Outside the Office Hours

-

A Scoping Review of Ecological Momentary Techniques to Assess and Treat Personality Disorders

Jonas Möller

1789422

Faculty of Behavioral, Management and Social Sciences, University of Twente

Examination Comité

Dr. Matthijs Noordzij

Dr.ing. Gert-Jan Prosman

25th of August, 2021

UNIVERSITY OF TWENTE.

Abstract

Introduction: Ecological momentary assessments (EMA) and interventions (EMI) have been successfully implemented in various fields of psychology for years to measure constructs of interest in their natural environment and in real-time. However, relatively little is known about new progressions in the field of personality disorders (PD) which are commonly recognised to be challenging to treat. This scoping review attends to six questions in total. Four primary objectives: opportunities for new interventions, evaluation of treatments, improving understanding of symptoms, and capturing and predicting emotional variability. Two additional secondary objectives were formulated to answer broad questions of descriptive nature: The prevalence of different personality disorders and the qualities of utilized techniques. The two databases Scopus and PsycINFO were therefore explored, appropriate content summarized and answers synthesised based on an influential paper by Myin-Germey and her colleagues. **Methods:** Papers were deemed eligible if they were available in English or German, if they were published in the last twenty years, if they included at least one participant who was diagnosed with an PD or qualified for a PD diagnosis, if a EMA or EMI design was utilized and acquired data through these methods was used in the treatment or assessment of the subjects. Extracted data was briefly summarized and integrated into two tables to present a quick overview. **Results:** All types of PDs were represented in the cumulative sample of the 15 papers, while borderline personality disorder (n = 8) and schizotypal (n = 4) constitute the most represented disorders. Data was mostly gathered through web-based internet sites (n = 11) and mobile applications (n = 4). In terms of capturing and predicting emotional variability several methods proved themselves useful, surpassing the objective by providing high accuracy in the diagnostic process. Improvements in the understanding of symptoms were represented, achieved through various means and tapped into different data when compared to other employed EMA measurements. While the evaluation of interventions was in no paper the main objective, methods that captured emotional variability were found to be suitable for this task. Opportunities for new interventions were also discovered, although they did not differ in their main objective, skill training, means of achieving this task differed among them. **Discussion:** Included papers illustrated their feasibility and demonstrated promising results which support their integration into everyday practice. Future research should establish experiments with randomized control trails, investigate their feasibility for wide implementation and develop interventions to include lesser represented personality disorders.

Table of Contents

1. A Scoping Review of Ecological Momentary Techniques to Assess and Treat Personality Disorders

Technology is increasingly more integrated into the research and clinical practices of psychologists allowing to measure behaviour and its predictors outside of controlled environments, therefore, gaining insight into the enfolding of emotions in the day-to-day life (Myin-Germey, Kasanova, Vaessen, Vachon, Kirtley, Viechtbauer, Reininghaus, 2018). One of the most promising methods is the Ecological Momentary Assessment (EMA), also known as the Experience Sampling Method (ESM).

Promising results have been published by integrating applications that repeatedly measure certain psychological constructs over long periods. Ecological momentary methods have been employed both in the treatment and the assessment of various mood and substance abuse disorders but relatively little is known about the efficacy with regards to PDs (personality disorders) (Morgenstern, Kuerbis & Muench, 2014). The ability to extend assessment and intervention beyond the hourly meetings in the office has the potential, to shape the future of psychological practice from predominantly in-person meetings to more time flexible interventions and assessments which do not require as much face to face contact.

Personality disorders are notoriously tricky to treat, problems ranging from trust issues that undermine rapport building to the comorbidity with a range of mental illnesses (Cloninger, Svrakic, Lester, Lecic-Tosevski, Koldobsky & Botbol, 2017). Employing new technological means might support treating and assessing PDs, the technologies in question could help to unravel the complex nature of emotional variability in personality disorders to give feedback on the intervention or help in the diagnostic process by showing the relations between psychopathological traits in the client. There seem to be numerous ways of how repeated sampling could be employed to support practitioners and their patients. But as far as the author was aware of, was no review conceived prior to this research that specifically deals with EMA techniques, their various ways of implementation and their technological qualities in the field of personality disorder.

Therefore, this scoping review focused on identifying the types of available evidence in the assessment and treatment of personality disorders that utilized ecological momentary techniques. The overarching goal of this paper was to utilize a broad approach to various papers with relatively few self-imposed restrictions to establish the current state of the art of this subject. Key characteristics and factors related to the treatment and assessment via EMA and EMI will be identified and mapped out to show how current research is developing concerning PDs.

1.1 Ecological momentary assessment

EMA is an intensive longitudinal research methodology aimed at the gathering of self-reports about experiences such as contexts, emotions, thoughts, behaviour and/or physiological states. Participants are usually instructed to complete questionnaires daily and possibly on multiple occasions across the day. Data can be gathered about certain events to find correlations between e.g. emotions and contexts, or data can be continually captured like the heart rate of a person which is especially useful in the studies concerned with physiology. This method is able to capture its data in a real-life setting and in real-time, meaning it diminishes the recollection bias that is prevalent in cross-sectional studies, therefore, providing high ecological validity (Versluis, Verkuil, Lane, Hagemann, Thayer & Brosshot, 2021). Cross-sectional studies such as traditional self-reports rely on memory and are, therefore, subjected to reconstructed experiences which can obscure data and cause a wide array of phenomena such as recalling dates, emotions or pain. However, ambulatory assessment is not only about diminishing biases but also assessing different types of knowledge. The introspective nature of ambulatory assessment in short-term recall reports allows measurements of the individual's semantic knowledge and, therefore, acts as a report of their experiences which is closer to the actuality than a longer-term retrospective report (Conner & Barrett, 2012).

The article by Myin-Germeys et al. (2018), enumerates a couple of reasons why this kind of technology is useful for clinical practice. For instance, it is helpful to improve the understanding of symptoms, being able to assess both mood and activities allows one to easily relate both constructs to another and uncover their emotional experiences throughout the day (Myin-Germeys et al., 2018). Capturing the emotional variability is another helpful utensil too, contexts can be linked to (mal)adaptive regulation strategies and daily assessments could be used to predict the future course of symptoms. The last definitive plus point EMA provides is the ability to evaluate treatments. It is argued that this methodology captures more sensitive data about the alteration in psychological, biological or social variables. This also allows for the detection of side effects earlier than conventional methods. Plotting the Baseline assessment of positive and negative affect against the changes captured by EMA predicted the treatment response in depressed clients who received antidepressant treatment in a time frame of six weeks (Myin-Germeys et al., 2018).

1.2 Ecological momentary interventions

Ecological momentary interventions (EMIs) just like EMAs are provided in the natural setting and during the day to day lives of their users. EMI in itself is not an intervention but provides a framework in which a wide array of different tasks or information can be delivered to the client at any time, such as providing recommendations for relaxation exercises when stress is detected (Heron & Smyth, 2010). EMI can provide real-time support based on explicitly identified situations and could act as a supplement for existing interventions or can be an intervention in its own right. Providing it as a supplement treatment alongside the main intervention has several benefits. The acquisition of new skills is fundamental for psychological interventions and with the help of EMIs clients are instructed to put them to use or test them in their natural environment. "EMI can reinforce the systematic use of treatment components in real-world settings, thus generalizing the intervention program's impact" (Heron & Smyth, 2010).

Additionally, it was reported that the continuous accompaniment of the EMI let the clients feel supported beyond their office visits, this positively influences the training of new skills and behaviour (Kazanzis & L'Abate, 2007). This technology also allows for the tailoring of content based on the client's needs, the timing of messages and the repetition of them. For instance, one alcohol use disorder study utilized this design, tracking their clients via their smartphones and provided them tailored support when they were in a high-risk situation such as closing into a bar (Gustafson et al., 2014). These tailored interventions delivered immediately in the natural setting are a form of adaptive treatments which are modified components based on the developments of prior stages in the intervention. This EMI is called just-in-time adaptive intervention (JITAI) which "adapts the provision of support (e.g., the type, timing, intensity) overtime to an individual's changing status and contexts, with the goal to deliver support at the moment and in the context that the person needs it most and is most likely to be receptive (Nahum-Shani et al., 2018).

1.3 Personality disorders

Personality is broadly defined as a relatively stable set of features such as one's interests and abilities that determine how one responds to situations and experiences (APA, 2013). This includes the style of coping, but also the adaption to life events and challenges. While a personality is mostly defined as enduring, most people will ultimately learn and evolve based on their experiences, changing behaviour and modifying parts of it. However, some people are

far less flexible and don't adapt to life's challenges, they maintain their behaviour despite the maladaptive consequences it is causing to them. This can lead to the disruption of everyday life and harms the individual on many different levels, troubles in the social circle and the work-life can arise, which is among other things characteristically in people who suffer under any PD (Davey, 2008).

The term PD has been often abused to describe individuals as untreatable or as especially difficult (Tyrer, Reed & Crawford, 2015). According to the American Psychological Association (APA) *personality disorders* are defined as "[…] a group of disorders involving pervasive patterns of perceiving, relating to, and thinking about the environment and the self that interfere with long-term functioning of the individual and are not limited to isolated episodes "(American Psychological Association, 2013). The definition of the diagnostic and statistical manual of mental disorders 5 (DSM-5) regarding personality disorders further adds that they are "[…] an enduring pattern of inner experience and behaviour that deviates markedly from the expectations of the individual's culture, is pervasive and inflexible, has an onset in adolescence or early childhood, is stable over time, and leads to distress and impairment" (DSM-5, 2013, p.654). The DSM-5 distinguishes between 10 different personality disorders and classifies them into three clusters which are A, the odd/eccentric Cluster, B, the dramatic/unpredictable cluster and C, the anxious/fearful cluster.

Cluster A personality disorders symptoms are similar to schizophrenia symptoms without the loss of reality or the experience of hallucinations, e.g. they experience delusions or poorly organized speech. This Cluster inhibits three subtypes which are paranoid personality disorder, schizotypal personality disorder and schizoid personality disorder (Davey, 2008). Personality disorders in cluster B have their erratic and self-interested behaviour in common, individuals suffering under these exhibit extreme behaviour and inflict great distress onto others. The last cluster C includes personalities that are characterized by anxious and fearful behaviour as a stable companion since the onset from late childhood (Davey, 2008).

Treatment is generally more impeded in comparison to disorders because of the difficultly of establishing a stable relationship with the therapist. In the cases of coexistent mental disorders, the prognosis is worsened because of the complex relationship between mental and personality disorders. The increased degree of disturbance requires more intensive treatment and the deeply ingrained behaviour is causing continuous problems in the future. In addition, the general distrust towards people, which is associated with many personality disorders, is damaging for the relationship the therapist is trying to build up with the client (Davey, 2008). According to McMurran, Huband and Overton (2010), 37% of individuals

diagnosed with a personality disorder fail to complete the treatment devised for them. Factors that contribute to this noncompliance are "[…] young age, lower education levels, unemployment, having juvenile convictions, and emotional neglect in childhood" (Davey, 2008, p.441). Furthermore, with an estimated lifetime prevalence rate of 14-15%, they are arguably one of the most prevalent psychopathologies in the mental health field (Zimmermann, Rothschild & Chelminski, 2005).

1.4 Research objectives

The overarching goal of this paper was to explore and consequently draw up recent scientific endeavors to create an overview of the current state of the art with regards to momentary ecological techniques and their respective technologies to assess as well as treat personality disorders. Primary research questions were formulated to capture the intricate matters of current technologies and processes. Therefore, inspiration was taken from the article of Myin-Germeys et al. (2018) which attended to several key points of new insights and technical developments in the experience sampling methodology. Here we applied these questions specifically pertaining to the treatment of personality disorders and addressed to what extent:

1) EMA methods can capture and predict emotional variability?
2) EMA methods are able to improve the understanding of symptoms?
3) EMA methods are helpful in the evaluation of treatments?
4) EMA methods open up new opportunities for new interventions?

In addition, secondary research questions were formulated to include general information, descriptive in nature. Regarding the second questions, the focus was laid on the technologies, what qualities they possessed, among them but not limited to validity, reliability, measuring accuracy and with regards to intervention studies the measurable skill gain of patients or their improved feelings etc., as well as the different ecological momentary methods and how they used repeated sampling for their cause. Therefore, following secondary research questions were formulated:

1) Which personality disorders were predominantly present in the studies?
2) What are the qualities of the employed techniques?

2. Methods

2.1 Search selection and criteria

The search for viable articles has been conducted in March 2021 on two databases, Scopus and PsycINFO. PsycINFO is conceived as a database strictly for the psychological discipline and is thought to yield high-quality articles, although with 4 million records it is not providing the same mass as other sites, which are related to the topic of this scoping review (Gusenbauer, 2018). Scopus, on the other hand, was chosen as a complimentary search, providing articles to a bigger extent. Other fields of expertise are also considered which was to a certain extend less possible in the PsycINFO search. With around 71 million articles is Scopus the biggest multidisciplinary database and is therefore viable for any research, especially to form an overview about how extensive a certain topic is, particularly useful for a scoping review, based on the amount of literature available (Gusenbauer, 2018).

Two different search strings with Boolean operators were utilized to discover articles that deal with the two core themes of this review, different utilization of EMA and EMI with regards to personality disorders (Table 1). The Boolean operator "OR" was used to include the variation of the key term in question, while "AND" was used to make sure that "personality disorder" is also included in the papers. The decision to utilize quotation marks for Scopus was only made because of their advanced search option. Scopus took the initial search literally and treated every word used in the search string as an "AND" term, the quotation marks forced the site to search for the grouped words individually. In addition, a couple of articles were retrieved through snowballing, these will be marked with an "S" accordingly throughout the span of this review to prevent confusion in the case of replication.

Several inclusion criteria were conceived beforehand to make the search overall narrower and more focused. Only articles in English and German were eligible in a twenty-year publishing period ranging from 2001-2021. According to an article by Shiffman, Stone & Hufford (2007), the concept of EMA is not a novel idea. Similar techniques like diary-keeping have been used in the 1940s but only since the mid to end of the 90s were more technical devices used, e.g. pagers, that bear resembles to smartphones that are currently in practice. Since the turn of the century, this technique attained more exposure and steady development incorporated more complex technical devices. To ensure this inclusion, a range of twenty years from the beginning of the 21 century was set and perceived as eligible with regards to the extensiveness of specificity. However, PsycINFO only provided articles in a ten-year period, 2010-2020, in the case of the second search string. In this case can one be sure about an extensive search, at least regarding this database.

Because of the fairly broad approach of this literature review and the relatively slim amount of suitable research only a couple of inclusion and exclusion criteria were conceptualized with regards to the matter of the research. First, articles were eligible when they included at least one participant with a PD. Second, papers were included when they utilized an ecological momentary approach in their study design, meaning information were gathered and/or provided in the natural environment of the subjects and are therefore tied to specific points in time. Assessment studies were eligible when they measured the same constructs over a period in a sample composed out of, but not restricted to, participants with any personality disorder. In addition, acquired data must have been used in the establishment of a possible diagnosis.

Intervention studies must either provide time unrestricted distribution of relevant information, support in the acquisition and implementation of skills or aid in the detection of treatment targets that support the main intervention. This inclusion criterion was established because at the time this review was conceived were, to the best of the authors' knowledge, no studies used an EMI as the only means of treatment for individuals suffering from any personality disorder.

After implementing the search string abstracts were firstly screened based on their title. If the title deemed itself intriguing enough, step 2 the thorough reading of the abstracts was initiated to determine the appropriateness of employed methods. Articles were then transferred into step 3 which meant to read the whole paper and consequently to conceptualize a list of articles that together constitute a broad spectrum of the recent state of the art scientific endeavours.

Table 1

Overview search strategies

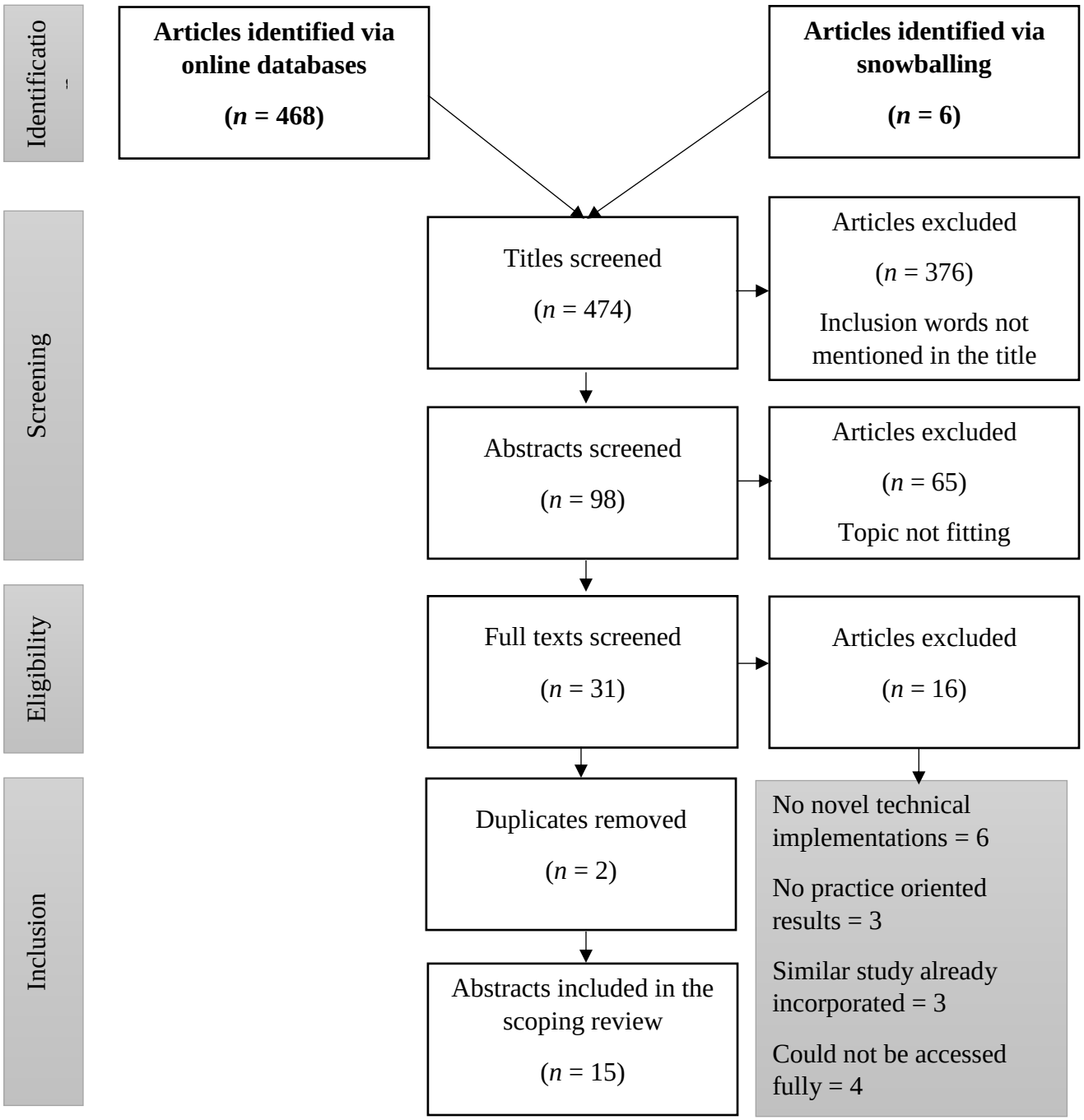
Search string	**1.** Ecological momentary assessment OR experience sampling method OR ambulatory assessment AND personality disorder		**2.** Ecological momentary intervention OR just-in-time adaptive intervention OR just-in-time AND personality disorder	
Database	PsycINFO	Scopus	PsycINFO	Scopus
Number of articles	192	91	48	143
(Snowballing)		(5)		(6)
Articles in total =	474			

To identify relevant articles a stepwise procedure was employed and used cross-platform. The first step on each platform after the implementation of the search string including the correct use of Boolean operators was to implement the inclusion criteria into the databases, English and German articles were picked out from 2001-2021. After the deduction of irrelevant characteristics, the rest of the existing articles were screened based on their titles. Step two involved identifying each article on the behalf of their choice of word in the title. If words used in the search string were also found in the title the article was transferred to the next step.

However, not every study models their title purely objectively based on the methods and technologies employed, articles that hinted at the use of EMI or EMA were also transferred to the next step. Articles were left out that directly showed no practice relevance, meaning, studies were included if they were employed in a working environment, clinic, office etc. Another relevant criterion for the inclusion was that the study had to be carried out with a clinical population who have been diagnosed with a personality disorder. Because of the high comorbidity with different mental disorders, studies were not excluded that possessed a mixed sample of subjects who suffered under both or either one of them. These characteristics were also searched for in the subsequent step which entailed a more detailed search in the abstract section of the articles that have passed step two. Figure 1 illustrates this process in a flowchart.

Figure 1

Flowchart of the search process of relevant articles



2.2 Procedure and analysis

All articles which were chosen to be incorporated into this scoping review were entirely read, analyzed and put into order by one researcher. Because this literature review is focused on how, by what means and which PDs are treated /assessed two tables were created. The first table served as an overview function regarding what mental disorders and personality disorders were present, how big the sample is and what population constitutes them. This supports the inquiry process regarding the focus in current research and how representative the study results were based on the sample size, as well as general insight concerning age, gender etc. All of these points were extracted in the methods section of each article.

The second analysis was executed to collect information on employed technologies, means of analysis, goals of the study, procedures and their classification, meaning to which key point they pertain to according to the article of Myin-Germeys (2018). Data were for the most part extracted from the method and result parts to show what advances were made and are to be expected in the future. Classifications were ascribed to the articles based on the clinical implications found in the respective discussion section to offer the main points of contribution to the field. Platforms were also noted to show how the treatment was administered, giving insight on how far contactless treatment has developed and how communication took place. However, two papers could not be properly classified based on the prior established criteria but were nevertheless deemed important for this review. The first paper in question by Eaton, South & Kruger (2009) was not a study but more a treatise about the cognitive-affective processing system (CAPS) and how the current conceptualization of stable patterns fails to accurately describe the concept of personality disorder. The qualification of that paper was argued, despite not having conducted experiments with individuals diagnosed with a PD, through the incorporation of the daily diary design which the authors described as a helpful tool in acquiring situational sensitivity which is according to them needed to treat maladaptive behaviour. Consequently, columns were marked with ‘n.a.’ if fitting information could not be found in the text.

Second, the other paper by Zimmerman et al. (2019) did not directly describe the diagnosed psychopathologies in their second experiment sample but later mentioned in the discussion section a specific case in their study where one subject was prior diagnosed with a personality disorder. Based on that information and them reporting on gathering their sample from an inpatient treatment centre, enough assurance was provided, indicating they were working with a sample that was at least partially made up of clients who were suffering from personality disorders.

3. Results

In total 15 out of 474, studies were thoroughly examined and included in the scoping review. Articles collected ranged from being published in 2009 – 2020 (see Table 2). Studies can be distinguished between assessment of PDs and treatment of PDs, the first 11 studies are conceptualized as assessment the remaining 4 as treatment of PDs.

3.1 Secondary findings: participants characteristics

The exact characteristics per study are summarized in table 2, for the sake of comprehensibility only explicit numbers of personality disorders were included, disregarding comorbidities such as mood or anxiety disorders which were also present but not important in this context. Additionally, personality disorders that were identified in less than 10% were grouped to make the table more transparent.

While not all studies reported extensively on their distribution of race and age, white/Caucasian middle-aged females were most represented. The exact age of the youngest and oldest participant was not decisively named, but the age range according to two studies stated that most likely the youngest participant was around 15 years old (Shelby et al., 2020) and the oldest at least 65 (Wenze, Armey, Weinstock, Gaudiano & Miller, 2016) at the beginning of the respected study. The most diagnosed personality disorders in these samples were borderline personality disorder (n = 8) and schizotypal (n = 4). Next to personality disorders were also depression (n = 7), anxiety (n = 6) and PTSD (n = 3) assessed, indicating the highly comorbid nature of personality disorders.

Considering the distribution of personality disorders among the articles, were two PDs predominant namely borderline personality disorder and schizotypal personality disorder. In total 1138 individuals took part in the 14 studies (subtracting Eaton, South & Kruger's paper), 259 participants or 19.5% were either diagnosed with BPD or reached a significant cut-off score which led to their inclusion. In contrast, individuals suffering under schizotypal personality disorder were not as numerous in terms of the included studies (n = 4 compared to BPD n = 8) but showed a higher number of included individuals, 269 participants or 20.1%. The stated numbers constituted the absolute minimum, and only included samples where the sample characteristics were sufficiently specified Taken all together, a great variability could be recognized, for two studies in particular (Lane, Gates, Pike, Beltz & Wright, 2019; Dotterer, Beltz, Foster, Simms & Wright, 2019) who acquired more of the lesser represented PDs, for instance, histrionic personality disorder.

3.2 Table 2

Participant characteristics

Authors	**Nationality or Ethnicity**	**Gender**	**Age range (if included) and mean age in years (SD)**	**Psychiatric diagnoses**
1. Selby et al. (2020)	18 (38%) White 7 (15%) African American 9 (19%) Asian 8 (17%) Hispanic 5 (11%) Multiracial	32 (68%) Female 14 (30%) Male 1 (2%) Transgendered	15-21 years of age, *M* = 19.1 (*SD* = 1.77)	42.5% moderate to higher levels of BPD
2. Zimmermann et al. (2019)	Not specified	1. Study: 224 Female (71.3%) 2. Study: 53 (68.8%) Female 3. Study: 8 (100%) Female participating Therapists, 95 (73.6%) Female nonparticipating Therapist & 26 (89.7%) Female participating patients	1. Study: 23.3 (*SD* = 5.3) 2. Study: *M* = 38.5 (*SD* = 13.3) 3. Study: participating therapists, *M* = 37.5 (*SD* =17.0) Nonparticipating therapists, *M* = 46.1 (*SD* = 10.3) Participating patients, *M* = 38.3 (*SD* = 11.1).	12% disorders of adult personality and behavior

3. Wright et al. (2019)	Not specified	53 (68.8%) Female	*M* = 38.5 (*SD* n= 13.3)	"All of the individuals had at least one personality disorder diagnoses, with many having a borderline personality disorder diagnoses (39%) and/or narcissistic personality disorder (20%)" (Wright et al., 2019, p. 505)
4. Eaton, South & Kruger (2009)	N.a.	N.a	N.a.	N.a.
5. Arribas, Goodwin, Geddes, Lyons & Saunders (2018)	Not specified	94 (69.2%) Female	Bipolar *M* = 38 BPD *M* = 34 Healthy Control *M* = 37	23.8% Borderline personality disorder
6. Minor, Davis, Marggraf, Luther & Robbins (2018)	46 (100%) Non-Hispanic Caucasian	High schizotypy: 16 (64%) Female Low schizotypy: 10 (47.6%) Female	High schizotypy: 18 – 27 years of age *M* = 19.92 (*SD* = 2.14) Low schizotypy: 18 – 25 years of age *M* = 19.81 (*SD* = 1.86)	54% high schizotypy 46% low schizotypy

7. Lane, Gates, Pike, Beltz & Wright (2019)	White 83 (82.2%) Black 15 (14.9%) Native American 3 (3%) Hispanic 5 (5 %)	66 (65.3%) Female	*M* = 44.9 (*SD* = 13.3)	53.5% avoidant personality disorder 50.5% obsessive-compulsive disorder 36.6% Borderline personality disorder 35.6% paranoid personality disorder 19.8% narcissistic personality disorder 16.8% schizotypal 13.9% schizoid personality disorder <10%: antisocial, dependent & histrionic personality disorder
8. Chun, Barrantes-Vidal, Sheinbaum & Kwapil (2017)	Not specified	163 (79.1%) Female	*M* = 19.8 (*SD* = 2.4)	189 with standard scores of >1.0 on positive or negative schizotypy

				150 randomly selected participants who had standard scores <1.0
9. Schiepek, Aichhorn, Gruber, Strunk, Bachler & Aas (2016)	Not specified	107 (70.4%) Female	*M* = 36.2 (*SD* = 11.6)	23.7 % Disorder of adult personality and behavior
10. Dotterer, Beltz, Foster, Simms & Wright (2019) (S)	76 (83.5%) White 12 (13.2%) Black 3 (3.3%) Native American Hispanic 4 (4.4%)	57 (62.6%) Female	*M* = 45.13 (*SD* = 13.37)	56% avoidant personality disorder 48.4% Obsessive-compulsive personality disorder 39.6% Borderline personality disorder 36.3% paranoid personality disorder 18.7% narcissistic personality disorder 16.5% schizotypal personality disorder 13.2% schizoid personality disorder

				<10%: antisocial, dependent & histrionic personality disorder
11. Ebner-Priemer et al. (2007) (S)	Not specified	100 (100%) Female	BPD patients, *M* = 31.3 (*SD* = 8.1) Healthy control group, *M* = 27. (*SD* = 6.8)	50% Borderline personality disorder
12. Carr et al. (2018)	Not specified	36 (27.9%) Male 93 (72.1%) Female	*M* = 36.4 (*SD* = 12.1)	24% Borderline personality disorder
13. Derks et al. (2019) (S)	Not specified	5 (100%) Female	18 – 49 years of age, *M* = 28 (*SD* = 11.82)	100% Borderline Personality Disorder
14. Schroeder et al. (2018)	Not specified	65 (89%)Female 7(9.6%) Male 1 (1.4%) Genderqueer	18 – 53 years of age, *M* = 37.3	39.2% Borderline personality disorder
15. Rizvi, Hughes & Thomas (2016) (S)	11 (69.8%) White 2 Hispanic 2 multiracial 1 Black	12 (75%) Female 4 (25%) Male	19-49 years of age, *M* = 27.5 (*SD* = 7.71)	100% met criteria for BPD

3.3 Secondary findings study characteristics

The sample size of the reviewed studies varied considerably from 8 to 314 participants. Utilized platforms were mostly web-based (n = 11) but mobile applications (n = 4) were also partly used as a form of data, the administration of daily questionnaires, distribution of information or to provide feedback directed to the patients. However, most web-based forms of assessment or treatment function were nearly identical to mobile applications. The main differences laid in the technical functions, e.g. internet sites needed constant internet access to function while some mobile apps had the feature to manually sent data which does not require internet access at every point in time. In addition, most of the websites were created with a mobile format in mind, restricting access through a personal computer. Other forms were less used like the paper format (n = 1; Zimmermann et al., 2019) to collect diary entries or handheld computers (n = 2; Chun et al., 2017; Ebner-Priemer et al., 2007)) for the participants to use.

In terms of assessment, the length of the studies ranged from once every day up to every 10 to 20 minutes. The studies that utilized one measurement per day were for the most part daily diary studies (n = 5) and were used in 50% of the assessment focused studies.

New noticeable means of analysis were also deployed. Several assessment studies made use of GIMME algorithms (n = 3) to quantify the relations of behavioural domains on several levels, a Bayesian modelling approach (n = 1) to predict facets of personality pathology or utilized other means of machine learning (n = 1) to analyze time-series data.

The EMI papers displayed a relatively broad field of application that in all but one case was conceived as an adjunctive component (n = 3) next to the main treatment. Predominantly, technologies were created which support the acquisition and the practicing of skills in the natural environment of the patients (n = 3). The other study focused on uncovering new potential treatment targets in borderline personality disorder through high-frequency monitoring of physiological data (n = 1).

3.4 Primary findings study characteristics

The results in this section pertained to the primary research question objectively, reporting on the robustness of included techniques and their various qualities. Each subheading was dedicated to one question and incorporates several papers belonging to the respective “RQ classification” which can be found in table 3.

3.4 Table 3

Study characteristics

Authors	Sample size	Objective(s)	Method	Analysis	Material	Platform(s)	Length	RQ Key Point
1. Selby et al. (2020)	47	Investigation of the emotional cascade model to predict the diagnostic status (BPD, MDD and PTSD), behavior and negative emotions.	Construction of a temporal Bayesian network model (TBN) from five EMA time assessments per day.	Modeling multiple feedback processes	Android smartphones K-fold cross-validation Machine learning	"Track It " mobile application	2 weeks	Mainly emotional variability Secondarily Evaluation of treatment
2. Zimmermann et al. (2019)	Phase: 1. 314 2. 77 3. 8 participating therapists, 129 nonparticipating therapists & 35 patients	The development and validation of a shortened version of the personality dynamics diary (PDD).	Phase: 1. Creating an item pool 2. Replicating psychometric properties with an inpatient sample.	Multilevel exploratory factor analysis Multilevel models with random intercepts applied to	Phase: 1. Online diary study 2. Online questionnaire version of the PID-5.	Phase: 1. Internet site 2. Internet site and paper booklet	Phase: 1. Two periods of, first seven days and then ten days 2. 9 months 3. At least 14 days	Understanding of Symptoms

			3. Exploring the feasibility and clinical utility in an outpatient setting.	compare scale means.	3. PDD questionnaire in an app format.	3. Mobile application and Internet sites		
3. Wright et al. (2019)	116	Using the group iterative multiple model estimation (GIMME) on longitudinal behavioral data to simultaneously model idiographic and nomothetic structures.	Online daily diary data was provided every evening of	Path structures are identified by relating behavioral variables, domains of personality, measures of stress and	gimmeSEM" function	Online daily diary via a secure website.	100 days of daily diary collection	Emotional variability
4. Eaton, South & Kruger (2009)	N.a.	Exploration of the cognitive-affective processing system	EMA to determine maladaptive behavior.	N.a.	Smartphones, Wireless sensors,	N.a.	N.a.	Opportunities for new interventions

		(CAPS) with regards to PDs.			GPS, audio and various physiological data assessment tools			
5. Arribas, Goodwin, Geddes, Lyons & Saunders (2018)	130	Application of a signature-based machine learning model on daily mood ratings to predict mood and classify individuals into the correct diagnostic group.	Smartphone app to rate their mood once per day on six categories, anxiety, elation, sadness, anger, irritability and energy.	Machine learning program that uses signatures, lend from the "rough part theory", to analyze data of sequential nature.	Smartphones Signature-based learning method.	Mobile application	Data collection lasted > 3 months for each participant	Mainly emotional variability Secondarily evaluation of treatment
6. Minor, Davis, Marggraf,	46	Application of the electronically activated recorder	Up to 12 times a day, five minutes of	Audio files were transcribed and analyzed based	IPod Touch EAR app	Mobile Application	Individuals were randomly recorded and	Understanding of symptoms

Luther & Robbins (2018)		(EAR) to periodically record individuals on the schizophrenia-spectrum to infer their high or low statutes based on the words they used.	audio recordings. Supplemental EMA data were captured, phone calls (five times a day) and a take-home social journal.	on ambient sounds and Linguistic Inquiry Word Count (LIWC) analysis (positive affect, negative affect, word count).	Take-home social journal Phone Calls		called for two consecutive days (waking hours).	
7. Lane, Gates, Pike, Beltz & Wright (2019)	116	Evaluation of the Subgrouping Group Iterative Multiple Model Estimation (S-GIMME) on ambulatory assessment data.	Daily dairies were filled out at the end of each day.	Path structure relations on the individual-, group- and subgroup-level. One-way MANOVA	S-GIMME algorithm Online daily diary:	Daily diary via a secure website	100 consecutive days	Emotional variability
8. Chun, Barrantes-Vidal,	206	Examination of cluster A personality disorder	Participants were instructed to	Hierarchical linear modeling	PDA Mplus6	Handheld computer	7 Days of EMA	Understanding of symptoms

Sheinbaum & Kwapil (2017)		traits. Uncover associations of traits with daily life outcomes.	complete a 32-item questionnaire 8 times a day.	Regression analysis Cross-level interactions Reconfiguration of the cluster A	Regression analyses Cross-level interactions			
9. Shiepek, Aichhorn, Gruber, Strunk, Bachler & Aas (2016)	151	Testing the feasibility of high-frequency monitoring of the psychotherapeutic process.	EMA through the Synergetic Navigation System (SNS). Answering of the Therapy Questionnaire (TPQ) once a day.	Pre-post differences assessed through ICD-10 symptom ratings and questionnaires corresponding to the diagnosis. Subscales of the DASS, ISR total score and	SNS- internet-based feedback system.	Web-based generic system (assessable through regular web-compatible devices, PCs, smartphones etc.)	8 weeks for patients in the day treatment unit of the department of Psychosomatics 12 weeks for patients in the Inpatient Psychotherapy	Understanding of Symptoms

				compliance rate were correlated.				
10. Dotterer, Beltz, Foster, Simms & Wright (2019) (S)	91	Combination of daily assessment and person-specific temporal network models to create individualized symptom-level structures of personality pathology.	Online dairy to measure daily domains of PD.	Path structure relations on the individual-, group- and subgroup-level. Multiple regression Multilevel growth models	GIMMIE-MS Regression Multilevel models	Online daily diary via a secure website	100 consecutive days	Emotional variability
11. Ebner-Priemer et al. (2007) (S)	100	Psychophysiological ambulatory monitoring of BPD patients to assess their affect regulation in the naturalistic environment.	Physical activity and ECG was measured by a portable device. Psychological assessment occurred every	Wilcoxon-Test assessing relative frequencies of emotions. Independent t-test.	MONITOR software Psion 3 a palm top computer Physiological digital recorder Vitaport 2.	Website	24 hours	Evaluation of treatment

			10 to 20 minutes.					
12. Carr et al. (2018)	129	Measurement and comparison of diurnal rhythms of individuals diagnosed with BD, BPD and a healthy control group to uncover potential monitoring and treatment targets.	10 times a day participants had to rate 6 mood items. Heart rate and acceleration were recorded, every five minutes on average.	Principal component analysis Linear interpolation Coherence analysis MESOR	Proteus patch: Heart rate, acceleration (3 dimensions) Mood Zoom (MZ) smartphone application	Online via Smartphone	1 week of high intensity monitoring	Opportunities of new interventions
13. Derks et al. (2019) (S)	5 Patients 4 health care professionals 3 experts	Development of an ambulatory feedback app that aids in emotional awareness and the recognition of emotional arousal.	The Sense-IT app went under three user experience cycles each with its own user group.	UX Questionnaires, interviews and task scenarios Inductive analysis.	Smartphone Smartwatch	In person	1. Cycle: 5 months 2. Cycle: 2 months 3.Cycle: 2 months	Opportunities for new interventions

				Quantitative analyzation of the SUS				
14. Schroeder et al. (2018)	73	Testing the feasibility of the mobile web app Pocket Skills.	Semi-personalized message or a non-personalized message encouraging them to use the app.	The qualitative data provided by open-ended questions was analyses through open coding.	Mobile devices Microsoft Windows Azure (Database)	Web-based for mobile devices	4 weeks	Opportunities for new interventions
15. Rizvi, Hughes & Thomas (2016) (S)	16	Testing the DBT Coach mobile application as a support alongside standard DBT.	Participants were encouraged to use the app freely and to transfer data weekly during skill group sessions.	Hierarchical linear modeling	IOS mobile devices	Mobile application	6 months in combination with standard DBT treating plus a 3-month follow-up period.	Opportunities for new interventions

3.5 Capturing and predicting emotional variability

Starting with the paper by Selby et al. (2020) who utilized a temporal Bayesian network (BN)-modelling approach which is subject to an alternative view about probability, as stated by the authors:" A BN model allows one to graphically represent one's knowledge about a set of uncertain variables" (Selby, 2020, p. 41). Data was gathered on the following eight "class variables" in total: negative emotion (Two-time steps), rumination (two-time steps), dysregulated behaviour (two-time steps), MDD and PTSD. They reported that their tool was able to predict the BPD variable with a general accuracy of 95% (220/232). While predicting the present status of BPD for an individual was only marginally lower with 92% (83/90) was the ability to detect the absence at 96% (137/142), therefore, being a bit more prone to false positives. Albeit, if the status of a patient is known, meaning if the patient is already diagnosed with BPD the model accuracy is overall higher. In this case, the average accuracy rose on all eight class variables to 91% if BPD was present compared to an average of 88% if BPD was not diagnosed. In the diagnosed case were the authors were able to predict the level of rumination with 81-92%, the level of negative emotion with 88-92% and dysregulated behaviours with 95-100% accuracy.

In the first study conducted by Wright et al. (2019) path structures were identified between behavioural variables, domains of personality, stress and functioning. On the group level, the general paths, one effect was strong enough to surpass the group-level cutoff, the path from stress to negative affect. This relation was significant in all but seven cases (92.5%) proposing that stress forecasts negative affect on the same day. Three groups were clustered together based on their shared path analysis, one sample displayed that affiliation regressed on negative affect indicating that some individuals withdraw from others when experiencing NA. It was also possible for them to relate subgroups to constructs like BPD and narcissistic personality disorder to assess the distribution of these personality disorders in the sub-group domain

Lane et al. (2019) applied EMA data into the S-GIMME and related the following symptoms to one another: Liability, impulsivity, anger, anxious, depressed, emptiness and urgency. One of the findings they reported on was the autoregressive effects which when included strengthen the recovery of lagged effects, meaning ambulatory assessment benefits because some processes surface only when the temporal resolution is high enough.

The study about signature-based machine learning by Arribas, Goodwin, Geddes, Lyons & Saunders (2018) asked for one rating per day but employed a lengthier data collection of at least three months per participant with some participants providing data up to twelve months.

One of the goals was to predict the mood of participants which was achieved by using their last twenty observations/days. Healthy individuals' mood scores (anxiety, elation, sadness, anger, irritability & energy) were predicted with an 89-98% accuracy, BPD participants mood scores dropped down to 82-90% and borderline constituted the bottom with 70-78%. Selby's et al. predictions of rumination and negative emotions are, therefore, rather close to the Arribas et al. prediction accuracy of mood in BPD patients being both in the range of ~80-90%. The same can be said about their ability to detect disorders, their model was able to classify with 93% accuracy BPD and HC participants. The obtained precision for BD and HC turned out to be lower with 84%. The general accuracy of the BPD variable (regardless of the status) with 95% is again close to Arribas et al. obtained result. However, the signature-based model performed significantly worse in the BD category which dragged the overall accuracy score down to 75%.

Concerning the personalized models of personality following results have been published by Dotterer et al. (2019) regarding the association between personality networks and symptomatology. The average explainable variance (R2) for models calculating each symptom count was reported as following: schizotypal = 0.12, obsessive-compulsive = 0.09, borderline = 0.17, avoidant = 0.24, narcissistic = 0.13 and paranoid = 0.30. Uniquely predictions of symptomatology were reported as following: paranoid predicted by lower within-internalizing degree, borderline by disinhibition autoregression, narcissistic by the degree of internalizing and externalizing domains and obsessive-compulsive by disinhibition plus hostility auto-regression in addition to the within-externalizing degree. When covariates were included in the prediction of symptomatology, the domain-level influence was nearly always reduced to not significant apart from two exceptions. Avoidant symptoms showed an association with levels of negative affect in a majority of the models ($r(91) = 0.32$; $p < 0.01$). The same was true for age which depicted significant associations with borderline ($r(91) = -0.23$; $p < 0.05$) avoidant ($r(91) = -0.31$; $p < 0.05$) and paranoid symptoms ($r(91) = -0.26$; $p < 0.05$).

3.6 Improving the understanding of symptoms

The Personality Dynamics Diary (PDD) was conceived according to the authors to incorporate:" […] the benefits of basic personality's dimensional approach into the assessment of personality pathology […] and (potentially) harmonizes the within- and between-person factor structures that remain confounded in traditional cross-sectional research designs" (Zimmermann et al., 2019, p.518). Based on 2,452 diary entries from 314 participants and a response rate of 87.2% was a questionnaire conceived which incorporated daily situation

characteristics (Positive events, social stress & workload) and daily behaviours (sociability, attention-seeking, externalizing symptoms, internalizing symptoms, daydreaming & perfectionism). Resulting reliabilities were tolerable with M = 0.67 (0.40 – 0.81) for within-person difference and M = 0.87 (0.71 – 0.97) for between-person differences. The constructed questionnaire was tested in an inpatient setting with 77 individuals, resulting in 3,079 diaries with a response rate of 51.4% over nine months. Similar reliabilities were found like in study 1, M = 0.73 (0.59 -0.82) for within-person differences and M = 0.88 (0.79 – 0.95) for between-person differences. Ratings (0-100) of the participants indicated no difficulties in using the PDD (M = 85.1, SD = 22.5), also in the digital form (M = 8.6), SD = 21.0) and moderate ratings in the comprehensiveness of the assessed aspects of daily life (M = 51.1, SD = 19.7). However, therapist differed in their opinions about how useful the PDD would be in the practice, new information for next sessions were barely obtained for them (M = 29.8, SD = 28.7) and new insights were also rather sparse according to them (M = 46.2, SD = 25.7).

Minor et al. (2018) combined several EMA techniques, namely the EAR, phone calls and take-home social journals, to explore affective states and corresponding behaviours. The phone calls and journals showed significant overlap in their measured constructs, $p(188) = 0.49$, $p < 0.001$ for positive affect and $p(188) = 0.60$, $p < 0.001$ regarding negative affect. The EAR turned out to be a measuring instrument that gathered data that only had a little overlap when compared to the other EMA measurements implying that EAR is assessing data that is inaccessible through conventional EMA methods. Results indicate that more negative affect affective words were used from individuals high in schizotypy, $y = -0.46$, $SE = 0.24$, $p = 0.035$, while they also display less social engagement, $y = 29.58$, $SE = 13.05$, $p = 0.014$, when compared to the low schizotypal group. Similar differences were shown in the other EMA methods, therefore, supporting the EAR results. Specific schizotypy traits were also captured to be predictive of negative affect, especially positive schizotypy traits, $y = 0.25$, $SE = 0.11$, $p = 0.02$ as measured by the EAR, while negative traits showed only a non-significant relation in the EAR case but the other EMA methods. Additionally, they were not able to sport adherence difference between the high and low group, participants wore the EAR device on average > 90% of the time while deleting around 1% of their recordings.

Shiepek et al. (2016) tested the feasibility of high-frequency equidistant time sampling monitoring in a day treatment and inpatient psychotherapy unit. The Therapy Process Questionnaire (TPQ) which consisted out of 42 items was given to participants and was expected to be answered daily. The length of this EMA experiment amounted to eight weeks for day treatment and twelve weeks for inpatient psychotherapy patients. Despite the relatively

high demand was the mean compliance rate at a decent 78.3% (SD = 26.0; median = 89.4%). Nearly half of them (49.3%) managed to complete >90% of the questionnaires and 25 patients (16.4%) were able to achieve a compliance rate of 100%. From the individuals who reached treatment completeness of >90% were only 32 (21.1%) who missed less than 80% of the daily questionnaires. Therefore, successful treatment completers were generally more likely to comply with the monitoring. Additionally, a positive correlation of 0.231 ($p < 0.004$) was also found between the length of the visit and the compliance. The authors also tested the relationship between diagnosis and compliance with expected results. Patients suffering from mood affective disorders had a mean compliance rate of 83.2%, while neurotic, stress-related and somatoform disorders averaged out at 81.1%. Disorders of adult personality and behaviour formed the bottom line with an average compliance rating of 66.1%. Severity was defined by the DASS (depression, anxiety, and stress), the Pearson correlation coefficients of the compliance rate indicated no significant relationship and, therefore, no correlation between the three subscales and the corresponding severity. Individuals diagnosed with a personality disorder were displaying less compliance and marginally higher, but still significant, rates of missing data 16.3% ($p = 0.051$).

The results regarding the reconfiguration of PD traits by Chun et al. (2017) indicated good interrater reliability (Cohens' k = 0.90; range: 0.79 to 0.95). The expression of schizophrenia-spectrum personality traits was split into positive, negative, disorganized and paranoid schizotypy based on their simultaneous regression analysis yielding the following significant results. Positive schizotypy was significantly associated with negative affect ($r = -0.103$; $SE = 0.052$; $p < 0.05$), difficulty coping ($r = -.165$; $SE = 0.07$; $p < 0.05$), and doing ones current activity ($r = -0.127$; $SE = 0.063$; $p < 0.05$). Negative schizotypy with no thoughts or emotions ($r = 0.273$; $SE = 0.101$, $p < 0.05$), difficulty coping ($r = -0.080$; $SE = 0.032$; $p < 0.05$), disliking ($r = -0.092$, $SE = 0.028$; $p < 0.01$) and having difficulty doing ones activity ($r = -0.094$; $SE = 0.038$; $p < 0.05$), as well as low positive situation ($r = 0.035$; $SE = -0.205$; $p < 0.220$). The disorganized type only with not feeling well in the moment ($r = 0.307$; $SE = 0.144$; $p < 0.01$). The Paranoid schizotypy was associated with feeling tired ($r = 0.065$; $SE = 0.032$; $p < 0.05$), unwell ($r = 0.064$; $SE = 0.026$; $p < 0.05$), in addition to paranoid ($r = 0.036$; $SE = 0.014$, $p < 0.05$) and psychotic-like experiences indices ($r = 0.013$; $SE = 0.006$; $p < 0.05$).

While the goal of Arribas et al. (2018) was not to study the symptoms of BPD and BD per se, obtained results could prove themselves useful in the differentiation between them. The employed signature-based machine learning model was able to categorize 74.85% (SD = 2.05) of partaking individuals into the correct diagnostic group based on a total of 20 consecutive

self-reports. In the case of mood prediction was the model was able to forecast the mood of the healthy control group with an accuracy of 89-98%. The accuracy concerning bipolar participants was 82-90% while it declined to 70-79% for borderline patients.

3.7 Evaluation of treatments

The psychophysiological ambulatory assessment study by Ebner-Priemer et al. (2007) was conducted with BPD patients to determine their affect regulation via high frequency psychological assessment of emotions (every 10-20 minutes) and the measuring of the heart rate of participants. Results indicated that medication influenced the heart rate of individuals, the comparison between BPD patients with and without medication showed significant variations on heart rate, additional heart rate, heart rate baseline and HF-HRV. They found a significant group difference ($p = 0.023$) in aHr between healthy participants, and BPD patients without medication. Age-matching the different groups revealed again a significant difference in aHr ($t = 2.21$, $df = 18$, $p = 0.040$).

3.8 Opportunities for new interventions

The pocket skills application developed by Schroeder et al. (2018) for BPD patients as an adjunctive treatment indicated promising results. Generally, the app was easy to handle as participants rated the app high on the SUS ($M = 85$, $SD = 16.01$). The unconditional growth curve models for PHQ-9 ($B= -0.79$, $p<0.001$; $B= 0.45$, $p<0.001$) and the OASIS ($B=-0.66$, $p<0.001$; $B= 0.13$, $p<0.05$) both indicated that participants improved over the course of the treatment. Scores on the PHQ-9 decreased with a rate of 0.79 each week, which slowed down to 0.45, the OASIS showed a similar pattern which started with a decrement of 0.66 and regressed to 0.13 each week. The intake and exit survey (DBT Ways of Coping Checklist) revealed significant results in the decrease of dysfunctional coping ($B= -0.06$, $p >0.001$), blaming others ($B= -0.04$, $p<0.001$) and a significant increase in their usage of DBT related skills ($B= 0.04$, $p<0.001$). According to the OASIS, 41.7% of the individuals were classified as recovered, while the PHQ-9 ranked 22.2% of participants as recovered. 8.3% were seen as improved according to the OASIS and only 4.2% with regards to the PGQ-9. The DBT skill use classified more individuals into the improved category with a percentage of 12.5% than to recover with 6.9%. Additionally, participants had a higher chance to recover when they were allocated to the semi-personalized message group, for PHQ-9 (27.0% vs 17.1%) and the OASIS

(45.9% vs 37.1%). However, the opposite was true for the DBT skill use which indicated (8.6% vs 5.4%).

Another skill application for BPD patients was conceived by Rizvi et al. (2016), named the DBT Coach. In comparison to the prior study, had this one a high dropout rate of 25% (4). Despite these dropouts, participants rated the app relatively high with 90% of them reporting that they would use the app outside of the study condition. However, frequency of use varied considerably, from two to 107 (Mdn = 11.5) over nine months, starting with 187 sessions at the beginning which at the end decreased to 70 sessions. Rate of use was only positively correlated to frequency of lifetime NSSI ($r(16) = 0.506$, $p < 0.05$) and the occurrence of NSSI episodes in the last year ($r(16) = 0.645$, $p < 0.01$). The same was true for the changes in NSSI which predicted with greater use of the DBT Coach application a greater decrease ($B = -2.49$, $SE = 0.56$, $t(38) = -4.43$, $p < 0.001$). These discoveries are further supported by their findings that distress was significantly reduced at the end of sessions; pre-Coach $M = 4.62$ ($SD = 2.67$), post Coach $M = 3.14$ ($SD = 2.19$), $B = -1.49$, $SE = 0.16$, $t(721) = -9.02$, $p < 0.001$. The same is true regarding the urges to self-harm; pre-Coach ($M = 1.75$, $SD = 2.67$); post-Coach ($M = 1.13$, $SD = 1.97$), ($B = -0.61$, $SE = 0.15$), $t(721) = -4.10$, $p < 0.001$. Overall sessions led to a decrease in distress 68% of the time, while 5% of the sessions led to an increase in distress. In terms of self-harm were 30% of the sessions were successful in reducing the urges, while 3% led to an increase.

Sense-IT, a mobile ambulatory biofeedback application designed for BPD patients invented by Derks et al. (2019) underwent three testing cycles to test initial usability. The first iteration showed that generally, participants rated the app as good ($M = 78.8$), they further mentioned the simplicity of use, likeable design and reported no problems wearing the accompanied smartwatch which measured physiological arousal. Regarding the received feedback, patients gave mixed reactions it was either too many or not enough. Also, the reliability suffered a bit because of that, rated as average to slightly unreliable. It appeared that the feedback lagged. Regardless, the app was overall rated as promising and the biosensor component was welcomed. According to usability experts, patients and therapists the app does hold potential but should strive for continuous improvement with regards to the usability of this application.

The study conducted by Carr et al. (2018) detected de-synchronization of diurnal function in BPD and BD patients but not in the healthy control group. In the case of BPD diagnosed individuals, a two-hour delay was present between activity and sleep, which further increased to three hours when compared to peak activity and peak HR. Furthermore, the general

level of HR was substantially higher compared to HC and moderately when compared to BD. At higher measuring frequencies a correlation between negative mood and activity became apparent in the BD and BPD group when compared to HC. HR was also greatly correlated to the irritable mood in BD and BPD in comparison to HC.

4. Discussion

This review was dedicated on exploring several study characteristics lend from the article of Myin-Germeys et al. (2018) which depicts how EMA can benefit numerous aspects of mental health research. The article itself inspired this approach of research, providing distinct key points on how this method contributes to this field by "[…] improving assessment of psychopathology and its underlying mechanisms" (Myin-Germeys et al., 2018, p. 123), and how EMA contributes to "[…] advancing and changing clinical practice by allowing a more fine-grained evaluation of treatment effects as well as by providing the opportunity for extending treatment beyond the clinical setting into real life with the development of ecological momentary interventions" (Myin-Germeys et al., 2018, p. 123). These themes were conceptualized in headings and reused as the four main research questions of this paper because they deemed themselves encompassing of all relevant points with regard to the EMA methods. Additionally, research questions were formulated of descriptive nature which served as a quick overview, and which will be discussed first. These questions were answered in detail in the results and will only be briefly summarized to synthesize the key points.

4.1 Which personality disorders were predominantly present in the studies?

Every PD currently listed in the DSM-5 were present in the study and are here listed as following: paranoid, schizotypal, schizoid, borderline, narcissistic, histrionic, antisocial, avoidant, dependent and obsessive-compulsive. As seen in the "secondary findings: participants characteristics" section, primarily PDs in Cluster B were present, consisting principally out of borderline PD and to a lesser extend out of narcissistic PD. The second most represented PD Cluster was Cluster A which mainly consisted out of individuals suffering under schizotypal PD. Therefore, Cluster C was the section least represented, exhibiting isolated cases of avoidant and obsessive-compulsive PDs.

4.2 What are the qualities of employed techniques?

Assessment oriented papers reported little to no problems related to a technical level, the use of gadgets appeared to be unproblematic and reliable. The paper assigned to the intervention category appeared to garner adverse reactions, generally, the technological aspects seemed to be approved of while the inherent usefulness of some applications were commonly scrutinized by practitioners because they could not discover new insights which could aid the intervention. The applications adjunctive to the main treatments were met with a lot of positivity from the patients themselves showing quantitative measurable improvements on e.g. skill use and remarking positive points on a qualitative level by praising certain aspects such as easy to understand instructions or a conversational agent which helped to increase the perceived credibility of the application.

The large proportion of borderline PD could be explained by the high prevalence of this disorder especially in mental health clinics, 10%, and psychiatric facilities, 20% (DSM-5, 2013, p. 665). Included papers also frequently reported drawing samples from these establishments, therefore, at least in the papers who purposefully sampled with greater variability, overrepresenting this personality disorder. Nevertheless, most included studies worked selectively with such individuals, developing interventions or screening processes especially because borderline PD is so widely represented.

However, based on the relatively large proportion of borderline PD general remarks about the whole field of personality disorders are harder to make. The essence appears to be that especially individuals suffering under borderline personality disorder are less easy to handle in the assessment and intervention process than any other individuals suffering under other PDs. One could argue that if these methods are showing at least some initial feasibility with a notoriously complicated disorder to treat, then one could generalize this success to other PDs. The relevance of such research is undeniable important, the high comorbidity to other disorders and PDs justifies the focus of the scientific community to develop new means of treatment and assessment.

4.3 To what extent EMA methods can capture and predict emotional variability?

As stated in the results section this scoping review incorporated five studies (Selby et al., 2020; Wright et al., 2019; Arribas et al., 2018; Lane et al., 2019; Dotterer et al., 2019) that predominantly focused on establishing means on how to capture and predict emotional variability with high accuracy. Selby et al. (2020) were able to reach accuracies as high as 96%

in terms of categorization of diagnostic groups, the study by Wright et al. (2019) showed how individuals with similar sub-group relations and, therefore, likely sharing the same diagnosis differ, this is generally important information and notifies the practitioner not only about the variability emotions but also how the individual will react to it, to the degree that the study by Dotterer et al. (2019) was able to predict the rates of change with regards to daily stress. While these methods were able to capture the intensity of emotions, moment-to-moment fluctuations, no advances were made to assess the normative state and therefore missed one point Myin-Germeys (2018) established in her paper which could potentially assess the emotional regulation strategies of patients.

4.4 To what extent EMA methods are able to improve the understanding of symptoms?

EMA methods are also suitable to aid the understanding of symptoms in the field of personality disorders, four papers in this review delivered conclusive evidence for this application. The EAR device (Minor et al., 2018) gave insight into social impairments, bearing high clinical value because they potentially indicate the beginning of a psychosis which is characterized by the declining ability to properly communicate and diminishing social connections, while at the same time attesting that EAR and other supplemental EMA data showed little overlap, meaning different constructs of the patient's daily lives were assessed in their study through different means.

The study conducted by Chun et al. (2018) further elaborated on Cluster A traits and their differences in symptoms and impairments, rearranging them Cluster into positive, negative, disorganized and paranoid schizotypy dimensions which ensued greater independence amongst them and support prior EMA findings (Barrantes-Vidal et al., 2013). In addition to these disorders, specific findings were Shiepek et al. (2016) able to show that EMA methods are suitable to every kind of disease pattern regardless of their restrictions, showing the general suitability of these techniques to study personality disorders. However, the PDD (Zimmermann, 2019) did not contribute much to the understanding of symptoms, especially therapists were reluctant to integrate it into their interventions based on the limited gain which seems fitting to the rather low reliability of the questionnaire.

4.5 To what extent EMA methods are helpful in the evaluation of treatments?

The evaluation of treatments was not the main objective in any of the incorporated studies, however, some clinical implications were formed as a byproduct apart from the assessment of PDs. The psychophysiological ambulatory assessment of affective dysregulation (Ebner-Priemer, 2007) primarily focused on trying to infer the diagnostic status based on their heart rate and self-reported emotions. However, continuous assessment has the potential to evaluate the treatment outcome by assessing emotional events, appraisal, the resulting behaviour and the emotional physiological response. The two different statistical methods, TBN by Selby et al. (2020) and the signature-based machine learning model by Arribas et al. (2018) were also conceived as a diagnostic tool in but like in the case of Ebern-Priemer et al. (2007), these methods are also suited for feedback process to evaluate the intervention developments from patients in treatment because they operate quite similarly. In the same way, as mentioned in the article of Myin-Germeys (2018) EMA measures were much more sensitive to change and reached beyond the means of conventional assessment, such as in the case of dose-response effects (Ebner-Priemer, 2007). The assessment of emotions and heart rate indicated significant differences on almost all physiological parameters between non-medicated and medicated patients, having the potential to not only yield psychological feedback on treatments but also physiological data which has the potential to aid the evaluation process.

4.6 To what extent EMA methods open up new opportunities for new interventions?

Generally, EMA methods were successfully utilized to support individuals in their acquisition of new skills and their translation into their daily life by sending a reminder to practice skills or make them aware of emotions which is a key point in BPD interventions. EMAs allowed Derks et al. (2018) to pursue a simple novel idea by combining physiological monitoring and skill training by supporting patients in detecting and regulating their emotions which does not solely rely on “talking cure”. In the case of Pocket skills (Schroeder et al., 2018) participants made substantial progress, they reported that the app supported them in learning the principles, practice them and subsequently implement them in real life. The app itself was perceived as credible which can be greatly attributed to the conversational agent “eMarsha”. Results from Rizvi et al. (2016) further underlined the feasibility and clinical utility of mobile interventions, while it was not clear to what degree the app was responsible for this reduction, the frequency of use was related to a greater decrease in NSSI episodes. As well as in the other intervention studies, the authors endorsed the generalizability of mobile studies to reach a larger

audience which is given through the ubiquity of smartphones and their widespread.

The treatise by Eaton, South & Kruger (2009) did not carry out an EMI but explained general points applicable to technology that is conceived to support PD patients. They argued, that there are potential benefits in exploring the situational factors to analyze the patient's cognitive-affective units (CAUs) and understand how the same person reacts to different situations. Knowing the dynamics between feelings/thoughts and their resulting behaviour which is tied closely to explicit situations will help to better understand the diversities of personality pathology. This approach also bears some resemblance to the GIMME algorithms by grouping individuals based on their individual's relations of traits instead of forming an overarching category of pathology. By utilizing two means of investigation, daily diary and behavioural challenge tasks, in their ecological valid environment could the therapist increase the insight from their patients via addressing their metacognition.

4.7 Links to previous research

A previous prevalence study of personality disorders in the general adult population in western countries (Volkert, Gablonski & Rabung, 2018) established results that were not in accordance with the distribution of PDs in the scoping review. The prevalence rate of Cluster A was according to their results the highest at 7.23% (95% Cl, 2.37-14.2%), Cluster B at 5.53% (95% Cl, 3.20-8.43) and Cluster C at 6.70% (95% Cl, 2.90-11.93). Furthermore, the prevalence for BPD was only the third-highest in Cluster B estimated at 1.90% (95% Cl, 0.85-3.34) which is not supported by the findings of this review. One feasible explanation could be the easy technological translation of DBT components, skill training and education may be better suited for conversion into the mobile format which allows for on-demand sessions in contrast to other interventions (Rizvi, 2016).

Whereas the establishment of links to every kind of employed method would certainly overextend the aim of this section, therefore, studies were examined which were able to draw up generalized statements about the feasibility of EMA methods to cover most of the results. Despite having mentioned in previous research that once-a-day diary entries may be insufficient to capture unstable symptoms (Ebner-Priemer & Sawitzki, 2007), such as shifts in mood, predictions of mood and diagnostic categories were established with only one daily entry a day, meaning minimal burden for the patient, in every study which utilized GIMME algorithms (Wright et al., 2019; Lane et al., 2019; Dotterer et al., 2019). Although having in every one of these studies a wide array of PD present and only one assessment point a day, results were

surprisingly accurate which can be attributed to two factors. First, the advanced method of computing that allowed to establish reliable results and, second, the extensive assessment period of 100 days, surpassing the average duration of 32 days and median duration of 14 days of EMA studies (Berkel, Ferreira & Kostakos, 2017). Looking at the number of misdiagnoses should a need for these kinds of methods be apparent. It appears that 40% of individuals diagnosed with BPD have been misdiagnosed with BD in comparison to the rate of 10% regarding patients who were not previously diagnosed with BPD (Ruggero, Zimmerman, Chelminski & Young, 2010).

4.8 Limitations and strengths

There are multiple points of concern regarding this review. The foremost important limitation was that the search and synthesizes was done by only one person. Several problems can arise because of that, for instance a selection bias could be apparent in the choice of the incorporated papers. Studies which contributed through novel ideas and technology were treated as more important than studies with designs who were more common. Studies which did not necessarily suited the author himself could have been more easily excluded based on personal preferences, how one is able to work with the presented data, and not as much based on the exclusion criteria. In addition to that were only two databases selected which could have contributed to a more narrow selection.

Another point are the very broadly defined inclusion and exclusion criteria which caused an inconsistent picture and impeded the formulation of uniformly answers in the discussion section. By incorporating multiple comprehensive research questions was the textual limit of this review exceeded and decisions had to be made on the topic of quality versus quantity. Ultimately, far reach was prioritized to underline the nature of a scoping review.

As mentioned beforehand were no RCTs were utilized in the intervention studies, therefore, limiting the ability to make concrete statements about the quality of the adjunctive treatments. Because of that results must be considered with care.

At last, the novelty of a great number of included papers limited the ability to make interferences with prior research, especially with regards to the assessment studies. As an example, the search for assessment oriented research already encompassed all easily available GIMME studies that utilized this approach to establish diagnostic interferences in PDs patients. While these limitations were caused by self-imposed guidelines future research could addressee these and take a more in-depth approach to appeal to only one research question that was formulated in this review.

Regarding the strengths, this paper mostly contributes by presenting a review which has not been done prior. Laying a foundation and a starting point for everyone who is interested in momentary ecological measurements with the latest technological means. As seen in table three, most of the papers were at the point of writing relatively current studies and therefore representative of the state of the art of EMA and EMI in the field of PD which was one of the key points in this study. At the same time, some limitations point out what steps have to be taken in the future to contribute to the improvement of such studies and to the adoption in general practice which is at least on a technological level feasible because most tools used are widely available.

4.9 Directions for future research

Apart from the obvious suggestion to implement RCTs, future research should consider including a greater variety of PDs into their research. As seen in this review mostly BPD was present in the samples, exclusively in the intervention studies and moderately represented in the sample of assessment studies. Of course, one must take into account how well intervention modules can be translated into the technological and naturalistic environment of the patients, for each PD these vary of course. Nevertheless, the widely applicable cognitive behavioural therapy is known to be viably translated into mental health smartphone apps, therefore, having the potential to be also tested in the framework of this scoping review (Bakker, Kazantzis, Rickwood & Rickard, 2016). Hence, future research has the potential to yield more evidence about the technological feasibility for lesser represented personality disorders as it was reported that severity and the diagnostic category of PD does not prevent engagement in such apps.

Another viable point for future research would be to uncover which incentives are needed for the faster adoption of included methods and techniques. Some studies quoted statements from practitioners which critiqued the relatively little gain from the implementation. This is especially important for the assessment studies, in this case little is known regarding the incentives and barriers for their implementation into widespread practice. Future studies could employ qualitative measures to assess the feasibility for everyday use, otherwise, adoption could be lessened or halted despite promising results. Are these methods suited for everyday practice? Are these methods too laborious and time consuming for practitioners? These questions have to be answered before one can speak about the future of assessments and interventions in general practice.

4.10 Conclusion

Although a lot of questions are still unanswered with regards to the feasibility of implementation, the current review illustrated the vast applicability of EMA and EMI in the field of PDs. Be it the capturing of emotional variability, understanding the relationship between symptoms or developing new kinds of interventions and their evaluation are all realizable. Ecological momentary techniques can be used for various means and the creativity of researchers seems to be the only limit.

5. References

American Psychology Association (2013). Mental Illness. Dictionary APA, https://dictionary-apa-org.ezproxy2.utwente.nl/mental-disorder

American Psychology Association (2013). Personality Disorder. Dictionary APA, https://dictionary-apa-org.ezproxy2.utwente.nl/personality-disorder

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

Arribas, I. P., Goodwin, G. M., Geddes, J. R., Lyons, T., & Saunders, K. E. (2018). A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational psychiatry*, *8*(1), 1-7.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR mental health*, *3*(1), e4984.

Carr, O., Saunders, K. E., Bilderbeck, A. C., Tsanas, A., Palmius, N., Geddes, J. R., ... & Goodwin, G. M. (2018). Desynchronization of diurnal rhythms in bipolar disorder and borderline personality disorder. *Translational psychiatry*, *8*(1), 1-9.

Chun, C. A., Barrantes-Vidal, N., Sheinbaum, T., & Kwapil, T. R. (2017). Expression of schizophrenia-spectrum personality traits in daily life. *Personality disorders: Theory, research, and treatment*, *8*(1), 64.

Cloninger, C. R., Svrakic, D. M., Lester, N. C., Lecic-Tosevski, D., Koldobsky, N., & Botbol, M. (2017). 31 Personality Disorders. *Person Centered Psychiatry*, 427.

Conner, T., & Barrett, L. F. (2012). Trends in ambulatory self-report: Understanding the utility of momentary experiences, memories, and beliefs. *Psychosomatic Medicine*, *74*, 327-337.

Davey, G. (2008). Psychopathology: Research, assessment and treatment in clinical psychology. British Psychological Society.

Derks, Y. P., Klaassen, R., Westerhof, G. J., Bohlmeijer, E. T., & Noordzij, M. L. (2019). Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: Multicycle usability testing study. *JMIR mHealth and uHealth*, *7*(10), e13479.

Dotterer, H. L., Beltz, A. M., Foster, K. T., Simms, L. J., & Wright, A. G. (2020). Personalized models of personality disorders: Using a temporal network method to understand symptomatology and daily functioning in a clinical sample. *Psychological medicine*, *50*(14), 2397-2405.

Eaton, N. R., South, S. C., & Krueger, R. F. (2009). The Cognitive–Affective Processing System (CAPS) approach to personality and the concept of personality disorder: Integrating clinical and social-cognitive research. *Journal of Research in Personality*, *43*(2), 208-217.

Ebner-Priemer, U. W., & Sawitzki, G. (2007). Ambulatory assessment of affective instability in borderline personality disorder. *European Journal of Psychological Assessment*, *23*(4), 238-247.

Ebner-Priemer, U. W., Welch, S. S., Grossman, P., Reisch, T., Linehan, M. M., & Bohus, M. (2007). Psychophysiological ambulatory assessment of affective dysregulation in borderline personality disorder. *Psychiatry research*, *150*(3), 265-275.

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, *118*(1), 177-214.

Gustafson, D. H., McTavish, F. M., Chih, M. Y., Atwood, A. K., Johnson, R. A., Boyle, M. G., ... & Shah, D. (2014). A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA psychiatry*, *71*(5), 566-572.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology*, *15*(1), 1-39.

Kazantzis, N., & L'Abate, L. (2007). *Handbook of homework assignments in psychotherapy*. Springer Science+ Business Media, LLC.

Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological methods*, *24*(1), 54.

McMurran, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, *30*(3), 277-287.

Mezzich, J. E., Botbol, M., Christodoulou, G. N., Cloninger, C. R., & Salloum, I. M. (2016). Introduction to person-centered psychiatry. In *Person centered psychiatry* (pp. 1-15). Springer, Cham.

Minor, K. S., Davis, B. J., Marggraf, M. P., Luther, L., & Robbins, M. L. (2018). Words matter: Implementing the electronically activated recorder in schizotypy. *Personality Disorders: Theory, Research, and Treatment*, *9*(2), 133.

Morgenstern, J., Kuerbis, A., & Muench, F. (2014). Ecological momentary assessment and alcohol use disorder treatment. *Alcohol research: current reviews*, *36*(1), 101.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*, *17*(2), 123-132.

Rizvi, S. L., Hughes, C. D., & Thomas, M. C. (2016). The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: A preliminary evaluation and challenges to client utilization. *Psychological services*, *13*(4), 380.

Ruggero, C. J., Zimmerman, M., Chelminski, I., & Young, D. (2010). Borderline personality disorder and the misdiagnosis of bipolar disorder. *Journal of psychiatric research*, *44*(6), 405-408.

Schiepek, G., Aichhorn, W., Gruber, M., Strunk, G., Bachler, E., & Aas, B. (2016). Real-time monitoring of psychotherapeutic processes: concept and compliance. *Frontiers in psychology*, *7*, 604.

Schroeder, J., Wilkes, C., Rowan, K., Toledo, A., Paradiso, A., Czerwinski, M., ... & Linehan, M. M. (2018, April). Pocket skills: A conversational mobile web app to support dialectical behavioral therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

Selby, E. A., Kondratyuk, S., Lindqvist, J., Fehling, K., & Kranzler, A. (2020). Temporal Bayesian Network modeling approach to evaluating the emotional cascade model of borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*.

Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *The Lancet*, *385*(9969), 717-726.

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)*, *50*(6), 1-40.

Versluis, A., Verkuil, B., Lane, R. D., Hagemann, D., Thayer, J. F., & Brosschot, J. F. (2021). Ecological momentary assessment of emotional awareness: Preliminary evaluation of psychometric properties. *Current Psychology*, *40*(3), 1402-1410.

Volkert, J., Gablonski, T. C., & Rabung, S. (2018). Prevalence of personality disorders in the general adult population in Western countries: systematic review and meta-analysis. *The British Journal of Psychiatry*, *213*(6), 709-715.

Wenze, S. J., Armey, M. F., Weinstock, L. M., Gaudiano, B. A., & Miller, I. W. (2016). An open trial of a smartphone-assisted, adjunctive intervention to improve treatment adherence in bipolar disorder. *Journal of psychiatric practice*, *22*(6), 492.

Wright, A. G., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology*, *125*(5), 641.

Wright, A. G., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*, *31*(4), 502.

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, *162*(10), 1911-1918.

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., ... & Wright, A. G. (2019). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a personality dynamics diary. *Psychological assessment*, *31*(4), 516.