have to be aware to select methods that are able to correct for repeated testing. While real-time methods such as SPC can handle this issue (Montgomery, 2012), it still has to be investigated what the effects hereof are on the false positive rates in this context. Third, of the 51 participants who completed the data collection period and had transition data available, we were only able to include 25 participants in our analyses. One of the main reasons for this was that the actual battery life of the used actigraphs was only half of the expected battery life, necessitating the use of two actigraphs

to cover the full data collection period of four months. As the used actigraphs could only be initiated by research staff, due to the required software not being available for participants, actigraph delivery through regular mail services was required. This caused missing data and required an additional merging step to merge each participant's data files after receiving the actigraphs back from the participants. The exclusion of half of the participants from the analyses due to data loss illustrates that there is still considerable room for improvement. Based on our experiences during this study, we would suggest working with devices with larger battery capacity or devices of which participants can change the battery themselves, and which also allow for real-time streaming of actigraphy data to certified protected servers. Notably, the latter suggestion is also conditional for any future development of real-time EWS calculation and immediate informing of the patient and their clinician.

Additionally, it could be worthwhile to investigate a sample of participants whose transition occurred in the opposite direction, that is, towards a state of decreased depressive symptoms. This process is likely to be seen in individuals treated for their depressive symptoms, as was done in the TRANS-ID Recovery study (Helmich et al., 2020). Repeating the analyses of the current study with data from the Recovery study may expand our knowledge of depressive symptom dynamics. Lastly, in this study demographics such as racial/ethnic identification, cultural/geographic background, or socio-economic status were not recorded as they were not expected to affect the investigated hypotheses.

To conclude, this is the first study in which we investigated whether transitions into increased depressive symptoms during tapering of antidepressant medication were preceded by actigraphy-based EWS. Though results were in line with the idea that EWS may precede transitions towards higher levels of depression in a small subset of participants, no evidence was found that changes in circadian rhythm variables or changes in the mean level of actigraphy preceded such transitions. Though actigraphy can be a relatively practical way to

obtain physical activity time-series data in clinical practice, due to the small difference between participants with and without a transition, clinical implementation of actigraphy-based EWS as a monitoring tool to inform individuals about their momentary risks of recurrence does not seem feasible in the near future.

Data availability

The data are not freely available in a public repository due to restrictions related to data containing information that could compromise the privacy of the participants. However, data are available upon request via the TRANS-ID Data Access Committee (info@transid.nl or h.riese@umcg.nl).

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-CoG-2015; No 681466 to M. Wichers). The actigraphs were kindly provided by the iLab of the department of psychiatry of the University Medical Center Groningen (UMCG, http://www.ilab-psychiatry.nl). The authors thank the participants for their time and effort, P. Harder for the assistance with the recruitment of the participants, M. Messchendorp and R. de Vries for data collection, and Marieke A. Helmich for participating in the conceptualisation of the study.

References

Biggs, R., Carpenter, S., & Brock, W. (2009). Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of The National Academy of Sciences*, *106*(3), 826-831. doi: 10.1073/pnas.0811729106

Bonett, D., & Wright, T. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, *65*(1), 23-28. doi: 10.1007/bf02294183

Buckman, J., Underwood, A., Clarke, K., Saunders, R., Hollon, S., 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. doi: 10.1016/j.cpr.2018.07.005

Burcusa, S., & Iacono, W. (2007). Risk for recurrence in depression. *Clinical Psychology Review*, *27*(8), 959-985. doi: 10.1016/j.cpr.2007.02.005

Buyukdura, J., McClintock, S., & Croarkin, P. (2011). Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(2), 395-409. doi:10.1016/j.pnpbp.2010.10.019

Cheverud, J. (2001). A simple correction for multiple comparisons in interval mapping genome scans. *Heredity, 87*(1), 52-58. https://doi.org/10.1046/j.1365-2540.2001.00901.x

Csikszentmihalyi, M., & Larson, R. (1987). Validity and Reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease, 175*(9), 526-536. doi: 10.1097/00005053-198709000-00004

Dakos, V., Carpenter, S., Brock, W., Ellison, A., Guttal, V., & Ives, A. et al. (2012a). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *Plos ONE*, *7*(7), e41010. doi:10.1371/journal.pone.0041010

Dakos, V., van Nes, E., D'Odorico, P., & Scheffer, M. (2012b). Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, *93*(2), 264-271. doi: 10.1890/11-0889.1

Difrancesco, S., Lamers, F., Riese, H., Merikangas, K., Beekman, A., & Hemert, A. et al. (2019). Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. *Depression and Anxiety, 36*(10), 975-986. doi: 10.1002/da.22949

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis, Series in Cognitition and Perception*. New York: Academic Press.

Esaki, Y., Obayashi, K., Saeki, K. et al. (2021). Association between circadian activity rhythms and mood episode relapse in bipolar disorder: a 12-month prospective cohort study. *Translational Psychiatry 11*, 525. https://doi.org/10.1038/s41398-021-01652-9

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. doi: 10.1016/j.patrec.2005.10.010

Geddes, J. R., Carney, S. M., Davies, C., Furukawa, T. A., Kupfer, D. J., Frank, E., & Goodwin, G. M. (2003). Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *The Lancet*, *361*(9358), 653-661. https://doi.org/10.1016/s0140-6736(03)12599-8

van Genugten, C., Schuurmans, J., Lamers, F., Riese, H., Penninx, B., & Schoevers, R. et al. (2020). Experienced Burden of and Adherence to Smartphone-Based Ecological Momentary Assessment in Persons with Affective Disorders. *Journal Of Clinical Medicine, 9*(2), 322. doi: 10.3390/jcm9020322

Glue, P., Donovan, M. R., Kolluri, S., & Emir, B. (2010). Meta-analysis of relapse prevention antidepressant trials in depressive disorders. *Australian & New Zealand Journal of Psychiatry, 44*(8), 697-705. https://doi.org/10.3109/00048671003705441

Gusmão, R., Quintão, S., McDaid, D., Arensman, E., van Audenhove, C., & Coffey, C. et al. (2013). Antidepressant Utilization and Suicide in Europe: An Ecological Multi-National Study. *Plos ONE*, *8*(6), e66455. doi: 10.1371/journal.pone.0066455

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.

Hamed, K., & Ramachandra Rao, A. (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

Helmich, M. A. , Smit, A. C. , Bringmann, L. F. , Schreuder, M. J. , Oldehinkel, A. J. , Wichers, M. & Snippe, E. (Submitted). Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression.

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

Hosmer, D. W., Lemeshow, S. (2000). *Applied Logistic Regression, 2nd Ed*. Chapter 5, John Wiley and Sons, New York, NY, pp. 160-164

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

Kunkels, Y., Knapen, S., Zuidersma, M., Wichers, M., Riese, H., & Emerencia, A. (2019). ACTman: Automated preprocessing and analysis of actigraphy data. *Journal of Science And Medicine in Sport.* doi: 10.1016/j.jsams.2019.11.009

Kunkels, Y., Riese, H., Knapen, S., Riemersma - van der Lek, R., George, S., & van Roon, A. et al. (2021). Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: An actigraphy study. *Translational Psychiatry, 11*(1). doi: 10.1038/s41398-021-01465-w

van de Leemput, I., Wichers, M., Cramer, A., Borsboom, D., Tuerlinckx, F., & Kuppens, P. et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, *111*(1), 87-92. doi: 10.1073/pnas.1312114110

Lemke, M., Puhl, P., & Broderick, A. (1999). Motor activity and perception of sleep in depressed patients. *Journal of Psychiatric Research*, *33*(3), 215-224. doi: 10.1016/s0022-3956(98)00067-3

Mandrekar, J. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316. doi: 10.1097/jto.0b013e3181ec173d

Minaeva, O., Booij, S., Lamers, F., Antypa, N., Schoevers, R., Wichers, M., & Riese, H. (2020). Level and timing of physical activity during normal daily life in depressed and non-depressed individuals. *Translational Psychiatry, 10*(1), 259. https://doi.org/10.1038/s41398-020-00952-w

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

Montgomery, D. C. (2012). Statistical quality control: A modern introduction (7 ed.). New York: John Wiley & Sons Inc.

Pratt, L. A., Brody, D. J., Gu, Q. (2017). Antidepressant Use Among Persons Aged 12 and Over: United States, 2011-2014. *NCHS Data Brief. 283*, 1-8. PMID: 29155679.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raoux, N., Benoit, O., Dantchev, N., Denise, P., Franc, B., Alliale, J., & Widlöcher, D. (1994).Circadian pattern of motor activity in major depressed patients undergoing antidepressant therapy: Relationship between actigraphic measures and clinical course. *Psychiatry Research*, *52*(1), 85-98. doi: 10.1016/0165-1781(94)90122-8

Schat, E., Tuerlinckx, F., Smit, A. C., De Ketelaere, B., & Ceulemans, E. (2021). Detecting mean changes in experience sampling data in real time: A comparison of univariate and multivariate statistical process control methods. *Psychological Methods*. doi:10.1037/met0000447

Scheffer, M., Bascompte, J., Brock, W., Brovkin, V., Carpenter, S., & Dakos, V. et al. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53-59. doi: 10.1038/nature08227

Servaas, M. N., Schoevers, R. A., Bringmann, L. F., Van Tol, M.-J., & Riese, H. (2021). Trapped: Rigidity in psychiatric disorders. *The Lancet Psychiatry*, *8*(12), 1022–1024.

Sim, K., Lau, W. K., Sim, J., Sum, M. Y., & Baldessarini, R. J. (2016). Prevention of relapse and recurrence in adults with major depressive disorder: systematic review and meta-analyses of controlled trials. *International Journal of Neuropsychopharmacology, 19*(2), pyv076. https://doi.org/10.1093/ijnp/pyv076

Smit, A. C., Schat, E., & Ceulemans, E. (2022). The Exponentially Weighted Moving Average Procedure for Detecting Changes in Intensive Longitudinal Data in Psychological Research in Real-Time: A Tutorial Showcasing Potential Applications. Assessment, 107319112210869. https://doi.org/10.1177/10731911221086985

Smit, A. C., & Snippe, E. (2022). Real-time monitoring of increases in restlessness to assess idiographic risk of recurrence of depressive symptoms. *Psychological Medicine, 1*-10.

Smit, A. C., Snippe E., Bringmann, L., Hoenders, & Wichers, M. (In Press). Transitions in depression: If, how, and when depressive symptoms return during and after discontinuing antidepressants. *Quality of Life Research*

Smit, A., 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. doi: 10.1159/000500594

Snippe, E., Smit, A.C., Kuppens, P., Burger, H., Ceulemans, E. (Under review). Recurrence of depression can be foreseen by monitoring mental states with statistical process control.

Sobin, C., Mayer, L., & Endicott, J. (1998). The motor agitation and retardation scale. *The Journal Of Neuropsychiatry And Clinical Neurosciences, 10*(1), 85-92. doi: 10.1176/jnp.10.1.85

van Someren, E., Swaab, D., Colenda, C., Cohen, W., McCall, W., & Rosenquist, P. (1999). Bright light therapy: Improved true positive rate to Its effects on rest-activity rhythms in Alzheimer patients by application of nonparametric methods. *Chronobiology International, 16*(4), 505-518. doi: 10.3109/07420529908998724

Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC press. https://doi.org/10.1201/9780429492563

Todder, D., Caliskan, S., & Baune, B. (2009). Longitudinal changes of day-time and night-time gross motor activity in clinical responders and non-responders of major depression. *The World Journal Of Biological Psychiatry*, *10*(4), 276-284. doi: 10.3109/15622970701403081

Wichers, M., & Groot, P., Psychosystems, ESM Group, EWS Group. (2016). Critical Slowing Down as a Personalized Early Warning Signal for Depression. *Psychotherapy and Psychosomatics*, *85*(2), 114-116. doi: 10.1159/000441458

Wichers, M., Smit, A., & 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

Wissel, C. (1984). A universal law of the characteristic return time near thresholds. *Oecologia*, *65*(1), 101-107. https://doi.org/10.1007/bf00384470

Witting, W., Kwa, I., Eikelenboom, P., Mirmiran, M., & Swaab, D. (1990). Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. *Biological Psychiatry, 27*(6), 563-572. doi: 10.1016/0006-3223(90)90523-5

World Health Organization. (2019). *Depression.* Retrieved 28/01/2020 from https://www.who.int/news-room/fact-sheets/detail/depression

Zanella-Calzada, L., Galván-Tejada, C., Chávez-Lamas, N., Gracia-Cortés, M., Magallanes-Quintanar, R., & Celaya-Padilla, J. et al. (2019). Feature extraction in motor activity signal: Towards a depression episodes detection in unipolar and bipolar patients. *Diagnostics, 9*(1), 8. doi: 10.3390/diagnostics9010008

Zuidersma, M., Riese, H., Snippe, E., Booij, S., Wichers, M., & Bos, E. (2020). Single-subject research in psychiatry: Facts and fictions. *Frontiers in Psychiatry, 11*. https://doi.org/10.3389/fpsyt.2020.539777

Risk Ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during discontinuation of anti-depressant medication

Yoram K. Kunkels[1*], Arnout C. Smit[1], Olga Minaeva[1], Evelien Snippe[1], Sandip V. George[1], Arie M. van Roon[2], Marieke Wichers[1], Harriëtte Riese[1]

Table of Contents

S1 ROC Curves

Receiver Operating Characteristics (ROC)-curves: In the calculated ROC-curves, the *p*-values of the modified MK-tests were used as scoring classifier to construct the ROC-curves. This was done to explore at what cut-off score of the *p*-values the quantifiers would have the best trade-off between false positive rate and true positive rate. Hence, when we choose a higher cut-off *p*-value of the MK-test(meaning that there can be a non-significant change in the EWS, if the cut-off value is above the critical value of the *Meff*-method), the test becomes more sensitive, at the cost of higher false positive rate (right side of the curve). When we choose a lower cut-off *p*-value, the test becomes more specific, at the cost of a lower true positive rate (left side of the curve). The false positive rate can be interpreted as the probability of false alarm of transition. In ROC-curves, the top-left corner denotes the "ideal classifier"; ROC-curves running close to the top-left corner can be interpreted as showing good to excellent performance. However, a curve running exactly along the diagonal line from the bottom left to top-right indicates that there is no predictive value at all. Additionally, the AUC (area under the curve) values were calculated which indicate how well the classifier we constructed performs. The AUC value ranges from 0 to 1 (Mandrekar, 2010), a value of 1 denotes a perfect accurate test, and a value of 0.5 denotes making decisions randomly. Below 0.5, reversing all decisions would improve the test result. Typically AUC values between 0.7 and 0.8 are considered to be acceptable, AUC values between 0.8 to 0.9 are considered excellent, while AUC values above 0.9 are considered to be outstanding (Hosmer & Lemeshow, 2000)[48]. ROC-curves were created in *SPSS* and p-values < 0.05 were considered statistically significant for testing the AUC's.

We used the p-value without the constrained of 'significance'. We used it as a scoring classifier for detecting a transition for p_α ranging from 0.0 to 1.0 and to construct the ROC-curve for an EWS. The performance of each EWS can be compared using the area under the ROC-curve. To implement a detection method for transition, a specific p_α has to be selected, corresponding with an acceptable false positive rate and/or true positive rate. We will not do this, since it depends highly on the costs of true and false positives, which are dependent on health care system and the preferences of the participants clinician.

For the EWS, both the one- and two-sided test, the ROC-curves are shown in Figure S2 and corresponding AUC-values are given in Table S3. From the ROC-curves, one can observe that the two-sided tests outperformed the one-sided tests, with a notable exception for IS.

Regarding individual EWS, the two-sided plot for kurtosis and IV run close to the top-left corner. This can be interpreted as that these EWS performed relatively well in preceding transitions while not suffering as much from giving false positives. Regarding the two-sided tests, the AUC-values of the autocorrelations at lag-1 (AUC = 0.708) and at lag-1440 (AUC = 0.736) are *acceptable*. The AUC-values for variance (AUC = 0.500) and kurtosis (AUC = 0.556) denote sub-par performance. AUC-values for IS was lower than 0.5 (AUC = 0.389), meaning that reversing all decisions based on IS improves the test result.

Post-hoc analyses on potential EWS combinations: Given that there is not much overlap between EWS results, we post-hoc explored combining them. We combined the results of the EWS to detect a transition in a logic OR function, and plotted the performance in a graph with true positive rate as function of false positive rate. We use a binary logistic regression analysis to combine all EWS together to detect transitions. In this analysis, the p-values of the different EWS are weighted to create a new scoring classifier. The result is plotted in the same graph. For example, while variance, kurtosis, and acf-1 all precede only three to four transitions each for different individuals, this small overlap was improved in a combination of EWS to precede transitions in eight out of sixteen individuals (50.0% true positives), while identifying false positives in two out of nine participants (22.2% false positives). Adding the mean to such an EWS combination would allow it to correctly classify transitions in twelve out of sixteen individuals (75.0% true positives), with a false positive rate of 55.6%. More detailed results and information regarding EWS combinations are given in the supplementary materials S7 and S8.

Next, to investigate the unique contribution of each EWS after adjusting for the others (and thus giving appropriate weight to each EWS in this way) binary logistic regressions were used to assess the association between EWS quantifiers and the binary outcome whether a transition was present or not. Results showed that the influence of ill-performing EWS was minimised, while those of well-performing EWS increased. Binary logistic regression can provide us with EWS combinations which should outperform EWS combinations which are designed manually. It is important to note that these results are explorative, and may overfit the current data due to the limited sample size. Hence, the same combinations of EWS can show differing results in a new sample. More detailed results and information regarding EWS combinations are given in the supplementary materials S7 and S8.

S2 Figure of ROC Curves

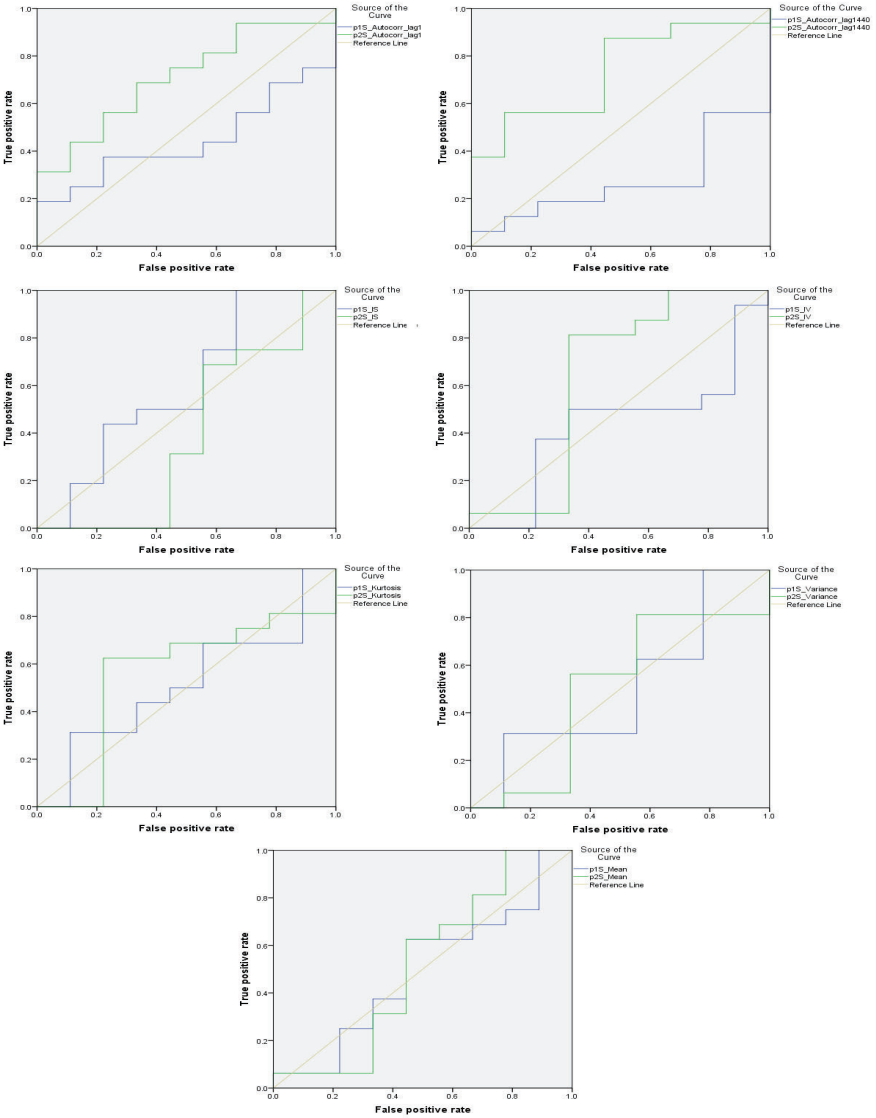


Figure S2: ROC-curves of respectively (from top to bottom and from left to right) variance (AUC = 0.500 p = 0.43), kurtosis (AUC = 0.556, p = 0.65), autocorrelation at lag-1 (acf-1;

AUC = 0.708, p = 0.089), Interdaily stability (IS; AUC = 0.389, p = 0.365), Intradaily Variability (IV; AUC = 0.632, p = 0.282), and autocorrelation at lag-1440 (acf-1440; AUC = 0.736, p = 0.054). ROC= receiver operator curve, AUC-area under the curve.

S3 Table of AUC values

One-sided tested quantifiers (starting with "P1S_") and two-sided tested quantifiers (starting with "p2S_") showing area under the curve (AUC) values.

Area Under the Curve

				Asymptotic 95% Confidence Interval	
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Lower bound	Upper bound
p1S_Autocorr_lag1	0.444	0.115	0.651	0.219	0.670
p1S_Autocorr_lag1440	0.271	0.102	0.062	0.071	0.471
p1S _IS	0.597	0.130	0.428	0.343	0.851
p1S _IV	0.431	0.123	0.571	0.189	0.673
p1S _Kurtosis	0.514	0.123	0.910	0.273	0.753
p1S _Variance	0.500	0.129	1.000	0.247	0.753
P1S_Mean	0.493	0.126	0.955	0.247	0.740
p2S_Autocorr_lag1	0.708	0.105	0.089	0.502	0.915
p2S_Autocorr_lag1440	0.736	0.102	0.054	0.536	0.937
p2S_IS	0.389	0.137	0.365	0.121	0.657
p2S_IV	0.632	0.142	0.282	0.353	0.911
p2S _Kurtosis	0.556	0.130	0.651	0.301	0.810
p2S _Variance	0.500	0.133	1.000	0.240	0.760

p2S_Mean	0.514	0.136	0.910	0.247	0.781

a. *Under the nonparametric assumption*

b. *Null hypothesis: true area = 0.5*

S4 Two-sided tests

Two-sided tests

The number of true positives for the 2-sides test was that 15 out of 16 unique transitions were preceded correctly by a significant change in at least one EWS (93.75%; instead of 14 out of 16 (87.5%) described for the 1-sided tests above).

Unfortunately, the false positive rate also increased when testing two-sided. Instead of 66.7% of false positives in the one-sided test, the false positive rate increased to 100% when testing two-sided. This meant that the increase in false positive rate was much larger than the increase in true positive rate, and testing two-sided was strongly detrimental to the accuracy of this method. Only in the variable acf-1440, did the true positive rate increase substantially more than the false positive rate, as the true positive rate increased by 31.3% compared to a 11.1% increase in false positive rate. However, this result should be interpreted with caution, as it was a single result from multiple post-hoc tests, and the significance of this difference could not be tested due to the limited sample size.

S5 Table of two-sided EWS

Two-sided performance of EWS, calculated over participants' actigraphy data, in identifying upcoming transitions in depressive symptoms. The upper panel shows the transition group wherein green cells indicate the true positives. The lower panel shows the non-transition group wherein red cells indicate the false positives. In both panels the yellow cells indicate zeroes (no results).

	Critical slowing down based EWS		**Context EWS**			**Mean**		
Transition group (N= 16)								
ID	Variance	Kurtosis	acf-1	IS	IV	acf-1440	Mean	**TOTAL**
1036	1	0	1	0	0	0	0	2
1045	0	0	0	0	0	0	0	0
1046	1	1	0	1	0	0	0	3
1052	0	1	1	0	0	1	0	3
1074	1	0	0	0	0	1	1	3
1075	0	1	0	0	0	0	1	2
1076	0	1	0	0	0	0	0	1
1077	0	0	1	0	0	0	0	1
1108	0	1	0	0	0	1	0	2
1133	0	0	0	0	0	1	0	1
1173	1	0	1	0	0	1	1	4
1181	0	0	0	1	1	0	1	3
1193	1	0	0	0	1	0	0	2

1224	1	0	0	0	0	0	1	2
1264	1	0	1	0	0	1	0	3
1293	0	0	0	0	0	0	1	1
TOTAL	7	5	5	2	2	6	6	
Percentage:	43.75%	31.25%	31.25%	12.50%	12.50%	37.50%	37.50%	93.75%

Non-Transition group (N = 9)

1041	1	0	0	1	1	0	1	4
1059	0	0	0	1	0	0	0	1
1067	0	0	0	0	0	1	0	1
1110	1	1	1	0	0	0	1	4
1178	1	0	0	0	1	0	1	3
1180	0	0	0	0	1	0	0	1
1255	0	0	0	0	0	1	0	1
1280	0	1	0	0	0	0	0	1
1295	0	0	0	1	0	0	0	1
TOTAL	3	2	1	3	3	2	3	
Percentage:	33.33%	22.22%	11.11%	33.33%	33.33%	22.22%	33.33%	100.0%

S6 Table of frequencies given a 7-day window

Frequency of significant increases in the period preceding a transition in depressive symptoms in individuals experiencing a transition and in individuals who did not have a transition into more severe depressive symptoms. Moving window size was set at 7 days, instead of 14 days in the analyses in the main part of the manuscript.

	acf-1	acf-1440	IS	IV	Kurtosis	Variance	Total
Transition group	2	3	2	2	2	3	14
No transition group	0	1	4	0	0	2	7

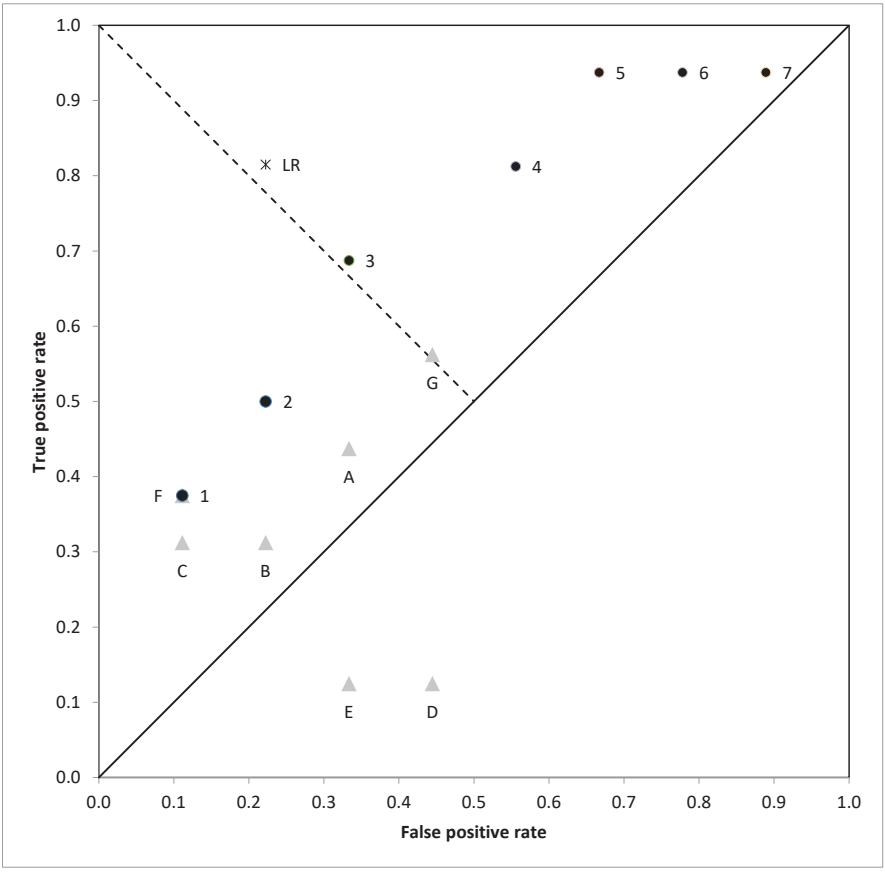
Note: acf-1, autocorrelation at-lag-1; acf-1440, autocorrelation at-lag-1440; IS, interdaily stability; IV, intradaily variability.

S7 Figure of true positive rate and 1-False positive rate-curve - two-sided

Introduction, method, and discussion

We combined the results of the EWS to detect a transition in a logic OR function, and plotted the performance in a graph with true positive rate as function of false positive rate. We use a binary logistic regression analysis to combine all EWS together to detect transitions. In this analysis, the p-values of the different EWS are weighted to create a new scoring classifier. The result is plotted in the same graph. In S7 and S8 we thus combine multiple risk quantifiers into an aggregate measure that increased the accuracy of the method. Binary logistic regressions were used herein to assess the association between EWS quantifiers and the binary outcome whether a transition was present or not. Results showed that the influence of ill-performing EWS was minimised, while those of well-performing EWS increased. Binary logistic regression can provide us with EWS combinations which should outperform EWS combinations which are designed manually. However, the current dataset was too small to apply cross-validation, meaning that the results likely overfit the current sample and it is thus yet unclear whether the same aggregate measures would also perform similarly in new samples. Therefore, larger studies are required to test the true potential of aggregate risk quantifiers.

Performance of the EWS and combinations using the two-sided tests. Gray triangles represent the individual EWS, whilst the black dots represent the cumulative total. The triangles correspond respectively to: A = variance, B = kurtosis, C = acf-1, D = IS, E = IV, F = acf-1440, G = mean. Dashed line: line of optimal solutions for equal costs for true and false positives. The results of our binary logistic regression are shown in this figure denoted by LR.

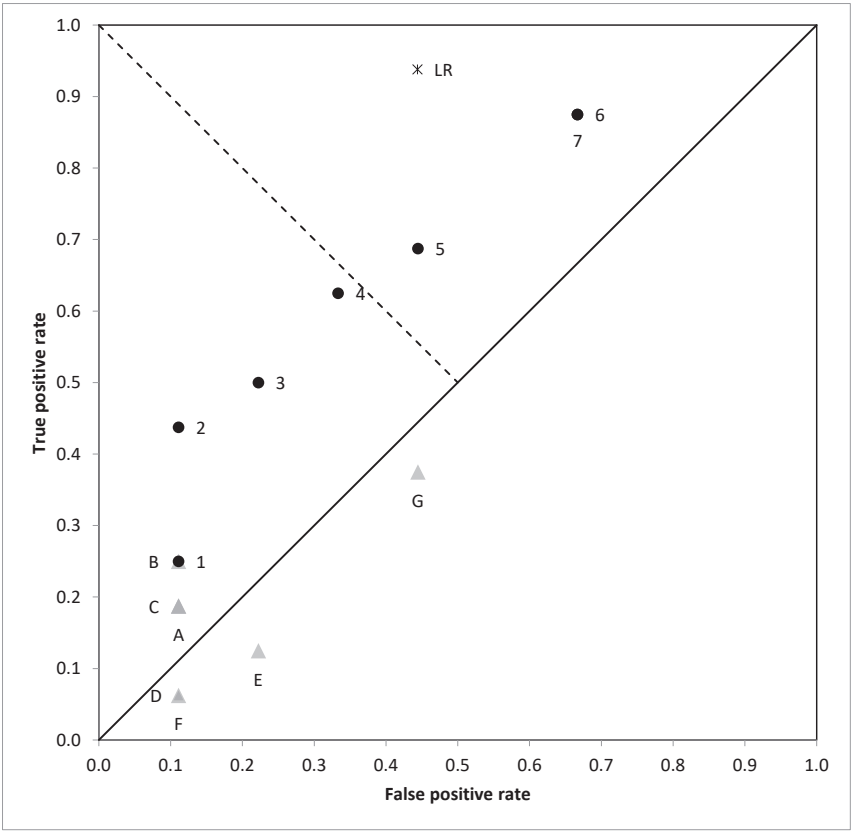


Note: A = variance, B = kurtosis, C = acf-1, D = IS, E = IV, F = acf-1440, G = mean. 1 = (acf-1440), 2 = (acf-1440 + acf-1), 3 = (acf-1440 + acf-1 + kurtosis), 4 = (acf-1440 + acf-1 + kurtosis + variance), 5 = (acf-1440 + acf-1 + kurtosis + variance + mean), 6 = (acf-1440 + acf-1 + kurtosis + variance + mean + IV), 7 = (acf-1440 + acf-1 + kurtosis + variance + mean + IV + IS). LR Logistic regression solution (cut off = 0.59, 80% correctly classified).

S8 Figure of true positive rate and 1-False positive rate-curve - one-sided

Introduction, method, and discussion

See S7. Performance of the EWS and combinations using the one-sided tests. Gray triangles represent the individual EWS, whilst the black dots represent the cumulative total. The triangles correspond respectively to: A = variance, B = kurtosis, C = acf-1, D = IS, E = IV, F = acf-1440, G = mean. Dashed line: line of optimal solutions for equal costs for true and false positives. The results of our binary logistic regression are shown in this figure denoted by LR.



Note: A = variance, B = kurtosis, C = acf-1, D = IS, E = IV, F = acf-1440, G = mean. 1 = (kurtosis), 2 = (kurtosis + acf-1), 3 = (kurtosis + acf-1 + variance), 4 = (kurtosis + acf-1 + variance + IV), 5 = (kurtosis + acf-1 + variance + IV + acf-1440), 6 = (kurtosis + acf-1 + variance + IV + acf-1440 + mean), 7 = (kurtosis + acf-1 + variance + IV + acf-1440 + mean + IS). LR Logistic regression solution (cut off = 0.5, 80% correctly classified).

5.

Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed.

This chapter has been published as:

George, S. V., Kunkels, Y. K., Booij, S., Wichers, M. (2021). Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-92890-w

Chapter 5: Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed.

Sandip V. George, PhD[1*], Yoram K. Kunkels, MSc[1], Sanne Booij, PhD[1,2,3], Marieke Wichers, PhD[1]

[1] Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), University of Groningen, University Medical Center Groningen (UMCG), Groningen, The Netherlands.

[2] Faculty of Behavioral and Social Sciences, Department of Developmental Psychology, University of Groningen, Groningen, The Netherlands.

[3] Center for Integrative Psychiatry, Lentis, Groningen, The Netherlands.

This chapter has been published as; George, S. V., Kunkels, Y. K., Booij, S., & Wichers, M. (2021). Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-92890-w

Abstract

While the negative association between physical activity and depression has been well established, it is unclear what precise characteristics of physical activity patterns explain this association. Complexity measures may identify previously unexplored aspects of objectively measured activity patterns, such as the extent to which individuals show repetitive periods of physical activity and the diversity in durations of such repetitive activity patterns. We compared the complexity levels of actigraphy data gathered over 4 weeks (~40000 data points each) for every individual, from non-depressed (n=25) and depressed (n=21) groups using recurrence plots. Significantly lower levels of complexity were detected in the actigraphy data from the depressed group as compared to non-depressed controls, both in terms of lower mean durations of periods of recurrent physical activity and less diversity in the duration of these periods. Further, diagnosis of depression was not significantly associated with mean activity levels or measures of circadian rhythm stability, and predicted depression status better than these.

Introduction

The association between physical activity and depression is well documented (1–3). Group-level studies on levels of physical activity have shown an inverse association between physical activity and depressive symptoms (4,5). Longitudinal studies, including intervention studies have shown that physical activity and exercise reduces symptoms and improves mood in individuals suffering from depression (6,7). Studies have also shown that the association between depression and physical activity, may be bidirectional (7,8). An unresolved question, however, is what precise characteristics of activity patterns are responsible for the impact of physical activity on mental health.

Currently, most interventions work under the assumption that it is only the pure level of activity that impacts mood. Other studies, however, argue that, in addition to mean levels of physical activity, diurnal rhythms in activity are relevant in explaining why activity levels are associated with depression (9–11). Healthy people are most active closer to the middle of the day and less active in mornings and evenings. Studies show that in depressed people, the activity rhythm peaks later than in healthy people (11–14). The shifted timing influences sleep quality and diminishes levels of positive affect during the day (15,16), and is hypothesized to contribute to depression. However, studies have been inconsistent and effect

sizes are small (17,18), suggesting that there may be other possible explanations for the association between activity patterns and depression. Currently, studies have only examined mean activity levels or activity rhythms with a constant periodicity (daily, weekly or annual activity rhythms) in association with depression (19–21) and were not able to extract other sorts of recurring activity patterns, with varying periodicity, which may be relevant.

A new approach to examine activity patterns is to use tools from complexity science (22–24). Complexity measures may provide new and complementary information on the nature of activity patterns in daily life as these measures are able to differentiate random activity spikes (noise) from patterns of activity that seem to repeat themselves, even if they vary in the timing of return (varying periodicity). Actigraphy patterns have been shown to be made up of a combination of intrinsic and extrinsic factors which give rise to activity time-series that have a complex nonlinear mechanism overlaid with random fluctuations (25). For example, restlessness behavior, which would generate movements at unpredictable moments is expected to show up as noise, rather than as specific repeating patterns of activity, whereas moments of sport, biking to work, or certain social activities that would involve elevated levels of physical activity would reveal themselves in repeating patterns of activity. Only these latter activity patterns add to the calculated complexity of the signal. Furthermore, not only the amount of repetition, but also the variety of repeating patterns adds to the calculated complexity of the measure. For example, differing durations of activities like biking, swimming or running would each cause a particular pattern of activity. When these activities are repeated in time, they constitute a diversity of recurrent physical activity patterns, which add to the complexity of the signal. Complexity measures would thus provide an objective way to measure to what extent these different types of physical activity (noise versus repeating activity patterns) are present in people with (risk for) depression. If we get a better understanding of what activity patterns differentiate depressed versus healthy people, this may not only provide more insight in how physical activity relates to depression, but, in the case that such patterns are causal to depression, it may also bring new possibilities for diagnostic tools to evaluate whether the patient exhibits healthy physical activity patterns. Moreover, complexity measures quantify an aspect of physical activity that is not captured by existing methods such as the mean activity levels or non-parametric circadian rhythm variables.

The study of physical activity patterns using small motion sensor detectors (accelerometers) that are encased in a unit about the size of a wristwatch and can be worn continuously for days to months, is called actigraphy (26). Studying actigraphy patterns of individuals has becomes increasingly popular (27). Actigraphs estimate levels of physical activity in an objective way, without recall bias (28,29). Furthermore, the measurement of activity patterns using these light-weight devices, is non-invasive, with low burden to the participants, and therefore allows for the possibility of long-term monitoring of physical activity patterns.

In the current study, we aim to understand the differences in the complexity of recurrent physical activity patterns of depressed and non-depressed individuals using actigraphy. In line with the argumentation given above we expect decreased levels of complexity (lower duration and diversity of recurrent activity patterns) in depressed people versus non-depressed people. For this purpose, we will use a unique sample of depressed and non-depressed people who were monitored for a month with accelerometers.

Materials and methods

Sample. The data used was collected as part of the Mood and Movement in Daily Life (MOOVD) study, which aims to study the dynamic association between mood and physical activity (30,31). All participants were aged between 20 and 50, and were monitored for 30 days using electronic diaries, actigraphy and saliva samples. In this paper we will use the actigraphy data and the diagnostic interview data on depression diagnosis. Data was obtained from 54 participants (depressed to non-depressed ratio 1:1) who were pair matched on gender, BMI, smoking status, and age. The participants were screened for their severity of depression based on their scores on the Beck Depression Index (BDI) (32). Scores below 14 are associated with minimal depression, while scores of 14 and above are associated with mild, moderate or severe levels of depressive symptoms. Participants scoring above 14 and participants scoring below 9, were invited for a diagnostic interview to establish whether they fulfilled the criteria for depression or were free of any mood disorders, respectively. For further details, we refer the reader to (30).

Measurements

Physical activity was measured using the ActiCal© (*Respironics,Bend,OR*) which is an omni-directional, water-resistant actigraph, which was worn on the non-dominant wrist. The activity counts were sampled at 1 minute-intervals and were used as the measure for physical

activity. Details of how activity measurements are conducted in ActiCal can be found in Heil, 2006 (33).

Statistical analysis

Our analysis is primarily focused on the recurrence quantification analysis (RQA) of actigraphy data. It quantifies the relative abundance, duration and diversity of recurrent patterns in a time-series. This kind of analysis has proved to be very useful in many different fields in science, including psychology (34–36).

Data pre-processing

Prior to the recurrence quantification analysis, we carried out two preprocessing steps. The first reduced the overall size of the data by resampling. This is achieved by averaging the data through 10-minute bins. This averaging or binning step gave us the average activity counts every ten minutes, which reduced the total length of the time-series, and computational time needed by the algorithm. To maintain uniformity, all datasets were constrained to a length of *4000* data points after binning, which gave us nearly *28* days or four weeks of data per participant. All datasets were ensured to have at least *3000* data points or about 21 days of data.

A second preprocessing step involved a rank transformation on the data (37–39). This analysis focused on methods that depend mostly on the ordering and rhythms in the time-series. Hence we rank transformed all the datasets initially, resulting in a uniform amplitude distribution. The resulting transformation preserved the rank and time ordering and consequently the dominant periodicities of the time-series. Using the rank transformation made sure that the quantifiers derived from the activity counts time-series are not affected by extreme events, such as sudden spurts in activity. In addition the transformation put all the time-series from different subjects onto an equal footing when it came to amplitude. This became especially useful in the context of the choice of a recurrence threshold, which we describe in the next subsection. The results from this rank transformed data, were then only related to the ordering of the time-series and not to the actual amounts of activity counts by a subject. This conversion was done by replacing a point in the time-series by its rank in the time-series. The resulting time-series of ranks was then divided by N, which was the total length of the time-series, which constrained the distribution between 0 and 1.

Recurrence quantification analysis

We initially conducted a recurrence quantification analysis (RQA) in all individuals in the MOOVD dataset. Then we compared the recurrence plot properties of the non-depressed and depressed groups with regard to complexity measures. Recurrence plots are simple binary plots that visualize the pattern of repetitions or rhythms in a time-series. Fourier transform based methods capture repetitions that are periodic, whereas circadian rhythm variables consider rhythms at a day level. RQA is free from these constrains and the exact patterns in a recurrence plot give us a deeper understanding about the nature of the underlying dynamics that the time-series is derived from.

A recurrence plot reveals the patterns a system makes when it revisits the same neighborhood of space. When the dynamics of a system is purely stochastic, the recurrence plot shows no discernable patterns. On the other hand, when the system shows deterministic behavior the recurrence plot shows distinct patterns in the form of horizontal and diagonal lines. These are quantified using RQA.

A recurrence plot is constructed in the following way. A recurrence threshold or distance is first chosen, say ε. The time-series is then scanned such that all points that fall within ε distance of each time-series point is identified. The recurrence plot is then generated as a planar plot of ordered time-series points along the x and y axes. A schematic describing this process is shown in Figure 1, where a region in blue, of size 2ε, is marked in the time-series in the upper panel to demonstrate this recurrence threshold. If two time-series points fall within ε distance of each other, the corresponding point is marked using a dark spot in the recurrence plot. In the time-series in Figure 1, all points within the blue rectangle in the upper panel are marked as black points within the shaded region in the lower panel. The recurrence quantification analysis in this paper is conducted using the free standalone software, TOCSY (40–42).

Prior to studying the structures in the recurrence plot, the threshold ε needs to be identified. In this study, we set the recurrence threshold by constraining the recurrence rate or density of dark points in the plot. The recurrence rate gives a probability that a specific state will recur. For our study we set the recurrence threshold as the distance where the density is 0.05, i.e. about 5% of the recurrence plot is made of dark points. A fixed density has been used previously in multiple studies to determine ε and is known to be useful in detecting finer

changes in the recurrence plot structure(43,44). A flowchart describing the process is shown in Figure 2.

Once the recurrence threshold is fixed, we quantify two main structures in the recurrence plot, the diagonal and vertical lines. The diagonal line structures in the recurrence plot are associated with the level of determinism in the time-series, since random processes will show these structures very rarely, whereas deterministic processes tend to show these structures more. It occurs either when a part of the time-series changes monotonically or when two parts of the time-series show similar local evolution or change. The vertical lines on the other hand indicate periods of "stasis" or very slow evolution. In a sense, it shows the length of the activity, with longer vertical lines suggesting an activity that lasts for longer. We are primarily interested in the mean and entropy, which represent the duration and diversity associated with the distributions of the diagonal and vertical lines(41). The average of the diagonal line distribution shows the average duration of recurring physical activity patterns in a time-series. The entropy quantifies the diversity associated with the diagonal structures in the recurrence plot. This provides a measure of the extend of time scales involved in the diagonal line distribution. Similarly the mean of the vertical line distribution shows the mean levels of stasis associated with the physical activity patterns (i.e how long an activity persists) and the entropy yields the diversity associated with the vertical line distribution(41). Another important quantifier that is associated with the diagonal line structure is called the determinism or DET measure. The DET measure reflects the ratio of points that form diagonal structures to the ratio of all recurring points. Thereby, it provides an estimate of how often different parts of a time-series co-evolve as a fraction of the total number of data point pairs in the plot. For a purely noisy process, with no underlying dynamics, this measure is very small, whereas for a process with underlying deterministic dynamics, the DET measure is high. Similarly, the laminarity or LAM measure reflects the ratio of points that form vertical structures to the ratio of all recurring points. This provides an estimate of how often slowly evolving processes occur, as a fraction of the total number of data point pairs. For instance, frequent periods of rest or physical activity that results in constant activity counts for an extended duration will lead to a higher LAM measure as opposed to cases when such patterns are rare. A further useful quantifier in this context is the ratio of the LAM to DET measure, which quantifies how often vertical structures appear in the system as a fraction of diagonal structures (45). While all the quantifiers mentioned above relate to the complexity of patterns found in the recurrence plot, the mean and entropy of the distributions relate directly

to the duration of recurrent activity patterns and the diversity of such patterns. A summary of the recurrence-based quantifiers used in this paper is given in Table 1.

To illustrate the difference between time-series data that is dominated by noise processes and one which is dominated by a periodic signal, we contaminated a sine wave with different levels of noise. Sample recurrence plots from these noise contaminated sine waves, along with a pure sine wave and a pure white noise signal, are shown in Figure 3. The simulation of sine waves contaminated with noise shows instances where a strong rhythm along with randomness is present, similar to daily rhythms which are prominent in actigraphy data. As the random component becomes larger, the recurrence plot becomes more diffused.

Missingness

Many datasets showed periods of non-wear in the beginning or the end of the collection period. Such periods were removed through visual inspection. Datasets that were left with less than 3000 measurement points, after resampling into 10-minute bins, were eliminated initially. After the recurrence plot construction, all datasets for which the density threshold of 0.05 was exceeded at very low recurrence thresholds, were eliminated. This happens when the dataset has considerable periods of inactivity, which leads to cluttering in the recurrence plot.

Hypothesis testing

Group differences in complexity were examined using a t-test. The t-test for independent samples checks if two independent groups have identical mean values. We use the Welch t-test which does not assume equal population variance, and generalizes to unequal sample sizes (47). All statistical analyses were performed using the scipy package in python (48).

Traditional actigraphy quantifiers

In order to check for differences in discriminative ability and overlap between the current complexity measures and more traditional measures such as mean levels of activity and circadian rhythm variables, the latter variables were extracted as well from the actigraphy data. Both the mean and circadian rhythm variables have been used in differentiating healthy and depressed individuals from actigraphy data. The mean activity was calculated as the average number of activity counts per individual. The circadian rhythm variables used were the interdaily stability (IS), intradaily variability (IV) and relative amplitude (RA) (49). The IS quantifies stability of the rhythm between days. It can vary between 0 and 1, with higher

values indicating a more stable daily rhythm. The IV indicates the fragmentation of the sleep-wake rhythm and varies roughly between 0 and 2. Higher values indicates higher fragmentation. The RA gives a description of how different the most active and least active periods in a day are. Further details about calculation may be found in (50). Circadian measures were calculated using the ACTman package in R (51). Using t-tests, we distinguish whether the means of the distributions of values between the two groups were significantly different. Furthermore, we calculated the correlations between the complexity and the above-mentioned variables, using the Spearman rank correlation coefficient. Apart from being robust to outliers, the rank correlation coefficients have the added advantage that they finds correlation even if the monotonic relationship between the covariates is nonlinear(39). Finally, we used logistic regression to predict the diagnostic status with traditional actigraphy quantifiers and recurrence quantifiers. The pseudo-R^2 values were then compared between the different quantifiers to quantify goodness of fit.

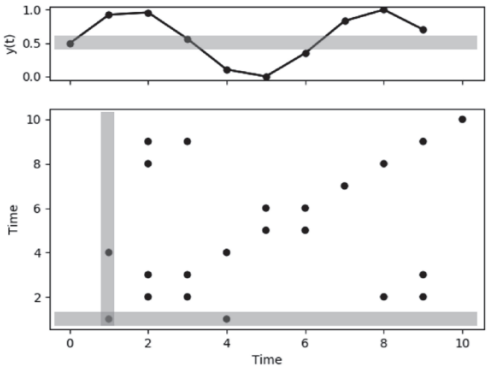


Figure 1. Schematic describing the construction of a recurrence plot. The upper panel shows the series of observations vs time. A region centered on the first data point, with width 2ϵ is shaded in blue. The lower panel shows the corresponding recurrence plot. The elements of the recurrence plot corresponding to the first point are shaded in blue, in the lower panel. When a point falls within the blue rectangle in the upper panel, it is shown as a black point in the lower panel. This analysis is repeated for every point in the time series resulting in the complete recurrence plot. The x and y axes of the recurrence plot represent the time of observation (x axis of the upper panel). Hence, when an observation $y(t_1)$ at time t_1 and $y(t_2)$ at time t_2 are within ϵ distance of each other, the point (t_1, t_2) is marked in black in the recurrence plot.

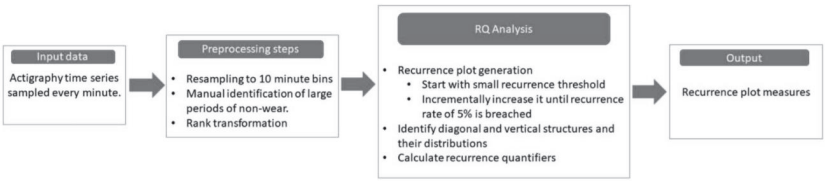


Figure 2. Flowchart indicating the analysis procedure, described in this paper, to extract recurrence plot measures from actigraphy data.

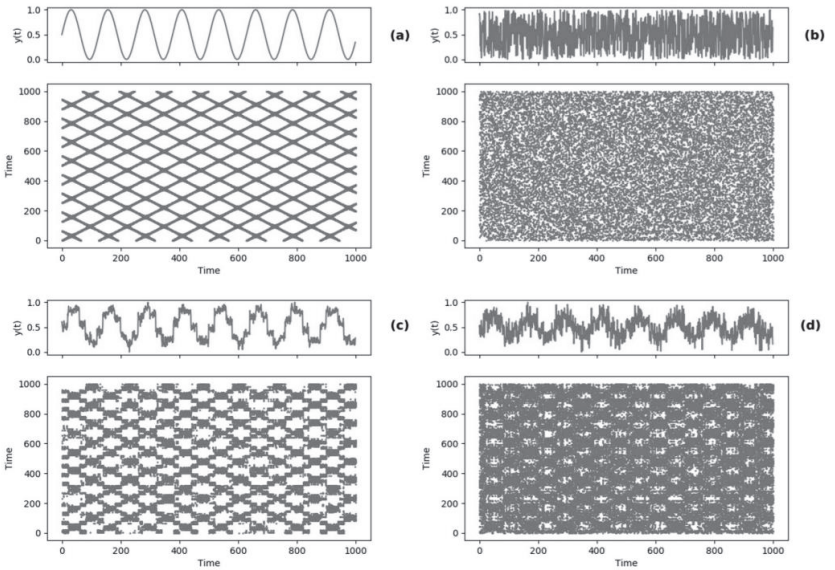


Figure 3. Sample time series and corresponding recurrence plots for (**a**) a pure sine wave (**b**) random noise (**c**) sine wave contaminated with additive white noise with signal to noise ratio (SNR) 5 and (**d**) sine wave contaminated with additive white noise with signal to noise ratio (SNR) 1. A higher SNR implies that the signal is more prominent as compared to the noise.

Quantifier	Calculation	Definition	Interpretation
DET	Ratio of diagonal structures to total recurrence points.	Level of deterministic activity in the data.	Lower levels indicate more randomness
LAM	Ratio of vertical structures to total recurrence points.	Level of slowly evolving processes in the time series.	Higher levels indicate more activities that linger
L_{avg}	Mean length of diagonal structures	Average duration of recurrent physical activity	Higher levels indicate longer recurrent physical activities
L_{ent}	Entropy of diagonal line distribution	Diversity of durations of recurrent physical activity patterns	Higher levels indicate recurrent physical activities of varying durations
V_{avg}	Mean length of vertical structures	Average duration of static activity patterns	Higher levels indicate lingering physical activities that last longer
V_{ent}	Entropy of vertical line distribution	Diversity of durations of static activity patterns	Higher levels indicate lingering physical activities of varying durations
$\frac{LAM}{DET}$	Ratio of LAM to DET measures	Level of statis as compared to deterministic structure	Changes in this ratio has been shown to be an indicator of change in stability[45,46].

Table 1. Definitions and interpretations in the context of activity data, for various recurrence plot quantifiers that are used in this work. DET: Determinism; LAM: Laminarity; L_{avg}: Average diagonal line length; L_{ent}: Entropy of diagonal line distribution; V_{avg}: Average vertical line length; V_{ent}: Entropy of vertical line distribution; $\frac{LAM}{DET}$: Laminarity to determinism ratio.

Results

We present the results of the between-group analysis comparing the RQA measures of the depressed group with the non-depressed group. One dataset was excluded initially due to insufficient data (< 3000 points). Another seven datasets were excluded due to cluttering in the recurrence plot which led to a recurrence rate larger than 0.05 even at very low values of the ε threshold. This left 21 depressed and 25 non-depressed participants. Sample recurrence plots from a non-depressed and a depressed individual are presented in Figure 4.

Descriptives and traditional actigraphy differences between the groups

Differences in demographic and clinical characteristics between the depressed and non-depressed subjects are shown in Table 2. In line with the fact that demographic variables are pair matched in the MOOVD study, we did not observe any significant differences. In Table 3 we showed the group differences between commonly used quantifiers of actigraphy, namely the mean activity counts and nonparametric circadian rhythm variables for actigraphy analysis proposed in (49). No significant differences for these quantifiers between the two groups were observed.

Results of recurrence quantification analysis

We then checked the mean differences in the recurrence plot parameters for the two groups. A significant ($p < .05$) difference between the two groups in the mean and entropy of the diagonal line length distribution and in the ratio of LAM to DET was found. The depressed group showed lower mean and entropy for the diagonal line distribution, whereas it showed a larger LAM to DET ratio, compared to the non-depressed group. The distributions of these three measures for the depressed and non-depressed groups are depicted in Figure 5.

Overall, it was found that almost all the complexity measures were either significant ($p<.05$) or borderline significant ($p < .1$) and in the expected direction. The means of the considered measures for the two groups and the corresponding t-statistic and p-value are shown in Table 4. A Bonferroni correction in the level of α for testing significance when conducting multiple tests causes all significant results to be lost.

A logistic regression was used to check whether the individual recurrence measures predicted the depression status better than the traditional actigraphy measures. The recurrence measures were found to have higher pseudo R^2 values than traditional actigraphy measures. Moreover, using a combination of traditional actigraphy measures and novel recurrence plot based

actigraphy measures yielded a higher R^2 value than each did individually, suggesting that they provide complementary information. For details, please refer to the supplementary material.

Correlations

Finally, the correlations of the recurrence plot parameters with the traditional variables considered earlier in this section and with each other were analyzed to examine the extend of overlap between these quantifiers. These are listed in Table 5. We see a perfect correlation between the mean and entropy measures of the diagonal line distribution. Correlations between average level of physical activity and diurnal rhythm variables on the one hand and complexity measures on the other hand were generally small. One weak but significant correlation was found between the interdaily stability and the LAM to DET ratio.

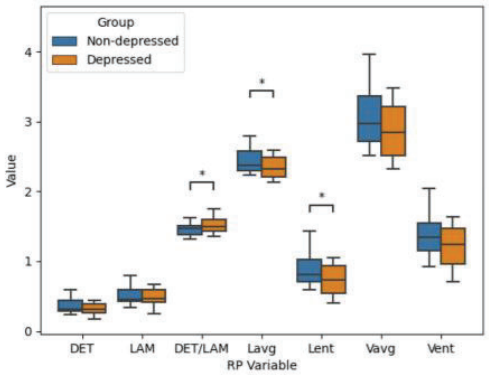


Figure 5. Box plots showing the differences between the non-depressed and depressed groups for the different recurrence plot variables. Significant differences ($p < .05$) are marked with asterisks ($*$).

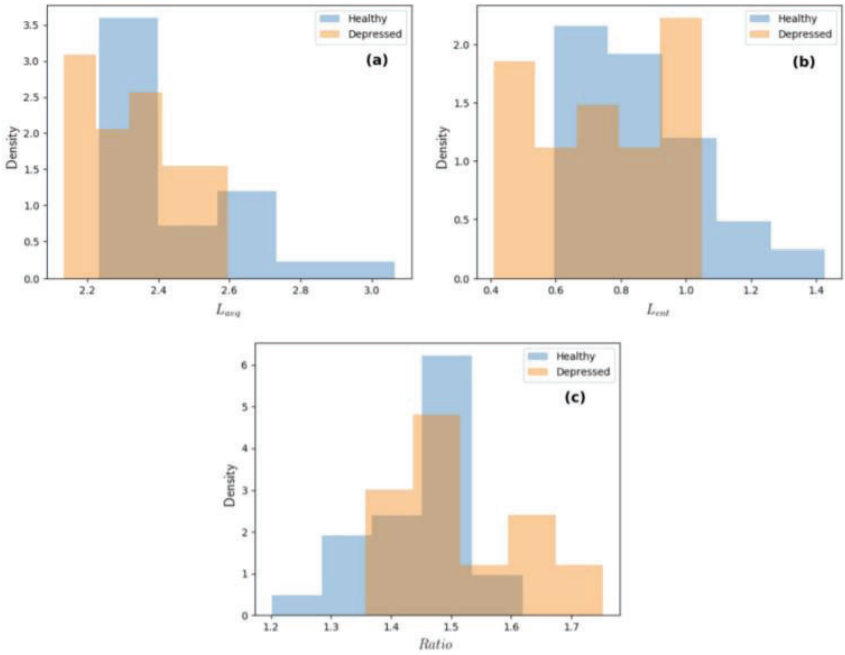


Figure 6. Histograms showing the difference in (**a**) mean diagonal length, (**b**) entropy and (**c**) ratio of determinism to laminarity between the non-depressed and depressed groups.

Property	$Mean_N \pm SD$	$Mean_D \pm SD$	t-statistic	p-value	Cohen's d
DET	0.359±0.11	0.308±0.087	− 1.725	0.091	0.511
L_{avg}	2.436±0.195	2.334±0.142	− 2.018	0.050*	0.594
L_{ent}	0.853±0.21	0.725±0.202	− 2.104	0.041*	0.634
LAM	0.511±0.13	0.466±0.119	− 1.168	0.249	0.353
$\frac{LAM}{DET}$	1.446±0.09	1.517±0.11	2.336	0.025*	− 0.721
V_{avg}	3.079±0.475	2.864±0.367	− 1.731	0.090	0.512
V_{ent}	1.352±0.28	1.206±0.257	− 1.707	0.095	0.520

Table 4. Differences in means for the different recurrence plot measures for the non-depressed($n = 25$) and depressed($n = 21$) groups. The measures that show statistically significant ($p < .05$) differences are indicated with an * DET: Determinism; L_{avg}: Average diagonal line length; L_{ent}: Entropy of diagonal line distribution; LAM: Laminarity; $\frac{LAM}{DET}$: Laminarity to determinism ratio; V_{avg}: Average vertical line length; V_{ent}: Entropy of vertical line distribution, $Mean_N$: Mean of the non-depressed group; $Mean_D$: Mean of the depressed group; SD: Standard deviation.

	L_{avg}	L_{ent}	$\frac{LAM}{DET}$	Avg Activity	IS	IV	RA
L_{avg}	1	0.999*	− 0.609*	0.091	0.073	0.031	0.201
L_{ent}		1	− 0.607*	0.107	0.073	0.035	0.206
$\frac{LAM}{DET}$			1	0.286	− 0.314*	− 0.098	0.013
Avg Activity				1	− 0.162	− 0.132	0.157
IS					1	0.355*	− 0.140
IV						1	− 0.185
RA							1

Table 5. Correlation between commonly used actigraphy measures and recurrence plot measures . Only the recurrence plot measures that showed significant differences between the two groups are listed. The table lists the Spearman correlation coefficient. Significant correlations($p < 0.05$) are marked with an *. IS: Interdaily stability; IV: Intradaily variability; RA: Relative Amplitude; L_{avg}: Average diagonal line length; L_{ent}: Diagonal line entropy; $\frac{LAM}{DET}$: Laminarity to determinism ratio.

Discussion

This work explored how complex recurrent patterns in physical activity are associated with depression. Significant differences between the non-depressed and depressed groups for multiple recurrence plot quantifiers, which were related to duration and diversity of physical activity patterns were observed. Using recurrence quantification analysis an overall lower level of the complexity of recurrent longitudinal patterns in depressed patients versus controls was shown. While the study does not conclusively prove differences in complexity between the two groups, especially after taking into account multiple testing corrections, it does leave room for cautious optimism about using these quantifiers to study depression using physical activity data, in future research.

The current work represents an important first step in multiple ways. The methodology used in this work goes beyond the way how classical approaches related activity patterns to depression. By using novel methods from complexity science, we were now for the first time able to capture other relevant aspects of recurrent temporal patterns of physical activity such as the duration and diversity of such activity patterns that have varying periodicity. We were also able to relate these novel aspects of physical activity to depression. Furthermore, this method allowed for discriminating noise patterns, which do not contribute to the complexity measures examined, from specific recurrent activity periods which do add to the calculated complexity measures.

Various resulting complexity measures significantly associated with a diagnosis of depression, whereas traditional measures such as mean level and diurnal rhythm measures did not. Moreover, complexity measures predicted the depression status better than the traditionally used actigraphy measures. This implies, first, that depressed people showed lower total duration of specific recurrent activities, such as walking, biking, or other sportive activities and less diversity in the durations of such activities. It may be these differences that mainly characterize how physical activity is different between depressed and non-depressed people. We should note, though, that in these actigraphy measures duration and diversity overlapped almost completely. Thus, people with longer duration of activities also showed higher diversity of activities. This means that in this study we are unable to differentiate between these two aspects of complexity. Second, the fact that the complexity measures performed better in discriminating the two groups than simply the mean value of physical activity is worth noting. Reduction in physical activity is known to be a defining characteristic of depression and studies using objective actigraphy have shown that depressed

individuals have a lower level of physical activity than individuals without depression(11,52). In this context, the lack of significant differences in physical activity between the depressed and non-depressed groups, which was also observed previously in a sub sample of the same study(31), is striking. One possible reason could be that the MOOVD study which focused on mood and movement attracted individuals with higher activity to begin with. Further, large variations in the BDI-II scores between baseline and follow-up were observed in the depressed group (see Table 2). Hence the mean physical activity per person is possibly averaged over periods with differing levels of depressive symptoms, resulting in less significant differences between the two groups. Taken together, this study makes a case for using more complex measures to understand reduction in physical activity in depression and suggests the need for within subject studies to understand the same. A previous study suggested that more complex dynamic measures of variables of interest in the field of psychopathology would not be able to contribute more information than the mean (53). This study, as well as other recent studies (54,55) suggest otherwise. Third, as there was minimal overlap between the complexity and the traditional measures as shown by the correlations in Table 5, findings suggest that the complexity measures provide complementary information regarding activity patterns over and above currently used indicators. This finding is supplemented with a linear regression analysis (Supplementary 1), which showed that the goodness of fit obtained using recurrence-based variables are consistently higher than those achieved by classically used actigraphy variables.

While this work was exploratory in nature, and should be considered as a first foray into studying the recurrence patterns of activity data in psychopathology, it is not unlikely that future studies would show that the reported differences in complexity of physical activity patterns between depressed and non-depressed people also contribute to the development of depressive symptoms. The current findings then have important implications for clinical practice. In that case the RQA complexity measure may be promising as a diagnostic tool in depressed patients to evaluate to what extent their physical activity patterns have a healthy level of complexity. Such a diagnostic tool could be implemented clinically with very low participant burden, as actigraphy data can be collected unobtrusively. Furthermore, automatic algorithms can be generated to provide conclusions to clinicians based on the patient's actigraphy patterns, similar to how medical specialists in other fields obtain objective information to aid their diagnostics, such as information on blood pressure or heart rhythm

(56–58). Recurrence quantification analysis itself has been recently proposed as a promising biomarker to identify autism spectrum disorders from EEG data (59).

We have some recommendations for future research. First, it is relevant to explore whether changes in the complexity of physical activity patterns actually precede the onset of symptom changes in individuals with depression. Recurrence quantification analysis has previously been shown to be successful in predicting upcoming transitions in many other scenarios including in epileptic seizures, stock market crashes and combustion noise (60–62). Therefore, it is interesting to examine whether changes in the current complexity measure can foresee the start of transitions towards higher levels of depression. The TRANS-ID study has collected unique personalized datasets in which people are followed intensively over the course of symptom transitions, including actigraphy measurements (55,63). This is therefore the ideal design to test the above hypothesis. If the observed change in complexity in physical activity as reported in this study, indeed precedes symptom transitions in depression, then this would, first, provide support for the causality of these patterns for developing symptoms. Second, this may suggest that the current complexity measure could be used to foresee clinically relevant increases in depressive symptoms. Another recommendation for future studies is to examine whether intervention on physical activity patterns in the direction of increased complexity in depressed patients would lead to a reduction in the level of symptoms.

Limitations

One methodological limitation is that 13% of the participants failed to show enough variation in actigraphy patterns to perform the RQA analyses. The latter analyses cannot be performed with the presence of too many zeros (perfect inactivity) in the data, as this would result in cluttering and consequent masking of information. Therefore, potential application in clinical practice should take into account that this method will not work for some patients who show too little activity. Another methodological issue is the sample size. Although, the power to calculate the complexity outcomes per person was high, as continuous actigraphy data were available for four weeks for each person leading to highly reliable values of complexity for every individual, the between-person power to compare group differences was lower. This may explain why none of the group comparisons regarding complexity outcomes were statistically significant after multiple testing corrections. Therefore, there is a need for replication of this finding with larger sample sizes.

Conclusions

This study explored the association between physical activity and depression by studying the recurrent activity patterns that are present in the actigraphy data of depressed and non-depressed individuals. It is concluded that the diversity and average duration of activities was significantly associated with depression, while mean levels in physical activity and circadian rhythm variables were not. This novel finding has important implications for understanding how physical activity relates to mood disorders like depression. If future studies will replicate this finding and show support that complexity patterns causally relate to development of symptoms, RQA measures may constitute an additional tool for personalized diagnostics and treatment strategies, in depression.

List of abbreviations

DET*: Determinism;* ***Lavg****: Average diagonal line length;* ***Lent****: Entropy of diagonal line distribution;* ***LAM****: Laminarity;* ***LAM/DET****: Laminarity to determinism ratio;* ***Vavg****: Average vertical line length;* ***Vent****: Entropy of vertical line distribution;* ***SE****: Standard error on the mean;* ***RQA****: Recurrence quantification analysis*

Data availability. The datasets generated and/or analysed during the current study are not publicly available due to the nature of the data (intensive longitudinal actigraphy data), which cannot be considered fully anonymous. However, data are available from the corresponding author on reasonable request. The codes used for data analysis in this paper may be found at github.com/sgeorge91/ComplexityInDepression.

Acknowledgements. This study was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programmme (ERC-CoG-2015; No 681466 to M. Wichers). Author contributions S.V.G. and M.W. designed the study and wrote the manuscript. S.V.G. and Y.K.K. analysed the data. S.B. designed the MOOVD study and collected the data. All authors jointly interpreted the results and revised and approved the manuscript.

Competing interests. The authors declare no competing interests.

Supplementary Information. The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-021-92890-w. Correspondence and requests for materials should be addressed to S.V.G.

References

1. Dunn AL, Trivedi MH, O'Neal HA. Physical activity dose-response effects on outcomes of depression and anxiety. In: Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]. Centre for Reviews and Dissemination (UK); 2001.

2. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev. 2015;9(3):366–78.

3. Teychenne M, Ball K, Salmon J. Sedentary behavior and depression among adults: a review. Int J Behav Med. 2010;17(4):246–54.

4. Weyerer S, Kupfer B. Physical exercise and psychological health. Sport Med. 1994;17(2):108–16.

5. Farmer ME, Locke BZ, Mościcki EK, Dannenberg AL, Larson DB, Radloff LS. Physical activity and depressive symptoms: the NHANES I Epidemiologic Follow-up Study. Am J Epidemiol. 1988;128(6):1340–51.

6. Strawbridge WJ, Deleger S, Roberts RE, Kaplan GA. Physical activity reduces the risk of subsequent depression for older adults. Am J Epidemiol. 2002;156(4):328–34.

7. Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. Prev Med (Baltim). 2008;46(5):397–411.

8. Roshanaei-Moghaddam B, Katon WJ, Russo J. The longitudinal effects of depression on physical activity. Gen Hosp Psychiatry. 2009;31(4):306–15.

9. Rawson MJ, Cornélissen G, Holte J, Katinas G, Eckert E, Siegelová J, et al. Circadian and circaseptan components of blood pressure and heart rate during depression. Scr Med. 2000;73:117–24.

10. Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. Hum Psychopharmacol Clin Exp. 2008;23(7):571–85.

11. Minaeva O, Booij SH, Lamers F, Antypa N, Schoevers RA, Wichers M, et al. Level and timing of physical activity during normal daily life in depressed and non-depressed individuals. Transl Psychiatry. 2020;

12. Levandovski R, Dantas G, Fernandes LC, Caumo W, Torres I, Roenneberg T, et al. Depression scores associate with chronotype and social jetlag in a rural population. Chronobiol Int. 2011;

13. Hori H, Koga N, Hidese S, Nagashima A, Kim Y, Higuchi T, et al. 24-h activity rhythm and sleep in depressed outpatients. J Psychiatr Res. 2016;77:27–34.

14. Teicher MH, Lawrence JM, Barber NI, Finklestein SP, Lieberman HR, Baldessarini RJ. Increased Activity and Phase Delay in Circadian Motility Rhythms in Geriatric Depression: Preliminary Observations. Arch Gen Psychiatry. 1988;

15. Vitale JA, Roveda E, Montaruli A, Galasso L, Weydahl A, Caumo A, et al. Chronotype influences activity circadian rhythm and sleep: Differences in sleep quality between weekdays and weekend. Chronobiol Int. 2015;

16. Miller MA, Rothenberger SD, Hasler BP, Donofry SD, Wong PM, Manuck SB, et al. Chronotype predicts positive affect rhythms measured by ecological momentary assessment. Chronobiol Int. 2015;

17. Burton C, McKinstry B, Szentagotai Tătar A, Serrano-Blanco A, Pagliari C, Wolters M. Activity monitoring in patients with depression: A systematic review. Journal of Affective Disorders. 2013.

18. Au J, Reece J. The relationship between chronotype and depressive symptoms: A meta-analysis. Journal of Affective Disorders. 2017.

19. Albert PS, Hunsberger S. On analyzing circadian rhythms data using nonlinear mixed models with harmonic terms. Biometrics. 2005;61(4):1115–20.

20. Shinagawa M, Otsuka K, Murakami S, Kubo Y, Cornelissen G, Matsubayashi K, et al. Seven-day (24-h) ambulatory blood pressure monitoring, self-reported depression and quality of life scores. Blood Press Monit. 2002;7(1):69–76.

21. Maes M, Meltzer HY, Suy E, De Meyer F. Seasonality in severity of depression: relationships to suicide and homicide occurrence. Acta Psychiatr Scand. 1993;

22. Lang M, Krátk\`y J, Shaver JH, Jerotijević D, Xygalatas D. Effects of anxiety on spontaneous ritualized behavior. Curr Biol. 2015;25(14):1892–7.

23. Martín-Martínez D, Casaseca-de-la-Higuera P, Alberola-López S, Andrés-de-Llano J, López-Villalobos JA, Ardura-Fernández J, et al. Nonlinear analysis of actigraphic signals for the assessment of the attention-deficit/hyperactivity disorder (ADHD). Med Eng Phys. 2012;

24. Parro VC, Valdo L. Sleep-wake detection using recurrence quantification analysis. Chaos. 2018;

25. Hu K, Ivanov PC, Chen Z, Hilton MF, Stanley HE, Shea SA. Non-random fluctuations and multi-scale dynamics regulation of human activity. Phys A Stat Mech its Appl. 2004;

26. Ong JC, Arnedt JT, Gehrman PR. Insomnia Diagnosis, Assessment, and Evaluation. In: Principles and Practice of Sleep Medicine. 2017.

27. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: A systematic review and meta-analysis. Journal of Affective Disorders. 2019.

28. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci. 2008;

29. Kaplan KA, Talbot LS, Gruber J, Harvey AG. Evaluating sleep in bipolar disorder: Comparison between actigraphy, polysomnography, and sleep diary. Bipolar Disord. 2012;

30. Booij SH, Bos EH, Bouwmans MEJ, van Faassen M, Kema IP, Oldehinkel AJ, et al. Cortisol and $α$-amylase secretion patterns between and within depressed and non-depressed individuals. PLoS One. 2015;10(7):e0131002.

31. Stavrakakis N, Booij SH, Roest AM, de Jonge P, Oldehinkel AJ, Bos EH. Temporal dynamics of physical activity and affect in depressed and nondepressed individuals. Heal Psychol. 2015;34(S):1268.

32. Beck AT, Steer RA, Brown GK. Beck depression inventory-II. San Antonio. 1996;78(2):490–8.

33. Heil DP. Predicting activity energy expenditure using the actical® activity monitor. Res Q Exerc Sport. 2006;

34. Marwan N. A historical review of recurrence plots. European Physical Journal: Special Topics. 2008.

35. Jenkins BN, Hunter JF, Richardson MJ, Conner TS, Pressman SD. Affect Variability and Predictability: Using Recurrence Quantification Analysis to Better Understand How the Dynamics of Affect Relate to Health. Emotion. 2019;

36. Lichtwarck-Aschoff A, Hasselman F, Cox R, Pepler D, Granic I. A characteristic destabilization profile in parent-child interactions associated with treatment efficacy for aggressive children. Nonlinear Dynamics Psychol Life Sci. 2012;

37. Pompe B. Ranking and Entropy Estimation in Nonlinear Time Series Analysis. In: Nonlinear Analysis of Physiological Data. 1998.

38. Conover WJ, Iman RL. The Rank Transformation as a Method of Discrimination with Some Examples. Commun Stat - Theory Methods. 1980;

39. Ziegel E, Press W, Flannery B, Teukolsky S, Vetterling W. Numerical Recipes: The Art of Scientific Computing. Technometrics. 1987;

40. TOCSY - Toolbox for Complex Systems (Recurrence Plots, Cross Recurrence Plots, System Identification, ACE, Nonlinear Wavelet Analysis, Nonlinear Regression Analysis, Adaptive Filtering, Coupling Direction) [Internet]. [cited 2019 Dec 19]. Available from: http://tocsy.pik-potsdam.de/

41. Marwan N, Romano MC, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. Phys Rep. 2007;438(5–6):237–329.

42. Thiel M, Romano MC, Read PL, Kurths J. Estimation of dynamical invariants without embedding by recurrence plots. Chaos. 2004;

43. Eroglu D, Marwan N, Prasad S, Kurths J. Finding recurrence networks' threshold adaptively for a specific time series. Nonlinear Process Geophys. 2014;

44. Javorka M, Trunkvalterova Z, Tonhajzerova I, Lazarova Z, Javorkova J, Javorka K. Recurrences in heart rate dynamics are changed in patients with diabetes mellitus. Clin Physiol Funct Imaging. 2008;

45. Rusinek R, Zaleski K. Dynamics of thin-walled element milling expressed by recurrence analysis. Meccanica. 2016;51(6):1275–86.

46. Rusinek R, Lajmert P, Krzysztof K, Kruszynski B, Warminski J. Chatter identification methods on the basis of time series measured during titanium superalloy milling. Int J Mech Sci. 2015;

47. Welch BL. The generalization ofstudent's' problem when several different population variances are involved. Biometrika. 1947;34(1/2):28–35.

48. Jones E, Oliphant T, Peterson P, others. SciPy: Open source scientific tools for Python. 2001;

49. Witting W, Kwa IH, Eikelenboom P, Mirmiran M, Swaab DF. Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. Biol Psychiatry. 1990;

50. Van Someren EJW, Swaab DF, Colenda CC, Cohen W, McCall WV, Rosenquist PB. Bright light therapy: Improved sensitivity to its effects on rest- activity rhythms in Alzheimer patients by application of nonparametric methods. Chronobiol Int. 1999;

51. Kunkels YK, Knapen SE, Zuidersma M, Wichers M, Riese H, Emerencia AC. ACTman: Automated preprocessing and analysis of actigraphy data. J Sci Med Sport. 2019;

52. Difrancesco S, Lamers F, Riese H, Merikangas KR, Beekman ATF, van Hemert AM, et al. Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. Depress Anxiety. 2019;

53. Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, et al. Complex affect dynamics add limited information to the prediction of psychological well-being. Nature Human Behaviour. 2019.

54. Kuranova A, Booij SH, Menne-Lothmann C, Decoster J, Van Winkel R, Delespaul P, et al. Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. BMC Med. 2020;

55. Schreuder MJ, Hartman CA, George S V., Menne-Lothmann C, Decoster J, Van Winkel R, et al. Early warning signals in psychopathology: What do they tell? Submitted. 2020;

56. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. npj Digit Med. 2019;

57. Insel TR. Digital phenotyping: Technology for a new science of behavior. JAMA - Journal of the American Medical Association. 2017.

58. Jacobson NC, Weingarden H, Wilhelm S. Using Digital Phenotyping to Accurately Detect Depression Severity. J Nerv Ment Dis. 2019;

59. Heunis T, Aldrich C, Peters JM, Jeste SS, Sahin M, Scheffer C, et al. Recurrence quantification analysis of resting state EEG signals in autism spectrum disorder - a systematic methodological exploration of technical and demographic confounders in the search for biomarkers. BMC Med. 2018;

60. Acharya UR, Sree SV, Chattopadhyay S, Yu W, Ang PCA. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. Int J Neural Syst. 2011;21(03):199–211.

61. Addo PM, Billio M, Guegan D. Nonlinear dynamics and recurrence plots for detecting financial crisis. North Am J Econ Financ. 2013;26:416–35.

62. Kabiraj L, Saurabh A, Nawroth H, Paschereit CO. Recurrence analysis of combustion noise. AIAA J. 2015;53(5):1199–210.

63. Trans-ID [Internet]. [cited 2020 Mar 27]. Available from: https://www.transid.nl/

Tables

Table 1: Definitions and interpretations in the context of activity data, for various recurrence plot quantifiers that are used in this work.

Quantifier	Calculation	Definition	Interpretation
DET	Ratio of diagonal structures to total recurrence points.	Level of deterministic activity in the data.	Lower levels indicate more randomness
LAM	Ratio of vertical structures to total recurrence points.	Level of slowly evolving processes in the time series.	Higher levels indicate more activities that linger
Lavg	Mean length of diagonal structures	Average duration of recurrent physical activity	Higher levels indicate longer recurrent physical activities
Lent	Entropy of diagonal line distribution	Diversity of durations of recurrent physical activity patterns	Higher levels indicate recurrent physical activities of varying durations
Vavg	Mean length of vertical structures	Average duration of static activity patterns	Higher levels indicate lingering physical activities that last longer

Vent	Entropy of vertical line distribution	Diversity of durations of static activity patterns	Higher levels indicate lingering physical activities of varying durations
LAM/DET	Ratio of LAM to DET measures	Level of statis as compared to deterministic structure	Changes in this ratio has been shown to be an indicator of change in stability(45,46).

DET: Determinism; LAM: Laminarity; Lavg: Average diagonal line length; Lent: Entropy of diagonal line distribution; Vavg: Average vertical line length; Vent: Entropy of vertical line distribution; LAM/DET: Laminarity to determinism ratio

Table 2: The means for demographic and clinical measures for the non-depressed (n=25) ($Mean_N$) and depressed (n=21) groups ($Mean_D$)

Demographic and clinical characteristics	$Mean_N \pm SE$	$Mean_D \pm SE$	t-statistic	p-value
Age	33.320±1.784	33.238±2.028	-0.030	0.976
BMI	22.334±0.527	24.274±1.377	1.315	0.200
Gender	24% male	29% male	-0.384 (z-statistic)	0.351
Pre-BDI-II	2.16±0.547	29.286±2.006	13.045	<0.001
Post-BDI-II	2.360±0.688	20.381±2.486	6.987	<0.001

The t-statistic and p-value are listed.. The pre and post BDI-II scores are the BDI-II scores before and after the data collection, respectively. SE:Standard error on the mean, $Mean_N$: Mean of the non-depressed group, $Mean_D$: Mean of the depressed group

Table 3: The means for traditional quantifiers of the actigraphy time series for the non-depressed (n=25) and depressed (n=21) groups

Quantifier	$Mean_N \pm SE$	$Mean_D \pm SE$	t-statistic	p-value	Cohen's d
Average activity	262.388±15.514	228.040±18.400	-1.427	0.161	0.435
IS	0.368±0.065	0.424±0.800	0.542	0.591	-0.166
IV	1.414±0.130	1.256±0.120	-0.891	0.378	0.265
RA	0.905±0.015	0.896±0.016	0.393	0.696	-0.117

The t-statistic, p-value and effect size measured using Cohen's d are listed. IS: Interdaily stability; IV: Intradaily variability; RA: Relative Amplitude; $Mean_N$: Mean of the non-depressed group; $Mean_D$: Mean of the depressed group; SE=standard error of the mean

Table 4: Means for the different recurrence plot measures for the non-depressed(n=25) and depressed(n=21) groups,

Property	**Mean$_N$ ± SE**	**Mean$_D$ ± SE**	**t-statistic**	**p-value**	**Cohen's d**
DET	0.359±0.022	0.308±0.019	-1.725	0.091	0.511
L_{avg}	2.436±0.039	2.334±0.031	-2.018	0.050*	0.594
L_{ent}	0.853±0.042	0.725±0.044	-2.104	0.041*	0.634
LAM	0.511±0.026	0.466±0.026	-1.168	0.249	0.353
LAM/DET	1.446±0.018	1.517±0.024	2.336	0.025*	-0.721
V_{avg}	3.079±0.095	2.864±0.080	-1.731	0.090	0.512
V_{ent}	1.352±0.056	1.206±0.056	-1.707	0.095	0.520

*The t-statistic, p-value and Cohen's d are listed. The measures that show statistically significant differences are indicated with an * DET: Determinism; Lavg: Average diagonal line length; Lent: Entropy of diagonal line distribution; LAM: Laminarity; LAM/DET: Laminarity to determinism ratio; Vavg: Average vertical line length; Vent: Entropy of vertical line distribution, Mean$_N$: Mean of the non-depressed group; Mean$_D$: Mean of the depressed group; SE=standard error of the mean*

Table 5: Correlation between commonly used actigraphy measures and recurrence plot measures

	Lavg	**Lent**	**LAM/DET**	**Avg Activity**	**IS**	**IV**	**RA**
L_{avg}	1	0.999*	-0.609	0.091	0.073	0.031	0.201
L_{ent}		1	-0.607	0.107	0.073	0.035	0.206
LAM/DET			1	0.286	-0.314*	-0.098	0.013
Avg Activity				1	-0.162	-0.132	0.157
IS					1	0.355*	-0.140
IV						1	-0.185
RA							1

*Only the recurrence plot measures that showed significant differences between the two groups are listed. The Table lists the Spearman correlation coefficient. Significant correlations($p<0.05$) are marked with a *($p>0.05$).*

IS: Interdaily stability; IV: Intradaily variability; RA: Relative Amplitude; Lavg: Average diagonal line length; Lent: Diagonal line entropy; LAM/DET: Laminarity to determinism ratio

Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring

This chapter has been published as:
Kunkels, Y. K., 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).
https://doi.org/10.1111/psyp.13898

6.

Chapter 6: Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring

Yoram K. Kunkels, MSc[1*], Arie M. van Roon, PhD[2], Marieke Wichers, PhD[1], Harriëtte Riese, PhD[1]

[1]University of Groningen, University Medical Center Groningen, Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), The Netherlands.

[2]Department of Vascular Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

This chapter has been published as; Kunkels, Y. K., 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). https://doi.org/10.1111/psyp.13898

Abstract

Wired ambulatory monitoring of the electrocardiogram (ECG) is an established method used by researchers and clinicians. Recently, a new generation of wireless, compact, and relatively inexpensive heart rate monitors have become available. However, before these monitors can be used in scientific research and clinical practice, their feasibility, validity, and reproducibility characteristics have to be investigated. Therefore, we tested how two wireless heart rate monitors (i.e., the Ithlete photoplethysmography (PPG) finger sensor and the Cortrium C3 ECG monitor perform against an established wired reference method (the VU-AMS ambulatory ECG monitor). Monitors were tested on cross-instrument and test-retest reproducibility in a controlled laboratory setting, while feasibility was evaluated in protocolled ambulatory settings at home. We found that the Cortrium and the Ithlete monitors showed acceptable agreement with the VU-AMS reference in laboratory setting. In ambulatory settings, assessments were feasible with both wireless devices although more valid data were obtained with the Cortrium than with the Ithlete. We conclude that both monitors have their merits under controlled laboratory settings where motion artefacts are minimized and stationarity of the ECG signal is optimized by design. These findings are promising for long-term ambulatory ECG measurements, although more research is needed to test whether the wireless devices' feasibility, validity, and reproducibility characteristics also hold in unprotocolled daily life settings with natural variations in posture and activities.

Introduction

In the last decennium, the number of heart rate monitors available for clinicians and researchers has increased steadily (El-Amrawy & Nounou, 2015). This opened opportunities for long-term monitoring of cardiac functions such as heart rate (HR) and heart rate variability (HRV) (Kemp & Quintana, 2013; Malik et al., 1996) in daily-life. When feasible and valid, long-term HR(V) monitoring may open-up possibilities for developing indicators of (mental) health processes complementary to those developed for the experience sampling method (ESM, or electronic dairy). ESM monitoring is a scientific method that has shown its potential in research and clinical practice (e.g. Schoevers et al., 2020; Shiffman et al., 2008; Vaessen et al., 2019). It nowadays typically involves filling-out short questionnaires, multiple time a day for weeks/months on a smartphone. The time-series data derived from ESM monitoring have been used to better inform diagnosis, intervention selection, and recently also for early predicting transitions in patients affect state (Kroeze et al., 2017; Smit et al., 2019). Long-term monitoring of HR(V) could potentially add to ESM's potential in clinic practice, as it might supplement a patients momentary affect data with physiological data.

The current study is performed within the scope of the TRANS-ID (TRANSitions In Depression*, www.transid.nl*) project, which aims to discover personalized signals that may indicate critical transitions in psychological and physiological symptoms. In TRANS-ID we investigate within single individuals which early warning signals precede depressive symptom change and thereby examine whether psychological symptoms behave according to the principles of complex dynamical systems (Scheffer, 2010). To gain insight into this, we use ESM to capture the micro-level changes of symptoms, emotions, behaviours and daily context over time (Kramer et al., 2014). Moreover, future TRANS-ID studies are planned to investigate whether monitoring patients' HR(V) data can support the study of dynamic processes, such as transitions from a healthy to a clinically affected state in patients. Such combined time-series

data collection is required for studies that aim to investigate the use of physiological measurements for predicting transitions in patients' affect state. We previously found support that ESM data can be used to calculate early warning signals to predict transitions in patients' affect state (Wichers & Groot. 2016; Smit, Snippe, & Wichers, 2019), and aim to investigate in future research the usefulness of HR(V) monitoring for predicting transitions in patients' affect state. As long-term continuous wearing of ECG-electrodes is not feasible (e.g. because of skin irritation), and invasive heart rate measurements are not possible in non-clinical settings, the second best option would be an intensive repeated measurements design. HR(V) is well known to fluctuate with changing posture and activities in ecological real-life monitoring designs (Riese et al., 2004; Vrijkotte, van Doornen, & de Geus, 2000). To account for this a highly controlled procedures for data collection in the laboratory, as well as in real-life, was used in the current study.

To test the potential of HR(V) time-series data, a specific study design and a HR(V) monitor suitable for long-term (i.e. four months) monitoring is required. Based on literature search, pilots that include analysis of the raw time-series data with the potential selected HR monitors, and our own expertise, we set the following criteria a monitor should fulfil; (i) feasibility of four months HR(V) monitoring, which requires participants to initialize and operate the monitors themselves on a smartphone; (ii) wireless monitoring, as multiple long wires attached to the electrodes can be accidentally pulled and detached from the monitor interrupting data collection (Shin et al., 2005; Winokur et al., 2013); (iii) sufficient battery and memory capacity to support long-term assessment; (iv) the ability to upload data to a protected server; (v) good validity and reproducibility of HR(V) measurements and; (vi) access to the raw data.

There are various types of heart rate monitors, for example in cardiology heart rate is typically monitored with an electrocardiogram (ECG) Holter system (Kennedy, 2013). Holter

monitors allow for 24 to 48 hours of continuous measurements with high accuracy. Such a higher degree of accuracy comes at a price though, as the large number of ECG spot-electrodes and wires increases the measurement burden. For research purposes, heart rate monitors with substantial less spot-electrodes and wires were successfully developed for robust continuous 24-48 hours ECG measurements (de Geus et al., 1995; Wegner et al., 2020). However, there are still a number of issues hampering long-term measurements (weeks/months) such as; limited data storage and battery capacity, wires between the monitor and the electrodes, skin irritation due to wearing ECG electrodes, and monitor costs.

Recently many innovative heart rate monitors were released. One could contend that there are many alternative consumer-grade monitors, including the well-known *Fitbit*, *Polar RS400*, or *Apple Watch*. Indeed, studies have shown agreeable accuracy of such monitors when compared to chest-strapped ECG monitors (Stahl et al., 2016). In other studies, however, wrist-worn monitors were found to provide non-consistent accuracy during motion when compared to a chest strap-based ECG monitor (Wang et al., 2017). In an effort to optimize accuracy, combined with the aforementioned essential criteria, such as access to raw data, we selected the Cortrium C3 ECG monitor (cortrium.com) and the Ithlete photoplethysmography (PPG) finger sensor (myithlete.com). Both ECG and PPG assess *interbeat interval* (IBI) time-series data from which HRV measures can be calculated.

The Cortrium is a wireless 3-lead ECG monitor, which is attached to the chest with three spot-electrodes. The signal is sent via a Bluetooth connection to the user's smartphone and data are saved in real-time. From the smartphone, data can be transferred to any protected server worldwide solving potential data storage issues. The Cortrium also has an internal memory. Such multiple data storage sites can act as a buffer against potential data loss. The renewed interest in heart rate monitoring has also reinvigorated interest in optically based methods, such as PPG. With PPG, blood volume changes are detected by illuminating tissue

and measuring changes in light absorption, from which the R-peaks are deduced. Especially for long-term monitoring the PPG method offers the advantage of being ECG spot-electrode free preventing skin irritation which can hamper feasibility and increase non-compliance. Earlier studies established substantial agreement (correlation coefficients between 0.81 and 0.99, and < 3% error rate) between PPG and ECG measures under controlled laboratory conditions (Lu et al., 2009; Teng & Zhang, 2003). The *Ithlete PPG finger sensor* is also controlled by a smartphone application, and offers data storage options via Bluetooth connection and on protected servers. The Ithlete finger sensor uses an infrared light emitting diode as a light source. Investigations into possible negative effects of body mass index (BMI) on accuracy of HR assessment with wrist-worn PPG devices have obtained evidence both for and against such a negative effect. We do not expect BMI to considerably hinder HR accuracy in our study with the Ithlete, as it measures at the tip of the finger, a location which was found to be most sensitive to blood volume fluctuations (Nardelli et al., 2020).

While the Cortrium and the Ithlete offer interesting features that can facilitate longer (e.g. months) intensive monitoring, their feasibility, validity, and test-retest reproducibility has not been established yet. Therefore, in the current study we aim to investigate the feasibility of these monitors during laboratory sessions and long-term ambulatory monitoring. Second, the validity of the wireless Cortrium and Ithlete monitors on HR and HRV measurements are tested against a standard wired ECG reference monitor under standardized laboratory conditions. Thirdly, test-retest reproducibility of HR and HRV assessed with the wireless monitors is tested under laboratory conditions over a period of two weeks.

Materials and methods

Participants

For this study, we recruited 64 participants (75% female, mean age = 26 years) from the University Medical Center Groningen (UMCG) and the University of Groningen, through recruitment flyers (see: osf.io/yanqd/). Participants were eligible to participate when the following criteria were met: being (i) 18 years or older, (ii) able to follow the study procedures, (iii) sufficiently proficient in Dutch to fill-out the ESM items and operate a smartphone, (iv) giving written informed consent, and (v) not suffering from cardiovascular diseases, diabetes mellitus, anemia, or using cardio-active medication. Participants received a €25 gift card when completing at least 80% of the measurements and a summary report of their personal data.

Of the original 64 recruited participants, 51 did successfully complete the study. One participant was excluded due to use of cardioactive medication, three other participants did initially agree with study participation but did not to show-up for the first appointment. Lastly, nine participants dropped out during the course of the study. The reasons for dropout were: skin irritation cause by the ECG electrodes (one time), fear caused by observing own heart rate (one time), time constrains (two times), not reacting any more to communication efforts (two times), and no reported reason (three times). The study protocol was submitted to the ethical review board of the UMCG, who confirmed that formal assessment was not required. The study is registered in the UMCG research register (no. 20160039).

Monitor specifications

As reference the *VU-AMS monitor* (www.vu-ams.nl) was used as its validity and reproducibility of measuring cardiovascular indices have been established and are on par with traditional non-ambulatory ECG monitors used in laboratories (de Geus et al., 1995;

Goedhart et al., 2007; Willemsen et al., 1996). Recorded signals were ECG (VU-AMS, Cortrium) or PPG (Ithlete). Sample rate was fixed by design at 250 Hz for both the Cortrium and Ithlete devices, and set to 250 Hz for the VU-AMS system to facilitate device comparison, although having a sampling frequency higher than 250 Hz would provide higher resolution data, as described in detail elsewhere, this sample rate is sufficient for the aims of the current study (Greaves-Lord et al., 2010), as the contribution of the rounding error at 250 Hz was found to be small (i.e., error variance = 1.3 ms^2, LF contribution = 0.4 ms^2, HF contribution = 1.4 ms^2). From the ECG and PPG signals R-peaks were triggered (details below) to obtain inter-beat intervals (IBI, in ms) between two successive heartbeats. HRV is calculated as its primary time-domain measure, the root mean square of successive differences (RMSSD, in ms) between two heartbeats (Malik et al., 1996).

Study design

A flowchart of the study is shown in Figure 1. Data were collected during two laboratory sessions and two weeks of ambulatory measurements. The laboratory measurements were designed to assess the cross-instrument and test-retest reproducibility of the monitors in a controlled laboratory setting. The ambulatory assessments were designed to assess monitor feasibility during ecological valid ambulatory settings. These ambulatory assessments took place in an ESM design; meaning that participants receive an ESM questionnaire five times a day at 3-hour intervals, after which they will conduct ECG/PPG measurements. The questionnaires were sent via text message, while a reminder text was sent after ten minutes if participants had not yet responded. Participants had to complete the ESM questionnaire within half an hour. Filling out the questionnaire took about two minutes. An overview of the included ESM items translated into English is available online (see: osf.io/e8vnh/).

Laboratory: During the first laboratory session participants started with a 15-minute intake, which included amongst others, questions about medication use, alcohol use, tobacco use, and contraception use (see osf.io/yanqd/ for more details). Next, the participant was attached to the *Cortrium, Ithlete*, and *VU-AMS*. Participants wore all three monitors simultaneously during the laboratory sessions. The first laboratory session took approximately 90 minutes and involved six standardized physical and mental tasks (see Figure 2). After the 14 days of ambulatory measurements, participants returned for the second laboratory session, which involved the same laboratory tasks and an additional structured evaluation interview.

Ambulatory setting: After the first laboratory session, participants continued with 14 days of monitoring themselves with the *Cortrium* and *Ithlete* during their normal daily life. Participants wore the *Ithlete* and *Cortrium* monitors simultaneously. Participants measured themselves five times a day by filling-out an ESM diary (two minutes) and subsequently conduct the ECG/PPG measurements (five minutes). ESM was used as a reminder for the heart rate measurement and as a timer for the acclimatization phase. ESM data were not used in the current study.

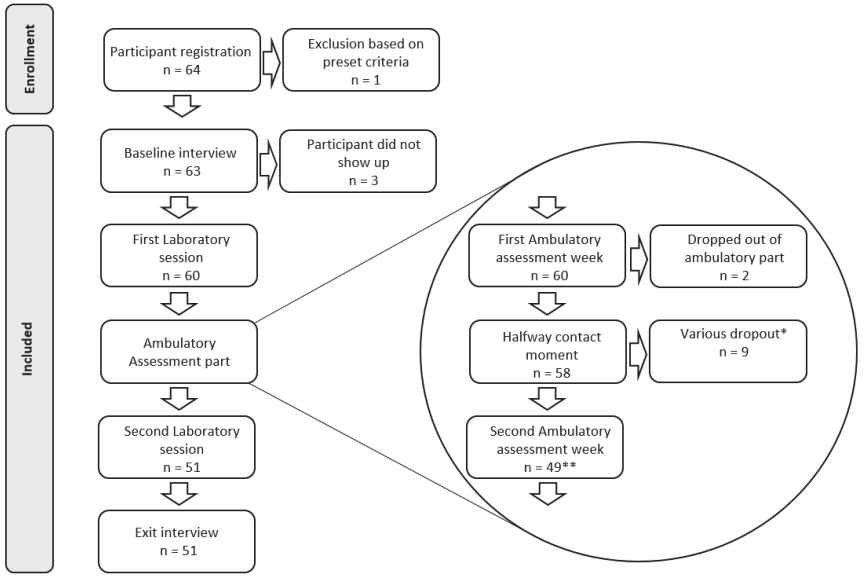


Figure 1: Flowchart of the study. Note: * A precise specification of these various dropout reasons is given in the method section. ** The second laboratory session included two more participants than the 49 that finished the second ambulatory week is because although they dropped out of the ambulatory assessment part they agreed to participate in the second laboratory session.

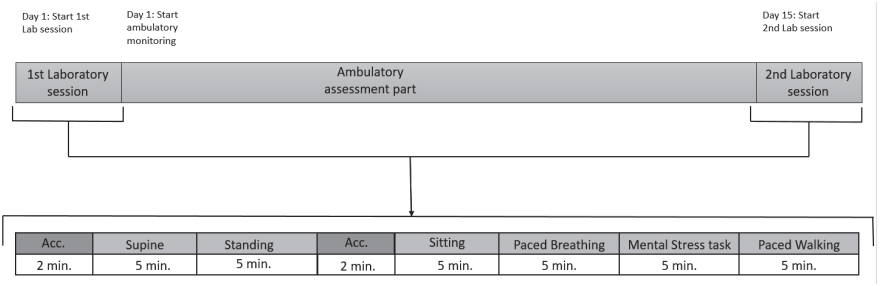


Figure 2: Visualization of the study design. Upper part: The study involved two laboratory sessions and ambulatory measurements. Lower part: Enlargement of the laboratory sessions. The two blocks labelled "Acc." indicate two minutes of acclimatization. The other blocks indicate the six laboratory tasks: supine, standing, sitting, paced breathing, a mental stress task, and paced walking.

Procedure

Laboratory:

The lab sessions involved tasks in the following preset order: acclimatization in supine position (two minutes), rest in supine position with eyes closed (five minutes); standing position with eyes open (five minutes); acclimatization in sitting posture (two minutes); rest in sitting posture (five minutes); paced breathing task in sitting position (five minutes); mental stress task (five minutes); and paced walking (five minutes). Acclimatization after posture changes were used to obtain stationary ECG signals. For the paced breathing task visual stimuli on screen guided participants to pace their breathing with a 0.25 Hz frequency. The mental stress task is a challenging Stroop task. The Stroop task is known to reliably elicit cardiac responses (Eliasson et al., 1983; Freyschuss et al., 1988). To increase mental stress the research assistant delivered critical feedback to the participant such as "That is not good enough". Paced walking was protocolled as walking with the research assistant in a constant normal walking speed through a preset walking route. After the second laboratory session participants participated in an in-house developed evaluation interview. The interview included 52 questions and took approximately 45 minutes. The difference with the first interview (intake) were the additional evaluation questions about feasibility, ease-of-use, and burdensomeness of the wireless monitors. Participants were also asked about the procedural fidelity, such as reasons for missing heart rate measurements during the 14-day ambulatory setting. An overview of the items used to assess feasibility, ease-of-use, and burdensomeness is available online (see: osf.io/4nuwg/).

Ambulatory setting: At the end of the laboratory session, the research assistant instructed participants to fill-out the ESM diary within 30 minutes when prompted by a text message on their smartphone in sitting posture. Participants were instructed to remain seated for two minutes to further acclimatize to ascertaining signal stationarity and preventing

changes in posture and motion artefacts. Then the ECG/PPG measurements were started in sitting posture while breathing spontaneously and refraining from talking. Assistance from a research assistant was available for participants during the full study period.

Data pre-processing

For labelling the VU-AMS data collected during the laboratory sessions, the VU-DAMS software (version 4.3) was used. Each task in the lab session and each assessment in the ambulatory situation was given a label. Labels indicate the start and end of a block of time-series data entered into the pre-processing procedure prior statistical analyses, and reported in the result section. Raw IBI time-series data were pre-processed in R-peak detection software. Data pre-processing steps included converting files, checking file integrity, and correcting for (motion) artefacts. Conceptually there were no differences between preprocessing in either the in-house developed Precar or the Drosan software, although the implementation logically differed due to inherent differences between raw ECG and PPG time-series data. Drosan version 2.52, (Zhang et al., 2019) was used for pre-processing the Ithlete data, and Precar version 3.83 (Greaves-Lord et al., 2010) for pre-processing the Cortrium and VU-AMS data. The CARSPAN program is an in-house developed software package for processing and analyzing IBI time-series (Mulder, van Roon, & Schweizer, 1995).

Data pre-processing involved checking the integrity of the time-series data. Missing data were interpolated up to a maximum of 10 sec. but in not more than 10% of the total block duration. Otherwise, time-series data in a block was set to missing due to poor data quality. Major reasons for unsatisfactory data quality were poor connection between spot-electrodes and the participants' skin, and motion artefacts. Both ECG and PPG methods are known to be vulnerable for such motion artefacts (Thakor & Zhu, 1991; Trivedi et al., 1997). Data analysts were first trained by analysing ten example files under supervision of an expert cardiology

analyst. Data analysts were allowed to work on the real time-series data files after sufficiently high *intraclass correlation coefficient* (ICC) values (ICC > 0.95) between the files processed by the analyst and those processed by the expert cardiology analyst were attained.

It was checked whether the data were not too noisy for analysis, whether the R-squared values were at least 0.30, and whether vcIBI values were above 20%. 1.03% of the Cortrium files and 11.00% of the Ithlete files were found to exceed these criteria. When such physiological implausible values were detected, these were followed-up up with a check in the raw data to make sure no R-peaks or artefacts were missed during data pre-processing.

Statistical analysis

All statistical analysis and plotting of the data were performed in the statistical programming language R (R Core Team, 2017). Prior analysis, data distributions were checked for dispersion and skewness by visually examining QQ-plots, density plots, and skewness-kurtosis plots (Cullen & Frey, 1999), and testing for normality with a Shapiro-Wilk test (Shapiro & Wilk, 1965). RMSSD values were natural log transformed to conform to assumptions of linear analyses (Ellis et al., 2008).

First, we described feasibility characteristics of the monitors, such as amount of data collected with each monitor. Descriptive statistics of the evaluation interview were calculated. Second, cross-instrument validity was assessed by comparing both the Cortrium and the Ithlete to the VU-AMS during the laboratory tasks. The variables of interest are mean IBI and ln(RMSSD). Intraclass correlation coefficients (ICC's) were calculated; values closer to one indicate closer adherence to the reference. ICC values were interpreted as follows: <0.40 as *poor*, between 0.40 and 0.59 as *fair*, between 0.60 and 0.74 as *good*, and between 0.75 and 1.00 as *excellent* (Cicchetti, 1994). Third, the test-retest reproducibility was tested. With paired

student's t-tests changes in mean IBI and ln(RMSSD) values obtained in the first and second laboratory sessions were tested. Absolute reproducibility, which shows the predicted trial-to-trial noise within participants, was assessed by calculating the *standard error of measurement* (SEM), also known as the within-subject standard deviation. Furthermore, the *coefficient of variation* (CV, in %) was calculated as an indication of reproducibility: lower CV values indicate higher reproducibility (lellamo et al., 1996). For instance, a CV of 20% indicates that around $^{2}/_{3}$ of test-retest differences can be found within 20% of the mean score (Atkinson & Nevill, 1998). Missing data were handled through list-wise deletion for the test-retest and cross-instrument parts separately. Bland-Altman plots (Bland & Altman, 1999) were used to visualize agreement between values obtained with the wireless monitors and the VU-AMS. In these plots, the differences of each couple of repeated measurements are plotted against the average of these two measurements. Third, the test-retest reproducibility was tested by comparing the measurements of the Cortrium and the Ithlete during the first laboratory session with the corresponding measurements during the second laboratory session. Additionally, a Welch t-test was performed on mean IBI and ln(RMSSD) values of the first and second laboratory sessions for each monitor as Bartlett tests indicated unequal variances.

Results

Descriptive statistics

Descriptive statistics for HR and HRV are given in Table 1 (see for a transposed version of Table 1, which allows for easy monitor comparison, https://osf.io/undy2/download). Visual indicators (i.e., QQ-plots, density plots, and skewness-kurtosis plots) and Shapiro-Wilk tests indicated that data were not normally distributed. Therefore, data were natural log-transformed prior statistical analysis. Inspection of the residuals versus fitted plots indicated that the assumption of equal variances was not violated.

TABLE 1
Descriptives of IBI and log-transformed RMSSD assessed during the first and second laboratory measurements showing all tested monitors and laboratory tasks. Means (SD) and [range] are given.

Monitor	Task and Lab session No.	IBI (in ms)	ln(RMSSD) (in ln(ms))
Cortrium	**Supine**	987.61 (180.40)	4.21 (0.68)
	1st Lab (*n*=39)	[657.18 - 1429.02]	[2.49 - 5.41]
	Supine	985.47 (167.11)	4.26 (0.72)
	2nd Lab (*n*=37)	[726.22 - 1413.60]	[2.79 - 5.43]
	Standing	765.54 (134.66)	3.31 (0.54)
	1st Lab (*n*=38)	[520.08 – 1113.66]	[2.26 – 4.61]
	Standing	746.27 (119.29)	3.27 (0.60)
	2nd Lab (*n*=37)	[538.96 – 1176.21]	[2.21 - 4.50]
	Sitting	895.38 (147.80)	3.81 (0.64)
	1st Lab (*n*=39)	[589.67 – 1242.34]	[2.20 – 4.95]
	Sitting	876.96 (134.33)	3.78 (0.62)
	2nd Lab (*n*=37)	[584.43 – 1255.89]	[2.05 - 4.99]
	Breathing	873.70 (147.02)	3.93 (0.66)
	1st Lab (*n*=39)	[555.63 – 1197.37]	[1.92 – 5.16]
	Breathing	845.61 (138.61)	3.85 (0.72)
	2nd Lab (*n*=37)	[570.91 – 1221.98]	[2.08 – 5.39]
	Mental Stress	882.86 (149.59)	3.91 (0.60)
	1st Lab (*n*=39)	[529.28 – 1260.23]	[1.92 - 5.03]
	Mental Stress	873.68 (129.91)	3.80 (0.68)
	2nd Lab (*n*=37)	[511.51 – 1183.36]	[1.50 - 4.98]
	Walking	694.08 (91.60)	3.25 (0.62)
	1st Lab (*n*=37)	[514.50 – 907.48]	[1.46 - 4.64]
	Walking	696.16 (82.20)	3.23 (0.55)
	2nd Lab (*n*=36)	[526.18 – 882.94]	[1.96 – 4.33]
	Supine	1006.91 (195.78)	4.37 (0.61)

	1st Lab (*n*=23)	[738.68 – 1433.23]	[2.86 - 5.37]
	Supine	1001.53 (159.51)	4.41 (0.63)
	2nd Lab (*n*=28)	[796.42 - 1414.95]	[2.78 – 5.40]
	Standing	769.59 (149.77)	3.37 (0.48)
	1st Lab (*n*=24)	[597.12 – 1245.97]	[2.73 – 4.49]
	Standing	740.24 (106.11)	3.39 (0.48)
	2nd Lab (*n*=28)	[601.94 – 1163.15]	[2.51 – 4.49]
	Sitting	898.00 (150.12)	3.91 (0.49)
	1st Lab (*n*=23)	[720.13 – 1305.79]	[2.87 – 4.67]
Ithlete	**Sitting**	882.26 (104.52)	3.92 (0.51)
	2nd Lab (*n*=28)	[722.15 – 1199.70]	[2.99 – 5.15]
	Breathing	879.51 (147.80)	4.05 (0.54)
	1st Lab (*n*=24)	[660.31 – 1226.03]	[2.99 - 5.18]
	Breathing	851.56 (112.09)	3.94 (0.58)
	2nd Lab (*n*=27)	[660.21 – 1219.78]	[2.28 - 5.04]
	Mental Stress	897.00 (154.40)	3.97 (0.46)
	1st Lab (*n*=23)	[676.07 - 1269.69]	[3.00 – 4.83]
	Mental Stress	888.90 (108.80)	3.93 (0.53)
	2nd Lab (*n*=27)	[730.72 – 1185.84]	[2.83 – 4.97]
	Walking	706.33 (67.21)	4.49 (0.52)
	1st Lab (*n*=6)	[604.43 – 814.47]	[3.76 – 5.16]
	Walking	679.22 (119.87)	4.39 (0.48)
	2nd Lab (*n*=5)	[539.61 – 859.38]	[3.82 - 5.05]
	Supine	991.00 (175.14)	4.22 (0.64)
	1st Lab (*n*=45)	[659.74 – 1432.45]	[2.51 – 5.39]
	Supine	1004.81 (173.88)	4.30 (0.71)
	2nd Lab (*n*=45)	[726.98 – 1431.02]	[2.59 – 5.42]
	Standing	761.78 (145.27)	3.22 (0.53)
	1st Lab (*n*=45)	[521.67 – 1251.65]	[2.27 – 4.46]
	Standing	750.28 (138.38)	3.22 (0.60)
	2nd Lab (*n*=45)	[541.23 – 1351.38]	[1.89 – 4.60]
	Sitting	896.43 (151.53)	3.79 (0.61)
	1st Lab (*n*=45)	[590.93 – 1301.51]	[2.20 – 4.94]
VU-AMS	**Sitting**	889.33 (147.44)	3.80 (0.59)
	2nd Lab (*n*=45)	[585.86 – 1437.04]	[2.05 - 4.96]
	Breathing	874.72 (146.82)	3.91 (0.62)
	1st Lab (*n*=45)	[557.01 – 1220.17]	[1.89 – 5.15]
	Breathing	851.37 (147.68)	3.83 (0.69)
	2nd Lab (*n*=44)	[572.56 – 1344.88]	[2.09 – 5.38]
	Mental Stress	883.39 (146.10)	3.84 (0.54)
	1st Lab (*n*=45)	[530.77 – 1271.88]	[1.90 – 4.83]
	Mental Stress	887.37 (142.97)	3.80 (0.64)
	2nd Lab (*n*=44)	[512.94 – 1371.76]	[1.52 – 4.99]
	Walking	704.34 (97.58)	3.09 (0.54)
	1st Lab (*n*=45)	[516.18 – 985.97]	[1.47 - 4.70]
	Walking	697.70 (80.62)	3.18 (0.51)
	2nd Lab (*n*=43)	[527.27 – 945.73]	[1.88 – 4.18]

Note: IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms).

Feasibility

Laboratory*:* Of the 51 participants who completed the study, 35 (69%) were able to obtain complete data with all three monitors in both laboratory sessions. During laboratory sessions technical difficulties leading to data loss was applicable for the: VU-AMS monitor in three participants (6%), Cortrium for six participants (12%), and the Ithlete in 15 participants (29%). In total, 45 participants (88%) completed both laboratory sessions with: the Cortrium (46 hours and 18 minutes of data), 36 participants (71%) with the Ithlete (29 hours and 40 minutes), and 48 participants (94%) with the VU-AMS (55 hours and 36 minutes).

Ambulatory setting: Two of the 60 participants who started with the ambulatory measurements stopped collecting data but agreed to participate in the second laboratory session. Nine participants dropped out completely during the ambulatory part of the study. The remaining 49 participants could maximally obtain 3430 measurements (49 participants * 14 days * 5 measurements). These 49 participants collected 2519 measurements (213 hours and 17 minutes of data, 73,44%) with the Cortrium and 2182 measurements (176 hours and 20 minutes of data, 63,61%) with the Ithlete. Three participants experienced technical difficulties (one participant with the Cortrium, two with the Ithlete) leading to a loss of more than 50% of their data.

Evaluation: All 49 participants reported to have missed at least one measurement due to non-adherence to the instructions. Participants specified seven reasons for missing measurements: (i) work (25 times), (ii) spare time activities (22 times), (iii) forgot monitor and/or measurements (16 times), (iv) technical difficulties with monitor(s) (10 times; four out of 49 participants (6.25%) with the Cortrium, six out of 49 participants (9.38%) with the Ithlete), (v) travelling (9 times), (vi) technical difficulties with smartphone or connection (five times), and (vii) skin irritation (four times). Participants reported four reasons to continue with

the measurements: (i) having agreed to participate in the study (23 times), (ii) wanting to support research (16 times), (iii) being interested in the study results (12 times), and (iv) being supportive to the researchers (3 times).

The Cortrium was given average scores of 66.78 (SD = 20.00) for user-friendliness, 66.39 (SD = 22.55) for social acceptability, and 51.65 (SD = 27.45) for burdensomeness. The Ithlete was given average scores of 74.29 (SD = 23.16) for user-friendliness, 69.65 (SD = 23.90) for social acceptability, and 38.51 (SD = 25.72) for burdensomeness. The Welch two sample t-test indicated that using the Ithlete was evaluated as less burdensome than the Cortrium ($t(100) = 2.49$, $p = 0.01$), and that the monitors did not differ in user-friendliness and social acceptability.

Cross-instrument validation

Cross-instrument performance of the Cortrium and the Ithlete against the VU-AMS was tested. Result obtained with the Cortrium are comparable to the VU-AMS (details given in Table 2). Best agreement was found for IBI during the supine task of the first laboratory session (ICC = 1.00, 95% CI = 1.00 – 1.00, SEM = 1.11 ms). Reproducibility, expressed as the CV was 0.11%, indicated that about $^2/_3$ of the differences are within 0.11% of the mean IBI values. Lowest agreement was found for ln(RMSSD) during the walking task of the first laboratory session (ICC = 0.53, 95% CI = 0.23 – 0.73, SEM = 10.81 ln(ms), CV = 35.20%). However, even in this latter case the agreement based on the ICC values should still be interpreted as *fair*.

Results obtained with the Ithlete during most tasks were are also comparable to the VU-AMS (details given in Table 3). Best agreement was found for the IBI during the mental stress task of the second laboratory session (ICC = 1.00, 95% CI = 1.00 – 1.00, SEM = 1.39 ms). Reproducibility expressed as the CV was 0.15%: indicating that about $^2/_3$ of the differences are

within 0.15% of the mean IBI values. However, during the walking tasks of both laboratory sessions the ln(RMSSD) values calculated from the Ithlete data did deviate substantially from those obtained by the VU-AMS. Lowest agreement was found during the first walking task (ICC = -0.08, 95% CI = -0.10 - 0.78, SEM = 22.93 ln(ms), CV = 32.12%). It should be noted, however, that in both walking tasks the sample sizes were very small (n = 6, and 5, respectively) as motion artefacts resulted in missing data.

In sum, in the walking tasks, the Cortrium outperformed the Ithlete. However, under circumstances without motion artefacts, differences between the Cortrium and the Ithlete were negligible. This is visualized in the Bland-Altman plots given in Figure 3 for IBI data collected during the first laboratory session with the Cortrium and VU-AMS, and the supplementary materials Figures S1 to S7 for the other variables and sessions. Although the absolute mean differences are small, the Bland-Altman plots showed that the Cortrium tended to underestimated IBI (range: between -2.6 and -0.4 ms). The Ithlete tended to overestimated IBI (range: between -1.3 and 10.7 ms) Again, with small absolute mean differences, both monitors overestimated ln(RMSSD), with values ranging between -0.004 and 0.198 for the Cortrium and values ranging between 0.055 and 1.377 for the Ithlete.

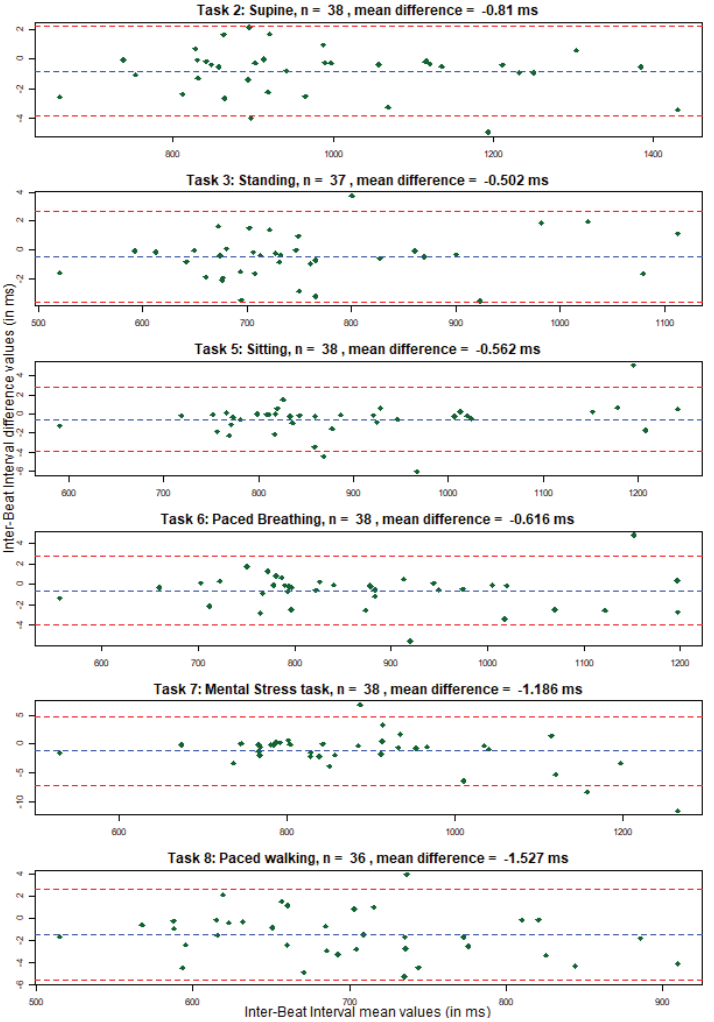


Figure 3. Bland-Altman plots of the Inter-Beat Interval data collected during the first laboratory session with the Cortrium versus the VU-AMS device (details on the laboratory session are described in the Method section and depicted in Figure 2). The blue dotted lines represent the mean difference between the Inter-Beat Interval values, while the red dotted lines represent the limits of agreement from negative 1.96 until positive 1.96 times the standard deviation of the differences. On the x-axis the Inter-Beat Interval mean values are given while the y-axis shows the differences between Inter-Beat Interval values obtained from the two devices."

TABLE 2
Cross-instrument reference method for all laboratory tasks during the first and second laboratory performance of the Cortrium when compared to the VU-AMS session.

Lab session	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
	Supine	ICC:	1.00	0.91
	(n = 38)	95% CI:	1.00 – 1.00	0.84 – 0.95
		SEM:	1.11	14.94
		CV:	0.11	17.12
		LOA:	[-3.88 – 2.26]	[-41.21 – 41.62]
	Standing	ICC:	1.00	0.89
	(n = 37)	95% CI:	1.00 – 1.00	0.80 – 0.94
		SEM:	1.13	5.71
		CV:	0.14	17.70
		LOA:	[-3.64 – 2.63]	[-13.12 – 18.52]
	Sitting	ICC:	1.00	1.00
	(n = 38)	95% CI:	1.00 – 1.00	1.00 – 1.00
		SEM:	1.22	0.65
		CV:	0.13	1.14
		LOA:	[-3.93 – 2.81]	[-1.42 – 2.17]
1	**Breathing**	ICC:	1.00	0.98
	(n = 38)	95% CI:	1.00 – 1.00	0.96 – 0.99
		SEM:	1.21	5.77
		CV:	0.13	9.02
		LOA:	[-3.97 – 2.73]	[-13.66 – 18.32]
	Mental	ICC:	1.00	0.81
	Stress	95% CI:	1.00 – 1.00	0.64 – 0.90
	(n = 38)	SEM:	2.18	13.41
		CV:	0.23	22.45
		LOA:	[-7.22 – 4.84]	[-29.91 – 44.43]
	Walking	ICC:	1.00	0.53
	(n = 36)	95% CI:	1.00 – 1.00	0.23 – 0.73
		SEM:	1.49	10.81
		CV:	0.20	35.20
		LOA:	[-5.65 – 2.60]	[-22.68 – 37.24]
	Supine	ICC:	1.00	0.94
	(n = 36)	95% CI:	1.00 – 1.00	0.89 – 0.97
		SEM:	1.54	14.20
		CV:	0.15	14.95
		LOA:	[-5.42 – 3.11]	[-41.47 – 37.22]
	Standing	ICC:	1.00	0.99
	(n = 36)	95% CI:	1.00 – 1.00	0.98 – 1.00
		SEM:	1.37	1.33

		CV:	0.17	4.25
		LOA:	[-4.51 – 3.10]	[-2.73 – 4.67]
	Sitting	ICC:	1.00	1.00
	(n = 35)	95% CI:	1.00 – 1.00	0.99 – 1.00
		SEM:	1.42	1.65
		CV:	0.15	3.13
		LOA:	[-4.38 – 3.51]	[-4.13 – 4.99]
2	**Breathing**	ICC:	1.00	0.98
	(n = 36)	95% CI:	1.00 – 1.00	0.97 – 0.99
		SEM:	1.92	5.32
		CV:	0.22	8.89
		LOA:	[-5.76 – 4.86]	[-12.87 – 16.64]
	Mental	ICC:	1.00	1.00
	Stress	95% CI:	1.00 – 1.00	0.99 – 1.00
	(n = 36)	SEM:	1.41	2.20
		CV:	0.15	3.92
		LOA:	[-4.83 – 2.97]	[-5.21 – 7.00]
	Walking	ICC:	1.00	0.88
	(n = 34)	95% CI:	1.00 – 1.00	0.74 – 0.94
		SEM:	2.27	4.03
		CV:	0.31	13.99
		LOA:	[-8.90 – 3.71]	[-8.39 – 13.95]

Note: IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms); *ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval of ICC; SEM: standard error of measurement in ms for IBI mean, ln(ms) for ln(RMSSD); CV: coefficient of variation in %; LOA: lines of agreement.*

TABLE 3
Cross-instrument performance of the Ithlete when compared to the VU-AMS reference method for all laboratory tasks during the first and second laboratory session.

Lab session	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
1	**Supine**	ICC:	1.00	0.96
	(n = 23)	95% CI:	1.00 – 1.00	0.89 – 0.99
		SEM:	5.01	8.42
		CV:	0.47	8.81
		LOA:	[-15.22 – 12.54]	[-16.69 – 29.99]
	Standing	ICC:	1.00	0.98
	(n = 24)	95% CI:	1.00 – 1.00	0.45 – 0.99
		SEM:	1.79	1.61
		CV:	0.22	4.77
		LOA:	[-4.47 – 5.46]	[-0.91 – 8.00]
	Sitting	ICC:	1.00	0.94
	(n = 23)	95% CI:	0.99 – 1.00	0.71 – 0.98
		SEM:	10.04	4.80
		CV:	1.06	8.42
		LOA:	[-24.76 – 30.89]	[-7.71 – 18.88]
	Breathing	ICC:	1.00	0.98
	(n = 24)	95% CI:	1.00 – 1.00	0.48 – 1.00
		SEM:	5.03	2.28
		CV:	0.54	3.35
		LOA:	[-14.22 – 13.67]	[-0.75 – 11.92]
	Mental Stress	ICC:	1.00	0.98
		95% CI:	1.00 – 1.00	0.95 – 0.99
	(n = 23)	SEM:	6.88	3.13
		CV:	0.72	5.14
		LOA:	[-16.87 – 21.29]	[-6.50 – 10.88]
	Walking	ICC:	0.97	-0.08*
	(n = 6)	95% CI:	0.78 – 1.00	-0.25 – 0.47
		SEM:	10.28	41.17
		CV:	1.38	51.82
		LOA:	[-17.75 – 39.22]	[-39.91 – 188.32]
	Supine	ICC:	1.00	0.98
	(n = 27)	95% CI:	1.00 – 1.00	0.94 – 0.99
		SEM:	3.58	8.04
		CV:	0.34	7.77
		LOA:	[-10.21 – 9.64]	[-17.22 – 27.38]
	Standing	ICC:	1.00	0.79
	(n = 27)	95% CI:	1.00 – 1.00	0.44 – 0.92

		SEM:	3.24	6.78
		CV:	0.41	20.81
		LOA:	[-9.08 – 8.91]	[-11.98 – 25.63]
	Sitting	ICC:	0.88	0.63
	(n = 27)	95% CI:	0.76 – 0.94	0.34 – 0.81
		SEM:	35.98	18.57
		CV:	3.83	31.54
		LOA:	[-95309 – 104.40]	[-42.23 – 60.71]
2	**Breathing**	ICC:	1.00	0.84
	(n = 26)	95% CI:	1.00 – 1.00	0.62 – 0.93
		SEM:	2.87	11.31
		CV:	0.32	18.54
		LOA:	[-8.26 – 7.68]	[-22.56 - 40.14]
	Mental	ICC:	1.00	0.82
	Stress	95% CI:	1.00 – 1.00	0.65 – 0.92
	(n = 26)	SEM:	1.39	13.24
		CV:	0.15	21.89
		LOA:	[-4.43 – 3.29]	[-30.63 – 42.78]
	Walking	ICC:	0.99	0.20*
	(n = 5)	95% CI:	0.70 – 1.00	-0.1 – 0.78
		SEM:	6.21	22.93
		CV:	0.87	32.12
		LOA:	[-4.44 – 29.98]	[-1.42 – 125.69]

*Notes: IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms). ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval of ICC; SEM: standard error of measurement in ms for IBI mean, ln(ms) for ln(RMSSD); CV: coefficient of variation in %; LOA: lines of agreement. *: ICC <= 0.40 indicating poor reproducibility between measurements from the monitor and the reference method (see method section for more details).*

Reproducibility

Data assessed during the first and second laboratory session were not different for the Cortrium, the Ithlete, and the VU-AMS (see Table 4 for test-retest statistics). For both the VU-AMS and the Cortrium, we did not find any differences in IBI and ln(RMSSD) during any of the tasks between the first and the second laboratory measurement. For the VU-AMS *good* to *excellent* reliabilities were found (ICC range = 0.64 - 0.88). For the Cortrium, ICC values indicated *fair* to *excellent* reproducibility (ICC range = 0.53 – 0.90). For the Ithlete, no difference was found for data assessed in the supine task (ICC range = 0.82 – 0.86). However, differences were found in the standing, sitting, paced breathing, and mental stress tasks (i.e. lowest ICC values were obtained in the paced breathing task for ln(RMSSD) (ICC = -0.21, 95% CI = -0.71 - 0.35, SEM = 35.21 ln(ms), CV = 54.22%). Not enough observations were attained in the walking task to test for any systematic change due to motion artefacts interfering with R-peak detection. While the Cortrium and VU-AMS did not show differences between the first and second laboratory sessions, the Ithlete did show differences in four tasks for IBI and ln(RMSSD).

TABLE 4
Absolute and relative test-retest reproducibility of IBI and log-transformed RMSSD assessed during the first and second laboratory measurements showing all tested monitors and laboratory tasks.

Monitor	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
	Supine	ICC:	0.86	0.82
	(n = 35)	95% CI:	0.75 – 0.93	0.67 – 0.91
		SEM:	61.55	21.76
		CV:	5.97	25.09
		LOA:	[-190.56 – 150.67]	[-70.59 – 50.04]
	Standing	ICC:	0.86	0.61
	(n = 34)	95% CI:	0.73 – 0.93	0.35 – 0.79
		SEM:	48.14	12.60
		CV:	6.03	37.65
		LOA:	[-122.71 – 144.19]	[-35.16 – 34.70]
	Sitting	ICC:	0.85	0.70
	(n = 35)	95% CI:	0.72 – 0.92	0.48 – 0.84
		SEM:	53.85	16.38
		CV:	5.78	30.05
		LOA:	[-142.61 – 155.91]	[-43.91 – 46.88]
Cortrium	**Breathing**	ICC:	0.86	0.71
	(n = 35)	95% CI:	0.74 – 0.93	0.49 – 0.84
		SEM:	52.83	21.94
		CV:	5.81	35.23
		LOA:	[-133.04 – 159.84]	[-61.67 – 59.95]
	Mental	ICC:	0.87	0.70
	Stress	95% CI:	0.76 – 0.93	0.47 – 0.83
	(n = 35)	SEM:	49.61	19.37
		CV:	5.38	32.75
		LOA:	[-146.39 – 128.65]	[- 52.70 – 54.71]
	Walking	ICC:	0.90	0.53
	(n = 32)	95% CI:	0.80 – 0.95	0.23 – 0.74
		SEM:	26.54	11.95
		CV:	3.62	38.09
		LOA:	[-73.84 – 73.32]	[-30.18 – 36.05]
	Supine	ICC:	0.87	0.70
	(n = 15)	95% CI:	0.66 – 0.95	0.33 – 0.89
		SEM:	55.60	29.42
		CV:	5.49	29.97
		LOA:	[-178.70 – 129.54]	[-99.51 – 63.61]
	Standing	ICC:	0.74	0.38*
	(n = 15)	95% CI:	0.38 – 0.91	-0.17 – 0.74
		SEM:	46.36	13.28

		CV:	5.97	40.70
		LOA:	[- 126.72 – 130.27]	[-38.20 – 35.44]
	Sitting	ICC:	0.72	0.08*
	(n = 14)	95% CI:	0.33 – 0.90	-0.41 – 0.56
		SEM:	40.76	25.54
		CV:	4.50	45.87
		LOA:	[-142.12 – 83.86]	[-8370 – 57.91]
Ithlete	**Breathing**	ICC:	0.30*	-0.21*
	(n = 15)	95% CI:	-0.27 – 0.70	-0.71 – 0.35
		SEM:	83.88	35.51
		CV:	9.31	54.22
		LOA:	[-224.40 – 240.62]	[-96.80 – 100.03]
	Mental	ICC:	0.81	0.34*
	Stress	95% CI:	0.49 – 0.94	-0.24 – 0.74
	(n = 13)	SEM:	55.68	22.96
		CV:	5.99	37.98
		LOA:	[-176.72 – 131.96]	[-71.41 – 55.68]
	Walking	ICC:	Not enough observations	Not enough observations
	(n = 1)	95% CI:	Not enough observations	Not enough observations
		SEM:	Not enough observations	Not enough observations
		CV:	Not enough observations	Not enough observations
		LOA:	Not enough observations	Not enough observations
	Supine	ICC:	0.86	0.80
	(n = 42)	95% CI:	0.75 – 0.92	0.64 – 0.89
		SEM:	66.73	22.73
		CV:	6.35	25.35
		LOA:	[-202.88 – 167.06]	[-73.98 – 52.01]
	Standing	ICC:	0.88	0.79
	(n = 42)	95% CI:	0.79 – 0.93	0.64 – 0.88
		SEM:	50.34	8.98
		CV:	6.30	29.03
		LOA:	[-134.67 – 144.41]	[-26.61 – 23.16]
	Sitting	ICC:	0.84	0.68
	(n = 41)	95% CI:	0.71 – 0.91	0.47 – 0.81
		SEM:	60.70	15.49
		CV:	6.48	29.39
		LOA:	[-172.07 – 164.42]	[-44.10 – 41.78]
VU-AMS	**Breathing**	ICC:	0.86	0.72
	(n = 40)	95% CI:	0.75 – 0.92	0.54 – 0.84
		SEM:	57.39	21.23
		CV:	6.28	34.53
		LOA:	[-146.44 – 171.71]	[-51.28 – 59.43]
		ICC:	0.83	0.80

Mental	95% CI:	0.70 – 0.91	0.65 - 0.89
Stress	SEM:	61.06	13.56
(n = 40)	CV:	6.51	24.39
	LOA:	[-177.73 – 160.74]	[-40.97 – 34.20]
Walking	ICC:	0.86	0.64
(n = 39)	95% CI:	0.76 – 0.93	0.41 – 0.80
	SEM:	32.00	6.50
	CV:	4.34	25.18
	LOA:	[-92.96 – 84.41]	[-21.14 – 14.89]

*Note: IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms). ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval of ICC; SEM: standard error of measurement in ms for IBI mean, ln(ms) for ln(RMSSD); CV: coefficient of variation in %; LOA: lines of agreement. *: ICC <= 0.40 indicating poor reproducibility, and significant differences between measurements during the first and second laboratory sessions. Comparison of data of Lab 1 and Lab 2 tested with paired Student's t-test showed all being non-significant (p > 0.05), except the walking task of the Ithlete which did not had enough observations to perform the t-test. The Cortrium and VU-AMS did not show significant differences between the first and second laboratory sessions. The Ithlete did show differences in four tasks for IBI means and ln(RMSSD).*

Discussion

In this study we tested the cross-instrument performance of two wireless heart rate monitors, the Cortrium ECG C3 and the Ithlete PPG finger sensor, against a standard wired reference method under controlled laboratory conditions. Moreover, we studied the test-retest reproducibility of these monitors over a period of 14 days, while their ambulatory feasibility was also investigated. We found that both the Cortrium and the Ithlete offer good to excellent cross-instrument agreement with the reference method under five standardized laboratory tasks, namely: supine, standing, sitting, paced breathing, and mental stress. The Cortrium did also perform well in a walking task, whereas the Ithlete showed inferior performance under such circumstances due to its higher sensitivity to motion artefacts. Test-retest analyses showed that results obtained with both the VU-AMS reference and the Cortrium monitor were comparable. Ithlete test-retest results were less robust, although IBI's during supine, standing, sitting, and the mental stress tasks showed good to excellent reproducibility.

Regarding feasibility, during ambulatory measurements, both the Cortrium and the Ithlete delivered at least two thirds of the maximum possible measurements. Participants reported that measurements were missed due to daily interferences, such as work obligations or leisure time activities. As all participants reported to have missed at least one measurement due to such non-adherence to instructions, we can identify non-adherence as an important contributing factor for missing data. Less often monitor related reasons were reported, such as technical difficulties and skin irritation due to ECG spot-electrodes. These findings can be interpreted as that HR(V) data collection with both wireless devices is feasible in highly protocolled ambulatory settings, although more valid data were obtained with the Cortrium than with the Ithlete. Main reasons reported for compliance were having agreed to complete the study and wanting to support research. It seems therefore worthwhile to invest in the participant- researcher relationship to reduce the amount of missing data in a study. Participants

reported no differences between the Cortrium and the Ithlete on user-friendliness and social acceptability.

We conclude from the current study that under most of the laboratory tasks, the Cortrium and the Ithlete showed good to excellent agreement with a standard wired ECG reference method when assessing HR(V). It was shown that for measuring HR(V) during tasks that do not involve gross body movements or physical activity, researchers are not limited to standard wired ECG monitors but can also opt for the modern wireless heart rate monitors investigated in this study. These wireless monitors offer a number of advantages of interest to researchers such as: online data storage, no need for battery replacement, giving access to the raw data, and lower monitor costs. There was, however, a difference between the Cortrium and the Ithlete in sensitivity to motion artefacts, which is associated with the larger amount of missing data obtained with the Ithlete (especially during tasks which include motion such as the walking tasks). Our findings indicate that the Cortrium recordings are fairly robust to motion. As such, the signal of the Cortrium is expected to be not as strongly affected by motion artefacts as consumer-grade wrist-worn PPG monitors, such as Fitbit monitors or the Apple Watch, whose signal is less robust under motion conditions than ECG monitors such as the Polar H7 chest-strap (Wang et al., 2017). Conversely, wrist-worn PPG monitors do offer their own set of characteristics, which could offer advantages regarding feasibility in some research designs. For example, such wearables can offer the ability to provide continuous recordings, as the device can be worn comfortably for long periods. This is due to such wrist-worn wearables often being designed to be worn as a bracelet or a watch. This could also prevent participants forgetting the monitor or the measurements. In our study, the reproducibility of the Ithlete monitor dropped considerably during motion, more so than various wrist-worn PPG monitors under the motion condition (Wang et al., 2017). Such dissimilarities could be due to differences in laboratory tasks, for instance, using a treadmill walking task versus walking a predetermined

route alongside a research assistant for a pre-set duration. A future study investigating both types of PPG monitors under similar conditions could elucidate whether wrist-worn PPG monitors definitively outperform finger-worn PPG monitors during motion. The robustness of the Cortrium signal during motion is similar to a chest-strapped ECG monitor, such as the Polar H7. Additionally, it avoids some disadvantages of chest-strapped monitors, such as wearability issues during long-term measurements. While a chest-strap ECG such as the Polar H7 does offer relatively robust signal and was thus considered for use in the current study, it was found less suitable for our long-term monitoring goals due to wearability issues such as obstructions of clothes while putting on the monitor. Additionally, these modern heart rate monitors offer lower prices compared to fully-fledged Holter ECG monitor, as for the price of one VU-AMS system researchers can acquire approximately three to four Cortrium C3's, or 100 Ithlete finger monitors. Such advantages offer researchers new opportunities for designing longitudinal studies wherein HR(V) data is monitored over weeks or months, in large samples within approximately the same budget. Longitudinal studies are necessary when studying dynamic processes, such as transitions from a healthy to a clinically affected state in patients, which unfold over timeframes longer than those studied in short-term research designs. The current study showed that the long-term ambulatory data collection required for such longitudinal studies is indeed feasible, although precautions are to be taken to minimize data loss and to improve adherence to instructions by participants.

Results showed that the PPG-based Ithlete did perform less well in conditions with higher risk of motion artefacts, such as walking. This finding is in line with earlier research showing the vulnerability of PPG measurements to motion artefacts (Trivedi et al., 1997), while extending these earlier findings to both controlled laboratory settings as well as ambulatory settings. Hence, when considering whether to choose the Ithlete or the Cortrium for a scientific study one should consider if heart rate measurements are under conditions free

of potential motion artefacts, and whether participants can be recruited easily and inexpensively. Under such conditions the Ithlete could be a sensible choice. However, under other conditions, for example in physical active situations, the Cortrium would be the more sensible choice. Moreover, the amount of data yielded from the Cortrium was higher than that of the Ithlete (83% and 69% of the maximum possible amount, respectively) during the ambulatory measurement period. Therefore, in scientific and clinical contexts wherein minimizing missing data is required, the Cortrium does hold the advantage. This advantage is grounded in the higher robustness of the Cortriums' ECG signal to motion related disruptions of stationarity when compared to the Ithlete's PPG signal. The correspondence between the Cortrium and the VU-AMS is hardly surprising as both monitors measure the ECG signal of lead II, thus delivering R-peaks which are relatively large and easy to detect. It should be noted though that the larger distance between the ECG spot-electrodes for the VU-AMS measurements allow for an even more robust signal, even during 24h monitoring in participants in physical active occupations (Riese et al., 2004; Vrijkotte, van Doornen, & de Geus, 2000).

While the current study showed agreeable performance of two wireless heart rate monitors in comparison to a wired ECG monitor, there are some limitations to be noted. First, reliability and validity characteristics of the HR(V) data were obtained from cross-instrument results under a controlled laboratory setting. These findings will thus only generalize to similar laboratory settings only. Reliability and validity of the Ithlete and Cortrium in ecological valid, unprotocolled ambulatory settings remains to be established as the two wireless devices were not tested against the ECG reference method and participants monitored themselves in real-life according to a highly standardized protocol (viz. after stabilization of the signal, in sitting posture). We did show that with both the Ithlete and the Cortrium HR(V) monitor data collection at home is feasible, although more valid data were collected with the Cortrium than

with the Ithlete. Second, we only assessed IBI and ln(RMSSD) calculated from the data yielded by the investigated heart rate monitors. There are, however, a multitude of other HR(V) metrics that are of interest to researchers, such as the *high frequency spectral power band* (0.15-0.40 Hz) of IBIs. Future studies could extend the current study to assess performance of the Cortrium and Ithlete regarding these metrics. Although no major differences should be expects as these metrics often correlate strongly with each other (Massin et al., 1999; Shaffer & Ginsberg, 2017). Third, in order to provide insights in reasons for missing data, for example what number of measurements is missed due to either monitor issues or non-adherence to instructions, a dedicated ESM question could be included in future studies. This is recommended as it would provide more detail on the amount and reasons for missing data. Fourth, as participants wore the heart rate monitors simultaneously, during laboratory as well as daily life measurements, some burden might have been experienced. However, none of the participants indicated during the evaluation interview that this negatively impacted feasibility, or that it was a reason for missing measurements. Fifth, a method-specific limitation of PPG is that circulation characteristics can result in a phase delay between R-peak and volume pulse start (Lu, Yang, Taylor, & Stein, 2009). However, in our sample of young and healthy participants such variations in delays can be assumed to be negligible (Drinnan, Allen, & Murray, 2001) and was not expected to have interfered with the conclusions of the study. Sixth, feasibility results have indicated that there were more technical difficulties with the Cortrium and Ithlete devices than with the reference method during the laboratory sessions. While this can be considered a limitation of feasibility, the reference method is unsuitable for our specific future study goals as, for example, it was not wireless, and not allowing participants to initialize the monitor through their smartphones. However, the devices tested in the current study were selected based on multiple criteria given in the introduction, instead of focusing solely on robustness to technical difficulties.

Lastly, it should be noted that while our study did include a mental stress task, we did not observe the expected cardiovascular response in any of the monitors, but instead showing effects similar to those in the sitting and breathing tasks (see Table 1). This suggests that the used task setup was insufficient to elicit the expected cardiovascular effects of the mental stress task.

In conclusion, two modern wireless heart rate monitors, the Cortrium and Ithlete, are able to provide data quality on par with a standard wired ECG reference method under controlled laboratory circumstances. Although both the Cortrium and the Ithlete performed similarly during non-motion tasks, the Cortrium was more robust during motion. Highly protocolled monitoring with the wireless devices in ambulatory daily-life setting is feasible. Participants highlighted work and spare-time activities as most common reasons to miss a measurement. Overall, we conclude that researchers can benefit from the advantages of modern wireless heart rate monitors such as online data storage and the absence of battery replacements without fully sacrificing data quality.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

cortrium.com. (2019). Retrieved from https://www.cortrium.com/ on April 11th 2019.

Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment : a handbook for dealing with variability and uncertainty in models and inputs*. Plenum Press: New York

de Geus, E. J. C., Willemsen, G. H. M., Klaver, C. H. A. M., & van Doornen, L. J. P. (1995). Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. *Biological Psychology*, *41*(3), 205–227. https://doi.org/10.1016/0301-0511(95)0137-6

Drinnan, M., Allen, J., & Murray, A. (2001). Relation between heart rate and pulse transit time during paced respiration. *Physiological Measurement*, *22*(3), 425-432. doi: 10.1088/0967-3334/22/3/301

El-Amrawy, F., & Nounou, M. I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare Informatics Research*, *21*(4), 315–320. https://doi.org/10.4258/hir.2015.21.4.315

Eliasson, K., Hjemdahl, P., & Kahan, T. (1983). Circulatory and sympatho-adrenal responses to stress in borderline and established hypertension. *Journal of Hypertension*, *1*(2), 131–139. https://doi.org/10.1097/00004872-198308000-00004

Ellis, R. J., Sollers III, J. J., Edelstein, E. A., & Thayer, J. F. (2008). Data transforms for spectral analyses of heart rate variability. *Biomedical Sciences Instrumentation, 44*, 392–397. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19141947

Freyschuss, U., Hjemdahl, P., Juhlin-Dannfelt, A., & Linde, B. (1988). Cardiovascular and sympathoadrenal responses to mental stress: Influence of β-blockade. *American Journal of Physiology - Heart and Circulatory Physiology*, *255*(6). https://doi.org/10.1152/ajpheart.1988.255.6.h1443

Goedhart, A. D., van der Sluis, S., Houtveen, J. H., Willemsen, G., & de Geus, E. J. C. (2007). Comparison of time and frequency domain measures of RSA in ambulatory recordings. *Psychophysiology*, *44*(2), 203–215. https://doi.org/10.1111/j.1469-8986.2006.00490.x

Greaves-Lord, K., Tulen, J., Dietrich, A., Sondeijker, F., van Roon, A., Oldehinkel, A., O… Huizink, A. (2010).

Reduced autonomic flexibility as a predictor for future anxiety in girls from the general population: The TRAILS study. *Psychiatry Research*, *179*(2), 187–193. https://doi.org/10.1016/j.psychres.2009.04.014

ithlete.com. (2019). Retrieved from https://www.myithlete.com/ at March 4th 2019

Kemp, A. H., & Quintana, D. S. (2013). The relationship between mental and physical health: Insights from the study of heart rate variability. *International Journal of Psychophysiology 89*(3), 288–296. Elsevier. https://doi.org/10.1016/j.ijpsycho.2013.06.018

Kennedy, H. L. (2013). The evolution of ambulatory ECG monitoring. *Progress in Cardiovascular Diseases*, *56*(2), 127–132. https://doi.org/10.1016/j.pcad.2013.08.005

Kramer, I., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., … Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry*, *13*(1), 68–77. https://doi.org/10.1002/wps.20090

Kroeze, R., van der Veen, D. C., Servaas, M. N., Bastiaansen, J. A., Oude Voshaar, R. C. O. V., Borsboom, D., … Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research*, *3*(1), 1–11. https://doi.org/10.17505/jpor.2017.01

lellamo, F., Legramante, M., Raimondi, G., Castrucci, F., Massaro, M., & Peruzzi, G. (1996). Evaluation of reproducibility of spontaneous baroreflex sensitivity at rest and during laboratory tests. *Journal of Hypertension*, *14*(9), 1099–1104. https://doi.org/10.1097/00004872-199609000-00009

Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of Medical Engineering and Technology*, *33*(8), 634–641. https://doi.org/10.3109/03091900903150998

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Massin, M. M., Derkenne, B., & von Bernuth, G. (1999). Correlations between indices of heart rate variability in healthy children and children with congenital heart disease. *Cardiology*, *91*(2), 109–113. https://doi.org/10.1159/000006889

Mulder, L.J.M., van Roon, A.M., & Schweizer, D. (1995). *CARSPAN*. Groningen: Iec ProGAMMA.

Nardelli, M., Vanello, N., Galperti, G., Greco, A., & Scilingo, E. P. (2020). Assessing the Quality of Heart Rate Variability Estimated from Wrist and Finger PPG: A Novel Approach Based on Cross-Mapping Method. Sensors (Basel, Switzerland), 20(11), 3156. https://doi.org/10.3390/s20113156

R Core Team. (2017). "R: A language and environment for statistical computing. R Foundation for Statistical Computing". Retrieved from: https://www.R-project.org/.

Riese, H., van Doornen, L. J. P., Houtman, I. L. D., & de Geus, E. J. C. (2004). Job strain in relation to ambulatory blood pressure, heart rate, and heart rate variability among female nurses. *Scandinavian*

Journal of Work, Environment and Health, *30*(6), 477–485. https://doi.org/10.5271/sjweh.837

Scheffer, M. (2010). Complex systems: Foreseeing tipping points. *Nature*, *467*(7314), 411–412. https://doi.org/10.1038/467411a

Schoevers, R. A., Van Borkulo, C. D., Lamers, F., Servaas, M. N., Bastiaansen, J. A., Beekman, A. T. F., … Riese, H. (2020). Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. *Psychological Medicine*. https://doi.org/10.1017/S0033291720000689

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/fpubh.2017.00258

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591. https://doi.org/10.2307/2333709

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

Shin, K., Hwang, H. T., Kim, Y. H., Kim, J. P., Yeo, H. S., Han, W., … Park, J. C. (2005). WHAM: A novel, wearable heart activity monitor based on Laplacian potential mapping. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, *7*, 7361–7364. https://doi.org/10.1109/iembs.2005.1616212

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

Stahl, S. E., An, H.-S., Dinkel, D. M., Noble, J. M., & Lee, J.-M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport & Exercise Medicine*, *2*(1), e000106. https://doi.org/10.1136/bmjsem-2015-000106

Teng, X. F., & Zhang, Y. T. (2003). Study on the peak interval variability of photoplethysmogtaphic signals. *APBME 2003 - IEEE EMBS Asian-Pacific Conference on Biomedical Engineering 2003*, 140–141. https://doi.org/10.1109/APBME.2003.1302623

Thakor, N. V., & Zhu, Y. S. (1991). Applications of Adaptive Filtering to ECG Analysis: Noise Cancellation and Arrhythmia Detection. *IEEE Transactions on Biomedical Engineering*, *38*(8), 785–794. https://doi.org/10.1109/10.83591

Trivedi, N. S., Ghouri, A. F., Shah, N. K., Lai, E., & Barker, S. J. (1997). Effects of motion, ambient light, and hypoperfusion on pulse oximeter function. *Journal of Clinical Anesthesia*, *9*(3), 179–183. https://doi.org/10.1016/S0952-8180(97)00039-1

Vaessen, T., Steinhart, H., Batink, T., Klippel, A., van Nierop, M., Reininghaus, U., & Myin-Germeys, I. (2019). ACT in daily life in early psychosis: an ecological momentary intervention approach. *Psychosis*, *11*(2), 93–104. https://doi.org/10.1080/17522439.2019.1578401

Vrijkotte, T. G., van Doornen, L. J. P., & de Geus, E. J. C. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension, 35(4)*, 880-886. doi: 10.1161/01.hyp.35.4.880

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology*, *2*(1), 104–106. https://doi.org/10.1001/jamacardio.2016.3340

Wegner, F. K., Kochhäuser, S., Ellermann, C., Lange, P. S., Frommeyer, G., Leitz, P., … Dechering, D. G. (2020). Prospective blinded Evaluation of the smartphone-based AliveCor Kardia ECG monitor for Atrial Fibrillation detection: The PEAK-AF study. *European Journal of Internal Medicine*, *73*, 72–75. https://doi.org/10.1016/j.ejim.2019.11.018

Wichers, M., & Groot, P. C. (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

Willemsen, G. H. M., de Geus, E. J. C., Klaver, C. H. A. M., van Doornen, L. J. P., & Carroll, D. (1996). Ambulatory monitoring of the impedance cardiogram. *Psychophysiology*, *33*(2), 184–193. https://doi.org/10.1111/j.1469-8986.1996.tb02122.x

Winokur, E. S., Delano, M. K., & Sodini, C. G. (2013). A wearable cardiac monitor for long-term data acquisition and analysis. *IEEE Transactions on Biomedical Engineering*, *60*(1), 189–192. https://doi.org/10.1109/TBME.2012.2217958

Zhang, D., Hu, X., Li, J., Liu, J., Baks-te Bulte, L., Wiersma, M., … Brundel, B. J. J. M. (2019). DNA damage-induced PARP1 activation confers cardiomyocyte dysfunction through NAD + depletion in experimental atrial fibrillation. *Nature Communications*, *10*(1), 1–17. https://doi.org/10.1038/s41467-019-09014-2

7.

Predicting recurrence of depression using cardiac complexity in individuals tapering antidepressants

This chapter is in press as:

George, S. V., Kunkels, Y. K., Smit, A. C., Wichers, M., Snippe, E., van Roon, A. M., & Riese, H. (in press). Predicting recurrence of depression using cardiac complexity in individuals tapering antidepressants. Translational Psychiatry

Chapter 7: Predicting recurrence of depression using cardiac complexity in individuals tapering antidepressants

Sandip V. George, PhD[1,2], Yoram K. Kunkels, MSc[1], Arnout Smit, PhD, Marieke Wichers, PhD[1], Evelien Snippe, PhD[1], Arie M van Roon, PhD[3], Harriëtte Riese, PhD[1]

[1]University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), Groningen, The Netherlands

[2] University College London, Department of Computer Science, London, United Kingdom

[3]University of Groningen, University Medical Center Groningen, Department of Vascular Medicine, Groningen, The Netherlands

George, S. V., Kunkels, Y. K., Smit, A. C., Wichers, M., Snippe, E., van Roon, A. M., & Riese, H. (*in press*). Predicting recurrence of depression using cardiac complexity in individuals tapering antidepressants. *Translational Psychiatry*.

Abstract

It is currently unknown whether the complexity and variability of cardiac dynamics predicts future depression and whether within-subject change herein precedes recurrence of depression. We test this in an innovative repeated single-subject study in individuals who had a history of depression and were tapering their antidepressants. In 50 individuals, electrocardiogram (ECG) derived Interbeat interval (IBI) time-series data were collected for five minutes every morning and evening, for 4 months. Usable data were obtained from 14 participants who experienced a transition (i.e., clinically significant increase in depressive symptoms) and 14 who did not. The mean, standard deviation, Higuchi dimension and multiscale entropy, calculated from IBIs, were examined for time-trends. These quantifiers were also averaged over a baseline period and compared between the groups. No consistent trends were observed in any quantifier before increases in depressive symptoms within individuals. The entropy baseline levels significantly differed between the two groups (Morning: p-value <0.001, Cohen's d=-2.185; Evening: p-value <0.001, Cohen's d=-1.797) and predicted the recurrence of depressive symptoms, in the current sample. Moreover, higher mean IBIs and Higuchi dimensions were observed in individuals who experienced transitions. While we found little evidence to support the existence of within individual warning signals in IBI time-series data preceding an upcoming depressive transition, our results indicate that individuals who taper antidepressants and showed lower entropy of cardiac dynamics exhibited a higher chance of recurrence of depression. Hence, entropy could be a potential digital phenotype for assessing the risk of recurrence of depression in the short-term while tapering antidepressants.

Introduction

Determining the risk of recurrence of depression, especially when tapering antidepressants is a challenging problem. Tapering of antidepressants can typically lead to a worsening of depressive symptoms (1–4). Hence, warning signs indicating the possibility of recurrence of depression or worsening of symptoms are of immense importance. Complex dynamical systems theory predicts the presence of early warning signals in the response of a system before many kinds of transitions (5). Recent studies based on this have shown promise in predicting depressive episodes from momentary affect data(6–8). Such changes in the dynamics of depression could lead to potential warning signals in cardiac dynamics as well.

These could include measures such as the heart rate, heart rate variability, and complexity, all of which have been shown to be altered in patients with depression (9–11) .

Since the response of the heart is well understood to be nonlinear, it is prudent to study its nonlinear dynamics when seeking warning signals for depression (12). These nonlinear dynamics can be quantified from the electrocardiogram (ECG) derived InterBeat-Interval (IBI) time series, using complexity measures such as the entropies, dimensions, and Lyapunov exponents. Disorders of various types including mental disorders (13–19), are associated with a reduction in the complexity of dynamics of the heart. The complexity of cardiac dynamics, as well as simpler measures such as the mean and variability of IBI have been shown to be reduced in individuals diagnosed with depression, as well as dysphoria (13,14,20–22), although there is debate about whether this reduction can be explained completely by the effect of antidepressants (11,23,24). For complexity measures of IBI time-series to be potentially used as an early warning indicator for upcoming increases in depressive symptoms, a reduction in the complexity of cardiac dynamics must occur in the period before transitions towards higher levels of depression. This has not been empirically studied yet.

To examine whether a reduction in the complexity of cardiac dynamics over time occurs just before patients transition towards higher levels of depressive symptoms, a single-subject design including IBI time-series data may be employed. A between-subject design, on the other hand, is appropriate if one wants to study average differences in complexity of cardiac dynamics that exist in the sample. The present TRANSitions In Depression (TRANS-ID) Tapering study employs a repeated single-subject design, where intensive longitudinal data of different types (momentary affect, physical activity and ECG) were collected for four months within formerly depressed individuals tapering their antidepressants, offering the possibility for both within-subject as well as between-subject studies (25–28).

We examine whether a decrease in the mean, standard deviation and complexity in IBI time-series data as captured with the Higuchi dimension and multiscale entropy precedes a depressive transition (i.e., recurrence of depressive symptoms) by 4 to 8 weeks, a timescale observed in previous studies (6,7,29). These complexity measures were chosen as they capture two different aspects of complexity. While the Higuchi dimension represents the number of variables required to capture the dynamics of the underlying process from which the time-series is derived, the multiscale entropy captures the information content in the time-

series. We conduct repeated single within-subject analyses, where we study whether decreases in these quantifiers over time precedes a transition towards depression for each individual separately in formerly depressed individuals who taper their antidepressant medication. Furthermore, to study average tendencies, we also conduct a between-subject analysis to test whether the baseline complexity (chosen as the first 4 weeks of assessments), is lower for individuals who experienced a transition towards higher depressive symptom levels during the study period versus those who did not.

Methods

Sample

Our sample consisted of participants of the TRANS-ID Tapering study, a study that aimed at examining early warning signals of increases in depressive symptoms during and after tapering of antidepressant medication (for details see (26)). In short, 69 individuals who had an earlier diagnosis with major depressive disorder (MDD) according to DSM-IV criteria monitored themselves for four months with weekly questionnaires, Ecological Momentary Assessment (EMA), actigraphy, and ECG sensors. These individuals made a shared decision with their mental health care provider to taper their antidepressant dosage (see SA1 for details) and did not meet the criteria for MDD at baseline.

The study was approved by the Medical Ethical Committee of the University Medical Center Groningen (UMCG, METc2016.443). All patients were informed that they could stop their participation at any time and provided written informed consent prior to participation.

Participants and procedures TRANS-ID Tapering ECG sub study

The flowchart of the TRANS-ID Tapering ECG sub study is shown in Figure 1. Out of 69 individuals, 50 individuals had usable ECG recordings. The presence or absence of transitions could be reliably calculated in 45 of these individuals. Among them 29 experienced a transition in depressive symptoms, while 16 did not. From those with a transition we excluded 7 individuals who did not have at least 3 weeks of data prior to the transition, to avoid spurious trends caused by too few datapoints. Also excluded from this group were 7 individuals whose transitions occurred outside of the ECG measurement period and 1 individual who had only morning assessments. Hence, of the original 69 individuals,

valid ECG data were available for 14 individuals with a transition, which formed the transition group in the current paper. Of the 16 who did not show any transitions, 2 individuals were excluded, as they had less than 3 weeks of data available for analysis, leaving a sample of 14 individuals who did not experience any transitions for analysis.

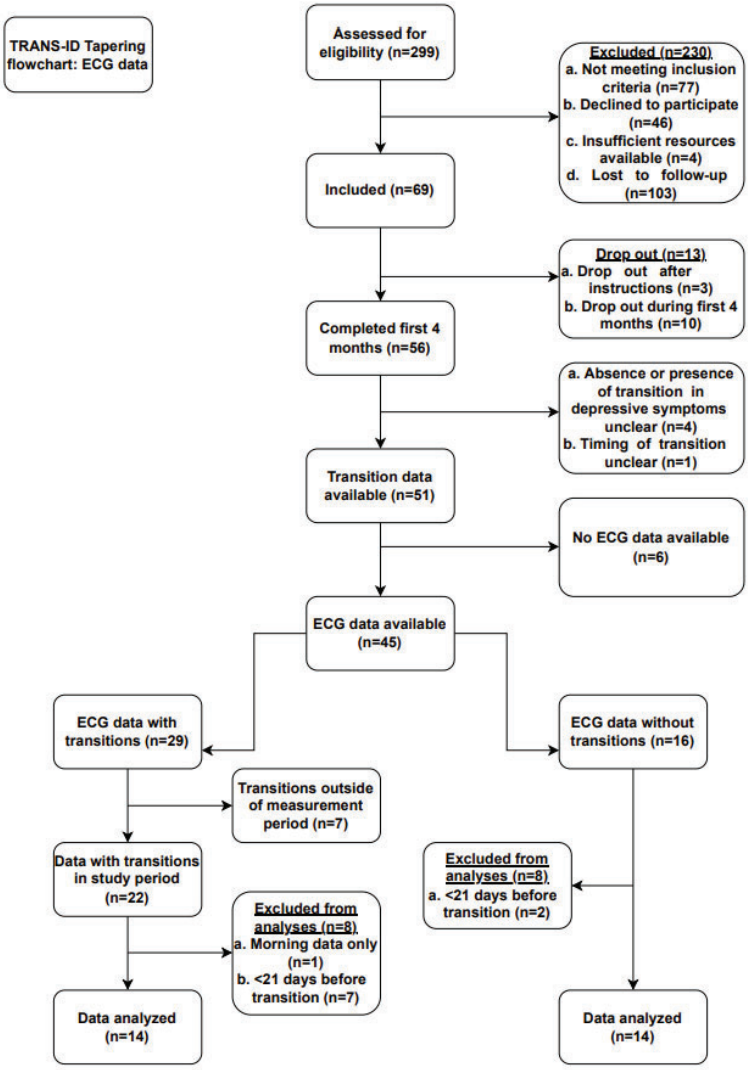


Figure 1: Flowchart describing the patient inclusion for the present study.

ECG assessments and pre-processing

Participants performed their ECG assessments at home after receiving a 15 minute step-by-step instruction on how to do so during the introductory session. Additionally, participants received a written manual (see https://osf.io/zbrxe/), ECG-electrodes, and contact details for 24/7 support (more details are given in SA2). The ECG files were processed meticulously to extract the InterBeat Interval (IBI) time-series data as detailed in SA3.

IBI data processing

Independent Variables: Four quantifiers were derived from each IBI time-series assessment; the mean and standard deviation, and two complexity measures, namely the Higuchi dimension and the multiscale entropy. All four quantifiers have been widely used to study cardiac dynamics (13,22,30–33).

Mean: This is the average of the IBIs representing the mean time between two R-peaks. A higher mean IBI is reflective of a lower heart rate.

Standard deviation: The standard deviation of the IBI time-series is referred to as the SDRR (standard deviation of RR intervals) and is a time-domain measure of the heart rate variability (33).

Higuchi dimension: The Higuchi dimension estimates the fractal dimension of a time-series directly, without any need for embedding in higher dimensions, which reduces the number of points required for a reliable calculation. It calculates the scaling behavior of the length of the time-series curves when two parameters, namely the delay time and initial time, are varied (34). In the present work, the maximum delay time is set to be 5, since the number of points used for calculating lengths reduces at higher delays.

Multi-scale entropy: The multiscale entropy measures the predictability of fluctuations in time-series, at different scales of measurement. A higher value of entropy indicates a higher complexity of the time-series. The multiscale entropy was estimated using the neurokit2 package (35) and the Higuchi dimension was calculated using the HDFA package (36) in Python v3.5.2.

Dependent Variable: The dependent variable in the study was a transition towards higher levels of depression based on the following criteria:1) a reliable increase (≥8.5 points) on the

weekly assessed SCL-90 depression subscale, 2) persistence of this increase for at least three weeks, and/or start or increase in treatment, and/or interruption of tapering, 3) a meaningful increase in depressive symptoms as experienced by participants based on a consensus rating of emails, telephone calls, open text fields, and the evaluation interview (see also (26,37)).

For the analyses in this paper, in order to compare the samples of individuals who experienced a transition with those that did not, a pseudo transition point was determined in the individuals who did not experience a transition. The transition times in the non-transitioning dataset were pair matched with the transition times of the transitioning dataset.

Statistical Analyses

The analyses are divided into within-subject and between-subject analyses, where the former identified changes in the quantifiers over time occurring within individuals, and the latter identified differences in average levels of the quantifiers during the 4 weeks of the study period between the individuals who experienced transitions and those who did not.

Within-individual analyses: To study changes at the level of an individual, each of the four quantifiers mentioned above was calculated for every IBI assessment, over a pre-transition period defined as 8 weeks before a transition. To avoid significant loss of data, all datasets were required to have a minimum of 3 weeks of data prior to the transition. These generated time-series of quantifiers were categorized into the morning and evening time-series. The Kendall correlation coefficient between these quantifiers and time was measured to determine the time-trends. Significant time trends, as well as the direction of such trends preceding transitions towards greater depressive symptoms and in patients who stayed in remission were studied, and the number of individuals with significant trends were quantified.

Between-individuals analyses: To study mean differences between individuals who experienced a transition in future and those that did not, we first took the mean values for each quantifier per individual, by averaging over the value for each measurement through the baseline period. The baseline period was considered as the first 4 weeks of measurement, with a minimum requirement of at least 3 weeks of data. These averaged quantifiers were compared between the groups by using the non-parametric Mann-Whitney U test. In addition to having multiple advantages over the more commonly used t-test, the Mann-Whitney U test

is more suitable for comparing small sample sizes and when the distributions are not normal, as in the case of heart rate variability measures (38,39). The effect size of the difference between the two groups was measured using Cohen's d (40).

In addition, to study how well each quantifier predicted an upcoming transition, we used logistic regression models to predict presence versus absence of a future transition using the calculated quantifiers at baseline. Since ECG variables are known to depend significantly on age, we also tested the models with age as a predictor. The goodness of fit was quantified using pseudo R2 values. The logistic regression was conducted in R version 3.6. (41).

Since the actual time of transition from baseline varied between individuals, a sensitivity analysis was conducted by averaging over the whole pre-transition period identified for the individual level study above. This controlled for the time elapsed between the assessments and the transition. The mean-differences and predictive capabilities of the different quantifiers was then studied for the data averaged over the pre-transition period.

Dependence among the quantifiers at baseline was measured using the Spearman correlation coefficient (Spearman's ρ) averaged at the level of an individual. Being a rank correlation coefficient the Spearman's ρ is both robust to outliers and can detect monotonic nonlinear trends. The p-value for significance was set at 0.05. The Mann-Whitney U tests, Kendall and Spearman correlations were performed using the scipy package in Python version 3.5.2 (42).

Results

Sample description

Data of 14 individuals who showed a transition during monitoring and data from 14 individuals who did not were analysed. The gender ratio did not differ between the groups (78% versus 71% women in the transition and non-transition groups respectively, p=0.66). Age was significantly higher in the transition group (M= 51.93 SD= 12.25, t = 2.19, p = 0.04) compared to the non-transition group (M = 41.79, SD = 11.30). The correlations between the different variables studied in our sample, averaged at baseline, with age and with each other are listed in Table 1. The mean and standard deviation of the IBI time series in the morning was related with age. Highest correlations were observed between the evening mean and standard

deviations, the morning entropy and morning dimension, and the evening mean and evening dimension measures.

Table 1: Correlations between the different ECG derived quantifiers used in this study for the baseline assessments.

	Age	$Mean_M$	SD_M	HD_M	MSE_M
Age	1	**0.568**	**0.440**	-0.146	-0.096
$Mean_E$	0.221	1 (0.188)	**0.646** (0.121)	-0.025 (0.037)	-0.238 (-0.193)
SD_E	0.182	**0.855** (0.111)	1 (0.090)	-0.258 (-0.033)	0.084 (-0.144)
HD_E	0.275	**0.709 (0.408)**	**0.634** (0.112)	1 (0.188)	**-0.742** (-0.361)
MSE_E	-0.299	-0.091 **(-0.416)**	-0.003 (-0.072)	**-0.438** (-0.329)	**1** **(0.540)**

Entries above the diagonal represent correlations between the morning assessments and entries below denote correlations between the evening assessments. Correlations of the morning assessments with evening assessments are given between the parentheses. The table lists the Spearman's ρ correlations. Significant correlations ($p<.05$) are listed in bold. SD = Standard Deviation, HD =Higuchi Dimension, MSE=Multiscale Entropy

Within-individual analyses

We started by examining the IBI quantifiers for each individual for significant trends over time using Mann-Kendall trend test. Few trends were found in the hypothesized negative direction, that is, a decrease over time for the Higuchi dimension and the multiscale entropy. Within the morning assessments, we observe negative trends (14%) for the Higuchi dimension in 2 out of the 14 individuals who experienced a transition and no trends among those who did not. No negative trends were observed for any individual with or without a transition in depression for the entropy quantifier. For the evening assessments, no negative trends were found for the Higuchi dimension, whereas one negative trend was found an individual without a transition (7%). No negative trends were observed for any of the individuals in either group for the entropy quantifier.

Positive time-trends were found in the morning assessments in 3 individuals (21%) for the Higuchi dimension and in 1 individual (7%) for the entropy among those with a transition. Three individuals in the non-transitioning group (21%) showed positive trends too for the Higuchi dimension, whereas no individuals in the non-transitioning group showed any trend for the entropy. In the evening assessments, the transitioning group showed no positive trends for the Higuchi dimension, whereas the entropy showed positive trends in 2 individuals (14%). The non-transitioning group showed positive trends in 4 individuals (29%) for the Higuchi dimension, and no trends for entropy. Detailed results for the within-individual analyses showing the trends for each individual and quantifier are presented in SA4.

Between-individuals analyses

Next, we studied group differences in the quantifiers averaged within individuals over the baseline period of 4 weeks. The mean differences for these averaged quantifiers between individuals who experienced a transition, and those who did not are listed in Table 2. Figure 2 shows the corresponding distributions for the two groups, as violin plots. For the morning assessments, individuals who experienced transitions showed a significantly higher mean IBI and Higuchi dimension, and a significantly lower entropy than individuals who did not experience a transition. For the evening assessments the individuals who experienced a transition showed a significantly higher Higuchi dimension and a significantly lower entropy than individuals who did not experience a transition.

Table 2: Differences in the person-averaged quantifiers between the group which experienced a transition and the group that did not.

Quantifier	Transition group (M ± SD)	Non transitioning group (M ±SD)	z-score	p-value	Cohen's d
$Mean_M$	**852.161±71.537**	**797.921±30.415**	**2.412**	**0.018***	**0.987**
SD_M	54.743± 8.976	58.788±1.752	0.253	0.133	-0.625
HD_M	**1.641 ±0.025**	**1.615 ±0.103**	**2.642**	**0.009***	**0.344**
MSE_M	**1.600 ±0.025**	**1.687 ±0.050**	**-3.469**	**<0.001****	**-2.185**
$Mean_E$	853.034 ±51.104	837.026 ±28.159	1.309	0.334	0.388
SD_E	45.475 ±8.362	43.738 ±4.907	0.804	0.525	0.253
HD_E	**1.698 ±0.061**	**1.635 ±0.096**	**2.550**	**0.012**	**0.785**
MSE_E	**1.464 ±0.046**	**1.569 ±0.069**	**-4.296**	**<0.001****	**-1.797**

The quantifiers, namely the mean, standard deviation, Higuchi dimension and Multiscale entropy, were calculated from IBI measurements, The assessments were taken every day and averaged over the baseline period. The subscripts (M or E) refer to the time of the day when the ECG measurement was carried out (morning or evening). SD = Standard Deviation HD = Higuchi Dimension, MSE = Multiscale Entropy. The differences between the groups were measured using a Mann-Whitney u test.

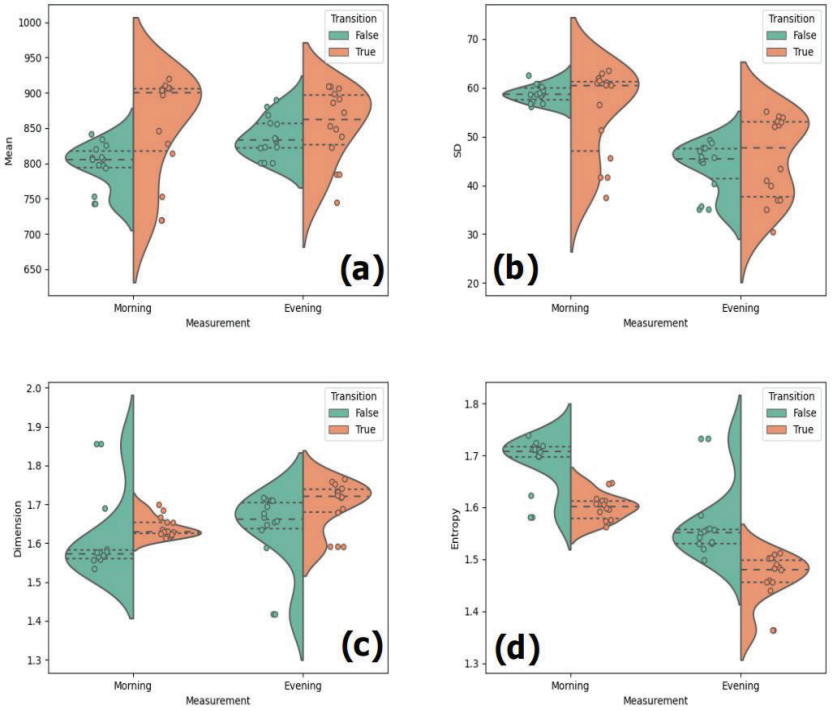


Figure 2: Violin plots showing the differences in the distributions of the person averaged quantifiers between individuals who experienced a transition and those who did not[a].

[a] The panels show (a) mean (b) standard deviation (c) Higuchi dimension and (d) Multiscale entropy. The distribution for individuals who experienced a transition are in orange and those who did not are in green. The circles represent the entropy values for each individual, scattered randomly along the x-axis. The quantifiers were averaged over the baseline periods.

The results of the logistic regression model used to predict whether the individual will undergo a transition or not, are given in Table 3. The model where the entropy alone predicts the transitions stood out with the highest explained variance among all the models

considered, with a lower baseline entropy significantly predicting a future transition towards higher depressive symptom levels.

In Figure 3, we show the values of baseline entropy and error for each individual, calculated using the morning and evening assessments. The individuals with a transition occupy a region in the lower left of the graph, pointing out once again that low values of entropies were largely associated with individuals who experienced transitions in the study period. A grid search on the entropy values found 1.67 as the morning entropy value (25/28 individuals correctly classified) and 1.51 as the evening entropy value that best discriminates the two groups (27/28 individuals correctly classified). Noticeably, in this sample, a smaller within-person standard deviation was observed in the entropies associated with the evening assessments, indicating that the entropy measurements during the morning were less stable than the evening.

Table 3: Logistic regression models predicting depressive transitions during study period.[a]

Predictor	Estimate	SE	z-value	p-value	R^2	Correctly predicted (%)
Transition~Age						
Age	**0.069**	**0.035**	**1.967**	**0.049**	0.203	67.9
Transition~$Mean_X$						
$Mean_M$	**16.803**	**7.774**	**2.161**	**0.031**	0.253	75.0
$Mean_E$	9.340	9.407	0.993	0.321	0.048	64.3
Transition~SD_X						
SD_M	-103.945	71.873	-1.446	0.148	0.122	60.7
SD_E	36.755	55.556	0.662	0.508	0.021	39.3
Transition~HD_X						
HD_M	4.706	5.407	0.870	0.384	0.039	75.0
HD_E	11.529	6.769	1.703	0.089	0.191	67.9
Transition~MSE_X						
MSE_M	**-37.75**	**12.49**	**-3.021**	**0.002**	0.635	82.1
MSE_E	**-150.73**	**73.41**	**-2.053**	**0.040**	0.880	96.4

[a] The models show how well transition status is predicted by the baseline quantifiers. The listed R^2 value is the Nagelkerke R^2.

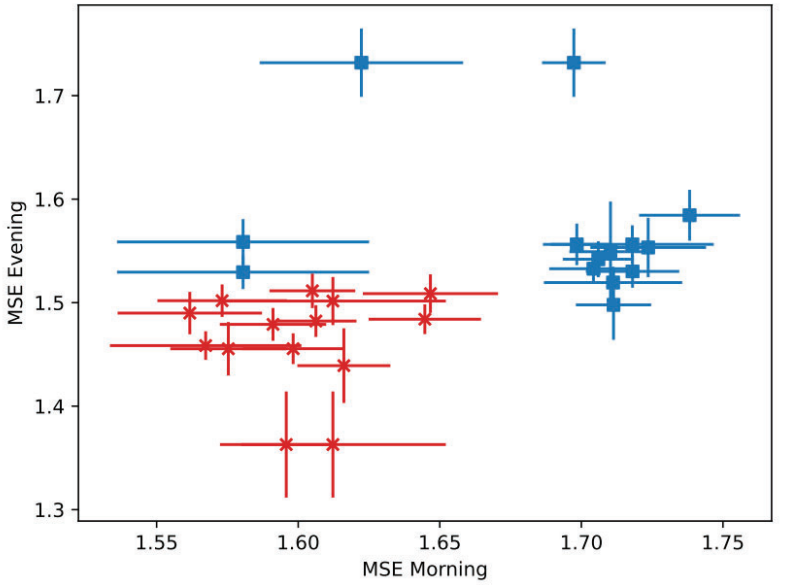


Figure 3: Scatter plot showing the entropy values for the morning and evening for each individual[a].

[a] Red crosses represent individuals who experienced a transition, and the blue circles represent individuals who did not. The error bars represent the standard error.

Sensitivity analysis

The sensitivity analysis, calculating the correlations, mean differences and predictive capacity of the quantifiers averaged over the pre-transition period instead of the baseline period, are presented in the supplements (SA6). Again, lower entropy over the pre-transition period was most strongly associated with the presence of a future transition in depressive symptoms, showing the highest significance in the Mann-Whitney U test and highest explained variance in a logistic regression.

Discussion

This work explored how the complexity of IBI time-series data behaves before a transition towards more severe depressive symptoms. While very few trends within individuals over time were observed in the different cardiac quantifiers before a transition towards depression,

we found that the baseline levels of entropy were significantly lower for individuals who experienced a transition compared to individuals who did not. In addition, we found higher mean IBIs and higher Higuchi dimension for individuals who experienced transitions. This seems to indicate that below a threshold level of complexity of cardiac dynamics, individuals who taper antidepressants are vulnerable for recurrence. Moreover, the combination of higher fractal dimension and lower entropy in individuals who experienced transitions suggests that these time series exhibit more noisy behaviour(43,44).

The present work is significant in multiple ways. First, it provides little evidence to support the existence of within individual warning signals in IBI time-series data that precede and predict an upcoming depressive transition, in line with similar studies using ecological momentary assessment and actigraphy data (25,37). Second, the current study shows that lower entropy values derived from IBI time-series indicate that individuals are more likely to experience an increase in depressive symptoms in the coming months, which may be helpful information when deciding on whether antidepressant medication should be tapered. Based on this we speculate that IBI time-series derived entropy quantifiers could become promising biomarkers for determining if antidepressants can be tapered with a reduced risk of recurrence of depression. Third, the current study answers an important question on how the complexity and variability of cardiac dynamics change before the recurrence of depression. While we find that the complexity of cardiac dynamics is significantly lower in individuals who experienced a transition towards increased depressive symptoms, no decrease in complexity over time was observed before transitions in most individuals. An explanation for the absence of this change is that the decrease in complexity may have taken place at a scale longer than the 8 weeks considered. An alternative reason could be that the loss of complexity is a stable vulnerability that persisted in some individuals from a previous episode of depression, since the sample consisted of individuals who experienced an episode previously. Based on past work, we expect that an earlier episode would have been associated with decreased complexity of cardiac dynamics (11,15), and individuals who experienced a transition in this sample possibly did not fully recover their complexity (45).

A major limitation of the study, originally designed for within-individual analysis, is the small sizes of the groups for the between-subject analyses. The current study explored multiple indicators in this small group of participants and may therefore be overfitting the

sample. Moreover, despite the large effect sizes observed for the entropy, inter-individual differences may not be fully captured. A second limitation is the tapering of antidepressants in the current sample. The cardiotropic effects of antidepressant medication on the dynamics of the heart are well documented, with many of them causing a reduction in the mean heart rate and its variability (22,46–48). This intake could dominate the effects of the upcoming transition, if any, on the cardiac dynamics. A considerable fraction of the present sample majorly tapered their antidepressants during the baseline period (17 participants reported tapering more than 2/3rds of their dosage in the baseline period). This could have changed the cardiac dynamics during the baseline period in both groups. A third limitation was the strict protocols set for the self-assessment of ECG. Apart from being burdensome, such strict assessment instructions could have resulted in cardiac dynamic signals with less noise, limiting generalization to other studies with ambulatory ECG assessments. However, with improvements in ambulatory ECG monitoring using less obstructive devices, it may be possible to monitor the complexity of the heart more easily for an extended time in normal daily life (49,50). Future studies are needed to estimate how well the results of the current study generalize to new samples.

The promising results of our study points to the need for a larger exploration of the use of cardiac complexity measures as a predictor for depression. If validated by future studies, patients who are planning on tapering their antidepressant medication may assess their ECG at home to assist decision making(51). Furthermore, we suggest including the IBI complexity measures used in this study in other models which predict recurrence of depression (52,53). While this study finds that complexity measures are lower in individuals who experience a recurrence of depressive symptoms, the results do not indicate whether this is true in individuals who experience a depressive episode for the first time, or in individuals who are not tapering antidepressant medication. We recommend exploring the variation of complexity of cardiac dynamics prior to transitions towards depression in other samples, where these drawbacks may not exist.

In conclusion, this study suggests that quantifiers of complexity of cardiac dynamics can serve as an indicator for future recurrence of depressive transitions. While the study failed to find any trends in these quantifiers preceding depressive symptom transitions, it suggests a strong possibility of using complexity-based quantifiers to identify individuals at risk for recurrence of depression (54). Though many challenges remain to be solved before a clinical

implementation is feasible, we believe that these indicators can greatly aid in decision making in the context of tapering antidepressants.

Acknowledgements

We would like to thank G. M. Bloem, D. Sloohof, G. Arts, and E. van den Kieboom for their work on (pre-)processing the physiological data. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-CoG-2015; No 681466 to M. Wichers). The Cortrium monitors were kindly provided by the iLab of the department of psychiatry of the University Medical Center Groningen (UMCG, http://www.ilab-psychiatry.nl).

References

1. Borges S, Chen YF, Laughren TP, Temple R, Patel HD, David PA, et al. Review of maintenance trials for major depressive disorder: A 25-year perspective from the US food and drug administration. Journal of Clinical Psychiatry. 2014.
2. Geddes JR, Carney SM, Davies C. Relapse prevention in antidepressant drug treatment in depressive disorders: a systematic review (vol 361, pg 653, 2003). Lancet. 2004;
3. Glue P, Donovan MR, Kolluri S, Emir B. Meta-analysis of relapse prevention antidepressant trials in depressive disorders. Australian and New Zealand Journal of Psychiatry. 2010;
4. Sim K, Lau WK, Sim J, Sum MY, Baldessarini RJ. Prevention of relapse and recurrence in adults with major depressive disorder: Systematic review and meta-analyses of controlled trials. International Journal of Neuropsychopharmacology. 2016.
5. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. Vol. 461, Nature. 2009.
6. Wichers M, Smit AC, Snippe E. Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: A confirmatory single-subject time-series study. Journal for Person-Oriented Research. 2020;
7. Wichers M, Groot PC, Psychosystems ESM, Group EWS, others. Critical slowing down as a personalized early warning signal for depression. Psychother Psychosom. 2016;85(2):114–6.
8. van de Leemput IA, Wichers M, Cramer AOJ, Borsboom D, Tuerlinckx F, Kuppens P, et al. Critical slowing down as early warning for the onset and termination of depression. Proc Natl Acad Sci U S A. 2014;
9. Carney RM, Freedland KE, Veith RC. Depression, the autonomic nervous system, and coronary heart disease. Vol. 67, Psychosomatic Medicine. 2005.
10. Schiweck C, Piette D, Berckmans D, Claes S, Vrieze E. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. Vol. 49, Psychological Medicine. 2019.
11. Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM. Impact of Depression and Antidepressant Treatment on Heart Rate Variability: A Review and Meta-Analysis. Biological Psychiatry. 2010;
12. Glass L. Using mathematics to diagnose, cure, and predict cardiac arrhythmia. Chaos. 2020;30(11).
13. Nahshoni E, Aravot D, Aizenberg D, Sigler M, Zalsman G, Strasberg B, et al. Heart Rate Variability in Patients with Major Depression. Psychosomatics. 2004;
14. Yeragani VK, Rao KARK, Smitha MR, Pohl RB, Balon R, Srinivasan K. Diminished chaos of heart rate time series in patients with major depression. Biol Psychiatry. 2002;51(9):733–44.
15. Leistedt SJJ, Linkowski P, Lanquart JP, Mietus JE, Davis RB, Goldberger AL, et al. Decreased neuroautonomic complexity in men during an acute major depressive episode: Analysis of heart rate dynamics. Translational Psychiatry. 2011;
16. Shekatkar SM, Kotriwar Y, Harikrishnan KP, Ambika G. Detecting abnormality in heart dynamics from multifractal analysis of ECG signals. Scientific Reports. 2017;7(1).
17. Hognon L, Heraud N, Varray A, Torre K. Adaptive Capacities and Complexity of Heart Rate Variability in Patients With Chronic Obstructive Pulmonary Disease Throughout Pulmonary Rehabilitation. Frontiers in Physiology. 2021;12.
18. Javorka M, Trunkvalterova Z, Tonhajzerova I, Javorkova J, Javorka K, Baumert M. Short-term heart rate complexity is reduced in patients with type 1 diabetes mellitus. Clinical Neurophysiology. 2008;119(5).
19. Perkiömäki JS, Mäkikallio TH, Huikuri H v. Fractal and complexity measures of heart rate variability. In: Clinical and Experimental Hypertension. 2005.

20. Zbilut JP, Thomasson N, Webber CL. Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. Medical Engineering and Physics. 2002;24(1).

21. Greco A, Benvenuti SM, Gentili C, Palomba D, Scilingo EP, Valenza G. Assessment of linear and nonlinear/complex heartbeat dynamics in subclinical depression (dysphoria). Physiological Measurement. 2018;39(3).

22. Noordam R, van den Berg ME, Niemeijer MN, Aarts N, Hofman A, Tiemeier H, et al. Antidepressants and heart-rate variability in older adults: A population-based study. Psychological Medicine. 2016;46(6).

23. Licht CMM, de Geus EJC, Zitman FG, Hoogendijk WJG, van Dyck R, Penninx BWJH. Association between major depressive disorder and heart rate variability in the Netherlands study of depression and anxiety (NESDA). Archives of General Psychiatry. 2008;65(12).

24. Licht CMM, Penninx BW, de Geus EJC. To include or not to include? A response to the meta-analysis of heart rate variability and depression. Vol. 69, Biological Psychiatry. 2011.

25. Helmich MA, Wichers M, Peeters F, Snippe E. Daily dynamics of negative affect: indicators of rate of response to treatment and remission from depression? Psyarxiv. 2021;

26. Kunkels YK, Smit AC, Minaeva O, Snippe E, George S v., van Roon AM, et al. Risk Ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during antidepressant discontinuation. Submitted. 2022;

27. Smit AC, Snippe E, Kunkels Y, Riese H, Helmich M, Wichers M. Transitions In Depression (TRANS-ID) Tapering. OSF [Internet]. 2020; Available from: https://osf.io/h75p9/

28. Smit AC, Snippe E, Wichers M. Increasing Restlessness Signals Impending Increase in Depressive Symptoms More than 2 Months before It Happens in Individual Patients. Psychotherapy and Psychosomatics. 2019.

29. Smit AC, Helmich MA, Bringmann LF, Oldehinhel AJ, Wichers M, Snippe E. Critical slowing down in momentary affect as early warning signal of impending transitions in depression. Submitted. 2022;

30. Garner DM, de Souza NM, Vanderlei LCM. Heart Rate Variability Analysis: Higuchi and Katz's Fractal Dimensions in Subjects with Type 1 Diabetes Mellitus. Romanian Journal of Diabetes, Nutrition and Metabolic Diseases. 2018;25(3).

31. Ho YL, Lin C, Lin YH, Lo MT. The prognostic value of non-linear analysis of heart rate variability in patients with congestive heart failure-a pilot study of multiscale entropy. PLoS ONE. 2011;6(4).

32. Norris PR, Anderson SM, Jenkins JM, Williams AE, Morris JA. Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 trauma patients. Shock. 2008;30(1).

33. Johnston BW, Barrett-Jolley R, Krige A, Welters ID. Heart rate variability: Measurement and emerging use in critical care medicine. Vol. 21, Journal of the Intensive Care Society. 2020.

34. Higuchi T. Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena. 1988;31(2).

35. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, et al. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods. 2021;53(4).

36. Kojima H. hfda 0.1.1 [Internet]. 2019 [cited 2022 Mar 29]. Available from: https://pypi.org/project/hfda/

37. Smit AC, Snippe E, Hoenders HJR, Wichers M. Transitions In Depression: If, how, and when depressive symptoms increase during and after tapering of antidepressant medication. Submitted. 2022;

38. Nachar N. The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. Tutorials in Quantitative Methods for Psychology. 2008;4(1).

39. Pinna GD, Maestri R, Torunski A, Danilowicz-Szymanowicz L, Szwoch M, la Rovere MT, et al. Heart rate variability measures: A fresh look at reliability. Clinical Science. 2007;113(3–4).

40. Cohen J. Statistical Power Analysis for the Behavioural Science (2nd Edition). Vol. 3, Statistical Power Anaylsis for the Behavioral Sciences. 1988.

41. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020.

42. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020;17(3).

43. Harikrishnan KP, Misra R, Ambika G. Combined use of correlation dimension and entropy as discriminating measures for time series analysis. Communications in Nonlinear Science and Numerical Simulation [Internet]. 2009;14(9–10):3608–14. Available from: http://dx.doi.org/10.1016/j.cnsns.2009.01.021

44. Costa M, Goldberger AL, Peng CK. Multiscale Entropy Analysis of Complex Physiologic Time Series. Physical Review Letters. 2002;89(6).

45. Servaas MN, Schoevers RA, Bringmann LF, van Tol MJ, Riese H. Trapped: rigidity in psychiatric disorders. Vol. 8, The Lancet Psychiatry. 2021.

46. Agelink MW, Boz C, Ullrich H, Andrich J. Relationship between major depression and heart rate variability. Clinical consequences and implications for antidepressive treatment. Psychiatry Research. 2002;113(1–2).

47. O'Regan C, Kenny RA, Cronin H, Finucane C, Kearney PM. Antidepressants strongly influence the relationship between depression and heart rate variability: Findings from The Irish Longitudinal Study on Ageing (TILDA). Psychological Medicine. 2015;45(3).

48. Licht CMM, de Geus EJC, van Dyck R, Penninx BWJH. Longitudinal evidence for unfavorable effects of antidepressants on heart rate variability. Biological Psychiatry. 2010;68(9).

49. Ikeda T. Current use and future needs of noninvasive ambulatory electrocardiogram monitoring. Vol. 60, Internal Medicine. 2021.

50. Kunkels YK, van Roon AM, Wichers M, Riese H. Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring. Psychophysiology. 2021;58(10).

51. Bentley KH, Kleiman EM, Elliott G, Huffman JC, Nock MK. Real-time monitoring technology in single-case experimental design research: Opportunities and challenges. Behaviour Research and Therapy. 2019;117.

52. Moriarty AS, Meader N, Snell KIE, Riley RD, Paton LW, Dawson S, et al. Predicting relapse or recurrence of depression: systematic review of prognostic models. The British Journal of Psychiatry. 2022 Jan 11;1–11.

53. Judd LL, Schettler PJ, Rush AJ. A brief clinical tool to estimate individual patients' risk of depressive relapse following remission: Proof of concept. American Journal of Psychiatry. 2016;173(11).

54. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a Lancet–World Psychiatric Association Commission. The Lancet. 2022 Mar;399(10328):957–1022.

General summary, discussion and conclusion

8.

Chapter 8: General summary, discussion and conclusion

General Summary

In this thesis I aimed to investigate the transitions in (depressive) mood symptoms from a complex dynamical system perspective. For this, early-warning signals (EWS) were calculated from physiological and behavioural time series data. In **chapter 1** the background and rationale for the following research questions are given: (1) are increases in actigraphy and electrocardiogram (ECG) derived EWS predictive of transitions in the severity of (depressive) mood symptoms? (2) are changes in context-driven indices predictive of transitions in symptoms? (3) are changes in actigraphy-derived mean activity levels predictive of transitions symptoms? Before I could investigate these questions, data had to be collected and pre-processed from its' raw form into statistical analysable time series data formats. In **chapter 2**, I showed the developed software suite for analysing raw actigraphy data, the *ACTman R* package. The *ACTman* package enables researchers to automate pre-processing and analyses steps when working with actigraphy data. By doing so in this structured way, the risk for human error is reduced and laborious pre-processing steps have been automated to expedite actigraphy research.

In **chapter 3**, I studied in individuals diagnosed with bipolar disorder, the ability of actigraphy-derived generic EWS (variance and kurtosis), context-driven warning signals (autocorrelation at lag-720), and spectral periodicity indices (to check whether participants 24h circadian rhythm changed) to predict upcoming transitions in mood episodes. EWS and spectral indices were able to detect upcoming changes in mood episodes in some of the bipolar patients, although not in all. Future studies into the false-positive rates of this method are required to assess its sensitivity and specificity characteristics. The studied EWS and spectral indices would often perform more akin to general instability markers than as transition markers, by, for example, showing a high degree of variability (rapidly increasing or decreasing EWS values) instead of the expected consistent increase in EWS values. Perhaps this issue is more pertinent in our sample of bipolar patients, who showed rapid switches between depressed, manic, and euthymic episodes. This is in contrast to patients suffering from unipolar depression, who do not experience manic episodes. Yet, this study showed the feasibility of using actigraphy data to calculate useable EWS and spectral indices, and thus laid the important groundwork for subsequent actigraphy studies on the TRANS-ID data.

In **chapter 4**, I studied the TRANS-ID Tapering sample of patients with a history of suffering from unipolar depressive symptoms, who were tapering the dosage of their anti-depressant medication. Here, a repeated single subject design was used to test our hypotheses. We expected to find within individuals with a transition into more severe depressive symptoms; (1) increased critical slowing down based EWS (variance, kurtosis, and acf-1), (2) increased IS and acf-1440, and decreasing IV, and (3) decreased mean levels of physical activity. Actigraphy data that were continuously collected by the participants for four months were used in the analyses. I applied a similar EWS analysis pipeline as developed for the analyses described in chapter 3; that is using a moving window strategy in which EWS (variance and kurtosis) are calculated and in this way creating new time series. Additional to generic EWS (variance, kurtosis, and autocorrelation at lag-1), more context-driven variables (interdaily stability (IS), intradaily variability (IV), and autocorrelation at lag-1440), were developed which were expected to provide unique information given their direct relation to the circadian rhythm of the participants. For this, instead of investigating averages of actigraphy time series data of mainly one minute, as used in chapters 2 and 3, I looked into autocorrelations of actigraphy data over 24-hours (lag-1440), given its relation to the 24-hour circadian rhythm. I found that in seven out of eight patients a significant change was observed in at least one of the three studied EWS, up till four weeks before an episode onset. However, while the obtained results were not fully in line with what was hypothesised (e.g., the direction of acf-1440 was reversed from what was expected), they provided a starting point for post-hoc analyses. For this, false positive rates were calculated and receiver operating characteristic (ROC) curves were constructed to assess specificity and sensitivity. We concluded that using EWS combinations can outperformed using single EWS, although this has to be further confirmed in a larger dataset. The results of this study indicate that the assumption that EWS are *generic*, and can thus be found in differing data types, such as actigraphy, can hold true.

In **chapter 5**, I investigated the complexity of recurrent physical activity patterns in a nomothetic study including participants with and without a diagnosis of depression. While such recurrence analyses are theoretically distinct from the previously studied EWS, they showed similar promise in predicting transitions in mood symptoms. Studied complexity markers included determinism, laminarity, and the ratio between these two. Physical activities such as walking, biking, working at a desk job, that are repeated over time produce a specific recurrent activity pattern (Lu & Tong, 2019). Such recurrent physical activity

patterns are expected to help to differentiate between a group of depressed and non-depressed participants Compared to the controls, the group diagnosed with depression did not differ on mean activity levels. However, as expected, lower levels of complexity were detected in the actigraphy data from the depressed compared to non-depressed, both in terms of lower mean durations of periods of recurrent physical activity and less diversity in the duration of these periods.

In **chapter 6**, feasibility, validity, and reproducibility of Interbeat Intervals (IBIs) gathered with two ambulatory electrocardiogram (ECG) monitors (Cortrium C3, cortrium.com, Ithlete finger sensor, myithlete.com) were studied. Both monitors were tested against a wired ECG reference monitor. I found that the two wireless ECG monitors delivered data with somewhat lower accuracy that the wired reference method, especially when participants were moving. Yet performance of the two tested monitors were found to perform at sufficient levels of feasibility, validity, and reproducibility, wherein the Cortrium device, given its robust data signal, was found to be the most suitable monitor for the data collection during our TRANS-ID studies.

In **chapter 7**, we investigated in a within-person study design whether the mean, EWS (variance) and complexity measures (Higuchi dimension and multiscale entropy) decreased in the period before a transition in depressive symptoms using the IBI data gathered in the TRANS-ID Tapering study. This was investigated as prior research has suggested that individuals suffering from depression might show decreased levels of cardiac complexity (Leistedt et al., 2011). Thus, we expected the IBI time series derived complexity indices to decrease in the period before the transition. In this study we did not find any evidence to support the notion that warning signals based on IBI time series data occurred preceded transitions in depressive symptoms. Moreover, we found that antidepressant tapering participants experienced lower entropy in cardiac dynamics, and had a higher risk for experiencing an recurrence of depressive symptoms. These complexity indices thus show promise for estimating risk for recurrence of depressive symptoms.

Based on the studies presented in this thesis, I would argue that we can answer the question of whether we can use EWS calculated on actigraphy and IBI time series data to foresee transitions in depressive symptoms with a resounding ”Maybe”. While not as clear cut as a “no” or a “yes”, I think that our investigations have shown that the predictive performance of such EWS hold some promise and warrant future studies herein. Hence, I

conclude that the hardware and software to collect and analyse intensive physiological time series data are available and feasible for those goals. However, it would be too much to say that the EWS we investigated are currently fit for clinical application in detection of transitions in depressive symptoms. For that, the predictive capabilities are not yet convincing enough. Additionally, there are currently still unexplored possibilities to tailor combinations of device settings, recording lengths, and statistical analysis choices that could potentially improve the predictive performance of EWS. In the following I will motivate and reflect on these statements.

Assessment of actigraphy time series data

The studies presented in this thesis sometimes faced unforeseen challenges in collecting ambulatory time series data by the TRANS-ID participants in their normal daily life, or in (pre)processing and analysing them. While the *ACTman* software suite streamlines actigraphy data pre-processing, the actigraphy research field as a whole might benefit from a more systematic and standardized way of formatting (raw) actigraphy files. For example, the two actigraphs currently supported by the *ACTman* software do still require the native software for initial converting proprietary raw data formats into general open-source formats, such as .csv or .ods. With the *ACTman* software suite, I enabled the automatic pre-processing of the collected actigraphy data used in the current thesis. For generalisability and comparability to other actigraphs, readily analysable data in a standardized format is required. Other researchers have also identified this issue and have pointed out the importance of using alternative and generalisable actigraphy algorithms, such as using the *Euclidean Norm Minus One* (ENMO) as a general algorithm to pre-process the data and separate the movement and gravitational forces from the data (Bakrania et al., 2016). However, as used actigraphy algorithms are often proprietary and widely differing components between manufacturers are used (using different filters, amplifiers, or frequencies), it seems worthwhile considering introducing an industry norm, such as the DIN-norms (*Deutsches Institut für Normung;* German Institute for Standardisation Registered Association) currently used for things as electrical outlets, or connectors. By doing so standardisation of actigraphy monitors can be applied more consistently. To give a practical suggestion, such standardisations can take the form of advising all actigraph manufacturers to always include a standardised time unit, say one second, and to assess in standardised units such as *g* instead of proprietary "step" or "count" units. While actigraphy *step* or *count* units seem an intuitive unit to use, this advantage is hindered by manufacturers use of self-designed – but highly similar – units, such

as the *MotionWatch count* instead of a normal *count*. As these counts are calculated differently it becomes very difficult to compare data derived from actigraphy monitors that use differing units.

In general, device specifications are often idealised by the manufacturer. Take, for example, the maximum boot space of a car. Car manufacturers like to advertise their large boot spaces to outdo their competitors. However, they all provide the boot space numbers in litres, which makes the boot seem large; but how often do you go about filling your boot with an actual liquid? The actual usable boot space will often be much lower as baggage in solid form will often leave space unused. When considering actigraphy hardware issues, these can be observed with physiological data monitors such as the *CamNTech* Motionwatch 8 used in our TRANS-ID study. Here product specifications mentioned possible recording length of over four months. When using the sample frequency needed to obtain data useful for scientific research only up to two months monitoring was feasible. The battery and data storage challenges with the monitor observed during the pilots needed protocolized attention in the main study. However, even with protocolled replacement of actigraphs for each participant, issues related to battery changes were a major contributor to missing data. Potential solutions may be using monitors that are chargeable by participants themselves, or come as a pair with a dedicated docking station wherein one device can be charged while the other is worn.

Assessment of IBI time series data

Whereas there were plenty of options to select a well-established and validated actigraphy device which were in line with our study requirements, the pool of suitable wireless ECG monitors was considerably smaller. Wireless monitoring was essential for our ambulatory TRANSID study as long wires attached to the electrodes can be pulled and detached from the monitor by accident, thus causing missing data, while they can also hinder participant movement, or provoke device mishandling errors as participants can accidentally attach the wrong wires to the wrong electrode (Shin et al., 2005; Winokur et al., 2013). Given the small number of suitable monitors, we deemed a dedicated validation study of ECG monitors necessary. In a validation study in chapter 6, the feasibility, validity, and reproducibility characteristics of two ECG monitors (Cortrium C3, cortrium.com, Ithlete finger sensor, myithlete.com) were investigated in both standardised laboratory- and ambulatory settings against a wired standard ambulatory device. Whereas the wireless ECG monitors did show somewhat lesser performance when compared to the wired reference

method – especially during activities wherein there was an increased risk of motion artefacts (e.g., walking) - they were found to perform at sufficient levels of feasibility, validity, and reproducibility. We found that the Cortrium outperformed the Ithlete on providing data with less noise, and being more robust against movement artefacts. Therefore the Cortrium was selected for monitoring in the main study. Participants in the TRANS-ID study collected their ECG data for four months, twice a day (morning, evening) in sitting position. ECG data were pre-processed with inhouse developed CARSPAN software (Mulder et al., 1995) to obtain IBI time series for the planned statistical analyses.

When considering ambulatory ECG hardware issues, we observed that a relatively large portion of data were lost due to monitor malfunctions or underperformance. For example, the Cortrium C3 device had three robust small metal and rubber-coated legs to which the ECG spot-electrodes were attached. However, we observed a multitude of devices being returned in a damaged state. The type of damage (the electrode legs connecting the device body to the ECG electrodes being broken off) did suspect that either too much force was applied when detaching the device, or that the electrode legs were not designed for the period of time we employed them. First indication of breakage was typically when the data were checked for distortions. These experiences were shared with the Cortrium company to that they were able to improve these points and incorporate these improved components in the next device version.

Actigraphy and IBI time series derived predictions of transitions in depression

When comparing our actigraphy derived EWS results with the limited existing literature, I concluded the following. First, application of EWS in highly complex real-life systems in psychiatry is still rather unique. While systems such as climate were investigated with EWS (e.g., Scheffer et al., 2009), applying these techniques to human physical activity time series data in order to study transitions in mood symptoms is truly novel. Second, some of the actigraphy data derived generic EWS were moderately successful in predicting transitions in depressive symptoms with comparatively low false-positive rates. However, in clinical practice finding a balance between false-positive and false-negative rates depends on many factors (i.e., how bad is it to miss a true alarm vs. how bad is it to have a false alarm, participant burden), and should be determined in cooperation with patients and their clinicians. Third, while reflecting on the theoretical starting point of the current thesis, namely the complex dynamic system theory from which we hypothesised that any time series

data set could pick up EWS for upcoming transitions, the findings of this thesis do not support the theory. In this thesis other time series derived measures to predicted upcoming transitions were explored also as well. Context-driven signals, such as acf-1440, could potentially also be useful in predicting transitions, as we found that when we released the direction criterion (i.e. expecting only increases) its predictive performance increased. This can be interpreted as suggestive evidence that the expected directions for context-driven EWS we had before commencing the study were inaccurate. Fourth, regarding our study investigating complexity measures in ambulatory ECG data, we were not able to find within-person evidence for the investigated EWS or complexity measures. However, we did find substantial differences between baseline levels of complexity within individuals who had experienced a transition and those who did not experience such a transition in depressive symptoms. Whether the lack of trends was due to anti-depressant medication use, or because such trends normally start well before the studied time period, remains yet unclear. Given the novelty of this study there is a need for replication studies hereinto. Moreover, given the limited sample size the current results are not expected to generalise well to other or larger samples. Hence, replication studies in more individuals are required to test this more thoroughly.

Nomothetic and idiographic research designs

Regarding future study designs, a next step could be to design studies balancing a nomothetic (group-based) and idiographic (single-subject) research approach (Zuidersma et al., 2020). Researchers and clinicians are currently typically trained from a nomothetic point of view and findings on the "average" patient which end up in clinical treatment guidelines. There is an increased need for an idiographic or single-subject approach as this approach is more in line with clinical practice were the clinician will aim for a personalized treatment for individual patients (Herrman et al., 2022). Indeed, in the studies presented in this thesis a balance between nomothetic and idiographic approaches was often employed. Consider, for instance, chapter 4 wherein we were interested in whether we could find EWS to support predicting transitions in individual patients, but were also interested in whether positive individual findings could also be found at the larger group level. Chapter 8 focused more on the nomothetic outcomes of the study, but would still carry a substantial idiographic component as data gathered and analysed at the individual level could potentially still help patients predict transitions in mood symptoms. Consistent with other TRANS-ID studies who investigated foreseeing transitions in depressive symptoms using Experience Sampling

Method (ESM, or electronic diary) derived EWS (Helmich, 2022; Smit, 2022), we found that EWS calculated from actigraphy or IBI time series data were not able to consistently foresee transitions in depressive symptoms. Moreover, in line with others (Bos et al., 2022) I also conclude from the findings of my thesis, that especially in study-designs centred around many replications of single-subject studies wherein we are essentially more interested in an individual's EWS performance instead of group EWS performance, individual differences are to be expected. Notably, there may also within a single-subject heterogeneity. That is, while at one point an EWS can have predictive value for a patient; this does not mean that this will remain true for any amount of time. Additionally, when considering time and transitions it can be elucidating not only to investigate for what period of time EWS can have predictive value, but also at what time scale we expect to find transitions; a point we will discuss in more detail in the next paragraph.

Time scale of transitions

In chapter 4, I also investigated in more detail at which time scale transitions in depression occurred. While autocorrelation at lag-1 is often used in EWS studies (Maturana et al., 2020; Wichers & Groot, 2016), it is not a standardised method and it should be noted that autocorrelation at lag-1's outcomes depends on the resolution at which the assessment device/software outputs the data. Lag-1 can mean that autocorrelations were calculated from data assessed at the minute level, as is the case of the actigraphy data studied in chapter 4, while lag-1 can also refer to data assessed at the day, hour or millisecond level. The latter is the case for ECG assessment, while ESM assessment intervals are typically hours apart. As already suggested above, it might be worthwhile to investigate if and how actigraphy assessments can be standardised if we would like to improve the interpretability and comparability of the yielded data in mood disorder research. There is no such thing as a gold standard amongst autocorrelation lags, and it might be worthwhile to explore alternatives which are ought to have a conceptual link with the topic in question. For example, our choice for studying autocorrelation at lag-1440 was based on our data binned at the minute level, such that lag-1440 represents a 24 hour or day cycle (as there are 1440 minutes in a day). Such chosen lag-sizes differed in some chapters (for example lag-720 and lag-1440) according to the research topic. For instance, in chapter 4, I chose lag-1440 (i.e., 24 hours) as it approximately corresponds to the circadian cycle. However, by doing so I have created an autocorrelation value that is different from the commonly used autocorrelation at lag-1, and thus does not strictly adhere to the dynamical system theory underlying lag-1

autocorrelations. Studying this circadian variant of autocorrelation offers possibilities for investigations into time scales we expect to find theoretically relevant changes. Consider, for example, the plots in figure 1. Here autocorrelation values are shown for lag-1 till lag-1500 based on actigraphy time series data of two individuals. Note the fluctuations between positive and negative autocorrelation values over the day; the autocorrelation for acf-720 (yellow line) is negative most of the times, in contrast to the autocorrelation for acf-1440 (red line) which is mostly positive. Both plots follow the same pattern one would expect to see from an autocorrelation function. That is, at lower lags autocorrelation is very high, as a measurement at this moment should be highly similar to a measurement a mere minute ago. However, the autocorrelation values keep decreasing, reaching negative values around acf-720. Thereafter the autocorrelation values rise again. This could intuitively make sense, as if you are awake now, chances are high you were asleep 12 hours ago. On the other hand, if you are awake now, chances are high you were awake 24 hours ago. By more systematically investigating the effects of using autocorrelations with different lags in actigraphy based time series data, it might be possible to study at which time scales transitions in depressive symptoms co-occur with changes in predictors derived from physical activity. Large longitudinal datasets, such as the TRANS-ID study which includes 4 months of data per participant, are a promising testing ground to investigate autocorrelation at a multitude of lags; from lag-1 (one minute) to, for example, lag-10080 (1 week) and every lag in between. While we do not yet know at what time scale to expect transitions in mood disorders, systematically exploring various lags in actigraphy data might help elucidate on what time scales changes may take place. Such findings could then form the basis for theories on the time scale of the course of mood disorders which in turn might inform future investigations into this topic. In the next paragraphs, I will discuss the clinical relevance of our investigations and I will comment on the development of a clinical tool based on the methods and techniques presented in this thesis.

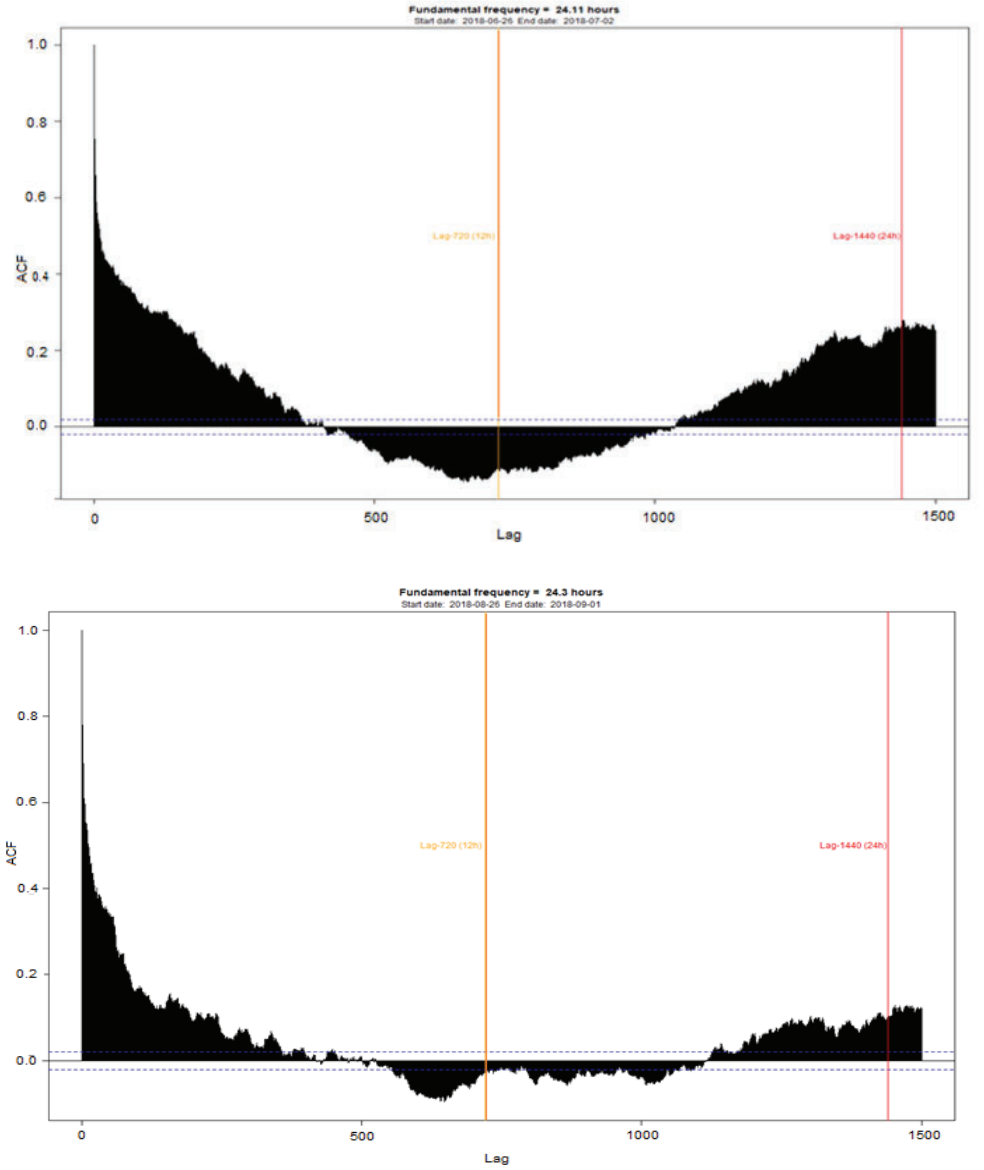


Fig. 1: Examples of a full auto-correlation function (ACF) from lag-1 till lag-1500 on actigraphy time series data of two individuals. The yellow and red lines show autocorrelation lags at respectively lag-720 (12h) and lag-1440 (24h). The x-axis shows the number of lags, while the y-axis shows the autocorrelation values (varying between 1: perfect

positive correlation, and -1: perfect negative correlation). Fundamental frequency given in the plot title refers to the length of that an individual's circadian rhythm within the measured time period.

Clinical relevance & clinical tool development

From my thesis results, I conclude that using EWS to predict upcoming transitions in mood for both unipolar and bipolar depression, be it via actigraphy or inter-beat intervals data, is currently not sufficiently effective for clinical implementation. Expectations based on complex dynamical system theory were not clearly supported and therefore plans for clinical implementation should be tampered. In this paragraph I want to further elaborate the clinical relevance of my studies and outline a potential road for clinical tool development.

In this thesis, studying EWS in patients suffering from mood disorders was performed with multiple goals in mind. One important goal is to facilitate the development of a helpful clinical tool to predict clinically significant changes in mood. The promise of such a clinical tool is relatively simple; it would allow clinicians and patients to receive early warnings of upcoming transitions in real-time, hopefully, these warnings would come early enough to be able to timely intervene and prevent negative transitions in mood. While sounding ambitious, we do have to remember that highly similar tools are already available for other types of time series data, such as financial market data. For example, simple and free-to-use websites such as www.tradingview.com allow users to study and analyse financial market data in real-time, using a plethora of validated and user-created indices. Websites such as www.tradingview.com show that the technical requirements for online real-time data analysis are already available (see figure 2).

Fig. 2: Graphical user interface showing real-time financial market monitoring and on-the-fly analyses on https://www.tradingview.com/chart/?symbol=spx. The upper panel shows price fluctuations of Bitcoin versus the Euro between 2018 and 2021. The middle and lower panels show real-time calculated indicators, respectively Note: MACD, Moving Average Convergence/Divergence; RSI, Relative Strength Index.

However, before this could become a reality for clinical applications, quite some future research is required. In my opinion, the most pertinent issues hampering the development of such a tool for (psycho)-physiological time series data are: (1) the need to further investigate EWS and related markers to use within such a clinical tool, in differing patients groups and settings; (2) online data and privacy protection conform current privacy regulations; (3) lack of available monitors capable of both real-time data collection, storage, and analysis, and; (4) evaluations of the specific needs of such a tool expressed by patients and their clinicians. I will elaborate on these four issues next.

First, while studies presented in this thesis provide suggestive evidence for a number of generic- and context-driven EWS which could be promising to include in such a tool, future studies are needed to see whether these EWS are also effective in (slightly) differing patient groups. For example, while we tested EWS in samples of both bipolar patients and individuals with a history of unipolar depression who were tapering their medication, other samples should also be investigated. By investigating EWS in more and differing samples we

could get a better idea of how EWS perform, and under which circumstances their predictive performance is best. Patients can experience a wide range of treatments, medications, and life-events, and we currently do not know well enough how these impact the predictive performance of EWS. In the current thesis, we did not investigate EWS in individuals who´s mood improved as a result of treatment for depression. Collecting (psycho-)physiological time series data before, during, and after treatment can be used to study these kind of transitions wherein the change is a positive one. Or to be more specific, from a depressed state to a non-depressed state instead of vice versa, as was done in the TRANS-ID Recovery study (Helmich et al., 2020). In the TRANS-ID Recovery study also actigraphy and IBI time series data were gathered but not analysed yet. Second, as personal data is being collected and streamed into online repositories, considerable attention is required to safeguard the associated servers against malicious attacks, such as storing privacy-sensitive data on certified and secured servers (Sytema & van der Krieke, 2013). Third, I would encourage deepening cooperation between university medical centres and medical device manufacturers. While both can have different goals (e.g., profit vs. non-profit), aligning both parties is needed to develop a device which would be advanced enough to allow for real-time data collection and analysis, while still being robust and reliable enough for long-term ambulatory assessments. Fourth and lastly, both clinicians and patients will have to be heard about their preferences for a clinical tool. For instance, the monitoring schedule should not hinder participants or be too much of a burden. Pilot studies and patient and clinician focus groups can help establish acceptable monitoring schedules and clinical needs for early warning signalling derived from gathered data.

In summary, while working on the ambition to develop a clinical tool for foreseeing transitions in depressive symptoms to support patients and clinicians, there is more research required to optimise (real-time) data collection and analysis, and to investigate and select the best performing EWS parameters. Additionally, having systematic evaluations with patients and clinicians on what would be acceptable false- and true alarm rates, is essential for all clinically relevant future EWS research in psychiatric settings. Notably, also the implementation of novel, and openly accessible information technologies may help accelerate such research, although this would require incorporating such *open science* practices even more in academic practice; a point I will discuss in the following paragraph.

Scientific relevance: Open Science

With the increased application of information technologies in the academic world over the last decades, technological thresholds for improved openness and transparency in academia have been taken away, for example by the introduction of globally accessible online repositories for research materials, documentation, and – where possible given privacy regulations – research data (van der Zee & Reich, 2018). The endeavours to improve the way we conduct good science are known as *Open Science* practices. Efforts to improve Open Science are often not part yet of researchers regular training, education, and daily work and might thus be better characterized as *academic citizenship* activities (Macfarlane, 2007). Starting with our work in the TRANS-ID project, we strived to archive and register the materials we could share, such as study protocols, materials, and other information on online repositories. Moreover, writing a PhD thesis in the Open Science emerging era influenced me as a researcher. Open Science can be defined as a movement striving to increase reproducibility by improving availability of research materials, and to increase transparency in academic work and the presentation thereof (McKiernan et al., 2016). I applaud and actively participated in the recent debate in the scientific community on how to effectively cooperate and disseminate knowledge in the light of substantial advancements in information technologies. Investigations revealing sub-optimal reproducibility rates of the scientific literature in psychology (Aarts et al., 2015) and cancer biology (Nosek & Errington, 2017) have added gravity to this matter and added to the motivation of the scientific community to critically reflect on itself and actively look for improvements. Developments such as increased awareness of open access possibilities and increased availability of cloud-based repositories for storing and sharing scientific data storage, have changed scientific workflow and practices. Given their importance, I have helped develop improved crowdsourcing strategies which allow for easier and more efficient research collaborations (Aczel et al., 2021). Moreover, I have participated in novel crowdsourcing strategies to investigate how many analyses results are conditional on choices made by the analysts (Bastiaansen et al., 2020). Also, I actively participate in an expert group within the Belgian-Dutch ESM Network for ESM Research in Mental Health (https://esm-network.eu/) that resulted in the *ESM Item Repository* (see www.esmitemrepository.com). The Network initiated a hackathon for our expert team to cooperate on this topic and we developed an online platform containing hundreds of publicly available ESM items. The code for the software tool ACTman (chapter 2; Kunkels et al 202) is publicly available in the Github code repositories (see for example:

https://github.com/compsy/ACTman). Another example which is increasingly ingrained in the scientific process is pre-registration of hypotheses and methods before conducting a study (Nosek et al., 2018). Among others, it is assumed that this procedure will reduce biases and allows for a clear distinction between exploratory and confirmatory research. During our TRANS-ID project, we pre-registered most of our studies, which allowed for a clear demarcation between confirmatory and exploratory research, and prevented hypothesising after results were known. Moreover, pre-registration can help make (pre-)processing steps of the raw data and research decisions clear at an early stage. However, for highly innovative studies, a perceived lack of flexibility when conducting exploratory investigations can feel restricting. Also, while the availability of pre-registration templates is supportive and reduces the time required to pre-register, it is still a time-consuming process.

Conclusions

The studies presented in this thesis support the notion that both actigraphy and cardiac assessments offer rich, high-resolution time series data that allows for advanced (predictive) analyses, and poses little to no burden to participants. Studied sensor derived EWS showed some promise in samples of bipolar and unipolar patients in prediction of an upcoming unfavourable mood transition before patients experience increasing severity of their symptoms. I reported that a number of generic actigraphy derived EWS were effective in predicting upcoming transitions in mood symptoms in some of the participants who tapered their antidepressants. Regarding the cardiac data of these participants, we found that complexity quantifiers may be supportive to identify individuals who are at risk for experiencing an increase in depressive symptoms. However, it is not clear if these results would generalise to other samples and even within the individuals currently studied. Moreover, it is also not yet clear what would be acceptable false alarm rates for patients and their therapists. As such more future research hereinto is required before we can start piloting with implementation of the studied markers in mental health care. Interestingly, preliminary evidence was found that aggregating actigraphy derived EWS and related indicators could improve prediction performance. When real-time analysis of collected data are established, the pioneering studies in this thesis may help facilitate the development of future clinical tools for mental health care.

References

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., … Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Aczel, B., Szaszi, B., Nilsonne, G., van den Akker, O. R., Albers, C. J., van Assen M. A., Bastiaansen, J. A., Benjamin, D., Boehm, U., Botvinik-Nezer, R., Bringmann, L. F., Busch, N. A., Caruyer, E., Cataldo, A. M., Cowan, N., Delios, A., van Dongen, N. N., Donkin, C., van Doorn, J. B., Dreber, A., Dutilh, G., Egan, G. F., Gernsbacher, M. A., Hoekstra, R., Hoffmann, S., Holzmeister, F., Huber, J., Johannesson, M., Jonas, K. J., Kindel, A. T., Kirchler, M., Kunkels, Y. K., Lindsay, D. S., Mangin, J. F., Matzke, D., Munafò, M. R., Newell, B. R., Nosek, B. A., Poldrack, R. A., van Ravenzwaaij, D., Rieskamp, J., Salganik, M. J., Sarafoglou, A., Schonberg, T., Schweinsberg, M., Shanks, D., Silberzahn, R., Simons, D. J., Spellman, B. A., St-Jean, S., Starns, J. J., Uhlmann, E. L., Wicherts, J., Wagenmakers, E. J,. (2021). Consensus-based guidance for conducting and reporting multi-analyst studies. *Elife*. *Nov 9;10*:e72185. Doi: 10.7554/eLife.72185. PMID: 34751133; PMCID: PMC8626083.

Bakrania K., Yates T., Rowlands A.V., Esliger D.W., Bunnewell S., Sanders J., et al. (2016) Intensity thresholds on raw Acceleration Data: Euclidean Norm Minus One (ENMO) and Mean Amplitude Deviation (MAD) Approaches. *PLoS ONE 11*(10): e0164045. https://doi.org/10.1371/journal.pone.0164045

Bastiaansen, J., Kunkels, Y., Blaauw, F., Boker, S., Ceulemans, E., & Chen, M. et al. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal Of Psychosomatic Research, 137*, 110211. https://doi.org/10.1016/j.jpsychores.2020.110211

Bos, F., Schreuder, M., George, S., Doornbos, B., Bruggeman, R., & van der Krieke, L. et al. (2022). Anticipating manic and depressive transitions in patients with bipolar disorder using early warning signals. *International Journal Of Bipolar Disorders, 10*(1). doi: 10.1186/s40345-022-00258-4

cortrium.com. (2022). Retrieved from https://www.cortrium.com/ on June 16th 2022.

Helmich, M. A. (2022). What's in a mood? looking for dynamic predictors of individual improvement in depression [Doctoral thesis, University of Groningen]. University of Groningen Research Portal. 10.33612/diss.200101721

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

Herrman, H., Patel, V., Kieling, C., Berk, M., Buchweitz, C., & Cuijpers, P. et al. (2022). Time for united action on depression: a Lancet–World Psychiatric Association Commission. *The Lancet, 399*(10328), 957-1022. doi: 10.1016/s0140- 6736(21)02141-3

Lu, J., & Tong, K. (2019). Robust single accelerometer-Based Activity Recognition Using Modified Recurrence Plot. *IEEE Sensors Journal, 19*(15), 6317-6324. https://doi.org/10.1109/jsen.2019.2911204

Leistedt, S. J. J., Linkowski, P., Lanquart, J. P., Mietus, J. E., Davis, R. B., Goldberger, A. L., & Costa, M. D. (2011). Decreased neuroautonomic complexity in men during an acute major depressive episode: Analysis of heart rate dynamics. *Translational Psychiatry*. https://doi.org/10.1038/tp.2011.23

Maturana, M., Meisel, C., Dell, K., Karoly, P., D'Souza, W., & Grayden, D. et al. (2020). Critical slowing down as a biomarker for seizure susceptibility. *Nature Communications, 11*(1). doi: 10.1038/s41467-020-15908-3

Macfarlane, B. (2007). Defining and rewarding academic citizenship: The implications for university promotions policy. *Journal of Higher Education Policy and Management*, *29*(3), 261–273. https://doi.org/10.1080/13600800701457863

McKiernan, E., Bourne, P., Brown, C., Buck, S., Kenall, A., & Lin, J. et al. (2016). How open science helps researchers succeed. *Elife*, *5*. https://doi.org/10.7554/elife.16800

myithlete.com. (2022). Retrieved from https://www. myithlete.com / on June 16th 2022.

Mulder, L.J.M., van Roon, A.M., & Schweizer, D. (1995). CARSPAN. Groningen: Iec ProGAMMA.

Nosek, B. A., & Errington, T. M. (2017). Making sense of replications. *ELife*, *6*. https://doi.org/10.7554/eLife.23383

Nosek, B., Ebersole, C., DeHaven, A., & Mellor, D. (2018). The preregistration revolution. *Proceedings Of The National Academy Of Sciences, 115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Scheffer, M., Bascompte, J., Brock, W., Brovkin, V., Carpenter, S., & Dakos, V. et al. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53-59. doi: 10.1038/nature08227

Shin, K., Hwang, H. T., Kim, Y. H., Kim, J. P., Yeo, H. S., Han, W., … Park, J. C. (2005). WHAM: A novel, wearable heart activity monitor based on Laplacian potential mapping. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, *7*, 7361–7364. https://doi.org/10.1109/iembs.2005.1616212

Smit, A. C. (2022). The prologue to depression: a tale about complex dynamics and simple trends [Doctoral thesis, University of Groningen]. University of Groningen Research Portal. 10.33612/diss.216681879

Smit, A., 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. doi: 10.1159/000500594

Sytema S, Van der Krieke L. Routine outcome monitoring: a tool to improve the quality of mental health care? In: Thornicroft G, Ruggeri M, Goldberg D, editors. Improving mental health care: the global challenge. 1st ed. Chichester: Wiley; 2013. p. 246–63.

Wichers, M., & Groot, P. (2016). Critical Slowing Down as a Personalized Early Warning Signal for Depression. *Psychotherapy And Psychosomatics, 85*(2), 114-116. doi: 10.1159/000441458

Winokur, E. S., Delano, M. K., & Sodini, C. G. (2013). A wearable cardiac monitor for long-term data acquisition and analysis. *IEEE Transactions on Biomedical Engineering*, *60*(1), 189–192. https://doi.org/10.1109/TBME.2012.2217958

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open, 4*(3), 233285841878746. doi: 10.1177/2332858418787466

Zuidersma, M., Riese, H., Snippe, E., Booij, S. H., Wichers, M., & Bos, E. H. (2020). Single-Subject Research in Psychiatry: Facts and Fictions. *In Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2020.539777

9.

Bijlagen

Chapter 9: Bijlagen

Nederlandstalige samenvatting

In de studies die ik hier in mijn proefschrift beschrijf was ons doel om te onderzoeken of vroege waarschuwingssignalen (*Early Warning Signals* ofwel *EWS*) berekend uit actigrafie en ECG data kunnen helpen om veranderingen in symptomen van depressie te kunnen zien aankomen. Hierbij is actigrafie een manier om beweging en waak/slaapcycli te meten met kleine bewegingsmeters die vaak aan de pols gedragen worden, zoals een *Fitbit*. Met een Elektrocardiogram (ECG, ook wel *hartfilmpje* genoemd) kan de elektrische activiteit van het hart gemeten worden. Ook hebben we onderzocht of de beschikbare apparatuur en software goed genoeg werkten voor deze taak.

Het eerste deel van mijn proefschrift (hoofdstukken 2 tot 5) is vooral gericht op het onderzoeken of EWS berekend vanuit actigrafie, en gerelateerde complexiteitsmaten, voorspellend kunnen zijn voor aankomende veranderingen in stemmingssymptomen. Daarnaast word in dit deel recent ontwikkelde software gepresenteerd die helpt om meerdere voorverwerkingsstappen van actigrafie data te automatiseren. In het tweede deel van mijn proefschrift (hoofdstukken 6 en 7) is onderzocht of vroege waarschuwingssignalen, berekend uit *interbeat intervals* (de tijd tussen individuele hartslagen, welke zichtbaar zijn in een ECG), voorspellend kunnen zijn voor aankomende veranderingen in (depressieve) stemmingssymptomen. Ik zal hieronder in meer detail de hoofdstukken van mijn proefschrift samenvatten.

In hoofdstuk 2 word de *ACTman* (*Actigraphy manager*) software geïntroduceerd. Deze software hebben wij ontwikkeld tijdens mijn promotieonderzoek om automatisch grote hoeveelheden actigrafie data te kunnen voorbewerken en te kunnen analyseren. Dit werd namelijk eerst vaak met de hand gedaan, wat veel tijd kost en kan leiden tot menselijke fouten. Deze software hebben we gebruikt om de actigrafie data die we tijdens mijn promotieonderzoek hebben verzameld te verwerken en te analyseren. De ACTman software kan gebruikt worden om meerdere relevante actigrafie- en EWS variabelen te berekenen, ook binnen een aanpasbare en voortschrijdende periode (de *moving window* methode).

In hoofdstuk 3 beschrijven we onze studie waarbij we herhalend, bij individuele patiënten met een bipolaire stoornis, onderzoeken of EWS en maten van spectrale periodiciteit kunnen helpen bij het aan zien komen van veranderingen in stemming. Een

bipolaire stoornis is een mentale aandoening waarbij patiënten periodes van depressieve gevoelens ervaren, welke verweven zijn met periodes van abnormaal verhoogde stemming en energie. Deze groep patiënten heeft unieke eigenschappen, zoals een verondersteld hogere kans om veranderingen instemming te kunnen observeren, vergeleken met patiënten met unipolaire depressie. De deelnemers in deze studie hebben hun fysieke activiteit gemeten voor 180 dagen, wat ons in staat stelde om te onderzoeken of we aankomende veranderingen in stemming konden identificeren op basis van actigrafie data.

In hoofdstuk 4 word het onderzoek beschreven waarin we onderzochten of EWS kunnen helpen bij het aan zien komen van veranderingen in depressieve symptomen, maar nu in een groep deelnemers die hun dosis antidepressieve medicatie aan het afbouwen zijn. Deze groep deelnemers is geworven en gemeten tijdens het TRANS-ID (*transities in depressie*) onderzoek. Deze TRANS-ID data zijn specifiek verzameld om te kunnen testen voor veranderingen binnen individuen, in tegenstelling tot het merendeel van wetenschappelijke studies waarin vooral word gekeken naar veranderingen binnen groepen. Binnen deze groep proefpersonen waren mensen die een verandering in stemmingssymptomen ervaarden tijdens het afbouwen, maar ook mensen die dit niet ervaarden. Daardoor konden we kijken of EWS anders presteerden in mensen die bijvoorbeeld wel een verandering ervaarden. Daarnaast hebben we gekeken of sensitiviteits- en specificiteitskenmerken uitmaakten voor het presteren van de onderzochte EWS.

In hoofdstuk 5 word een studie gepresenteerd waarbij we data onderzoeken die verzameld is door deelnemers van de *Mood and Movement in Daily Life (MOOVD)* studie. Deze deelnemers hebben dertig dagen onder andere hun ECG gemeten. Deze data hebben we onderzocht op groepsverschillen tussen depressieve en niet-depressieve deelnemers. Daarbij keken we naar variabelen zoals de gemiddelde hoeveelheid activiteit, circadiane ritmiek, en complexiteitsmaten. Nieuwe complexiteitsmaten gebaseerd op zogenaamde *herhalingsgrafieken* (recurrence plots) worden binnen dit hoofdstuk gepresenteerd en hun effectiviteit in het detecteren van stemmingsveranderingen onderzocht. Gegeven de niet-rechtlijnigheid van IBI-data kunnen dergelijke *herhalingsgrafieken* helpen om belangrijke patronen in de data te ontrafelen, welke met andere technieken minder goed zichtbaar zijn.

In hoofdstuk 6 hebben we twee nieuwe, draadloze, en draagbare ECG monitors (de *Cortrium C3* en de *Ithlete Finger Sensor*) onderzocht om hun haalbaarheids-, validiteits-, en reproduceerbaarheidskenmerken weer te kunnen geven. Deze twee monitors werden getest

tegenover een ECG referentie monitor (het *VU-AMS*) om te zien of de door hun verzamelde data valide was onder verschillende geprotocolleerde condities. Daarnaast werd onderzocht of deelnemers zelf in staat waren thuis hun ECG te meten met deze twee ECG monitors.

In hoofdstuk 7 werd onderzocht of complexiteit- en variabiliteitsmaten voor hart dynamiek afnamen in de periode voor een verandering in depressieve symptomen in een groep TRANS-ID deelnemers die hun antidepressieve medicatie afbouwden. Een afname hierin werd verwacht omdat we weten dat deze maten vaak substantieel lager zijn in individuen die lijden aan depressie.

Samengevat vonden we dat de onderzochte apparatuur en software geschikt was voor het meten en analyseren van ECG en actigrafie data. Echter, na het analyseren van de verzamelde data konden we nog niet zeggen dat vroege waarschuwingssignalen berekend uit deze data geschikt zijn om in de klinische praktijk te gebruiken om veranderingen in depressieve symptomen bij patiënten te voorspellen. Daarvoor moet eerst nog meer onderzoek gedaan worden naar hoe ECG en actigrafie patronen eruit zien bij zowel gezonde mensen als mensen met een depressieve stoornis.

Mijn promotietraject was voor mij een voorproefje hoe het is om in de wetenschap te werken. En hoe het is om samen te werken met alle bijzondere, intelligente, en sympathieke mensen die mijn werkveld rijk is. Deze ervaring was niet mogelijk geweest zonder enkele speciale mensen die ik hiervoor graag wil bedanken.

Marieke, het was fantastisch om te zien hoe enthousiast je je inzette voor het wetenschappelijke werk en het TRANS-ID project in het bijzonder. Je hebt al je PhD's met volle overgave begeleid om hun werk naar een hoger niveau te tillen. Bedankt dat ik mocht meewerken aan zo'n bijzonder project!

Harriëtte, wat heb ik ontzettend veel van jou mogen leren, zowel op gebied van de fysiologie, waar jij duidelijk expert op bent, alsmede op sociaal en communicatief niveau. Dat ik jou niet menigmaal tot waanzin heb gedreven bij het afronden van mijn proefschrift, waarin ik zeker niet de makkelijkste of bereikbaarste was, getuigt van jouw kracht als promotor. Je was er altijd, zelfs als ik het voor mijn gevoel even had laten zitten of ergens in gefaald had. En dat was juist wat ik nodig had, iemand die er ongeacht deze problemen toch was. Daarnaast was ik de eerste PhD-student die je op in deze rol hebt mogen begeleiden, ik weet zeker dat de volgenden alleen maar makkelijker zullen zijn. Dank je wel voor alles!

Arie, wat fijn dat je mee hebt willen helpen bij het succesvol afronden van mijn promotietraject. Ook jij bent duidelijk expert op gebied van de fysiologie. Daarnaast heb je een sterk analytisch inzicht en aanzienlijke vaardigheden in het programmeren en ontwikkelen van software. Met jouw praktische en gegronde adviezen hebben we veel hindernissen in de dataverwerking kunnen overwinnen. Daarnaast was jouw feedback, zowel op de geschreven papers, alsmede op de uitvoering van het project, altijd erg down-to-earth, kort, maar vooral ook krachtig.

Prof. dr. D. Borsboom, **Prof. dr. I. Germeys** en **Prof. dr. R. C. Oude Voshaar**, bedankt voor het lezen en beoordelen van dit proefschrift.

Ook wil ik graag alle **deelnemers aan het TRANS-ID onderzoek** bedanken. Er is veel tijd en energie van jullie gevraagd om tot deze resultaten te komen. Jullie hebben dit onbaatzuchtig gedaan voor een betere wetenschap en om anderen te helpen. Zonder jullie was dit allemaal niet mogelijk geweest.

Esther en **Margo**, dank jullie voor alle ondersteuning op de afdeling. Jullie inzet en kennis van dit soort zaken zorgt ervoor dat wij ons op het onderzoek konden focussen.

Gerda Bloem, dank voor alle ondersteuning met alle apparatuur, een cruciaal punt in ons onderzoek. Door jouw focus en oog voor detail hebben we gebruik kunnen maken van goedwerkende apparatuur.

Daniël, Evi, en Gijs, dank jullie voor de hulp bij het verwerken van de hartslagdata en de verdere ondersteuning bij het project. Jullie hebben allen unieke en sterke vaardigheden, en ik ben er zeker van dat jullie het ver kunnen schoppen hiermee.

Evelien, wat heb jij je ontzettend ingezet om van dit project een succes te maken. Je hebt de taken die daarbij horen vol overgave aangepakt en je er volledig voor ingezet deze goed af te ronden. Jouw gedrevenheid hierin is naar mijn inziens enkel overschaduwd door jouw intelligentie en kennis van de onderwerpen die wij in dit project bestudeerden.

Maurits, **Vera**, en alle andere **open science enthousiastelingen**, wat was het fijn om met jullie samen te werken, en te debatteren over iets wat we allen belangrijk vinden; open science. En niet alleen het transparant maken van wetenschap staat bij jullie hoog op de agenda, maar ook het verbeteren van de wetenschap in de breedste zin. Dankzij jullie inzet maken we ons werkveld stapje voor stapje beter en integerder.

Sandip and **Olga**, it was a pleasure working together on quite a few papers and projects. You are already experts in your respective fields, and I'm sure you will make an even bigger impact on the scientific community in the future.

Olivia, **Martine**, **Anu**, **Davinia**, and **others from the ESM Item Repository**, hi you guys, I never would have thought that one hackathon would result in such a cool, long-lasting, and relevant project. There is almost no project that I have enjoyed working more on, in such a stress-free manner. I feel this is mostly because of how complementary our skills are, with such people-skills, charisma, tech-savviness, and just plain interest in helping to improve science and people's lives, we are bound to keep improving the Repository into the open science gem it is.

YAM FARM, alleen terugkijkend kan ik beseffen hoezeer ik het wel niet met jullie getroffen heb als directe collega's. Het was fijn om gezamenlijk koffie te gaan drinken en het werk te kunnen bespreken. Ook de borrels, etentjes, en schrijfweken die we gezamenlijk hebben meegemaakt waren mooie avonturen.

Arnout, als er op werk lol te beleven was, of als er iemand was die op een vriendelijke wijze in de maling genomen kon worden, dan waren we er vaak samen al snel bij. De manier hoe we elkaar daarbij aanvulden vond ik erg bijzonder. Naast de grappen en grollen is het mij ook duidelijk geworden hoe sterk jij bent in jouw vakgebied, en hoe goed je deze kennis kunt overdragen aan anderen. Ook heb je een erg empathisch kant die jou een erg prettige collega en mens maken.

Marmar, jij creatieve geest. Wat ben jij goed in het ontwerpen en creëren van dingen! Maar daarnaast ben je ook iemand die heel goed kan luisteren en eigenlijk nooit een snel een slecht oordeel over een ander zal vellen. Jouw persoonlijkheid schept een gevoel van veiligheid en geborgenheid die mij erg geholpen heeft.

Fionneke, naast jouw kennis van de wetenschap en de studies waar je je voor inzet was het ook gewoon fijn om met jou over andere dingen te kunnen spreken. Je bent nuchter maar ook empathisch en invoelend. Zo schep je de mogelijkheid voor de mensen om je heen om open van alles te kunnen praten.

Anouk, ik vond het altijd bijzonder om te zien hoe sterk je je kunt maken voor dingen waar je in gelooft en hoe fijn praktisch je oplossingen vind voor ogenschijnlijk complexe problemen.

Robin, jij bent denk ik een geboren leider, iemand die haar eigen plan trekt en anderen kan motiveren daarbij te ondersteunen. Je hebt een sterk gevoel voor hetgeen wat om je heen gebeurd en weet dit vlot naar actie te vertalen. Gecombineerd met een overtuigend intellect kun je het ver schoppen in de werkrichting die je ambieert.

Marieke, je bent een bijzondere combinatie van iemand met een zeer ontwikkelde intelligentie maar ook een neiging om heel basaal en zonder filter te kunnen reageren. Dit maakt je tot een heel eerlijk en authentiek persoon. Je enthousiasme over sommige onderwerpen werkt vaak aanstekelijk.

Elwin, **Guus**, en **Stef**, volgens de theorie van Robin Dunbar kan een persoon maximaal vijf goede vrienden hebben. Ik heb er echter maar drie nodig gehad, en dat zijn jullie. Het is altijd fijn om met jullie even mijn zinnen te verzetten en even ergens anders aan te denken dan het afronden van een promotietraject of proefschrift. **Elwin**, jij bent altijd diegene geweest waaraan ik mijn sleutel van mijn kluisje gaf als ik dacht dat ‘m zelf kwijt zou raken. Je bent betrouwbaar en down-to-earth, maar ergens ook empathisch en proberend om een goede

middenweg te vinden. Ik hoop dat we nog veel biertjes mogen drinken. **Guus**, je bent altijd de doerak van de groep geweest, diegene die iedereen uitdaagde, maar daardoor ook op zichzelf laat reflecteren. Je hebt vaak een sterk gevoel over dingen, wat je met passie uitdraagt. Daarnaast heb je enorme creativiteit waar je een succesvolle business omheen hebt gebouwd. Ik hoop dat we nog vaak mogen proosten. **Stef**, ik ken je al een ontzettend lange tijd, sinds we samen als tieners op scouting zaten. Ik ben blij dat we zo veel mooie dingen hebben mogen meemaken, al hebben we ook het nodige verdriet gedeeld. Je hebt een passie voor muziek en je kent originele manieren om hier vorm aan te geven. Ik hoop dat we nog vaak mogen toosten op de goede dingen in het leven.

Cheyenne en **Jenita**, / lieve **zussen**, ook al is er een leeftijdsverschil tussen ons ben ik ontzetten blij jullie als zussen te hebben. Zeker nu we voor mijn gevoel de laatste jaren meer naar elkaar toe gegroeid zijn. **Cheyenne**, dank je wel voor je hulp bij het organiseren van de promotie. Je bent van nature iemand bij wie het regelen van dingen goed af gaat. Je bent secuur en doortastend en ziet de dingen het liefst goed geregeld. Je bent ontzettend betrokken bij dingen die je belangrijk vind. Je bent erg goed op weg met een mooie baan en een woning. Ook ben je jezelf steeds verder aan het ontwikkelen met cursussen en trainingen. Super dat je het allemaal zo goed doet, ik ben blij dat jij mijn zus bent! **Jenita**, dank je wel voor je inzet bij het grafische ontwerp van mijn proefschrift. Je bent een creatief en ondernemend persoon. Daarnaast ben je geïnteresseerd om ook de andere kant van het verhaal te horen. Inmiddels heb je een baan gevonden die aansluit op je studie en ben je serieus bezig met jouw ambities om een creatieve onderneming op te zetten. Geweldig om te zien hoe je hierin groeit, ik ben blij dat je mijn zus bent!

Lia, **Leo**, **Gerda**, **Eric**, / lieve **ouders**, ik kan niets op papier zetten wat geheel mijn dankbaarheid naar jullie beschrijft. Jullie hebben mij altijd met raad en daad bijgestaan. We hebben mooie maar ook moeilijke periodes meegemaakt, en volgens mij zijn we hier allen erg in gegroeid. **Mam**, dank je wel dat je er altijd voor mij geweest bent en dat je altijd voor mij gevochten hebt. Ik vind het erg fijn zo veel dingen met jou te kunnen bespreken en gevoelens te kunnen delen. **Pa**¸ dank je wel voor de momenten die we samen door konden brengen en jouw humor. Ik kijk er naar uit samen over interessante onderwerpen, wetenschappelijk, sciencefiction, enzovoorts, te kunnen blijven praten. **Gerda**, dank je voor alle gesprekken waarin je mij een spiegel voor kon houden zodat ik op dingen kon reflecteren. **Eric**, dank je voor alle steun en jouw grondige en nuchtere kijk op dingen, zonder dat zou ik hier niet kunnen zijn.

Lieve **Berber** en **Sarah**, wat hebben wij veel meegemaakt, mooie maar ook uitdagende momenten. **Berber** dank je voor jouw steun de afgelopen jaren die geholpen hebben dit proefschrift af te ronden. Samen vervelen we ons geen moment en kunnen we bijzondere dingen meemaken. Fijn hoe je altijd hebt ondersteund bij het beter maken van dingen zo ook dit proefschrift. Je steun, aanmoediging en liefde hebben het verschil gemaakt tijdens dit veeleisende proces. Bedankt dat je er altijd voor me bent geweest en me hebt aangemoedigd om door te gaan, zelfs als het zwaar werd. Ik ben ontzettend dankbaar voor jouw steun. **Sarah**, mijn lieve dochter, je bent nog zo jong maar ik hoop dat jouw al een mooie en gelukkige toekomst staat te wachten. Ik hoop dat je veel gezondheid, geluk, en plezier mag treffen op jouw levenspad. Of je nu wel of niet ook de academische kant op gaat, volg je hart, wees eerlijk naar jezelf, en weet dat ik altijd van je hou.

Curriculum Vitea

Yoram Kevin Kunkels was born on March 6th 1986 in Haarlem, the Netherlands. He finalised the first three years of senior general secondary education (HAVO) at Lyceum Sancta Maria in 2004. Thereafter, he finalised the first three years of middle-level vocational education (MBO) at Nova College Hoofddorp, attaining the diploma *MBO car technician level 3* in 2008. He worked as a wheel-alignment and tire specialist for a number of years before starting work as a car detailing and polishing entrepreneur within the *Sonax car care* franchise. While working in the automotive industry did offer many ways to develop one's practical sense of engineering, logistics, and sales, it did not offer as many intellectual challenges at the position Yoram was working in then. During these years Yoram thus focussed on trying to gain access to higher education. While early attempts to enrol in such educational institutions were hampered by a lacking educational requirements, in 2011 a study advisor at the *University of Amsterdam* (UvA), Mrs. J. M. de Vries, supported Yoram in getting math tutoring and the chance to enrol in university by successfully completing an admission test, the so-called *colloquium doctum*. Yoram did successfully complete the test and gained admission to the bachelor study *Psychology* at the UvA, completing his bachelor in 2013 with honours and a specialisation in psychological methods. During this time he got not only interested in science in general, but also after watching a presentation by Brian Nosek, in *Open Science*, a movement to make scientific research more transparent and reproducible. Together with a co-student, Bennett Kleinberg, he participated in the *Reproducibility Project: Psychology*, reproducing one of 100 published psychological studies. The final article, to which 270 international authors contributed, identified a number of issues with reproducibility in psychological studies and was published in *Science* in 2015. In 2016 Yoram graduated as a research master student from the UvA, and was honoured to be invited to join the TRANS-ID team where he started heart rate and physical activity data collection, and academic reporting thereof, under the supervision of prof. dr. M. Wichers, dr. H. Riese, and dr. ir. A. M. van Roon. For his work to promote *Open Science* practices Yoram received the SHARE Open Science award in 2019. Software written by Yoram during his PhD was placed in public repositories on Github, which were part of the 2020 *GitHub Arctic Code Vault* project, which snapshotted this software, amongst software from many other developers, and physically stored in a decommissioned coal mine beneath an Arctic mountain in Svalbard, Norway, in order to create a long-term software archive.

List of publications

Published

Kunkels, Y. K., Knapen, S. E., Zuidersma, M., Wichers, M., Riese, H., & Emerencia, A. C. (2020). ACTman: Automated preprocessing and analysis of actigraphy data. *Journal of Science and Medicine in Sport, 23*(5), 481-486. https://doi.org/10.1016/j.jsams.2019.11.009.

Kunkels, Y., Roon, A., 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). https://doi.org/10.1111/psyp.13898

Kunkels, Y. K., Riese, H., Knapen, S. E., Riemersma - van der Lek, R. F., George, S. V., van Roon, A. M., 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). https://doi.org/10.1038/s41398-021-01465-w

Kunkels, Y. K., Smit, A. C., Minaeva, O., Snippe, E., George, S. V., van Roon, A. M., Wichers, M., & Riese, H. (2023). Risk Ahead: Actigraphy-Based Early-Warning Signals of Increases in Depressive Symptoms During Antidepressant Discontinuation. *Clinical Psychological Science, 0*(0). https://doi.org/10.1177/21677026221148101

Bastiaansen, J., Kunkels, Y., Blaauw, F., Boker, S., Ceulemans, E., & Chen, M. et al. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal Of Psychosomatic Research, 137*, 110211. https://doi.org/10.1016/j.jpsychores.2020.110211

George, S. V., Kunkels, Y. K., Booij, S., & Wichers, M. (2021). Uncovering complexity details in actigraphy patterns to differentiate the depressed from the non-depressed. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-92890-w

George, S. V., Kunkels, Y. K., Smit, A. C., Wichers, M., Snippe, E., van Roon, A. M., & Riese, H. (*in press*). Predicting recurrence of depression using cardiac complexity in individuals tapering antidepressants. *Translational Psychiatry*.