



Gaining and Visualizing Mental Health Insights from Self-Report Data
Presentation of Insights from ESM Data into Client Conditions for Practitioners

Kasper van Maasdam¹

Supervisor(s): dr.ir. Willem-Paul Brinkman¹, drs. Esra Cemra Su de Groot¹

¹EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Kasper van Maasdam

Final project course: CSE3000 Research Project

Thesis committee: dr.ir. Willem-Paul Brinkman, drs. Esra Cemra Su de Groot, prof.dr.ir Inald Lagendijk

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Abstract

ESM is an important step towards improving mental health care and its efficiency. Most research in this field has focused on the client as its end user. However, mental health practitioners can also use the data gathered using Experience Sampling Method (ESM) to gain insights into their clients. To discover what methods of visualization practitioners find most insightful and intuitive to identify mental health conditions, and why, a user evaluation containing mock-ups based on existing literature has been performed. In total, 8 people participated in the study, 6 of which were psychology students, 1 was a psychology researcher and 1 was a mental health practitioner. The user evaluation has shown that the use of spider plots comparing the average and variability of the ESM data of a client to that of a cohort with a certain mental health condition is an intuitive visualization method to identify mental health conditions in clients.

1 Introduction

Mental health is a topic of increasing importance, as mental health disorders are increasing in prevalence, especially in adolescents [Ten Have et al., 2023; Netherlands, 2022]. This creates increasing pressure on mental health institutions [OECD and European Observatory on Health Systems and Policies, 2023]. One proposed method of cost-effectively alleviating the pressure on mental health care professionals is by using mobile health (mHealth) to augment treatment [Khosravi and Azar, 2024]. mHealth is a scalable and accessible method of achieving health objectives by using mobile devices [Price et al., 2014]. However, care must be taken to ensure such applications have a positive affect on both the practitioner and the patient [Stefancic et al., 2022].

mHealth tools focusing on mental health already exist. Some provide interventions to those with mental health problems or those at risk, for example Grow It! [Dietvorst et al., 2022]. Others focus on complementing clinical treatment by providing a better picture of the client to the practitioner, for example Therap-i [Riese et al., 2021]. And some tools allow to be used for both purposes, for example PsyMate [Myin-Germeys et al., 2011]. A common element in the mHealth tools described is that they measure individual differences in emotional dynamics using the Experience Sampling Method (ESM).

ESM is a momentary assessment technique that collects contextual mental health data by prompting participants with a short questionnaire multiple times a day at (semi-)random intervals [Larson and Csikszentmihalyi, 2014]. This is often done intensively over a period of a few days, but longer periods have also been studied [Van Os et al., 2017; Bringmann et al., 2021]. ESM is a promising tool for use in e.g., psychology and psychiatry [Van Os et al., 2017]. The aim of ESM is partly to prevent biases associated with retrospective assessments in clinical practice [Van Roekel et al., 2019] and partly to capture the temporal and context-dependent dynamics of mental state [Van Os et al., 2017].

Tools that make use of calculated collection aim to provide insights into the user's mental well-being. This can be used by the user's own initiative, or this can be requested by a practitioner before, during or after therapy [Van Os et al., 2017; Dietvorst et al., 2022]. As described by Riese et al. [2021], there is a research gap in the use of ESM-monitoring tools to assist treatment in clinical practice. Most mHealth apps in use today are designed to assist and provide information to the user filling in the surveys only [Anjeer Alshehhi et al., 2023]. However, one can wonder: How can applications that collect ESM data, be used to benefit psychologists or psychiatrists?

Existing research has investigated the use of ESM data in clinical practice, revealing the idea that ESM could be useful in at least some degree in all phases of care [Bos et al., 2019]. Despite its perceived usefulness by practitioners, there is no consensus on how it is best used and instead, focus on specific application purposes is warranted [Bos et al., 2019]. For instance, practitioners agree that the way to visualize ESM data would depend on the specific clinical application [Weermeijer et al., 2024].

This research provides feedback from practitioners on the presentation of specific insights that can be derived from the ESM data. This is achieved by performing a qualitative questionnaire-based user evaluation with mock-ups of example visualizations. A thematic analysis is then performed on the responses. The specific application considered in this research is the presentation of meaningful insights from ESM data to identify mental health conditions. For example, the insights should be able to inform a practitioner how a person relates to other people with a certain mental health condition. This application is considered because this is where a knowledge gap resides. Not much research into the clinical applications of ESM data exist [Riese et al., 2021] and those that do exist focus on the use of an individual's data [Riese et al., 2021; Kroeze et al., 2017]. However, ESM is a method that allows collecting the same type of data from many individuals [Van Os et al., 2017]. Comparing one such individual to a cohort, might provide insights that prove useful to practitioners in practice..

Previous literature has performed user evaluations on the visualization of ESM data both in a more general sense and for other specific insights [Bringmann et al., 2021; Bell et al., 2018; Kroeze et al., 2017; Daniëls et al., 2019]. Kaiser and Laireiter [2018] for example, who visualize process-symptom-bridges using an idiographic network, or Weermeijer et al. [2023], who perform a pilot evaluation study on the general use of ESM in specialized mental health care. This research builds upon the established techniques and further explores their usefulness in identifying mental health conditions.

There is also literature not specific to visualizations of ESM data, but techniques to visualize comparisons of other health data between patients [Schaaf et al., 2021; Kreienkamp et al., 2025]. For example, Chari et al. [2019] created a cohort analysis browser that visualizes the comparison between a patient and a cohort. This is a source of inspiration for this research, because these papers provide a baseline for how to measure and visualize a comparison between clients. This research extends it to mental health data gathered by ESM.

The following research question has been explored: What methods of visualization do practitioners find most insightful and intuitive to identify mental health conditions, and why? To answer this, apart from the user evaluation and thematic analysis, existing literature is consulted to find out what information practitioners find useful and insightful for identifying mental health conditions and how a solution to the problem this research investigates would fit into the workflow of practitioners. This can advise future work in how to interpret and use the results of this research.

In the writing of this paper, no generative AI has been used.

2 Related Work and Design Choices

To find out what practitioners find most insightful and intuitive, existing literature on three topics have been investigated. First, literature on ESM in general and what is desired from a practitioner's perspective has been explored. Next, literature that includes examples of ESM data presentation methods and potential feedback from practitioners was studied. Finally, literature non-specific to ESM, but to visualizations to compare an individual to a cohort based on health data was explored.

Based on this literature, design choices are defined which have been used as a basis to shape the mock-ups. There are design choices of two types. The first type is to anchor the visualizations to previous research by using features that are suggested by literature. The other type is to build upon the existing literature by designing underexplored ideas that this research investigates within ESM visualization context. The second type of design choices have allowed the user evaluation to elicit practical insights on what a practitioner would like to see when comparing their client to others with a certain mental health condition.

2.1 Previously Used Design Choices

Flexibility in terms of content personalization and dynamic data visualizations are important topics in the literature [Weermeijer et al., 2024, 2023]. These topics are part of making the tool more personalized towards clients [Bos et al., 2019]. One way to achieve these ideas is by allowing the practitioner to select what to visualize. Existing literature has already implemented this feature [Bos et al., 2022; Chari et al., 2019] with practitioners scoring it high as a feature of a health data comparison visualization [Schaaf et al., 2021]. This results in the design choice that inclusion factors are selectable.

Research has also shown that practitioners can be overwhelmed by visualizations of ESM data and require practice or training [Weermeijer et al., 2023]. Such training, however, is not always successful or sufficient partially because it is too much effort to read the manual or because information from training sessions get lost over time [Hall et al., 2025; Weermeijer et al., 2023]. These issues were mainly identified in network diagrams. So, based on this information, no further investigation was done into network diagrams. Additionally, a small usage guide was introduced with the intention that both the practitioner can learn to understand while working with the tool and the practitioner can potentially retain their knowledge on the tool after potential training. These ideas

are akin to those that substantiate the minimal manual [Carroll et al., 1987].

2.2 Experimental Design Choices

Chari et al. [2019] created a tool that allows medical practitioners to customize and visualize health data of cohorts and compare with patients using a spider chart. However, their chart has overlapping colors, four lines, two of which dashed, and overlapping text. Complexity like this in the visualization could cause practitioners to be overwhelmed [Weermeijer et al., 2023]. So, this study created a simplified spider chart, inspired by Chari et al. [2019] with the idea that spider charts provide a clear view of all the information the practitioner wants to compare a client to a cohort.

There is not yet a definitive answer to what the most essential information required to effectively compare ESM data of a client to a cohort is. Comparing each survey of all of the people in the cohort and each survey of the client individually would require practitioners to compare potentially hundredths of values. When comparing the confidence of the conclusions derived from data, the analysis of aggregated data outperforms the analysis of individual data points [Nguyen et al., 2020; Heer and Shneiderman, 2012]. Based on the results of their research, this study proposes a way to represent ESM data of an individual by using the average values of the numerical responses to every survey in the individual's data-gathering period. This reduces ESM to a method of getting a snapshot of a participant's mental state as opposed to a method of capturing the dynamics. More research should be done to determine if such a reduction can capture a mental health condition when averaging it over many people diagnosed with the same condition and under what conditions.

To reduce the loss of the temporal nature of ESM data by such a reduction, another complementary visualization has been proposed: A visualization of the variability of the ESM-responses throughout the data gathering period. The design choice was to use two of the same type of graph, namely the spider chart. Combining the average and the variability is also part of the approach that pre-existing tools use [Lenaert et al., 2019]. With that in mind, this study investigated whether practitioners think that the combination of average values and their variability is sufficient for the practitioner to effectively compare their client to a cohort.

Two types of questions common in ESM-surveys are questions about a participant's mental state and questions about a participant's context [Van Os et al., 2017]. The former is often captured with numerical values and the latter with categorical values [Van Roekel et al., 2019]. As described above, the design choice of visualizing the average and variability with spider charts only works with numerical values. One way to capture more abstract measures of one's mental state is by visualizing aggregated values which are calculated based on the survey responses. Responses both numerical and categorical can be used as long as the final calculated value is numerical. Examples of such calculated values are positive and negative affect [Weermeijer et al., 2023; Torkamaan, 2022], consummatory pleasure [Van Roekel et al., 2017] and avoidance of social situations [Kaiser and Laireiter, 2018]. The final one of the examples also includes contextual informa-

tion.

Because contextual information can be encoded in calculated values, the design choice has been made to include calculated values as inclusion factors that can be selected to be added to the spider chart. By adding this feature, this study has explored practitioners' opinions on the necessity of directly visualizing contextual measured values to compare clients.

Besides a spider chart, this study considers another type of visualization. Kreienkamp et al. [2025] introduced a novel idea: applying dimensionality reduction to present ESM data in a scatter plot. They used this technique to visualize clusters. In this study, however, for the purposes of comparing a single client to a cohort, clustering is not needed. This study proposes the design choice of displaying the individuals of the cohort with a certain mental health condition as dots on a scatter plot and the client as a distinct dot. The expectation was that this makes it easy to spot how similar the client is to those in the cohort.

However, a scatter plot of ESM data with its dimensionality reduced to 2, likely loses its interpretability. The axes have no longer a well defined meaning and neither the average nor the variability of the ESM-survey are identifiable.

3 Method

Based on the design choices derived from literature and described in the previous section, two mock-ups have been made. This section explains how these were made. This section also provides information about the participants of the user evaluation, it describes the user evaluation and explains the approach to the thematic analysis.

3.1 Creating Mock-Ups

To explore the ideas based on literature described in the previous section, input from practitioners is used. One way to get this is by using mock-ups [Becerra-Suarez et al., 2022]. To create the mock-ups, the online tool Figma has been used. Figma is often used for its prototyping purposes [Borysova et al., 2024].

Based on the design choices, two mock-ups have been created: one of a spider diagram (Figure 1) and one of a scatter plot showing ESM data with reduced dimensionality (Figure 2).

Both visualizations have common elements. One of these elements is the selectable inclusion factors. These have been explicitly described as design choices.

To alleviate concerns from practitioners about the complexity and validity of these diagrams [Weermeyer et al., 2024] and because displaying underlying uncertainty within visualizations is essential to build user's trust [Sacha et al., 2016], a band of mean ± 1 standard deviation is introduced. This is not the optimal method to visualize uncertainty [Correll and Gleicher, 2014]. It is, however, well-established [Correll and Gleicher, 2014]. Such visualizations are not necessary for the scatter plot, because all the individual data is shown.

Furthermore, both mock-ups have a condition selection box, allowing practitioners to select to what condition they want to compare their client. Both mock-ups also share a section containing a usage guide.

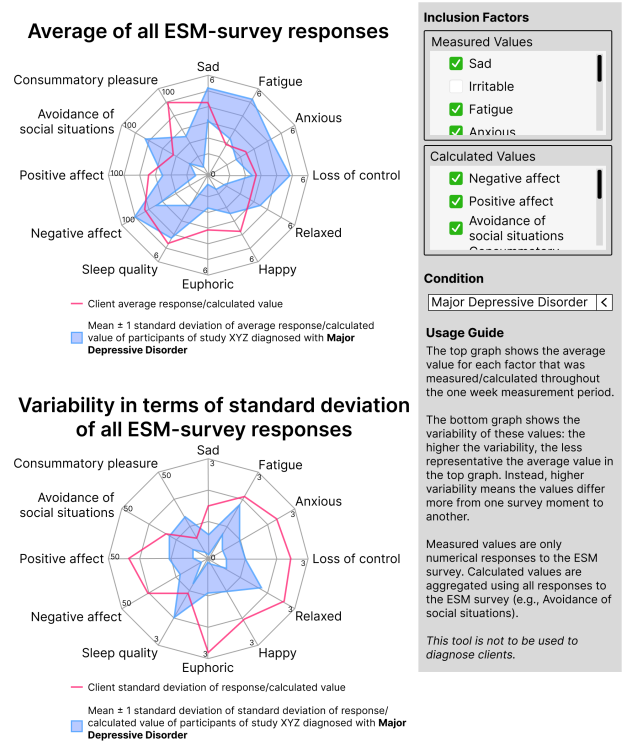


Figure 1: Mock-up of spider diagram to visualize ESM data comparison.

Additionally, the mock-ups share the colors in the plots. These plots have been created with people with color-vision deficiency in mind by ensuring the colors are distinguishable by those with the most common types of color-vision deficiency and introducing a second dimension. The second dimension for the spider chart it is the line compared to the area and for the scatter plot it is the size of the dots. These methods to make the plots more accessible are in line with existing literature [Stevens et al., 2024].

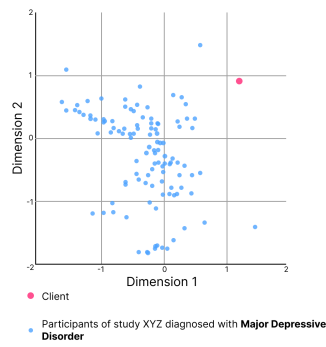
The mock-up with the spider chart has two plots. One showing the average and one showing the variability of the collected ESM data. This was also an explicit design choice mentioned before.

The inclusion factors are split into two sections: measured values and calculated values. Additionally, a small explanation on what measured values are, is added to the usage guide. This has been done to ensure clarity for the practitioners.

3.2 User Evaluation

To find out what visualization methods practitioners find most insightful and intuitive for identifying mental health conditions, a survey has been made. The choice of using a survey rather than, for instance, an interview or focus group has to do with the time constraints of this research and ease of distribution. Distribution is made easy because the survey has been made with Qualtrics. Practitioners can participate in the research by simply clicking a link. However, it is important to note that the user evaluation is preceded by screen requiring

Dimensionality reduced scatter plot



Inclusion Factors

Measured Values

☒ Sad
 ☐ Irritable
 ☒ Fatigue
 ☒ Anxious

Calculated Values

☒ Negative affect
 ☒ Positive affect
 ☒ Avoidance of social situations

Condition

Major Depressive Disorder

Usage Guide

Each dot in the plot is a person that participated in an ESM-data collection period. The dimensionality of all their data from the selected inclusion factors has been reduced to 2. This allows for visual comparison.

Measured values are only numerical responses to the ESM survey. Calculated values are aggregated using all responses to the ESM survey (e.g., Avoidance of social situations).

This tool is not to be used to diagnose clients.

Figure 2: Mock-up of dimensionality reduction to visualize ESM data comparison.

participants to provide informed consent before being able to continue.

Next, the participants are presented with questions about their professional background. Here, they must choose either clinical psychologist/psychiatrist, psychology researcher, medical student, psychology student or no clinical background. Additionally, their age is requested in five-year bins and they can optionally provide their gender as either male, female or other. This information is requested because it allows their answers to be put into context.

All responses to the survey were anonymized by not storing any identifiable information such as IP, email address, location, browser data, names and more. This was all done using the built-in features of Qualtrics.

After the initial questions is an explanation of what this study is about, the definition of ESM, the intended purpose of the tool of which mock-ups have been shown, how participants are expected to answer the questions, a scenario guide the participants, a list of example questions of the ESM-questionnaire and the explanation that the condition (Major Depressive Disorder) has already been selected, but that another condition or a control group could have been selected too. See Appendix A for the full text.

The survey, then, shows participants the mock-ups and asks them corresponding questions (seen in Table 1); one question per page is asked and each time the corresponding mock-ups are shown. These questions have been made with the idea that their answers not only challenge the design choices, but also explain their reasoning. All the questions are open ended and the total expected duration of the survey is around 15 minutes.

The order of the questions has been taken into account. First, only the spider chart is shown and a general question is posed that forces the participant to take a look at the mock-

up. During this phase, the hope is that the participant tries to understand the visualization and reads the usage guide. After this, more specific questions are posed.

The fourth question was asked because practitioners expressed that actual use of ESM data visualization tools translate to interactive use with clients [Weermeijer et al., 2024]. The choice to provide participants with an example use case, namely showing these graphs to their client as a way to guide or augment one or more therapy sessions, is intended to encourage them to think not only about how they will use it in practice, but also about what features of the visualization encourages or discourages their envisioned use cases.

At the fifth question, the scatter plot is introduced. The participant should now have an understanding of what they would want to see from such a plot. This and to reduce the time required led to the choice of posing less questions for the scatter plot, expecting participants to provide a complete answer to questions 5 and 6 after being inspired by the previous questions. Question 5 and 6 are essentially the same questions they had before, this is intended to encourage comparison between the two methods of visualization.

The participants also had the opportunity to see the questions in Dutch. This was done because the survey was performed in the Netherlands.

3.3 Participants

Participants have been recruited in three ways. The first way is through distribution among psychology master students, psychology PhD students and psychology researchers. The second way is through distribution among psychology bachelor students. Finally, the user evaluation has been distributed among practitioners in the field of psychology. The distribution has been done by sending a web link to the user evaluation hosted on Qualtrics. The web link was accompanied by a short description of what the research entails and what is expected of the participants.

In total, 8 people participated in the user evaluation. 6 out of 8 of the participants indicated they were psychology students. 1 participant indicated they were a clinical psychologist/psychiatrist and 1 participant indicated they were a psychology researcher. 3 out of 8 participants indicated they were younger than 21 years old, 2 indicated they were between 21 and 25, 2 indicated they were between 26 and 30, and 1 indicated they were between 51 and 55. 4 out of 8 of the participants indicated they were female, 3 out of 8 indicated they were male and 1 indicated other.

3.4 Thematic Analysis

To derive insights from the responses to the user evaluation survey, a thematic analysis has been conducted as described by Braun and Clarke [2006]. For this research, the choice has been made to create a rich description of the data set as opposed to a detailed account of a single particular aspect. According to Braun and Clarke [2006], this method is useful of research into an under-investigated area, which matches the purpose of this study. Furthermore, a theoretical approach to the thematic analysis is chosen, because this approach results in themes that adhere to the research question more closely opposed to an induced approach. Additionally, the were identified at the semantic level. This means themes were identi-

Table 1: User Evaluation Survey Questions

Nr.	Survey Question
1.	Consider the spider plots. Explain how these graphs can or cannot be useful to you when trying to compare your client to a mental health condition in practice
2.	The measured inclusion factors do not contain client responses to questions about context (e.g., Who are you with?). These are only used indirectly in some calculated values. How would this influence your capability to effectively compare your client? And how would you use the calculated values to gain insights related to context?
3.	The information visible is the average (upper diagram) and the variability (lower diagram). What information would you like to see additionally, if any?
4.	Why would or why wouldn't you show these graphs to your client as a way to guide or augment one or more therapy sessions or how would you use it instead?
5.	Consider the scatter plot. Explain how this graph can be useful to you when trying to compare your client to a mental health condition in practice.
6.	Compare the differences in how easy it is to compare the similarity between your client and a mental health condition using the spider chart and the scatter plot. How would you use the scatter plot differently in practice compared to the spider chart and why?

fied within the explicit meanings of the data, showing patterns withing, from which the content is summarized and interpreted [Braun and Clarke, 2006]. This contrasts the latent level where one would look for intrinsic or extrinsic motivations beyond what a participant of the survey has written [Braun and Clarke, 2006]. Finally, an essentialist/realist approach is taken. This assumes that which a participant wrote down represents their experiences and meaning, opposed to a constructionist approach, which also considers contexts such as social and cultural ones that may be part of what shapes the answer of an individual [Braun and Clarke, 2006]. The chosen approach complements the choice for semantic themes [Braun and Clarke, 2006].

Braun and Clarke [2006] provide a six-phase guide to performing a thematic analysis. This guide has been followed. First, all the responses were read, taking notes of initial ideas. Next, initial codes were generated by reading through all the data again. Then, codes were compared and contrasted to group them and all relevant data into potential themes. After that, a thematic map is generated to review the themes to check if they make sense within both the coded extracts and the dataset as a whole. Next, the potential themes were concretely defined by naming them and refining the specifics of the themes and the overarching narrative. Finally, extracts are selected that bare significance to the narrative, research question and related literature to produce the report of the thematic analysis.

During the thematic analysis, all of the survey responses were reviewed and included in the dataset. However, a small subset of responses, namely 4, were excluded from thematic coding, because they did not have interpretable or relevant semantic content.

In addition to the method described by Braun and Clarke [2006], an assessment of interrater reliability is used. For this, the procedure described by McHugh [2012] was followed. It uses two coders and calculates Cohen's Kappa [Cohen, 1960] along with a 95% confidence interval.

During the thematic analysis of this research, both coders

were computer science students that had previous experience with ESM data and academic research, but not with thematic analysis specifically. One coder created the codes and sent them to the other coder. Both applied them to the user evaluation responses. The interrater reliability was determined to have a Cohen's Kappa of $\kappa = 0.702 \pm 0.097$. This 95% confidence interval is within what Landis and Koch [1977] propose to be the range describing a substantial strength of agreement. During the calculation of the Cohen's Kappa, empty responses and excluded responses were left out.

4 Results

Three themes were identified: Application in Therapy, Data Representation and Visualization Methods, and Improvements. Table 2 contains some important quotes from the responses of the user evaluation.

4.1 Theme 1: Application in Therapy

First, all but two participants wrote that they would like to use the spider graph interactively with their client. They proposed doing this in numerous ways. Two psychology students proposed to ask their client about their values when their values look extreme or use the spider graph as guidance to discuss what has been going on lately. For example, one of them presented a scenario where the client scored high on sadness: They would ask whether their client knew what made them sad in a conversation with their client. Another student has a different approach and would let their client decide if they want to discuss the data. Yet another student indicates that the spider chart could not only give them, but also give the client a good indication of whether they fit a disorder. That student would discuss the graphs with their client. The mental health practitioner had a clear opinion on the ESM data visualizations. They stated the following: "If I make a client fill something out, then I will use that information, otherwise I don't understand why I would make a client fill it out."

One student, however, indicated that they would not show the visualizations to their client. Rather, they expressed the intention of using the visualizations indirectly with their

Table 2: Quotes related to themes (Application in Therapy, Data representation and Visualization Methods, and Improvements) of responses to the user evaluation about ESM data visualization to identify mental health conditions

Participant	Quote
Theme 1: Application in Therapy	
Mental health practitioner	If I make a client fill something out, then I will use that information, otherwise I don't understand why I would make a client fill it out.
Psychology student	With an extreme score, you could ask about it, explore this research area, or perform further tests for a diagnosis, like in this case the BDI-II for depressive symptoms.
Psychology student	You can use these graphs in a session to discuss what has been going on lately.
Psychology researcher	These plots might be a bit hard to understand if you are not already familiar with them.
Psychology student	Would use it to discuss these factors together with the client and deepen our understanding of them. To research in which situation these occur and in which these don't. But I don't know if I would do this for multiple sessions.
Theme 2: Data representation and Visualization Methods	
Psychology researcher	I don't know if these graphs would be particularly useful for the patient, as I could only tell them "Hey, on average you are more depressed than the rest of the population." This is not helping in therapy.
Psychology researcher	I think the scatter plot is more easy to understand initially, but the spider plot is more useful overall.
Psychology student	Can actually be used to get insights into what the client suffers from most.
Psychology student	The idea of a scatter plot, I believe, is more intuitive than a spider diagram for the client.
Mental health practitioner	There is no explanation of what dimensions 1 and 2 are, so I don't know what I can read from this.
Theme 3: Improvements	
Psychology student	Without context, you don't know if a factor is disproportionate on that moment.
Psychology researcher	It would be important to see, how the emotions changed from day to day, as this only gives you the values for the entire week. it could be important to see if there were outlier days, if they were more happy on the weekend etc. If you want to go into more detail, you could also include a graph about emotions at each timepoints to compare e.g. morning vs. evening.

client. The student mentioned that they would only discuss the insights from the graphs verbally in simple language with their client.

The reason that the student did not want to show the spider charts to their client directly is because they believed that the spider charts would be too hard to understand for a client. The mental health practitioner also wrote that they believed that clients might have a hard time understanding the spider charts. Their reason was that they would not already be familiar with them.

Next, one student mentioned that they were unsure whether they would discuss the spider charts with their client in more than one session.

Additionally, the mental health practitioner introduced the notion of phases in their responses. They explained that the spider plot would be of use during the intake phase, the treatment phase and the evaluation phase.

Furthermore, there was the opinion on the usefulness of the spider chart in how it would guide or encourage further treatment or investigation. For example, the idea of using the spider chart as a guide to perform further tests for a diagnosis was posed by a psychology student. The student provided an example within the scenario provided: "You could [...] explore this research area, or perform further tests for a diagnosis, like in this case the BDI-II for depressive symptoms." The same student also mentioned later that the spider charts

could introduce talking points, that may previously not have been addressed, for during a session with the client. One student noted that besides the spider chart, manual investigation to get insights beyond a general picture of the client is required. They stated that checking in what situations certain behavior occurs could be of practical use. Later, they wrote that this deepening of the understanding could be done in discussion with the client.

The opposite opinion exists for the scatter plot. The psychology researcher explained that the scatter plot does not help in guiding therapy approaches. They wrote that only comparing a client to peers does not help discover which specific strategies could help the client.

Finally, when asked about how the participants would use the scatter plot differently in practice compared to the spider chart, some mentioned they would either use the spider chart more or that the scatter plot is less useful. One participant, however, wrote that they would show the scatter plot to their client.

4.2 Theme 2: Data Representation and Visualization Methods

When asked about how the spider chart could be useful to the participants to compare their client to a mental health condition, the variability statistic was praised. They explained that the variability adds to the average statistic by providing a better insight into the client's feelings throughout the week.

The psychology researcher, however, deemed the graph showing the variability as less useful. They explained: “It is much harder to interpret how the variation in these emotions compared to their peers is relevant for a clinical diagnosis.”

Next, the calculated values were considered useful as a way to get insights into the context. One student stated that they would use the calculated values to gain context by comparing them with the measured inclusion factor, for example by looking at these side by side. They would then look at how they match. The researcher mentioned that they could use measures like sleep quality and avoidance of social situations to deduce some context.

Two opinions were expressed specifically on the average statistic. The psychology researcher thought the spider chart showing the average could be helpful to compare a patient’s average emotions. However, in a later response, they stated that they were unsure the spider charts would be useful, writing: “I could only tell them [clients] ‘Hey, on average you are more depressed than the rest of the population.’ This is not helping in therapy.” One student also mentioned that comparing using the average statistic is not useful to identify whether their client has a mental health condition and how severely.

Furthermore, one opinion is shared among all participants. They all liked how specific the spider charts are with regard to the values it displays. Some write that identifying for which values the client scores high is useful or that it “can actually be used to get insights into what the client suffers from most.” One student says it “gives a good indication of whether something does or doesn’t fit the disorder.”

Additionally, two psychology students identified that the spider charts provide an intuitive overview at a glance. One states that one would not require prerequisite knowledge to understand how well something fits a disorder.

The same is written about the scatter plot. That too is deemed to give an intuitive overview at a glance. One student explained that the scatter plot “shows at a glance whether the client is anomalous.” They follow this up by stating: “The idea of a scatter plot, I believe, is more intuitive than a spider diagram for the client.” Another student wrote that the scatter plot “would provide some insight into the severity of the client’s depression.”

However, the mental health practitioner noted that they do not find the comparison between the client and the cohort in the scatter plot interesting. The psychology researcher puts it like this: “I think the scatter plot is more easy to understand initially, but the spider plot is more useful overall.”

The reason for this, provided by the participants, is that the scatter plot has unclear dimensions and is not specific enough. The practitioner wrote: “There is no explanation of what dimensions 1 and 2 are, so I don’t know what I can read from this.” Another stated the scatter plot does not say anything about emotions. The researcher mentioned that the scatter plot was not useful because it was limited to two dimensions.

4.3 Theme 3: Improvements

When asked about the spider charts, both psychology students and the psychology researcher mentioned that it would be useful to have control over the temporal granularity of the visualizations. They mentioned, for example, that they would

like to be able to see the spider charts for specific days, or even parts of days such as morning or evening. Such temporal granularity, they explained, would help them understand the dynamics of their patient, making it more specific and therefore be of more use in sessions with the client. Some students also mentioned that being able to access individual measurements would be more insightful than seeing just an average with the variability.

Related to the request for more granularity, some expressed a desire to see what days their client had reported the most extreme values. One student also wanted to see the ESM survey entry that contained the highest or lowest measured values. How they would specifically use this information was omitted.

A recurrent remark across the participants of the user evaluation was that context is essential information to get a better understanding of their client. They not only mentioned that mood could be dependent on context, but also that this could explain certain specific behavior. One student stated: “Without context, you don’t know if a factor is disproportionate on that moment.” Some also explained that they would like to see explicitly what the data that is visualized looks like in specific contexts. The psychology researcher expanded this with: “It would be more helpful to directly be able to see the client’s social context, as that can have quite a bit of explanatory value on the emotions that they experienced during the day.”

Next, the mental health practitioner explained that they would like to see more specific selection criteria for the research study cohort. They wrote that they would be interested in differences in age categories.

Also, the psychology researcher introduced the idea of potentially including a network model to show the relationships between emotions. They provided an example scenario where a participant could be showing a high level of anxiousness at the same time as showing high levels of fatigue. They noted that this would not be necessary, however, for a basic clinical overview.

Finally, a student and the practitioner mentioned that the ESM visualization would be a good way to see improvement of the patient over time. The practitioner suggested having a client perform ESM measurement periods in different phases of care to compare the client to themselves. They suggested this for both the spider chart and the scatter plot.

5 Responsible Research

First of all, this study was approved by the Human Research Ethics Committee of Delft University of Technology (ID = 5405). When considering what affect the research of this study can have on society and to determine if this is responsible, those who are affected must be made explicit.

This study can both directly and indirectly affect future researchers that explore areas such as HCI or mental health mHealth; directly is meant in the sense of future work based on this paper and indirectly is meant in the sense that more influential research based on this study can affect future researchers. The future researchers have been taken into consideration when writing this paper. For one, effort has been put into describing the methods used in this research such

that future researchers can reproduce and replicate this study. For example, reproducibility is improved by explaining in detail the assumptions and decisions made during the thematic analysis. The extensive argumentation supporting the design choices for the mock-ups are meant to improve replicability.

Additionally, the participants of the user evaluation can be affected by this study, both directly and indirectly. Indirectly, they may be affected the same way practitioners in the mental health field are affected. Directly, the participants have entrusted the researchers of this paper with their responses. This has been taken seriously and steps have been taken to handle their data responsibly. The most important step was to require informed consent by the participants before they could continue. Another important step is the choice to anonymize the responses by not requesting names, email addresses, etc. and not storing IP-addresses, location, browser data, etc. Also, the personal questions at the start of the user evaluation survey have been made to make it hard to identify. This has been done to select age in bins of five years and make gender selection optional. Despite the efforts that were just described, it would still be unfavorable if their data were to get leaked: Qualtrics was used to create and host the surveys, meaning some control over the data has been delegated to the company that owns that tool; if they lied about what data they record and store or if there is a bug in their systems, the participants could still have a risk of identifiable data being leaked that could be linked to their responses. However, it had been decided that the benefits of using Qualtrics outweigh the potential threat due to giving up a bit of control. This decision was made based on the reputation of Qualtrics as a reliable survey tool.

Current and future clinical psychologists and psychiatrists can also be indirectly affected by this study. A potential tool which uses the insights gained from this study could reach the stage where it is ready to be used in clinical practice. This means that the way they do their jobs in the future might be shaped however so slightly by this study. The best way to ensure this study serves its role responsibly is to try and make the results of this study as representative of the current and future clinical psychologists as possible. Despite having only limited responses to the survey, responses have been collected from current practitioners and future practitioners, the latter of which is in the form of psychology students.

Finally, future clients could be indirectly affected by this study. Because this study advances the field technology within clinical practice, clients could be exposed more to such tools in the future. Any tool that is inspired by this study can affect those clients. Because this study is not written by experts in the field of clinical psychology or psychiatry, the approach that was taken to be part of this future responsibly is by drawing as much inspiration from existing literature as possible, relying on experts indirectly in such a manner.

Note that this study has not used any real ESM data, avoiding any ethical implications of such actions. The choice of synthesizing the data for the visualizations may, however, have reduced the representativeness and therefor generalizability of the mock-ups that were created.

6 Discussion

This study aimed to explore what methods of visualization practitioners find most insightful and intuitive to identify mental health conditions, and why. The themes identified provide nuanced insights into the design choices presented by this study.

First of all, it is important to note that spider charts were generally liked as a way of visualizing the ESM data. The first reason is that it allowed practitioners to gain insights at a glance. This was the idea behind the adapted spider chart inspired by Chari et al. [2019]. The level of complexity of the spider chart did not seem to overwhelm the practitioners as the study by Weermeijer et al. [2023] warned about. The second reason that spider charts were liked is because of how specific they are. These two properties combined make the spider chart intuitive and have a high information density.

Next, there was a critique that the spider chart may be too complicated to be shown directly to clients, but practitioners seemed to be generally interested in using these graphs interactively with their client and to use it as a way to guide or augment one or more therapy sessions. Also, the notion of using these graphs in multiple phases of care matches what was discovered in the study by Van Os et al. [2017]. This was expanded by the idea that these visualizations could be generated for multiple ESM data gathering periods to see a trend of improvement for a client.

Additionally, the design choice to display the average and the variability of the measured values over the entire ESM data measurement period was not as positively received. The average statistic did not get much direct positive feedback. The variability was, however, seen as an addition to the average, albeit perhaps harder to interpret. The general feedback on the spider chart showed that the combination does provide useful insights to practitioners. No alternative aggregate statistic to be used in the spider charts came up during the study. The largest point of improvement is allowing practitioners to specify the temporal granularity. Practitioners seemed to dislike the choice of averaging over the entire ESM data measurement period. Future implementations should consider adding visualizations with data aggregated over shorter time periods or even allowing the practitioners to define the temporal granularity themselves. This confirms that flexibility in terms of content personalization and dynamic data visualizations as described by Weermeijer et al. [2023] is important for ESM data visualizations, also for identifying mental health conditions.

Furthermore, the design choice of adding calculated values as inclusion factors was well received. Practitioners saw them as useful information, also to gain insights into context. However, they are not sufficient to replace direct information about context. Practitioners would like to see visualizations that show the relationship between emotions and contextual measured values or between emotions themselves.

Finally, the scatter plot seemed to have performed nearly exactly as expected. At first sight, it looks like an intuitive visualization from which insights can be extracted at a glance. However, after further inspection, the scatter plot won't be useful. The concept of dimensionality reduction is complex to understand, which results in a mis- or nonunderstanding of

the meaning of the axes, despite the small usage guide in the mock-up.

7 Limitations and Future Research

As part of this study, a way to represent ESM data of an individual by using the average values and variability of the numerical responses to every survey in the ESM data-gathering period is proposed. This way of representing the ESM data reduces it to a format that loses any trends or complex dynamics in the client's mental state. It compresses that information to merely a snapshot. This study assumes that the data in such a reduced form can still capture the features that describe mental health conditions. More research should be done to determine if this assumption is valid, especially when averaging such an aggregate over many people diagnosed with the same condition and under what conditions this assumption is valid.

Future studies could investigate the feasibility of the visualization methods presented in this study. All data presented in the visualizations in the mock-ups has been synthesized. This assumes that creating visualizations is feasible. This does not mean that the results of this study are only valid if the visualizations turn out to be feasible: The results have a more general meaning. However, it does mean that direct implementation of the visualizations in the mock-ups may not be possible. By synthesizing the data, this study leaves the research into how to actually implement the algorithms required to present the data in the format required for the visualizations to future researchers.

Additionally, future studies could use the results of this study when designing a tool that allows practitioners to identify mental health conditions in their clients based on ESM data. The findings of this study provide a starting point on what practitioners would like to see in such a tool. The design choices made for the mock-ups in this study can be implemented or adapted according to the provided feedback. For instance, a tool can be made that contains spider charts or similar visualizations that display aggregates such as average and variability. Future researchers can choose to implement the proposed improvements as well, such as allowing for a finer temporal granularity. The tools that implement the findings of this study can be used to research in more depth the requirements of practitioners for a tool to identify mental health conditions.

8 Conclusion

From the user evaluation responses, it was concluded that spider charts are an intuitive way to visualize ESM data to compare specific emotions and moods of a client to a mental health condition. Practitioners especially praised the spider charts as intuitive and specific. Using the average and variability of all the data collected is adequate, but a finer temporal granularity is preferred. Additionally, context is deemed to be crucial information and aggregate values that contain information about context are not sufficient to compare clients. Visualizations that directly show dependency on context are preferred by practitioners. Furthermore, a scatter plot showing a dimensionality reduced representation of ESM data is not useful for identifying mental health conditions in prac-

tice, because the meaning of the axes was misunderstood or the axes were not considered useful.

A User Evaluation Introductory Text

The following quote is the full introductory text that is shown before the first question of the user evaluation.

This survey is a part of a research that will try to explain what methods of visualization practitioners (psychologists/psychiatrists) find most insightful and intuitive to identify mental health conditions from ESM-data.

ESM is a momentary assessment technique that collects contextual mental health data by prompting participants with a short questionnaire multiple times a day at (semi-)random intervals.

You will be shown two mock-ups, each of a tool that presents the ESM-data to identify mental health conditions. You will be asked open questions about the mock-up. Please try to explain any answer you provide. The majority of the questions is related to the first mock-up. Any data you will see in this survey is fabricated with the clarity of the visualizations in mind.

Imagine the following scenario: You are a practitioner treating a client. Your client has just performed a week of ESM-data gathering by answering ten questionnaires per day. The response to the questions is either a slider from 0 (disagree) to 6 (agree) or multiple choice. Example questions of the questionnaire are:

- I feel anxious
- I feel stressed
- I feel sad
- I feel excited
- I feel relaxed
- What am I doing? (Leisure, school, work, etc.)
- Where am I? (At home, at a friend or family member's home, at work, etc.)
- Who am I with? (Nobody, family, friends, colleagues, etc.)

As a practitioner, you now want to find out how your client relates to other people that have been diagnosed with some condition. In this example: Major Depressive Disorder.

To answer this question, you use a tool that visualizes this comparison. See the mock-up of such a tool containing fabricated data on the next page. Major Depressive Disorder has already been selected, but it is possible to select another condition or a control group.

In this scenario, you, as the practitioner, are assumed to know all the questions of the ESM-questionnaire of your client.

References

- Anjeer Alshehhi, Y. A. A., Ahmad, K., Abdelrazek, M., and Bonti, A. (2023). Data-Task-Visualisation Analysis based on Top mHealth Apps. In *Proceedings of the 35th Australian Computer-Human Interaction Conference*, pages 231–244, Wellington New Zealand. ACM.
- Becerra-Suarez, F. L., Villanueva-Ruiz, D., Tuesta-Monteza, V. A., and Mejia-Cabrera, H. I. (2022). Usability Evaluation of Mobile Application Software Mockups. In Arai, K., editor, *Intelligent Computing*, volume 506, pages 321–331. Springer International Publishing, Cham. Series Title: Lecture Notes in Networks and Systems.
- Bell, I. H., Fielding-Smith, S. F., Hayward, M., Rossell, S. L., Lim, M. H., Farhall, J., and Thomas, N. (2018). Smartphone-based ecological momentary assessment and intervention in a blended coping-focused therapy for distressing voices: Development and case illustration. *Internet Interventions*, 14:18–25.
- Borysova, S., Borysov, V., Kochergina, S., Spasskova, O., and Kushnarova, N. (2024). Involvement of interactive educational platforms in the training of graphic design students using the Figma platform as an example. *International Journal of Education and Information Technologies*, 18:119–132.
- Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M., and 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.
- Bos, F. M., Von Klipstein, L., Emerencia, A. C., Veermans, E., Verhage, T., Snippe, E., Doornbos, B., Hadders-Prins, G., Wichers, M., and Riese, H. (2022). A Web-Based Application for Personalized Ecological Momentary Assessment in Psychiatric Care: User-Centered Development of the PETRA Application. *JMIR Mental Health*, 9(8):e36430.
- Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.
- Bringmann, L. F., Van Der Veen, D. C., Wichers, M., Riese, H., and Stulp, G. (2021). ESMvis: a tool for visualizing individual Experience Sampling Method (ESM) data. *Quality of Life Research*, 30(11):3179–3188.
- Carroll, J., Smith-Kerker, P., Ford, J., and Mazur-Rimet, S. (1987). The Minimal Manual. *Human-Computer Interaction*, 3(2):123–153.
- Chari, S., Qi, M., Agu, N. N., Seneviratne, O., McCusker, J. P., Bennett, K. P., Das, A. K., and McGuinness, D. L. (2019). Making Study Populations Visible through Knowledge Graphs. Version Number: 1.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Correll, M. and Gleicher, M. (2014). Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151.
- Daniëls, N. E. M., Hochstenbach, L. M. J., Van Bokhoven, M. A., Beurskens, A. J. H. M., and Delespaul, P. A. E. G. (2019). Implementing Experience Sampling Technology for Functional Analysis in Family Medicine – A Design Thinking Approach. *Frontiers in Psychology*, 10:2782.
- Dietvorst, E., Aukes, M. A., Legerstee, J. S., Vreeker, A., Hrehovcsik, M. M., Keijsers, L., and Hillegers, M. H. J. (2022). A Smartphone Serious Game for Adolescents (Grow It! App): Development, Feasibility, and Acceptance Study. *JMIR Formative Research*, 6(3):e29832.
- Hall, M., Lappenbusch, L. M., Wiegmann, E., and Rubel, J. A. (2025). To Use or Not to Use: Exploring Therapists’ Experiences with Pre-Treatment EMA-Based Personalized Feedback in the TheraNet Project. *Administration and Policy in Mental Health and Mental Health Services Research*, 52(1):41–58.
- Heer, J. and Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54.
- Kaiser, T. and Laireiter, A.-R. (2018). Process-symptom-bridges in psychotherapy: an idiographic network approach. *Journal for Person-Oriented Research*, pages 49–62.
- Khosravi, M. and Azar, G. (2024). A systematic review of reviews on the advantages of mHealth utilization in mental health services: A viable option for large populations in low-resource settings. *Global Mental Health (Cambridge, England)*, 11:e43.
- Kreienkamp, J., Agostini, M., Monden, R., Epstude, K., De Jonge, P., and Bringmann, L. F. (2025). A Gentle Introduction and Application of Feature-Based Clustering with Psychological Time Series. *Multivariate Behavioral Research*, 60(2):362–392.
- Kroeze, R., Van Der Veen, D. C., Servaas, M. N., Bastiaansen, J. A., Voshaar, R. C. O. V., Borsboom, D., Ruhe, H. G., Schoevers, R. A., and Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research*, 3(1):1–11.
- Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159.
- Larson, R. and Csikszentmihalyi, M. (2014). The Experience Sampling Method. In *Flow and the Foundations of Positive Psychology*, pages 21–34. Springer Netherlands, Dordrecht.
- Lenaert, B., Colombi, M., Van Heugten, C., Rasquin, S., Kasanova, Z., and Ponds, R. (2019). Exploring the feasibility and usability of the experience sampling method to examine the daily lives of patients with acquired brain injury. *Neuropsychological Rehabilitation*, 29(5):754–766.

- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3):276–282.
- Myin-Germeys, I., Birchwood, M., and Kwapil, T. (2011). From Environment to Therapy in Psychosis: A Real-World Momentary Assessment Approach. *Schizophrenia Bulletin*, 37(2):244–247.
- Netherlands, S. (2022). Mental health has worsened among young people. Last Modified: 2022-06-01T15:00:00+02:00.
- Nguyen, F., Qiao, X., Heer, J., and Hullman, J. (2020). Exploring the Effects of Aggregation Choices on Untrained Visualization Users’ Generalizations From Data. *Computer Graphics Forum*, 39(6):33–48.
- OECD and European Observatory on Health Systems and Policies, editors (2023). *Netherlands: Country Health Profile 2023*. State of Health in the EU. OECD Publishing, Paris.
- Price, M., Yuen, E. K., Goetter, E. M., Herbert, J. D., Forman, E. M., Aciermo, R., and Ruggiero, K. J. (2014). mHealth: A Mechanism to Deliver More Accessible, More Effective Mental Health Care. *Clinical Psychology & Psychotherapy*, 21(5):427–436.
- Riese, H., Von Klipstein, L., Schoevers, R. A., Van Der Veen, D. C., and Servaas, M. N. (2021). Personalized ESM monitoring and feedback to support psychological treatment for depression: a pragmatic randomized controlled trial (Therap-i). *BMC Psychiatry*, 21(1):143.
- Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G., and Keim, D. A. (2016). The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249.
- Schaaf, J., Sedlmayr, M., Prokosch, H.-U., Tegtbauer, N., Kadioglu, D., Schaefer, J., Boeker, M., and Storf, H. (2021). Visualization of Similar Patients in a Clinical Decision Support System for Rare Diseases – A Focus Group Study. In Röhrig, R., Beißbarth, T., Brannath, W., Prokosch, H.-U., Schmidtmann, I., Stolpe, S., and Zapf, A., editors, *Studies in Health Technology and Informatics*. IOS Press.
- Stefancic, A., Rogers, R. T., Styke, S., Xu, X., Buchsbaum, R., Nossel, I., Cabassa, L. J., Stroup, T. S., and Kimhy, D. (2022). Development of the First Episode Digital Monitoring mHealth Intervention for People With Early Psychosis: Qualitative Interview Study With Clinicians. *JMIR Mental Health*, 9(11):e41482.
- Stevens, H. P., Winegar, C. V., Oakley, A. F., and Piccolo, S. R. (2024). Identifying images in the biology literature that are problematic for people with a color-vision deficiency. *eLife*, 13:RP95524.
- Ten Have, M., Tuithof, M., Van Dorsselaer, S., Schouten, F., Luik, A. I., and De Graaf, R. (2023). Prevalence and trends of common mental disorders from 2007-2009 to 2019-2022: results from the Netherlands Mental Health Survey and Incidence Studies (NEMESIS), including comparison of prevalence rates before vs. during the COVID-19 pandemic. *World Psychiatry*, 22(2):275–285.
- Torkamaan, H. (2022). *Health Recommender Systems for Mental Health Promotion*. PhD thesis, DuEPublico: Duisburg-Essen Publications online, University of Duisburg-Essen, Germany.
- Van Os, J., Verhagen, S., Marsman, A., Peeters, F., Bak, M., Marcelis, M., Drukker, M., Reininghaus, U., Jacobs, N., Lataster, T., Simons, C., ESM-MERGE Investigators PhD, Lousberg, R., Gülöksüz, S., Leue, C., Groot, P. C., Viechtbauer, W., and Delespaul, P. (2017). The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice: van Os et al. *Depression and Anxiety*, 34(6):481–493.
- Van Roekel, E., Keijsers, L., and Chung, J. M. (2019). A Review of Current Ambulatory Assessment Studies in Adolescent Samples and Practical Recommendations. *Journal of Research on Adolescence*, 29(3):560–577.
- Van Roekel, E., Vrijen, C., Heininga, V. E., Masselink, M., Bos, E. H., and Oldehinkel, A. J. (2017). An Exploratory Randomized Controlled Trial of Personalized Lifestyle Advice and Tandem Skydives as a Means to Reduce Anhedonia. *Behavior Therapy*, 48(1):76–96.
- Weermeijer, J., Kiekens, G., Wampers, M., Kuppens, P., and Myin-Germeys, I. (2024). Practitioner perspectives on the use of the experience sampling software in counseling and clinical psychology. *Behaviour & Information Technology*, 43(3):540–550.
- Weermeijer, J. D. M., Wampers, M., De Thurah, L., Bonnier, R., Piot, M., Kuppens, P., Myin-Germeys, I., and Kiekens, G. (2023). Usability of the Experience Sampling Method in Specialized Mental Health Care: Pilot Evaluation Study. *JMIR Formative Research*, 7:e48821.