Integrating the use of sensing technology to detect early warning signs of relapse for those with lived experience of bipolar disorder

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy, March 2023

Shazmin Majid (BSc, MSc)

Supervised by:

Dr Stuart Reeves

Professor Richard Morriss

Dr Grazziela Figueredo

Acknowledgements

I would firstly like to thank every user that contributed toward this research. Thank you for your guidance and enthusiasm throughout. I would especially like to thank the two organisations Bipolar Lift CiC and Bipolar UK who were a huge help in user recruitment. I would especially like to thank Matthew Moore and Rob Mason who immensely shaped this research to be more closely aligned to real lives of those with bipolar disorder. This was the upmost aim of this research.

I would like to thank my incredible supervisory team Dr Stuart Reeves, Professor Richard Morriss and Dr Grazziela Figueredo who have consistently supported me throughout this Ph.D. journey. Thank you for always believing in me, even when I forgot to do so on occasions. I would like to thank the Horizon Centre for Doctoral Training for the opportunity to undertake this work and contribute towards the scientific discourse of how we harness technology for a better tomorrow. I would like to thank the NIHR Biomedical Research Centre for supporting this research, especially Dr Jennifer Martin who has always advocated this project and Dr Elvira Perez Vallejos who has provided incredible guidance.

On a personal level, I would like to firstly and foremostly thank my father Salim Ahmed. You taught me to be strong and driven, always making any dream achievable. You taught me a world with no limitations and that I can do whatever I put my mind to. By achieving this dream, I owe it all to you. I would like to thank my grandparents Abdul Majid Ahmed and Roshan Ahmed who have always been my backbone, and held me together in the final year of my Ph.D. Your love and kindness have been overwhelming. My drive to finish this Ph.D. has always been to honour your name and the sacrifices you have made for us to have a better life. I would like to thank my stepmother Roshni Ahmed, who will always be both like my mother and my best friend. You are my role model and have taught me to be better in an infinite number of ways. I would also like to thank my siblings, Maariyah, Amelia, Alayna and Ayaan who motivate me to become a better role model for them. I hope I make you all proud.

I would like to thank all my dear friends who have laughed and cried with me during this journey, what an incredible experience it has been! Namely I would like to thank the extraordinarily talented women in my life, Dr Claire Yau, Dr Natalie Leesakul and Dr Vanja Ljevar. The three of you represent my inner compass. I would also like to thank Dr Joseph Dib, Dr Harleen Rai and all my peers at Horizon CDT, especially those who are members of "GDPR". Finally, I would like to thank the love of my life Cameel Rishmawi, who taught me that true love does exist in the final year of my PhD.

Abstract

In the world of pervasive mobile technology, it is inevitable that novel technological solutions have been leveraged to understand symptoms of bipolar disorder (BD). Increasingly, these technologies use a combination of passive and active sensing techniques. BD is a complex condition where the sense of *self* is consistently in "flux". There are questions of how much this sense of *self* is currently reflected in self-tracking technology for BD. Upon investigating this, we found that user involvement in self-tracking technology is variable, where high-level involvement is seldom seen in the literature. Furthermore, this technology is being developed without reference to clinical guidelines, best practice principles and with a lack of high-quality research evidence. Using a combination of participatory design methods from healthcare and human-computer interaction, the overall aim of this doctoral research is to bring the user's personal and lived experience of BD to the forefront in order to design and assess a mobile self-tracking tool which uses passive and active sensing techniques to understand early warning signs (EWS) in BD – a clinically validated framework in understanding relapse. The research was organised into three work packages: Concept Generation and Ideation, Prototype Design and Deployment and Evaluation. In the first work package, the everyday practices of self-tracking were explored in two user-led workshops (n=18 users). The findings revealed a high degree of complexity and individual variability in self-tracking where over 50 methods of tracking were described. In the next phase, the findings were built upon using follow up interviews (n=10) to guide the redesign of a self-tracking tool to be closely aligned to users' needs and preferences. During the Evaluation phase, the final prototype was enrolled for a 6-month beta test in a real-world context with eight users. The findings revealed that the tool was useful in understanding EWS from both a subjective (i.e., user led) and statistical viewpoint. Frequencies in the passive data were connected to EWS via the active data, however, there were inconsistencies in how users interpreted the data compared to our statistical analysis – proving that "no one size fits all" in technology for BD. Overall, the tool was demonstrated good usability and acceptability from users, with constructive suggestions for improvement.

Table of Contents

List of Abbreviations

BD	Bipolar disorder
EWS	Early Warning Signs
PPI	Patient and Public involvement
HCI	Human Computer Interaction
SMI	Severe Mental Illness
DSM	Diagnosis and Statistics Manual
BD1	Bipolar I Disorder
BD2	Bipolar II disorder
UCD	User-centred design
RADAR-BASE	Remote Assessment of Disease and Relapse
aRMT	Active data collection app
pRMT	Passive data collection app

List of Tables

List of Figures

Chapter 1: Introduction

Chapter summary

- Bipolar disorder (BD) is a complex and serious mental health condition. The process of self-tracking has been proven to be helpful in bipolar disorder, and it makes experiences comparable provides insights in mood changes over time.
- With the advent of pervasive technology, novel technological solutions have been leveraged to promote self-tracking, often using a combination of active and passive tracking to tackle the inherent challenges of self-tracking in this condition.
- There are overarching gaps and limitations within current literature in the domain of self-tracking and BD in relation to user involvement, design, clinical and development guidelines as well as questions as to what constitutes best practice in this intersection between *healthcare* and *technology.*
- The current works aims to contribute towards the literature for self-tracking technology using a combination of passive and active sensing methods to provide insight into relapse of BD, focusing on early warning signs which have clinical validity in prevention.
- The overall aim of this doctoral work is to bring user's personal and lived experience of BD to the forefront in order to design and assess to assess a mobile self-tracking tool which uses passive and active sensing techniques to understand EWS for relapse for those with BD.

1.1. Background and motivation

Bipolar disorder is a complex, cyclical and serious mental health condition which is recognised as being in the top 20 causes of disability worldwide (Ferrari et al., 2016). People with BD experience extreme mood shifts between depression and mania. Depression is characterised by feeling low and lethargic, whereas mania is defined by feeling very high and overactive (*Overview - Bipolar Disorder*, 2021). Long-term monitoring of BD, particularly by self-tracking behaviours and symptoms is central to managing this condition, and the foundation of proven psychological therapy for BD such as psychoeducation (Kendall et al., 2014; Morriss et al., 2016). The process of self-tracking has shown to have dramatic effects for people with BD, making experience relatable and comparable, and providing insights into mood changes over time. In her book, Emily Martin describes some of the underlying mechanisms of self-tracking practices. She states that many people are encouraged to keep a "mood chart" and the process of filling one out, a small act of discipline, can have dramatic effects for people, as by assigning a mood a name or number can make experiences comparable and understand changes over time (E. Martin, 2009). Early Warning Signs (EWS) often referred to as “relapse signatures” are a set of 63 unique symptoms and signs that arise 2-4 weeks before relapse to either mania or depression have been proven useful to self-track. Understanding EWS by self-tracking has proven clinical validity such as improve time to episode, decrease percentage of hospitalisation and improve functioning (Lobban et al., 2011).

Unfortunately, the very nature of BD poses issues in self-tracking. Previous research suggests that personal insight for people living with this condition is often dependent on emotional state and most impacted during mania when self-assessment is often erroneous (Goodwin & Jamison, 2007). The lifelong nature of this condition coupled with extreme mood changes make self-tracking problematic, as subjective recall is difficult during affective episodes where insight is likely to be limited. The difficultly with manual tracking has been acknowledged in the literature and opened a pathway to more automated self-tracking approaches. These approaches have the potential to passively monitor behaviour, detect changes in mood and to provide feedback about EWS (Murnane et al., 2016), with minimal input from the user. Research suggests that individuals are open to technology that automatically collects data in the background and many have used wearables to track factors such as sleep, activity and heart rate in order to support the understanding how these behaviours impact their mood (Matthews, Murnane, & Snyder, 2017a).

In the advent of pervasive mobile technology, it is inevitable that novel technological solutions have been leveraged to understand symptoms of BD. Increasingly, these technologies use a combination of active and passive sensing techniques to understand the condition. The active aspect of these technologies still relies on self-input of data about mood and wellbeing. However, passive sensing techniques objectively and automatically collect daily data to reflect behavioural activities. There is a growing body of research in smartphone sensing methods such as capturing the number of text messages (Beiwinkel et al., 2016; Faurholt-Jepsen et al., 2015; Puiatti et al., 2011), phone calls (Faurholt-Jepsen et al., 2015; Gruenerbl et al., 2014), emails (Puiatti et al., 2011), GPS data (Gruenerbl et al., 2014; Osmani et al., 2015a), voice features (Gideon et al., 2016; Karam et al., 2014a; Maxhuni et al., 2016; Muaremi et al., 2014; Vanello et al., 2012), accelerometer data (Bardram et al., n.d.-a; Garcia-Ceja et al., 2018; Vanello et al., 2012), and app usage (Alvarez-lozano et al., 2014a) and exploring their utility for promoting self-tracking in BD, particularly in understanding severity of symptoms. In this way, objective behavioural data can be collected automatically, in the background, without the need for user input minimising the level of obtrusiveness and user burden (Faurholt-Jepsen et al., 2018).

However, there are questions whether the innovative developments in this field are being mindful of the holistic experience of the user living with this condition. The very nature of the condition means that users are living with essentially two contrasting psychiatric conditions, and their sense of self is consistently in “flux”. The experiences of the condition are fraught with created experiences of confusion, contradiction and self-doubt which impinges on the ability to establish continuity in their sense of self (Inder et al., 2008). The lives of those are unfortunately characterized by disruption and discontinuity and by definitions of the self that are external and based on their illness. The complex nature of the *self* must be understood when considering self-tracking technology. When examining the current literature for self-tracking technology, a number of overarching gaps and limitations were found which will be addressed in this research:

A) Currently the sense of self is not understood as users are not being involved in the design and evaluation of self-tracking technology for BD, despite recommendations of involvement (Majid et al., 2021; Sanches et al., 2019). This may explain why individual variability of illness expereince was concluded as not being considered in a recent review of current self-tracking technology for BD (Ortiz et al, 2021)

This was also reflected in another large review of the past decade on HCI and affective health (including BD) which found that out of 76 studies involved, only 16 studies reported clinical evaluations which was descried as those users who use mental health services or who meet the formal criteria for a specific mental health problem (Sanches et al., 2019). This finding specifically to BD was recently replicated in a recent review by Majid et al (2021) who found variable user involvement in the design of self-tracking technology for BD, often falling towards the lower end of the spectrum of involvement.

B) There is an apparent issues when it comes to the design of self-tracking technology. The current app marketplace is fraught with unmet needs as indicated by the number of user wish lists requests, a lack of adherence to clinical guidelines and best practice principles (Nicholas et al., 2017a), lack of supporting feasibility and efficacy studies and in some cases misleading and harmful content (Lagan et al., 2020)

C) When exploring best practice principles of self-tracking technology, which essentially sits at the intersection of *healthcare* and *technology,* there are suggestions that methods of user involvements should be drawn from these distinct yet overlapping disciplines (Majid et al., 2021; Sanches et al., 2019). For example patient and public involvement (PPI) is a method of involvement in healthcare, whereas in technology there is an established vocabulary of involvement as seen in the field of human-computer interaction (HCI) (Baumer, 2015). The hybridisation of these methods is described as useful in the complex exploration of users' needs, preferences and individual variability (Majid et al, 2021)

This research aims to tackle the aforementioned limitations to design and develop a self-tracking tool for BD which using a combination of active and passive techniques to understand EWS – a proven clinical framework to understand relapse in BD (Lobban et al., 2011). By being a proven clinical framework, this responds to limitation B in relation to the lack of clinical guidelines currently seen in BD mobile apps (Nicholas et al, 2017a). As to response to limitation A, the design and evaluation of this tool will closely involve users who have lived experience of BD. To respond to limitation C, the research will combine participatory methods seen in HCI and PPI to involve and empower users in these processes, and to ensure the technology reflects user's needs, preferences and individual variability. The self-tracking tool will be enrolled for field testing with users with lived experience of BD to explore usability and acceptability of this tool in a real-world context. The gathering of research data from field testing responds to limitation B in relation to lack of feasibility and efficacy studies (Lagan et al., 2020). The research process will be closely designed to integrate the 7 best practice recommendations for the design and evaluation of self-tracking technology for BD as outlined by Majid et al (2021) and outlined in Table 1.

1. Involve users in all stages of design and evaluation, including concept generation and ideation, prototype design and deployment, and evaluation stages with the goal of creating user empathy and empowerment. This process should have an adequate number of users to welcome diversity in thought. Equal representation is also a crucial consideration that needs to be considered when recruiting users.
2. Ensure early involvement as this will be cost-effective in the long run (avoid redesign and problems with use and implementation in the later stages).
3. Combine principles of PPI and HCI to not only have users to assist in designing technology but also in designing and running research (eg, users cofacilitating design workshops) and use end-to-end user involvement.
4. For academic and industry sectors to establish better mechanisms to access target user groups with lived experience of mental health issues, for example, by building relationships with existing patient-directed organizations such as charities and patient-led community groups.
5. Increase awareness of HCI and design communities in PPI principles and practices and increase awareness of PPI community in HCI and design methods or skills.
6. Encourage use and mixing of formal scientific or design methods with informal experiential and empathic practices to capture richness in understanding the dynamic requirements of technology users, which are cognizant of use in context.
7. Keep the user informed at all stages of the process, including final outcomes, future use, and next steps, which are often forgotten about

Table 1 Best practice recommendations for the design and evaluation of self-tracking technology for BD (Majid et al, 2021)

As user involvement is a key topic investigated in this doctoral work, the process of this research will be organised according to the user-centred design process which is typically outlined as an iterative process through various phases of design such as concept generation and ideation, prototype design and deployment, evaluation and deployment. The process of which is outlined in Figure 2. UCD is not a research methodology but a framework that can contribute to the research which allows us to explore the design space, generate requirements, identify design challenges and tensions, and allow us to understand the potential use and usefulness of new types of systems. As these are the main premise of the thesis, the chapters of this doctoral work will be organised according to this framework and associated aims and objectives will be outlined in the subsequent section.

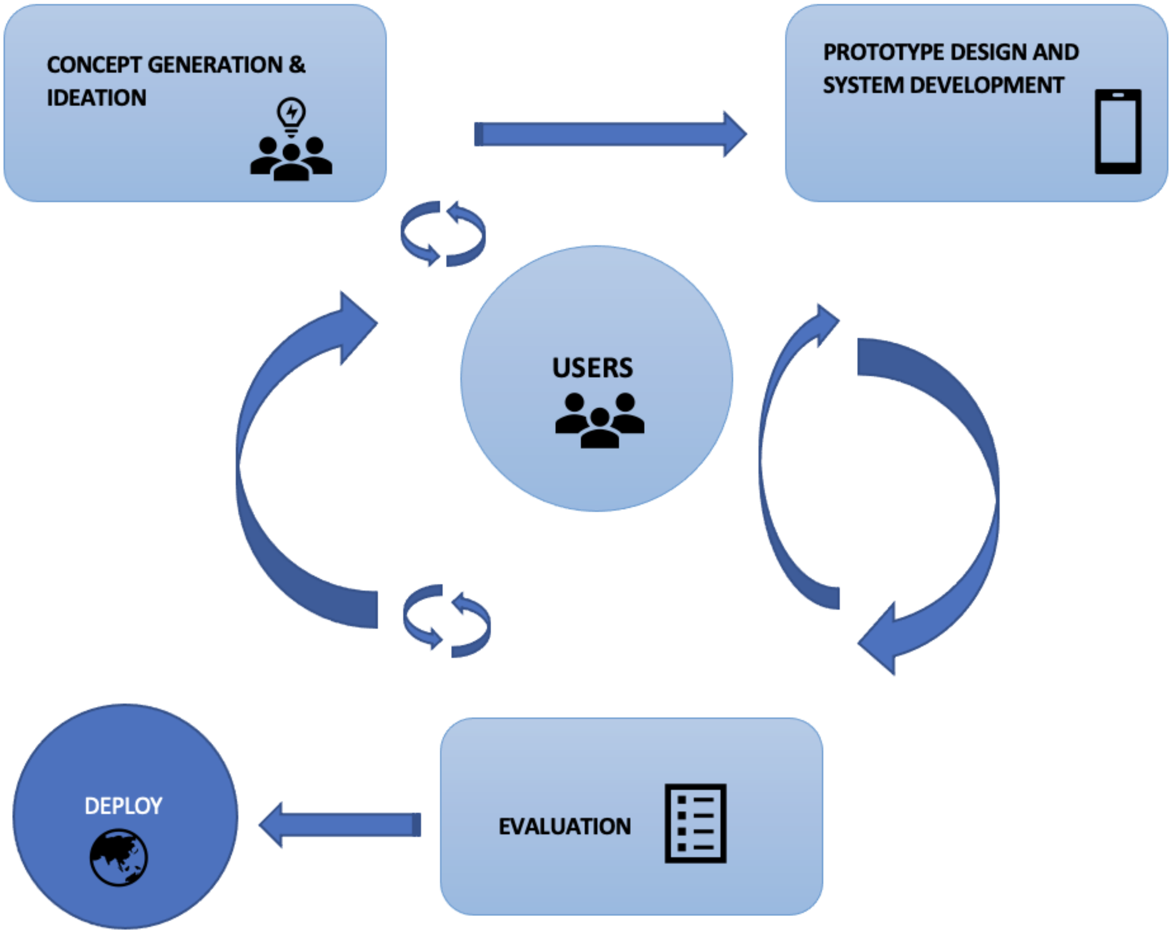


Figure 1 User-centred design process from McCurdie et al (2012)

The findings of this research aim to promote the understanding of one's own individual signs of BD through close monitoring of one's moods, activities, warning signs and behavioural patterns which has been described as a key aspect of staying well (Johansson & Werbart, 2020). The key contributions of this thesis are in relation to bringing users perspectives to the forefront as the research will be the first of its kind in applying a hybridised methodology of HCI and PPI to explore self-tracking in BD, considering best practice principles. As applying this hybridised methodology of HCI and PPI is novel and complex, a personal reflexive account of the researchers personal expereince will be another key finding of this research.

1.2. Research aims, objectives and questions.

The overall aim of this doctoral work is to bring user's personal and lived experience of BD to the forefront in order to design and assess to assess a mobile self-tracking tool which uses passive and active sensing techniques to understand EWS for relapse for those with BD. The research process of this doctoral work will be organised according to the user-centred design framework outlined in Figure 1 as this framework can contribute towards the research by enabling us to explore a design space, generate requirements, identify design challenges, and better understand the potential use and usefulness of this tool. The chapters of the thesis will be organised according to the different phases of the design specifically Concept Generation and Ideation, Prototype Design and Deployment and Evaluation

1.2.1. Concept Generation and Ideation

1.2.1.1. Aims and Objectives:

To combine principles of HCI and PPI to explore real-life self-tracking practices for those with lived experience of BD and relate this to the design of self-tracking technology. Specifically, the reasons, methods and challenges of self-tracking will be explored, as well as opportunities for technology.

1.2.1.2. Research Questions:

RQ1: What are the reasons and methods for self-tracking in those with BD?

RQ2: What are the challenges and opportunities for self-tracking technology for BD?

RQ3: How does using this hybridised approach of HCI and PPI impact the findings of this research?

The findings from these research questions aim to inform the design of mobile self-tracking apps for BD in general.

1.2.2. Prototype Design and Development

1.2.2.1. Aims and Objective:

The findings from the Concept Generation and Ideation section aim to provide an overall view on the experience of self-tracking in BD. The current section will build on these findings to explore how they can be used to design of a mobile self-tracking application for BD. This section will aim to develop a working prototype which will be subject to alpha testing with a small group of target users to explore usability and functionality.

1.2.2.2. Research Questions*:*

RQ4: How do the findings from the Concept Generation and Ideation phase impact the design of the mobile self-tracking application for BD?

RQ5: How can the mobile self-tracking application be used to explore EWS for those with BD?

RQ6: What is the general user experience of using the prototype?

RQ7: What, if any, changes are required the working prototype before deployment in the evaluation phase?

The findings from these research questions aim to inform the design of the specific app investigated.

1.2.3. Evaluation

1.2.3.1. Aims and Objectives

The Prototype Design and Deployment stage aimed to direct the findings on the experience of self-tracking in BD from the Concept Generation and Ideation phase to guide the design of a novel self-tracking tool in BD. The Evaluation stage aims to develop and enrol a working prototype for a 6-month period of beta testing with those with BD to explore usability and acceptability. To explore how much of the data collected from prototype relates to EWS, which will be explored primarily using qualitative methodologies.

1.2.3.2. Research Questions*:*

RQ8: How much of the data collected by the sensing system (both passive and active data) relates to EWS?

RQ9: Is this type of technology usable, feasible and acceptable for people living with this condition?

RQ10: What are the researchers experience of applying the overall methodology employed in research thus far?

The findings from these research questions (RQ8 and RQ9) aim to inform the design of the specific app investigated. The findings from these research questions (RQ10) aim to inform the design of mobile self-tracking apps for BD in general.

Due to the novel and complex nature of the hybridized methodology employed throughout this study, there will be reflexive account of the researchers' personal experience of this process throughout the course of this doctoral work.

1.3. Outputs of the current work

The findings of this work aim to contribute to the following gaps in the literature which have been identified via the exploration of the following specified research questions.

1. Understanding the role of individual variability for self-tracking technology in BD generally and how this impacts the design and deployment of such technology $^{\text{RQ1, RQ2,}}$
2. Provide a best practice example of the involvement of users in the design and deployment of technology for BD, with the aim of empowering users to be the experts in defining how they manage their own condition $^{\text{RQ3}}$
3. Providing rigorously collected research data to contribute towards the knowledge base of the impact of self-tracking technology in BD management, particularly by integrating the EWS which has high clinical validity. $^{\text{RQ1, RQ2, RQ5}}$

1.4. Contributions and connections with research disciplines

The primary contributions of this research are towards bringing the users perspectives to the forefront when designing and evaluating self-tracking tech for BD. We attempt to describe best practice towards how to involve users, specifically those suffering with severe mental health issues such as BD, in the design and research in relation to mental health technology. We aim to follow best practice principles for the design, evaluation and research of the self-tracking tool, and provide the first case example of combining involvement approaches seen in the healthcare sector well as methods seen in human-computer interaction to design and assess a novel self-tracking tool for BD. Specifically, we will reflect on how the close involvement of target users not only in the design of technology but also the design of the research impacts our findings, and how this compares against similar literature in this domain. The personal reflexive account of applying this methodology from the researcher’s perspective aims to guide future researchers navigating in this complex space of health and HCI.

The contribution of this work aims to span across psychiatry, psychology and computer science. By assessing a novel self-tracking tool which uses a combination of active and passive sensing technology to specifically understand EWS, we hope to further contribute towards the research findings in relation to self-tracking, passive sensing and the EWS respective to the health and HCI domain.

By leveraging proven clinical frameworks (such as the EWS checklist) and then testing this within the context of technology in a longitudinal field trial we aim to contribute findings in the medical and psychiatric field towards the acceptability and usability of this tool. Connections with communities

This research was closely worked upon with respective charity and patient lead organisations such as Bipolar Lift CiC (Nottingham), Bipolar UK (UK Based) and MindTech Patient and Public Involvement Team (Nottingham). The consistent effort of these organisations ensures that the perspectives and priorities of those living with BD were integrated into the condition, problem and tools we are seeking to explore and improve as a result of this research. The work was closely developed in conjunction with these organisations who assisted in research design, user recruitment and development of findings to ensure that this research has real user benefit.

1.5. Paper and conferences

Below are the papers and conference talk that arose as part of the doctoral thesis

1.5.1. Papers

1. **Majid, S.**, Morriss, R., Figueredo, G., & Reeves, S. (2022, June). Exploring self-tracking practices for those with lived experience of bipolar disorder: Learning from combined principles of Patient and Public Involvement and HCI. In *Designing Interactive Systems Conference* (pp. 1907-1920).

2. Polhemus, A., Novak, J., **Majid, S**., Simblett, S., Morris, D., Bruce, S., ... & Wykes, T. (2022). Data Visualization for Chronic Neurological and Mental Health Condition Self-management: Systematic Review of User Perspectives. *JMIR mental health*, *9*(4), e25249.

3. **Majid, S**., Reeves, S., Figueredo, G., Brown, S., Lang, A., Moore, M., & Morriss, R. (2021). The Extent of User Involvement in the Design of Self-tracking Technology for Bipolar Disorder: Literature Review. *JMIR Mental Health*, *8*(12), e27991.

4. **Majid S** (September 2020) Designing for the User, With the User: Self-tracking Technology for Bipolar Disorder. *Doctoral Consortium EAI Pervasive Health 2020 - 14th EAI International Conference on Pervasive Computing Technologies for Healthcare (Atlanta, USA)*

1.5.2. Conferences

1. **Majid, S** (October 2019) The Extent of User Involvement in the Design of Self-Tracking Technology for Bipolar Disorder: Literature Review. *Mixed Reality Lab*

Talk at the Department of Computer Science, University of Nottingham (Nottingham, UK)

2. **Majid, S** (September 2020) Integrating the use of sensing technology to detect early warning signs for bipolar disorder. *Industry Day Talk at the Department of Computer Science, University of Nottingham (Nottingham, UK)*

3. **Majid, S** (January 2022) Integrating the use of sensing technology to detect early warning signs for bipolar disorder. *Delivered to the Psychiatry Team at Nottinghamshire Healthcare NHS Foundation Trust*

4. **Majid, S** (April 2022) Combining principles of human-computer interaction (HCI) and patient and public involvement (PPI) to understand patient lived experience: a case study in self-tracking technology for bipolar disorder. *Full talk at UX Healthcare Europe 2022 (London, UK)*

1.6. Overview of the thesis

1.6.1. Chapter 2: Related work

The related work section consists of a literature review of bipolar disorder, self-tracking, current self-tracking technology and discusses the gaps and opportunities that led to the current research.

1.6.2. Chapter 3: Methodology

The methodology chapter outlines the process of hybridising HCI and PPI methods and organised according to the different phases of this design research doctoral study. This chapter describes the self-tracking mobile application to be used in the study.

1.6.3. Chapter 4: Concept Generation and Ideation

This chapter explores the everyday practices of self-tracking for those with BD in a study which closely follows the best practice principles of design and research of mental health technology as outlined by (Majid et al., 2021). This chapter consists of the findings of a series of workshops that were conducted in a UK with n=18 users. Specifically, the workshops explored the reasons, methods and challenges for self-tracking as well as opportunities for future technology.

1.6.4. Chapter 5: Prototype Design and Deployment

This chapter explores the findings from the Concept Generation and Ideation phase and convert these into tangible design changes for the technology that will be deployed in this study, so that it is tailored for users in the context of BD. It outlines the findings of the alpha testing phase which was conducted with n=2 users with lived experience of BD where issues around functionality and usability were explored. The final prototype is also described.

1.6.5. Chapter 6: Evaluation

This chapter presents the process as well as the findings of the 6-month longitudinal field-testing phase of the research where a working prototype was used in a real-world context with a group of n=6 users. This chapter reports on the quantitative and qualitative findings of the research and reports on usability, accuracy and validity of the novel digital self-tracking tool, as its relationship with uncovering EWS for mania and depression.

1.6.6. Chapter 7: Discussion

This chapter reports on the key contributions of this doctoral work towards self-tracking technology, bipolar disorder and the methodologies of involvement of the user. This chapter also reports on the implications of applying this methodology in personal reflexive account.

Chapter 2: Related Work

Chapter summary

- There is some, albeit limited, proven utility of self-tracking technology in understanding relapse in BD. However, this is fraught with a lack of emphasis on the individual variability in the experience of this condition.
- A method to explore individual variability and gain insight into users' needs and preferences are user-centred design approaches, as recommended in the literature.
- Our literature review demonstrates that currently the adoption of these approaches is highly variable, with a significant proportion of studies reporting no-user involvement in the development and design of technology. Demonstrating that currently, individual variability is not being consistently explored for.
- It was concluded that technology should be tailored according to users' needs and preferences, in order to maximize the potential that technology has in order to improve the lives of those living with this condition.

2.1 The growing concern of severe mental illness

The impact of severe mental illness (SMI) is universal and significant, affecting the ability to function socially, academically, and occupationally. Such disorders include depression, bipolar disorder, schizophrenia and other psychoses, dementia and developmental disorders including autism (Evans et al., 2007). It can also affect their physical health where the presence of conditions such as obesity, metabolic syndrome, diabetes mellitus, symptoms of cardiovascular disease, and respiratory disease all exceed that of the general population by at least two times for those with SMI, and HIV prevalence may be increased by as much as eight times (Scott & Happell, 2011). This high rise of chronic disease may largely be responsible for the increased risk of premature mortality that can be seen in this population. A recent report by the Office of Health Improvements and Disparity in the UK (2022) found that in England overall individuals with SMI are statistically significantly more likely to die prematurely (before the age of 75) compared to individuals who do not have SMI (*Office of Health and Improvement Disparities, 2022)*. Additionally, individuals with severe mental illness may also experience stigma and discrimination, which can further complicate their lives. In a study by Morgan et al., 2018 individuals with severe mental illness experience significant discrimination and stigma in various areas of their lives, including employment, education, and healthcare. Research into treatment and management of severe mental illness is highly important as it can help individuals to manage symptoms, improve their functioning and lead fulfilling lives.

In this research, there is a specific focus on BD. The global impact of BD is significant, as it is a prevalent mental health condition that affects people of all ages, races, and socioeconomic backgrounds. According to the World Health Organization (WHO), BD affects around 60 million people worldwide, and is a leading cause of disability worldwide (Colombo et al., 2012). The condition has a significant impact on a person's quality of life, and can lead to a range of social and economic consequences, including loss of productivity, increased healthcare costs, and increased risk of suicide (Ketter, 2010). The carers of those with BD who provide emotional and practical support are also impacted by BD, as it has been reported that some carers experience higher levels of depression, anxiety and increased mental health service usage (Steele et al., 2009). In terms of the economic impact, bipolar disorder is a major contributor to the global burden of disease, as it is associated with decreased productivity, increased absenteeism, and a higher risk of disability. In the UK specifically, the population-level economic burden was estimated at £5.1 billion in 2010-2011 and £6.43 billion at 2018-2019 where the majority of economic burden result from productivity losses and informal care (Simon et al., 2021). Despite the prevalence and impact of bipolar disorder, access to appropriate treatment and care remains limited in many parts of the world, particularly in low- and middle-income countries (Demissie et al., 2018) where as few as 10% receive care over their lifetime (Fekadu et al., 2015; Kebede et al., 2006). This highlights the need for increased awareness, funding, and research to improve the diagnosis, treatment and overall management of bipolar disorder.

2.2 Bipolar Disorder: an overview

The concept of bipolar disorder, also known as manic-depressive illness, has been recognised for centuries. The ancient Greeks and Romans described symptoms of manic and depressive states. Specifically, these two extremes of mood were first systematically described by Hippocrates (460-337 BCE) (Mason et al., 2016). His work on this topic was based on both contemporary explorations of Pythagoras, Alcmaeon and Empedocles of Crotana and his own clinical observations of his mood (Angst & Marneros, 2001). He assumed the brain as the organ of mental functions, mental disturbances and mental disorders. To illustrate, in his famous work 'On the Sacred Disease' (Hippocrates) he wrote:

"The people ought to know that the brain is the sole origin of pleasures and joys, laughter and jests, sadness and worry as well as dysphoria and crying. Through the brain we can think, see, hear and differentiate between feeling ashamed, good, bad, happy … Through the brain we become insane, enraged, we develop anxiety and fears, which can come in the night or during the day, we suffer from sleeplessness, we make mistakes and have unfounded worries, we lose the ability to recognise reality, we become apathetic and we cannot participate in social life … We suffer all those mentioned above through the brain when it is ill …"

Early classification of the disease was first put forth by Karl Kahlbaum in 1874 (Kahlbaum, 1874) and later build upon in the 19th century by German psychiatrist Emil Kraepelin to characterise dysfunction of mood and bring together mania, depression and all psychotic states into what was termed "dementia praecox" and "manic-depressive insanity", now known as schizophrenia and bipolar disorder respectively (Angst, 2002) – the process of which was of critical importance to development of psychiatry. In 1952, the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-I) first attempted to categorise and standardise mental illness and classified bipolar disorder as a psychotic disorder (DSM-1, 1952). Two versions later in 1980 where the DSM-2 was released, there was a characterisation of illness with a specific classification criteria and shapes the modern definition of bipolar disorder, which is termed as such (Angst, 2002).In the latest version of the manual (DSM-5), bipolar disorder is divided into several subtypes, including bipolar I disorder, bipolar II disorder, cyclothymic disorder, and other specified and unspecified bipolar and related disorders (A. P. Association, 2013). The classification of these subtypes of bipolar disorder are explored in the next section. Research on bipolar disorder has advanced significantly in recent years, and new treatments have become available. However, much remains unknown about the causes of the condition and how best to treat it. Studies on the genetic, neural, and environmental factors are ongoing to better understand the disorder.

2.3 The bipolar spectrum: symptoms and presentation

Bipolar disorder is understood on a spectrum with marked and severe mood disturbances to milder mood variations. The Diagnosis and Statistics Manual-5 (DSM-**5** **A**. P. Association, 2013) defines mania and depression as follows:

Mania

Mania is a distinct period of abnormally and persistently elevated, expansive, or irritable mood, energy, or activity. During this period, three (or more) of the following symptoms have been present to a significant degree:

- Inflated self-esteem or grandiosity
- Decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
- More talkative than usual or pressure to keep talking
- Flight of ideas or subjective experience that thoughts are racing
- Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
- Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation

- Excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

Depression

Depression is a period of at least two weeks during which there is either depressed mood or loss of interest or pleasure, and at least four other symptoms that reflect a change in functioning, such as problems with sleep, eating, energy, concentration, or self-worth. The following symptoms must also be present:

- Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). In children and adolescents, can be irritable mood
- Markedly diminished interest or pleasure in nearly all activities nearly every day (as indicated by either subjective account or observation)
- Significant weight loss when not dieting or weight gain or decrease or increase in appetite nearly every day.
- Insomnia or hypersomnia nearly every day
- Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
- Fatigue or loss of energy nearly every day
- Feelings of worthlessness or excessive or inappropriate guilt nearly every day (not merely self-reproach or guilt about being sick)
- Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
- Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

It's worth noting that for a diagnosis of bipolar disorder, a person must have had at least one manic or hypomanic episode and may also have had one or more major depressive episodes.

In DSM-V, there is a distinction drawn between bipolar 1 disorder (BD1), in which a person experiences a full manic episode and bipolar 2 disorder (BD2), in which a person experiences depression episodes with less severe manic symptoms, known has hypomanic episodes. The third type of classification is cyclothymic disorder which is characterised as recurrent hypomanic episodes and subclinical episodes of depression. The fourth classification is otherwise specified and unspecified bipolar and related disorders: These subtypes are used for individuals who do not meet the criteria for bipolar I, II, or cyclothymic disorder, but still have symptoms consistent with the disorder. As BD1 and BD2 will be the focus on this study, the full diagnostic criteria according to DSM-5 is outlined in Table 2.

Bipolar I Disorder (BD1)
During the mood disturbance and increased energy or activity, ≥3 (or 4 if irritable mood only) of the following: • Inflated self-esteem or grandiosity • Decreased need for sleep • Pressured speech • Racing thoughts or flight of ideas • Distractibility • Increased goal-directed activity • Excess pleasurable or risky activity
Marked impairment not due to a substance or medical condition. In addition, these symptoms: • Do not meet criteria for a mixed episode • Cause functional impairment, necessitate hospitalisation, or there are psychotic features • Are not related to substance misuse • Are not due to a general medical condition • Are not caused by somatic antidepressant therapy
Criteria for mixed episode:

- Criteria are met both for a manic episode and for a major depressive episode during at least a 1-week period
- Causes functional impairment, necessitates hospitalization, or there are psychotic features
- Symptoms are not due to substance misuse, a general medical condition, or somatic antidepressant therapy

Bipolar II disorder (BD2)

During the hypomanic episode, ≥3 (or 4 if irritable mood only) of the following:

- Inflated self-esteem or grandiosity
- Decreased need for sleep
- Pressured speech
- Racing thoughts or flight of ideas
- Distractibility
- Increased goal-directed activity
- Excess pleasurable or risky activity

During the major depressive episode, ≥5 of the following symptoms are present during the same 2-week period, and represent a change from previous functioning. At least one of the symptoms is either depressed mood or loss of interest or pleasure

- Depressed mood most of the day, nearly every day
- Markedly diminished interest or pleasure, nearly every day
- Significant weight loss when not dieting or weight gain, or decrease or increase in appetite, nearly every day
- Insomnia or hypersomnia, nearly every day
- Psychomotor agitation or retardation, nearly every day
- Fatigue or loss of energy, nearly every day
- Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional), nearly every day
- Diminished ability to think or concentrate, or indecisiveness, nearly every day
- Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation with or without a specific plan.

In addition, these depressive symptoms:

- Cause functional impairment (e.g., social, occupational)
- Are not better explained by substance misuse, medication side effects, or other psychiatric or somatic medical conditions.

Table 2: DSM - V diagnostic criteria for BD

Though mania and hypomania are characterised as the defining symptoms of bipolar disorder, it has been found that depressive symptoms are more common than manic symptoms throughout the course of the illness. In a 12-year prospective longitudinal study, in which n=146 with BD1 completed weekly mood ratings, it was found that depressive symptoms were three times more likely than manic symptoms (Judd et al., 2003). This was found to be significantly higher in those with BD2, where users experience depressive symptoms 50% of the time (Judd & Akiskal, 2003). On average, it was found that patients spend 33% of the time depressed and a large proportion (60%) had four or more mood episodes in a year (Post et al., 2003). In more severe forms of mania, people can experience racing thoughts which can render incoherent and is also characterised by increased activity, productivity and creativity. These are usually described as being pleasurable symptoms towards the early stages of mania, especially after large periods of depressive symptoms. However, after a period when the episodes progress this can turn into feelings of severe distractibility, restlessness, and difficulty concentrating. There might also be an increase in risk-taking behaviour during period of mania, with behaviours such as shoplifting, increased spending, risky-sexual behaviours and motor accidents (Blanco et al., 2008); (Griswold & Pessar, 2000). Nearly half of BD patients have a lifetime history of psychotic symptoms (Keck et al., 2003). Among these, the following symptoms may occur: hearing thoughts spoken aloud; hearing voices referring to himself/herself, made in the third person; auditory hallucinations in the form of a commentary, etc (Dell'Osso et al., 2017). From a clinical perspective, the presence of psychotic symptoms in BD has been associated with a higher severity of illness and increased morbidity in the long-term (Coryell et al., 2001).

Individuals can also experience mixed affective episodes, where the symptoms of depression and mania or hypomania are presented at same time, which has significant impact on functioning (Table 1). In DSM-V, for a mixed episode to occur an individual must present with a manic or hypomanic episode with a proportion of depressive symptoms (3 out of 6) at the same time of mania or hypomanic episode. Alternatively, a major depressive episode must be present with a proportion of manic or hypomanic symptoms (3 out of 7). There is large variation in how people experience episodes where some experience discrete episodes which occur rarely (e.g., no more than one episode per year,) a subset of individual also experience “rapid cycling” in which individuals experience at least 4 episodes within a 12 month period. Ultra-rapid and ultra-ultra rapid cycling has also been identified in the population where mood variation can occur from week-to-week or even within the course of a day. In most people, the periods of rapid cycling resolved within 2 years so this is not a long-term pattern over a course of a lifetime (National Institute of Health Research, 2014).

In conclusion, BD is a complex condition which is classified into subtypes with the major classifications being BD1 and BD2 which are characterised by the severity of manic and depressive symptomology, with the former characterised having more severe manic symptoms and the latter having more severe depressive symptoms. BD1 and BD2 are the focus on this research as there are the major diagnostic criteria for BD, with characteristics of symptom presence that have the potential to be sensed by active and passive sensing. Minimising future mood episodes is the key objective for individuals living, and the essence of management of this condition and underpins the large body of research in this topic – including this research. Not only does this impact functioning and quality of life but also it has been found that multiple episodes cause neurological damage in the brain, related to progressive functional deficits (Deicken et al,2003).

2.4 Incidence and prevalence

The incidence and prevalence of BD vary depending on various factors, including age, gender, race, and socioeconomic status. Incidence refers to the number of new cases of a particular condition that occur within a specific time period. Prevalence, on the other hand, refers to the total number of cases of a condition that are present in a population at a given point in time (Ford, 2020).

BD is the fourth most common mental health problem worldwide after depression, anxiety and schizophrenia (Vos et al, 2015), and also recognised as one of the 10 most debilitating conditions worldwide. This disorder affects 45 million people worldwide (James et al., 2018) within community based epidemiological studies reporting a lifetime prevalence rate in European studies varying from 0.1%-2.4% (James et al., 2018) (Carta et al., 2012). A recent study conducted in 2021 with a large sample of n=36187 explored the prevalence of BD in the UK and found that prevalence doubled from 0.150% in 2001 to 0.357% in 2018. The pooled lifetime prevalence of BD1 was 1.06% and for BD2 it was 1.57%, meaning the prevalence of BD2 is higher over the course of a lifetime (Clemente et al., 2015).

In terms of gender, BD1 is more frequent amongst women compared to men and this is thought to be related to postnatal onset of bipolar disorder (Dell'Osso et al., 2017). Likewise BD2 is more common amongst women who usually tend to have a predominance of depression (López Zurbano et al., 2015). The Aesop Study (Fearon et al., 2006), examined the incidence of BD in people in three cities in the UK and reported a higher incidence in black and minority ethics groups, compared to a white population and this finding has been consistent with other UK-based studies (Lloyd et al., 2005). The age of onset of BD is early, with the first episode of BD occurring before the age of 30 years, however there is a smaller peak between the ages of 45-54 years for both BD1 and BD2. The highest peak of onset occurs between the ages of 15 and 19 years according to large cohort studies in the US and the Netherlands respectively (Kroon et al., 2013); (Merikangas et al., 2007). Furthermore, there is on average an 8-year delay between the first recollection of a mood episode to receiving a BD diagnosis (Mantere et al., 2004).

BD is also associated with a higher risk of comorbidity, which is the presence of other conditions alongside the primary condition. These comorbid conditions can include anxiety disorders, substance abuse, and eating disorders. In a more recent study it was found that BD2 patients were significantly more likely (odds ratios ranging from 1.4 to 2.1) to have comorbid psychiatric conditions (anxiety disorders, ADHD, eating disorders, personality disorders) than BD1 after adjusting for gender and age, the only exception being substance use disorders (Karanti et al., 2020).

2.5 Treatment and management of bipolar disorder

2.5.1 Pharmacological treatment

Treatment and management of bipolar disorder typically involves a combination of medication, therapy, and lifestyle changes. The specific treatment plan will depend on the individual's symptoms and needs. Medication is the primary treatment for bipolar disorder and the most prescribed medications are mood stabilizers, such as lithium and valproate. These medications are effective in preventing manic episodes and reducing the severity of symptoms. Antipsychotic medications, such as quetiapine and olanzapine, may also be used to manage manic or mixed episodes. Antidepressants are not recommended as first-line treatment for bipolar disorder, as they can trigger manic or hypomanic episodes, but they can be used in combination with mood stabilizers to treat depressive episodes.

Lithium medication has the strongest evidence for long-term relapse prevention (Geddes & Miklowitz, 2013). The metal, was introduced by John Cade in 1949 (Cade, 1949), has been in clinical use for over 50 years and the evidence base for its efficiency comes from randomised controlled trials, where lithium was compared to a placebo drug. A meta-analysis of 5 placebo-controlled lithium randomised controlled trials (n=770) showed that lithium reduces the risk of manic relapses by 38% and depressive relapses by 28% (Geddes et al., 2004). Lithium is also one of two known anti-suicidal treatments which has shown a reduction in the risk of suicide by 50% (Cipriani et al., 2005). However, the benefits of lithium are restricted by its adverse effects which include polyuria, polydipsia, tremor, weight gain, cognitive impairment, sexual dysfunction and skin lesions (e.g. acne, psoriasis). Other risks involve hypothyroidism, increased blood calcium and slightly increased risk of renal failure (British National Formulary (BNF) 2020) . Other epilepsy drugs are also used as mood stabilisers in the treatment of BD including lamotrigine and sodium valproate (Bowden, 1998). A recent systematic review in 2021 indeed found that valproate was more effective than placebo in preventing new BD episodes and not significantly different than lithium (Yee et al., 2021). There is also consistent evidence of the effect of lamotrigine in the treatment of specifically depressed symptoms of BD, where the overall effect was modest but significant over placebo (Geddes et al., 2009).

For manic or hypomanic episodes, the National Institute for Health and Care Excellence (NICE) recommends anti-psychotic medication including haloperidol, olanzapine, quetiapine and risperidone (National Institute of Health and Care Excellence, 2014). Antipsychotic treatments are also increasingly used as a maintenance treatment, either as a single therapy or adjunct. Their pharmacological properties, efficacy and tolerability in bipolar disorder are varied. This is particularly the case for prevention of depressive episodes, where their clinical effects appear less pronounced (Jauhar & Young, 2019). They also have undesirable side effects to mood stabilisers include weight gain and sedation (Ertugrul & Meltzer, 2003).

2.5.2 Psychological treatment

Treatment guidelines for BD suggest that the most effective management of BD includes the integration of psychopharmacological treatment, with adjunct, targeted psychological treatment. Evidence-based models of psychotherapy include cognitive-behavioural therapy, family-focused therapy, interpersonal social rhythm therapy, group psychoeducation and systemic care management. Though these evidence based models are substantially different in terms of methods, assumptions and structure (Miklowitz et al., 2008), they have common objectives.

Common objectives for psychological treatment for BD include (Geddes & Miklowitz, 2013):

- Improve the ability to identify and intervene early with warning signs of recurrences via self-awareness
- Increase acceptance of condition and drug adherence
- Enhance ability to cope with environmental stressors associated with symptoms
- Psychological treatments also aim to stabilise sleep and wake rhythms and other daily routines
- Re-engage with social, familial and occupational roles
- Enhance family relationships and communication
- Reduce drug or alcohol misuse

Improving skills around self-awareness in order to recognise and respond to early warning signs is a large underpinning for psychotherapy for BD, typically delivered in the form of psychoeducation. In a recent review in 2021, it was shown that psychoeducation as an adjunct to pharmacological treatment leads to a reduction of new episodes, length of hospital stay and adherence to drug therapy proving the positive nature of self-awareness in management of BD (Rabelo et al., 2021).

In summary, treatment and management of bipolar disorder typically involves a combination of medication, therapy, and lifestyle changes. Medication is the primary treatment, and proven psychological therapy, such as psychoeducation, can help individuals with bipolar disorder learn to identify and manage their symptoms and improve their ability to function. The essence of psychoeducation is to improve self-awareness of their condition, so that patients can identify signs of relapse and act accordingly. This research focuses on this concept of self-awareness, particularly on the role of technology in cultivating this via self-tracking.

2.6 The role of self-tracking in BD

Self-tracking is core component in cultivating self-awareness in BD, which is the aim of proven psychological modalities for the treatment of BD. Self-tracking, also known as self-monitoring, is a process of collecting and analysing data about one's own health and behaviour. This can include tracking symptoms, medication adherence, sleep patterns, and other relevant information. In the context of bipolar disorder, self-tracking can play an important role in managing the condition by providing individuals with more insight into their symptoms and how they may be affected by different factors (Matthews, Murnane, & Snyder, 2017a).

In her book, Emily Martin (Martin, 2009) describes some of the underlying mechanisms of self-tracking, stating that many people are encouraged to keep a “mood chart”. Martin argues that the process of doing so can have dramatic effects for people by making experiences comparable, and to thus understand changes over time. Relatedly, understanding the factors that are being tracked and their relationship to self-management is also important. Research suggests that individuals with BD look towards life events or life stress as a factor to track however there is a lack of a relationship between tracking this and mood episode initiation (Johnson & Roberts, 1995). As the individual has more frequent bipolar episodes, the relationship between a life event or stress and a relapse into mania and depression tends to become weaker (Hlastala et al., 2000). The main issue is that monitoring stress or life events may be useful but many episodes are not preceded by life events or stress so it would be incomplete. Particularly useful in capturing when self-tracking is clinical EWS. EWS arise prior to full scale episodes, and EWS interventions aid individuals to detect these, allowing action to be taken to prevent a full episode (Lobban et al., 2011). The evidence suggest that a minimum of three of four early signs can be used but more than six are likely to be necessary for an effective intervention (Morriss, 2004). EWS are particular to the individual but—from a clinical view—are reliably similar from episode to episode in the same person. In a Cochrane systematic review in 2010, it was found that time to first recurrence of any type (RE, hazards ratio 0.57, 95% CI 0.39 to 0.82), time to manic/hypomanic episode, time to depressive episode, and percentage of people hospitalized, and functioning favoured the group who received an EWS intervention compared to control. Depressive nor hypomanic symptoms differed between intervention and control groups. Furthermore, more recent evidence suggests that tracking EWS show positive results in detecting transitions to mood episodes (Bos et al., 2021).

In 2011, Lobban et al. developed a EWS checklist which describe a set of 63 unique symptoms and signs that may arise approximately 2-4 weeks before a full manic or depressive episode and are listed in checklist form (Lobban et al., 2011). Prior to this, identification of EWS required a clinical interview where most EWS were idiosyncratic. The checklist was developed to aid in this process and resulted in patients recognizing a greater number of EWS, increasing the median number from one to sixteen for depression and to eleven for mania; thereby increasing the chances of recognizing relapse. The most identified early signs for depression were low motivation, low energy, feeling tired/listless; most endorsed for mania were feeling energetic/very active, more talkative, and feeling emotionally high.

Self-tracking EWS have proven utility in the management of BD, as indicated by the research (Lobban et al., 2011; Morriss, 2004; Morriss et al., 2007), as well as using the EWS checklist which proves a useful aid in recognizing EWS which moves beyond individual recall to a systematic understanding of these relapse signatures (Lobban et al., 2011). Given its clinical validity, EWS prove to be a useful mechanism to self-track in order to aid better self-management of BD. There is limited research on how individuals with BD currently self-track. A core study by Murnane et al. (Murnane et al., 2016) explored self-monitoring practices, attitudes and needs of individuals with BD using a survey with 552 users. They found that individuals reported that they primarily self-tracked items such as mood, sleep, finances, exercise and social interactions using modalities such as social feedback, mentally, technology and paper to aid in the process. Over 63% of respondents reported that tracking via any method had impacted their approach to maintaining their health, 11% reported no impact and 26% were unsure of the impact on their health. When looking at the benefits of self-tracking, in line with the aims and objectives of EWS interventions and psychological interventions in general, it was found that self-tracking made health management easier, promoting self-awareness and supported interactions with healthcare providers. Matthews et al. (Matthews, Murnane, & Snyder, 2017a) takes this one step further and used user interviews (n=10) to explore how and why users engage in self-tracking, asking how the experience of BD influences self-tracking practices and what role technology plays in supporting this. It was found that users used a combination of methods to self-track, with the goal of maintaining "*normalcy and self-stabilisation*". It was also found that individuals seek to identify negative patterns early on to anticipate mood episodes, performing their own personal analytics to make connections between their collected data between behaviours, symptoms and warning signs. Again, echoing the core elements of psychoeducation.

Both studies also report on several considerable challenges that individuals face in their attempts to self-tracking. Users experience considerable challenges in self-tracking as their sense of self and emotional state is in "*flux, uncertain and unreliable*". In relation to tracking factors, users described there is a need for methods (such as technology) to sufficiently capture BD-specific behaviours (such as EWS) at the level of granularity needed for BD management due to the nature of the illness (e.g., reporting multiple moods in one day). The very nature of BD poses issues as self-tracking becomes inherently difficult during mania or depressive episode where they simply feel too ill to engage in self-tracking. Furthermore, fluctuating mood states in turn can cause "*confusion, contraction and self-doubt*" (Inder et al., 2008) which in turn can impact the reliability of self-assessments. In addition, research indicates that personal insight is dependent on emotional state in BD, which is more error prone during periods of mania (Goodwin & Jamison, 2007). In the study by Murnane et al (Murnane et al., 2016), there was a specific call for increased passive, automatic self-tracking methods which reduce the burden that users can experience during such times, and supports the continuous collection of data.

In conclusion, self-tracking can play an important role in the management of bipolar disorder by providing individuals with more insight into their symptoms and how they may be affected by different factors. Particularly useful in tracking are EWS, which have proven clinical utility in improving the management of BD. As well as using checklists which can aid individuals to recognize more EWS, when compared to relying solely on subjective recall. Despite the known benefit of self-tracking, little research has been done in understanding the processes and mechanism of self-tracking in BD. Two core studies report on this and describe the subjective experiences of self-tracking to promoting self-management, especially understanding the connection between behaviour, symptoms and early warning signs (Murnane et al., 2016; Matthews, Murnane, & Snyder, 2017a). However, the very nature of BD poses issues around self-tracking as the sense of self is consistently in "*flux*" where mood variations can create erroneous self-report, especially during stares of mania. There is call for more automated, passive methods of self-monitoring which account for these challenges and create an avenue for the continuous collection of data.

2.6.1 Self-tracking technology for BD

Given the evolution of pervasive mobile technology, it is inevitable that novel approaches have been leveraged to aid individuals with bipolar disorder to monitor their symptoms and manage their condition. This includes wearable devices, such as smartwatches or fitness trackers, that track physical activity, sleep patterns, and heart rate, as well as mobile apps that allow individuals to track mood, medication adherence, and other aspects of their condition. Additionally, some self-tracking tools can be integrated with telehealth platforms, allowing individuals to share their data with their healthcare providers in real-time (Batra et al., 2017).

More generally, smartphone apps focused on mental health are increasing in number (Anthes, 2016). There are approximately 10,000 mental health and wellness apps available for download for mental health diagnosis, treatment, and support. Self-monitoring apps are predominant, was found that most applications for serious mental illnesses fall into this category (Batra et al., 2017). It is increasingly common for such technology to use a combination of active and passive sensing methods to aid in self-management. Active sensing refers to self-report measures in the form of mood monitoring, validated symptomology questionnaires, diary entries etc where the users are asked to share such information at specified time intervals. Whereas passive, automatic sensing methods have the potential to monitor behaviour, detect changes in mood and to provide feedback about EWS without the need for users input. Research suggests that individuals are attracted to technology that automatically collects data in the background, as described by Murnane et al (Murnane et al., 2016) as this method responds to the challenges that individuals experience during the course of active sensing, where the nature of BD poses significant issues in self-report.

A recent systematic review by Ortiz et al (Ortiz et al., 2021) highlights the following types of passive mobile data collection which are being leveraged to promote self-tracking and prevent relapse in bipolar disorder: such as sleep duration and efficiency; number of steps walked; number of drinks per day; speech features; number and duration of calls; number of texts sent; number of characters in a text; number of social interactions per day; number of times screen was on; location (GPS); light exposure; battery use; typing kinetics; metadata. Active measures were described as the following: reported quality or changes in sleep; mood; anxiety, stress; energy, activity levels; early warning signs of relapse; irritability; speed of thoughts, impulsive thoughts or actions; adherence to medications; cognitive problems. Validated self-report measures have also been leveraged such as: ASRS: Altman Self-Rating Mania Scale (E. G. Altman et al., 1997), HDRS: Hamilton Depression Rating Scale (Hamilton, 1960), PHQ: Patient Health Questionnaire (Kroenke et al, 2001), QIDS: Quick Inventory of Depressive Symptomatology (Rush et al., 2003) and YMRS: Young Mania Rating Scale (Young et al, 2000). Table 3 summarises the current mobile technology that has been leveraged for self-management of BD as adapted from Ortiz et al, 2021.

Active measures (A) and passives measure (P) collected via mobile devices	**P:** speech-based classifiers	**P:** call duration, number of speakers turns; and speech features	**A:** mood, energy; 5 social rhythmic core activities. **P:** text/calls log, ambient light, screen unlocks, GPS	**A:** mood, energy **P:** number and duration of calls, number of texts per day, GPS	O: speech recognition	**A:** Mood **P:** GPS	**A:** Mood **P:** GPS, activity levels, light exposure, duration of phone calls, length of text messages	**A:** Mood, energy, rapidity of thinking, impulsive actions and impulsive feelings **P:** typing kinetics	**A:** mood	**P:** Keystroke entry date and time, accelerometer displacement (collected via a custom keyboard)
Rating scales used	HDRS	Not specified	Social Rhythm Metric Scale	HDRS, YMRS	HDRS, YMRS	QIDS	QIDS	HDRS, YMRS	QIDS, ASRS	HDRS YMRS
Duration of follow up (months)	6-12	3	1	Up to 12	7	2	4	0.5	2.5	1
N	6	6	7	13	37	22	20	18	146	9
Main study objectives	To investigate whether electronically collected data can predict clinical change	To investigate whether electronically collected data can predict episodes	To investigate whether electronically collected data can be used to infer stability and rhythmicity using the Social Rhythm Metric Scale	To investigate whether electronically collected data can predict clinical change	To investigate whether electronically collected data can predict clinical change	To investigate whether electronically collected data can predict clinical change	To analyse whether group-personalized regression models for predicting mental health scores can be harnessed from objective smartphones data streams	To investigate whether electronically collected data can predict clinical change	To understand whether electronically collected data can predict manic and depressive symptoms	To investigate whether electronically collected data can predict clinical change
Author and year of publication	(Karam et al., 2014b)	(Muaremi et al., 2014)	(Abdullah et al., 2016a)	(Beiwinkel et al., 2016)	(Gideon et al., 2016)	(Palmius et al., 2017)	(N et al., 2018)	(Stange et al., 2018)	(Vazquez-Montes et al., 2018)	(Zulueta et al., 2018)

A: Mood **P:** light exposure, activity, sleep and heart rate	**A:** Mood, energy, speed of thoughts, impulsivity and physical pain, sleep duration, number of awakenings, time out of bed	**A:** Mood and anhedonia **P:** mobility distance and radius, call duration and number of texts and lengths (in 24	**A:** Mood, activity, sleep duration, medication taken, mixed mood, irritability, anxiety, cognitive problems and	**A:** mood **P:** light exposure, activity, sleep and heart rate
Not specified	Not specified	PHQ	HDRS YMRS	
24	0.5	0.5	9	12
37	10	93	84	36
To investigate whether electronically collected data can predict clinical change	To understand whether variability in mood, cognition, energy and impulsivity and their relationship to sleep, pain and social and task-based stress	To investigate whether electronically collected data can predict clinical change	To investigate the feasibility of predicting daily mood by objectively collected data	To evaluate the effectiveness of an app in conjunction with a wearable to predict and prevent mood episodes based on circadian rhythm data
(C. H. Cho et al., 2019)	(H. Li et al., 2019)	(Pratap et al., 2019)	(Busk et al., 2020)	(Cho et al., 2020)

Table 3 Review of current technology for self-tracking using active and passive data adapted from Ortiz et al, 2021 (Ortiz et al., 2021)

The quantitative aspects of the studies report on the associations between the active (self-report) and passive (automatically collected data) as well as report on the predictive quality of the combination of data on clinical change or episode onset. For the former, quantitative data is usually presented in the form of correlation coefficients between the data sources, which measures the strength of a linear relationship between them. In absence of this, coefficients are taken from any available regression or multilevel models, which indicate the degree to which the passive data explain the changes in the active data. For prediction of clinical changes or episode onset, performance metrics are taken from machine learning models trained to the passive data to predict clinical changes or onset of relapse. The study found four commonly reported metrics, which are outlined below:

- Sensitivity: measures the model's capability to predict the symptom presence
- Specificity: measures the model's capability to predict symptom absence
- Accuracy: measures the proportion of correct predictions from the total number of predictions)
- Area under the curve (AUC): which measures the model's ability to distinguish between the presence and absence of clinical symptom

When looking at the impact of these technologies in promoting self-management in BD, the results largely vary and it was summarised that predicting mood episodes for this condition remains a difficult challenge (Ortiz et al., 2021). Looking at this is more detail, it was revealed that studies to date were largely heterogenous, which measured different variables using different methods with small sample sizes and brief length of follow up. From a statistical perspective, the predictive models had higher AUC when it comes to predicting manic or hypomanic episodes. Certain passive features, such as keyboard accelerometer displacement was a predictor or both manic and depressive symptoms whereas distance travelled as a negative predictor (indicating the probability a person is not having an episode) of mania.

It was noted that BD is a heterogenous condition where the individual experience of the condition varies from person to person and that future studies should consider this, especially in the context of statistical analysis. Individual variability was simply not considered, even though it acts as a strong potential predictor for relapse. To bring this into the context of previous section of this research, one should consider EWS and how heterogenous they can be. They are 63 items that an individual can choose from, which act as “relapse signature”. These signatures are particular to the individual, they largely vary from person to person (Lobban et al., 2011) indicating the complex nature of understanding relapse in individuals. It opens up the question of essentially how much of individual variability is accounted for in the whole ecosystem of this type of technology for BD? Is individual variability being accounted for in the design, research or implementation of this type of technology? Are we truly understanding the everyday experiences of this condition and tailoring technology accordingly? The answers to these questions will be explored in the next section of this thesis.

2.6.2 Tensions within self-tracking technology for BD

As indicated in the review by Ortiz et al (Ortiz et al., 2021), individual variability is not considered within current self-tracking technology for BD. It was concluded that technology should be designed for the tailored monitoring of symptoms of BD. The current section of this thesis explores this further to truly understand the role of individual variability in the design of self-tracking technology. To do this, one must simply focus on the user of the technology, the patient, and understand the holistic experience of using this type of technology. As a first step, Nicholas et al (Nicholas et al., 2017a) explored consumer perspective of the current marketplace for mobile apps for BD. The findings from this study echo the overarching argument from Ortiz et al (Ortiz et al., 2021) that the individual experience of BD is not being considered. Firstly, they found that a large percentage of users’ needs being met, which was indicated by the proportion of reviews which contained user wish lists. This opens up the question as to whether users are involved in the design prior to the enrolment of such technology. Secondly, it was reported that apps are being developed independently of research data, and without reference to clinical guidelines. When exploring this in relation to the above, we found that only a few of the technologies mentioned in the study had reference to EWS, not only being a clinical framework within proven utility (Lobban et al., 2011) but also a framework which accounts for the individual variability present when understanding the complex nature of relapse. Developing technology in this way carries significant risk. Without understanding the individual experiences of BD, nor following good clinical practice, we fall into a realm where technology may create more harm than good. For example, how much is current technology accounting for the individual challenges that someone may face when undertaking self-tracking activity. In the key paper by Matthews et al (Matthews, Murnane, & Snyder, 2017a), it was found that too much tracking can be counterproductive for some, which was similarly echoed by Murnane et al (Murnane et al., 2016) who described the downward spiral that the overuse of technology can lead to. Murnane et al (2016) took this one step further to describe that there must be caution exercised on the emotional and behavioural patterns that self-tracking technology can present, and how they can be triggers for some individuals. For example, telling one individual about a social "*off*" week may result in positive behavioural change, or it may fuel rumination, depression and anxiety in another. The summary is that in the advent of this issues in the landscape of developing technology for BD, these highly important individual differences are not being accounted for, undermining the potential power of technology in improving the lives of those with BD.

The next question is how do we as researchers and/or designers account for this individual variability in the experiences of BD in technology? The first step is to look at how we are developing this technology. Murray et al (Murray et al., 2016) have argued that to establish and optimise the use health interventions, principles of user-centred design (UCD) are required. The application of participatory approaches, such as user-centred design principles and activities (ISO 9241-210:2010) are becoming standardised more increasingly, an example being by the National Health Service (NHS) in the UK; but also, internationally in private and public health and the industries who provide them, such as medical technology and pharmacological companies. Given that UCD is a commonly adopted approach, we first need to explain how it is described to understand this in relation to other approaches of user involvement adopted. The principle is outlined below and further illustrated in Figure 1.

1. The design is based upon an implicit and explicit understanding of users, tasks, environments and interactions in context
2. Users are involved throughout design and development
3. The design is driven and refined by user-centred evaluation
4. The process is iterative
5. The design addresses the whole user experience
6. The design team includes multidisciplinary skills and perspective

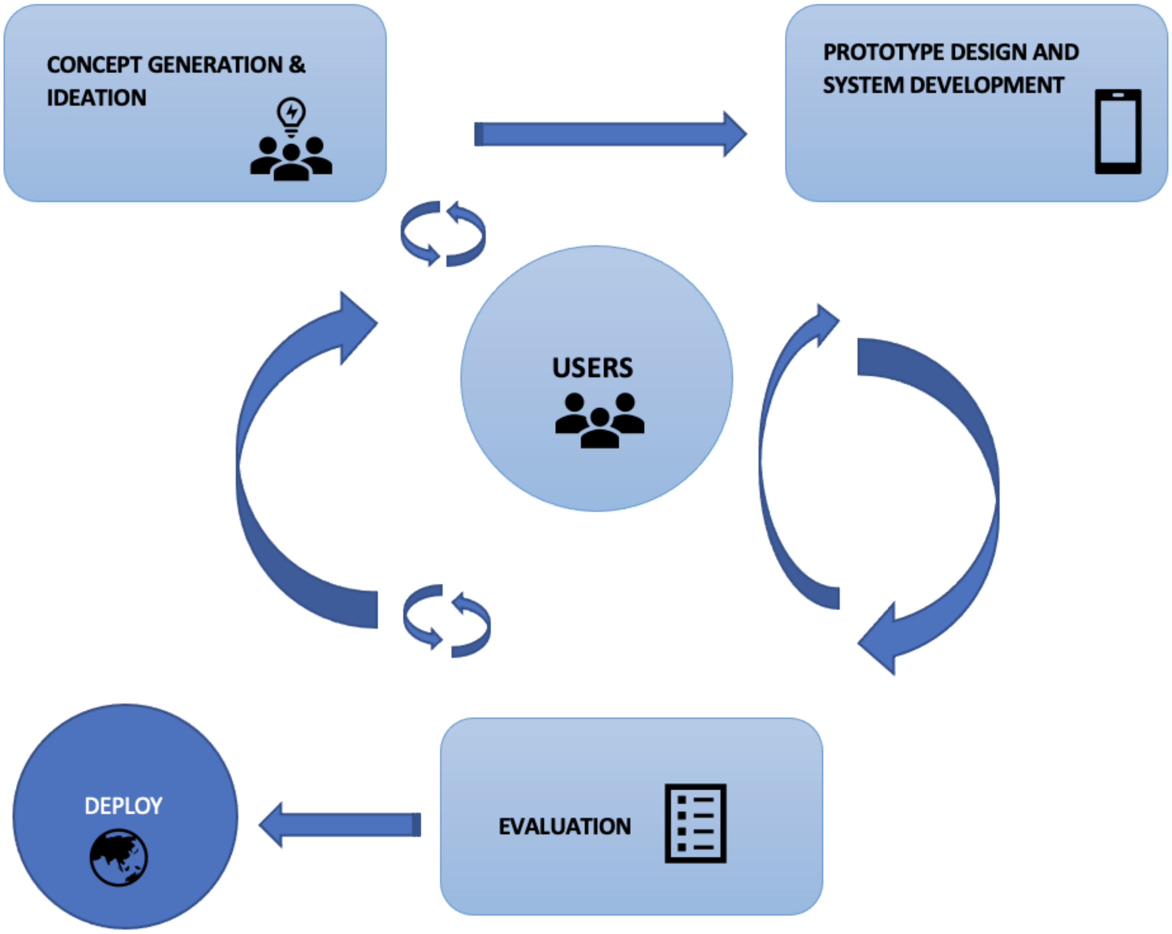


Figure 1: User-centred design process (adapted from McCurdie et al., 2012

To measure the level of individual variability being accounted for, a first step is to look at the landscape of user involvement approaches in the design and evaluation of self-monitoring applications for BD. This is investigated via three specific questions a) are users being involved in the design and evaluation of technology? b) if so, how is this happening? and lastly c) what are the best practice 'ingredients' regarding the design of mental health technology? The next stage on this thesis explores these questions in the form of a literature review.

2.7 User involvement in the design and evaluation of self-monitoring applications for bipolar disorder: literature review

2.7.1 Introduction

This section explores the landscape of user involvement approaches in the design and evaluation of self-monitoring applications for BD. This is investigated via three specific questions a) are users being involved in the design and evaluation of technology? b) if so, how is this happening? and lastly c) what are the best practice 'ingredients' regarding the design of mental health technology?

In regards to question a), Goodwin et al. (Goodwin et al., 2016) states that there is a lack of parity of user involvement in the design of physical and mental health applications, where for mental health, users are involved less frequently when compared to physical health. A recent review (Sanches et al., 2019) examined the last decade of studies in affective health (including bipolar disorder) and human-computer interaction (HCI). User-involvement was considered, and it was identified that more ethically sensitive design practices, including the voices of people living with affective disorders need to be integrated. Out of the n=139 publications included in this study, only n=16 of the studies reviewed reported "clinical evaluations" described as involving service users of mental health services, or who meet the formal criteria for a specific mental health problem. This lack of user involvement is reflected in the quality of mental health apps for BD, as Nicholas et al (Nicholas et al., 2017b) established that a significant proportion of apps contained wish list requests, indicating that users' needs are not being met by current app design.

Understanding how a serious mental health condition like BD has an impact on daily experiences is important and helpful when designing technology in order to create a meaningful technological experience that account for individual variability. A study which looked into the pathology of bipolar disorder (Matthews et al, 2015) described the following unique design considerations for mobile technology which have been reported in the literature: 1) the side-effects of medication (such as Lithium) on a user's ability to read on-screen text 2) the impact of medication non-adherence in BD and how this may impact engagement 3) sensitivity to reward-based stimuli in BD and how this can be used for novel interventions and 4) the association of increased creativity in BD and the suggestion that such a user group could contribute greatly compared to other users in involvement orientated type design. In terms of creativity, there are inherent neurobiological similarities between individuals with mood disorders and those engaged in creative pursuits, supporting the concept of temperament contributing to heightened creativity in people with bipolar disorders (Santosa et al, 2007)

In regard to question b) it is important to understand that user involvement in the design of such technology involves an intersection between healthcare and technology development, both of which traditionally have different approaches when involving the user, though there are exceptions to the rule. Patient and public involvement (PPI) dominates as concept for 'involvement and engagement' within healthcare studies (Jennings et al., 2018) and improvement. Whereas UCD dominates within HCI or service design more broadly and integrates participatory approaches to achieve involvement and engagement of end users. Both of these approaches look to provide a voice to patient end users within applied development projects. It is important to consider these approaches and how they are applied, paying particular attention to where they overlap and where the tensions lie, some of which is outlined in Figure 2. This is especially important when looking into the future towards best practice, as per question c).

2.7.2 Current research

There is a growing body of research on the design of mental health technology with user input, but little has been done to compare approaches of user involvement to understand best practice. This literature review specifically focuses on understanding the current landscape of user involvement in the design of self-tracking technology. Specifically, we aim to understand what practices are being used and to what extent they are involving the user. We acknowledge that there can be multiple stakeholders, users in and "users" of mental health technology but as self-reflection and awareness are the main aim of self-tracking systems, in this instance we define “users” as those who have a diagnosis of bipolar disorder who are the primary user of this technology. To explore these practices, we review available literature for self-tracking technology for bipolar disorder and make an overall assessment of the level of user-involvement in design. The findings from this review are used to form best practice “ingredients” for design of mental health technology. This combines the already existing practices of PPI and HCI in order to evolve from the generic guidelines of user-centred design and to ones that are tailored towards mental health technology.

2.7.3 Background

Healthcare based design approaches

PPI is described as the involvement of patients, carers and the public as “active partners” in the design, delivery and dissemination of research to ensure its relevance and usefulness. In other words, “*research is carried out ‘with’ or ‘by’ members of the public rather than ‘to’, ‘about’ or ‘for’ them” (National Institute of Health Research, 2012)*. Unlike the design of mental health technology, guidelines exist on best practices for PPI involvement and measuring its effectiveness. In terms of practices, there is a continuum of PPI involvement, which has close links to Arnstein’s Ladder of Citizen Engagement (Arnstein, 1969). Involvement occurs at different levels and each level has a corresponding level of effort, commitment and potential impact or outcome. This ladder stretches from tokenism to being full embedded, where patients are the more dominant voice, delivering and managing the research themselves. The lower ends of this ladder consist of researchers asking for users’ views, which are then used to refine key study documents, such as recruitment materials or to inform research decision making but does not go so far as to involve the PPI representatives as partners in research activities. In line with best practices, PPI users should be remunerated for time and effort on projects, though this is variable depending on the research and stage of project (Jennings et al., 2018). PPI organisers, often PPI users themselves, can also be listed as co-authors on published works. In terms of effectiveness, a systematic review of PPI in health and social research identified the following as benefits of PPI involvement: enhanced quality and appropriateness of research, development of user-focused research objectives, user-relevant research questions, user-friendly information, questionnaires and interview schedules, appropriate recruitment strategies for studies, consumer-focused interpretation of data and enhanced implementation and dissemination of study results (Jennings et al., 2018).

PPI processes are often used in health technology assessment studies (Abelson et al., 2016), where the most commonly used approach involves patients and/ or members of the public directly participating on committees on the agency involved in technology assessment, which involves the systematic evaluation of the properties, effects and/or impacts of health technology. PPI is also being increasingly used in mental health technology development in recent years (Liverpool et al., 2019; Hall et al., 2019), with some more citizen focussed approaches. PPI is a more passive ex(ercise held at timed points in the research rather than continuing involvement with direct interaction with technology on an ongoing basis when compared to more technology-based approaches in HCI which are more iterative and cyclical, as described below.

Human-Computer Interaction design approaches

Technology-based approaches in the discipline of Human-Computer Interaction have developed a "*powerful vocabular*y" (Baumer, 2015) when it comes to involving the user in design evolving from user-centred design (UCD) to more participatory democratic forms, considerations of non-use, design fictions, critical engagements and various other approaches. In particular, Orlowski (2016) (Orlowski et al., 2016) touched on two specific methods when designing mental health technologies with users, which were utilised in two case studies. First is Participatory Design (Spinuzzi, 2005), a Scandinavian-born practice which is characterised by a 3-stage iterative design process aimed an understanding user's implicit knowledge: (1) exploration of work (2) discovery processes

(3) prototyping. Each stage is organised and carried out with users. Another method mentioned by Orlowski (2016) (Orlowski et al., 2016) stemming from UCD is Design Thinking (M. Altman et al., 2018). Developing empathy for users is at the heart of Design Thinking as well as working in collaborative multidisciplinary teams and using "action-orientated rapid prototyping" of solutions. Like Participatory Design, Design Thinking is an iterative process which includes several rounds of "need-finding", "ideation" and "implementation". The interesting part of this cycle is the "need-finding", which is focused on developing empathy for the users and asking questions, such as "who are we trying to help and what are the social, political and economic contexts?". When comparing these approaches in HCI to healthcare-based approaches, such as PPI, participatory design is often referred to as "co-design", similar to medium-level involvement as described in PPI. Design thinking takes this one step further by focusing on the development of empathy and achieving that parity in voice which is often described being absent in PPI. Both approaches are also iterative, which is needed in the development of technology, rather than the linear process in PPI. Also, by being technology-based approaches, there is already acknowledgement of the various tensions, such as cost and regulation and how to embed these into the process of development which are detailed in relevant regulations (ISO 9241-210:2010 and ISO/IEF 62366).

2.7.4 Methods

2.7.4.1 Study Design

Search strategy

For this review, we used two methods of search. The papers included arose from the hits of searches from an ongoing systematic review that doctoral student SM was working on which was looking into user preferences for self-monitoring technologies for central nervous system disorders which includes bipolar disorder as a central nervous system disease. This systematic review is registered on the international prospective register of systematic reviews (PROSPERO 2019; CRD42019139319) and used the following search terms: ehealth OR mhealth OR 'digital health' OR telehealth OR telemonitoring) AND ('remote patient monitoring' OR 'remote monitoring technology' OR application OR wearable OR app OR device) AND ('central nervous system' OR psychiat* OR neurolog* OR neurodegen* OR 'mental health' OR chronic) AND (prefer* OR evaluation OR feedback OR usability OR design OR visual*) AND patient AND (data or symptom OR UI OR "user interface") AND (disease OR disorder OR condition)) on the following databases: Association of Computing Machinery, PubMed, EMBASE, IEEE Explore and Web of Science and the Cochrane Library for papers published in the English language between 2007 and 2019. During abstract search, those papers related to bipolar disorder and self-monitoring technology form part of the current review. A further check of the literature was done on Google Scholar to review for missing papers using the following search for papers published in the English language between 2007 and 2019: (bipolar AND app*)

Inclusion criteria

The following paper inclusion criteria was used:

a) Sample of users with bipolar disorder
b) Feature novel self-monitoring technology

Information extraction

Papers were screened and information was extracted and forms part of Table 4: number of users with BD, description of self-monitoring technology, description of user-involvement methods, model of user-involvement and further description of model of user-involvement. This information was assessed according to criteria of involvement, as described below, the results of which are also outlined in the table.

2.7.4.2 Assessment criteria for user-involvement

The assessment criteria of user-involvement are based on the best-practice model of user-involvement which describes that users should be involved in the concept generation and ideation stage, prototype design and deployment stage and evaluation stage, with mechanisms for iteration as described earlier in (Figure 1). To also embed PPI processes, we explained the empowerment of decision making and/or creating empathy as part of high user-involvement as this was described as being high on the PPI continuum of involvement (Jennings et al., 2018). The criteria are further detailed in Table 4.

No user-involvement mentioned [a]	No mentioned of user involvement in design and evaluation
Low user-involvement [a]	Users were only involved in one stage of design and evaluation with or without iteration
Medium user-involvement [a]	Users were involved in more than one stage of design and evaluation with iteration
High user-involvement [a]	Users were involved in the concept generation and ideation stage, prototype design and deployment stage and evaluation stage with iteration, likely to have explicit mention of empowering decision making and/or creating empathy

Table 4 Criteria for assessing user involvement in selected papers

[a] The categories used here have been developed specifically for this paper are different to the degrees of involvement in designing and running a research study as used in PPI practices

2.7.4.3 Expert panel review

The expert panel consisted of an individual with 7+ years of lived experience of bipolar disorder and expertise in PPI, Professor of Psychiatry, Assistant Professor of Human-Computer Interaction, Assistant Professor in Human Factors, Research Fellow in Involvement and Implementation, Senior Research Data Scientist and doctoral student SM. Included papers and criteria of assessment were subject to discussion with the expert panel via bi-monthly online meetings with doctoral student SM over a period of a year. As the inclusion criteria for the study was only two-fold, there were no disagreements over included papers. There were also no disagreements over the criteria of involvement as the papers fell distinctly into these based on extracted information, which is limited by what was reported in the studies. Information extracted from the papers was reviewed and findings were discussed at meetings in order to make inferences over what constitutes best practice in this context which was based on each individual members subjective opinion of what constitutes best practice based on the experience in this field.

2.7.5 Results

The results from the literature search are presented in Table 5 including number of users, description of remote monitoring technology (RMT), assessment criteria for user-involvement, further description of methods employed and model of user involvement. Overall, n=3 studies fell into the category of no user-involvement mentioned, n=3 fell into the category of low user-involvement, n=4 fell into the category of medium user-involvement and n=2 in high user-involvement.

2.7.5.1 Expert panel review

For question C, which is what are the best practice ‘ingredients’ regarding the design of mental health technology, the findings of this paper have been discussed within a group of individuals who offer academic and clinical expertise in this area as well as an individual with lived experience of bipolar disorder who also has a good level of experience and participation in PPI. The discussions revolved around the findings from this review, as well as recommendations for the future based on the subjective experience of each panel member. Based on these discussions, the following recommendations were put forward:

1) Involve users in all stages of design and evaluation including concept generation and ideation, prototype design and deployment and evaluation stages with a goal to create user empathy and/or empowerment. This process should have an adequate number of users in order to welcome diversity in thought. Equal representation is also a crucial consideration that needs to be taken into account when recruiting users.

2) Ensure early involvement as this will be cost-effective in the long run (avoid re-design and problems with use and implementation in the later stages)
3) Combine principles of PPI and HCI to not only have users to assist in designing technology but also in designing and running research (e.g., users co-facilitating design workshops). Utilising end-to-end user involvement.
4) For academic and industry sectors to establish better mechanisms to access target user groups with lived experience of mental health issues for example by building relationships with existing patient directed organisations such as charities, patient-led community groups etc.
5) Increase awareness of HCI and design communities in PPI principles and practices and increase awareness of PPI community in HCI and design methods/ skills
6) Encourage use and mixing of formal scientific/ design methods with informal experiential and empathic practices to capture richness in understanding dynamic requirements of technology users which are cognizant of use in context
7) Keep the user informed at all stages of the process, including final outcomes, future use, next steps etc which is often forgotten about

Study reference	**N of Users**	**BD- specific?**	**Description of self-monitoring technology**	**Assessment criteria for user-involvement**	**Further description of methods employed**	**Model of user-involvement**	**Further description of user-involvement model**
No user-involvement mentioned (n=3)							
(Palmius et al., 2017)	N= 22		Combination of True Colours Monitoring system and customised app that records geographical location	No mention of user-involvement in design and evaluation			
(Tsanas et al., 2016a)	N= 48	No – borderline personality disorder also included	Mood Zoom (MZ) smartphone questionnaire	No mention of user-involvement in design and evaluation			
(Schärer et al., 2015)	N=118	Yes	Personal Life-Chart App: electronic diary-based smartphone application	No mention of user-involvement in design			
Low user-involvement (n=3)							
(Niendam et al., 2018)	N= 76[a]	No – Psychosis also included	Ginger.io: smartphone based mental health tracking application	Users involved in evaluation stage	○ Users completed non-standardised measures for satisfaction and perceived impact on clinical care	None mentioned	
(Fortuna et al., 2018)	N = 1	No – other serious mental illnesses also included	PeerTECH: self-management smartphone application	Users involved in evaluation stage	○ Feasibility and acceptability of PeerTECH was established by user capacity to use the app and complete self-management tasks	None mentioned	
(Bush et al., 2015)	N = 18[a]	No - other serious mental illnesses also included	Virtual Hope Box: smartphone application	Users involved in evaluation stage	○ Questionnaires were administered to a clinical population for ease of set up, usage, helpfulness, benefits and whether they would recommend the app, employing a combination of Likert-type	Agile development process	Performed agile development due to minimal risk involved. Performed iterative and incremental usability tests of developing features and components using a small group of target users

					rating scales and open-ended questions. ○ Feedback was used to modify the design of the Virtual Hope Box		
Medium user-involvement (n=4)							
(Schuurmans et al., 2016)	N/A – this is a protocol	No – other severe mood disorders	eCare@Home (ECH): tablet-based self-monitoring tool	Users involved prototype design and evaluation stage with iteration	**Prototype design and deployment stage:** ○ Three rounds of interviews with 8 users where interactive demo materials and screenshots were provided as stimuli and feedback was used to iterate design **Evaluation stage:** ○ Credibility and Expectancy Questionnaire (CEQ), System Usability Scale (SUS), Client Satisfaction Questionnaire (CSQ-8) administered to measure system usability, user experiences and client satisfaction.	Co-creation approach	Aim of the approach was to create a product that would be usable for the specific target population and move away from traditional rigid 'waterfall' methods which only have a single round of assessment or iteration. This approach impacted the tool by uncovering usability requirements which were implemented
(Abdullah et al., 2016b; Matthews et al., 2016)	N= 7	Yes	MoodRhythm: smartphone application that can track social rhythms	Users involved in protype design and evaluation stage with iteration	**Prototype design and deployment stage:** ○ Users used the app and shared feedback, design insights and suggestions for improvement at least once a week. Wireframes were sent back to users which incorporated this feedback where further feedback was given **Evaluation stage:** ○ Post study usability scale using System Usability Scale (SUS)	Participatory design process	During the design process, users used the MoodRhythm app in their daily lives and shared their feedback, design insights, and suggestions for improvements to the app. This process allowed users to provide feedback on an ongoing basis during the design process and helped to identify and address concerns that users might have about these technologies, ensuring the app was effective for daily use.

(Bardram et al., n.d.-b) (Alvarez-lozano et al., 2014b) (Osmani et al., 2015b, Faurholt-Jepsen et al, 2016)	N= 42 (all papers combined)	Yes	MONARCA system: combination of passive and active self- monitoring smartphone application	Users involved in prototype development and evaluation stage with iteration	**Prototype design and deployment stage** ○ Three-hour workshops were held for design and iterative prototyping where feedback was incorporated into design **Evaluation stage:** ○ System Usability Scale was administered in a field trial. A non-standardised questionnaire for usefulness and perceived usefulness was also developed and administered	Patient-Clinician Designer Framework (PCD) using principles of user-centred design	Through this design process, users were *involved* in making decisions about system features using collaborative design workshops. The design of the MONARCA system employs a mobile phone application as the main component
(Hidalgo-Mazzei et al., 2016, 2018) (Hidalgo-Mazzei et al., 2017)	N = 303 (all papers combined)	Yes	OpenSIMPLE: smartphone-based psychoeducaton programme	Users involved in prototype development and evaluation stages with iteration	**Prototype design and deployment stage:** ○ Users were involved in focus groups, interviews and surveys with research teams. Unclear how findings were used to iterate the prototype **Evaluation stage** ○ Engagement was calculated based on weekly percentage of completed tasks. Usability was calculated using the System Usability Scale (SUS) and satisfaction and perceived helpfulness Likert scales.	User centred design	Using the user centred design approach, suggestions were incorporated based on feedback from the users during the feasibility study as well as modifications to adapt the platform for an open study. Several features were added to OpenSIMPLE based using this approach
High user-involvement (n=2)							

(Buitenweg et al., 2019)	N = 59 [a]	No – other serious mental illnesses also included	QoL ME: smartphone based, personalised quality of life (QoL) assessment app	Users were involved in concept generation and ideation stage, prototype design and deployment stage and evaluation stage with iteration with a goal to empower patient decision making	**Concept generation and ideation stage** ○ 10 users were to share their experiences with smart devices, apps and quality of life questionnaires and to ideate about QoL ME in a focus group **Prototype design and deployment** ○ Paper sketches (wireframes) were presented and were gradually refined, expanded and made function where a first protype was developed. 25 users were involved in this stage **Evaluation stage:** ○ Prototype was subjected to usability testing and systematically assessed using the System Usability Scale (SUS) with a total of 25 users. ○ Goal to empower patient decision making ○ No information was found in relation to this	Co-creation approach	The QoL-ME was co creatively developed in an iterative development process with groups of people with severe mental health The process consisted of 6 iterations divided over 3 stages: (1) brainstorming stage, (2) design stage, and (3) usability stage. The development process was described as fitting in the framework of participatory design. Feedback was used to make several changes to QoL ME
(Bauer et al., 2017)	N=25 [a]	No – also included post-traumatic stress disorder	Spirit App: smartphone self-monitoring application that allows patients to undertake modules and complete questionnaires for mental health assessment	Users were involved in concept generation and ideation stage, prototype design and deployment stage and evaluation stage with iteration with a goal to	**Concept generation and ideation stage:** ○ One focus group was run with users to propose the SPIRIT app and a second focus group was ran to refine SPIRIT app concept **Prototype design and deployment stage:** ○ Focus groups were run with user to elicit feedback on storyboard and	Human-centred design process, participatory design process and Principle of Digital Development	Target users and domain experts were engaged in a participatory design process throughout development via focus groups and usability testing with national consumer advocacy groups and providers and patients in rural clinics. The process also adhered to the Principles of Digital Development which includes:

				empower patient decision making	prototype and prototypes was refined based on feedback **Evaluation stage:** ○ Developed a usability testing framework which was conducted with 5 users where feedback was incorporated into the app ○ Goal to empower patient decision making and/or creating empathy ○ The study had an advisory group called CAB which consisted of eight “consumers” and “consumer advocates” who met monthly to advise the SPIRIT scientific team on all aspects of trial design and conduct, which was resulted in changes to the app and study.		1. Design with the user 2. Understand the existing ecosystem 3. Design for scale 4. Build for sustainability 5. Be data driven 6. Use open standards, open data, open source, and open innovation 7. Reuse and improve 8. Address privacy and security 9. Be collaborative

Table 5 Results table for literature search

2.7.6 Discussion

2.7.6.1 Principal Results

This research presents the current landscape regarding user led design and evaluation of self-monitoring applications for bipolar disorder. This was investigated via three specific questions a) are users being involved in the design and evaluation of technology? b) if so, how is this happening? and lastly c) what are the best practice 'ingredients' regarding the design of mental health technology?

For question a) in total 19 papers were included in this review, which resulted in the evaluation of 13 novel smartphone applications for self-monitoring for bipolar disorder. In total, six of these papers have been grouped together as they have been published in respect to the same smartphone application, which is highlighted in the results table. In regard to question a), the results from this review indicate that users are being involved in design and evaluation, but this is highly variable in terms of level of involvement. In total, 3 self-monitoring applications (n=4 papers) were classified as having no mention of user involvement in design, 3 self-monitoring applications (n=3 papers) were classified as having low user-involvement, a further 4 self-monitoring applications (n=9 papers) were classified as having medium user involvement and 2 self-monitoring applications (n=3 papers) were classified as high user involvement, which was the lowest frequency of user-involvement category in this review.

In respect to question b), there is variability in the models of user involvement in design and evaluation where the following have been described: agile development process, co-creation approach, participatory design, patient-clinician designer framework, user and human-centred design and principles of digital development. The key characteristics of these models are described in Table 6. A standout method mentioned was "agile development process" which is a software development process aimed at producing outcomes fast in relation to market constraints and the ability to accommodate changes during the software development cycle (Cao et al., 2009). It's use case in mental health technology is sparse and there are questions as to how a model aimed at quickly meeting market constraints considers the users voice and needs during the process of designing mental technology, which sheds light on why it was ranked low in terms of involvement as it only considered users in the evaluation stage. Finally, this large amount of variability sheds further light on the need for quality guidelines in the reporting of user-involved development of mental health technologies.

Model of user-involvement	**Key characteristics**
Agile development process	An iterative approach to project management which is aimed at product fast outcomes in relation to market constraints and the ability to accommodate changes during the software development cycle
Co-creation approach	A process which is aimed at creating *with* users/stakeholders to ensure results meet their needs and are usable
Patient-clinician designer framework	A process which uses the key principles of user-centred design to be applied in the context of mental health. A framework which aims to involve patients and clinicians in the process of design through collaborative design workshops and iterative prototyping
Human-centred design	A process which is based on designing based on characteristics and intricacies of human psychology and perception which is considered to carry out a deeper analysis that user-centred design
Principles of digital development	A process which focuses on the following nine principles during digital development: 1. Design with the user 2. Understand the ecosystem 3. Design for scale 4. Build for sustainability 5. Be data driven 6. Use open data, open standards, open source and open innovation 7. Reuse and improve 8. Address privacy and security 9. Be collaborative

Table 6 Key characteristics of user-involvement

The descriptive section of the table, which describes how models of involvement were implemented as outlined in selected papers, also uncovers pertinent findings to question b). It was found that those papers that were classed as high user involvement displayed increased level of detail on how they implemented their chosen user involvement model. Additionally, those papers that were classed as high user involvement not only described their methods as mainly participatory design, but also described a combination of methods such as co-creation, participatory design and human-centred design. This highlights some early suggestions that it is the combination of methods which could be the driving force of ideal user-involvement which underpins points 5 and 6 of our recommendations of best practice as outlined in the results section. These points describe that there should be a mixing of methods to capture the unique and dynamic requirements of mental health technology users and that there should be an increased awareness of these methods in both the HCI and PPI design communities.

Focus groups were the method of choice during concept generation and ideation stages. For prototype design and deployment stage, the following methods were shown to have been employed: focus groups and workshops, sharing of wireframes and interviews. For evaluation stages, there was a combination of the use of standardised and unstandardized questionnaires to measure factors such as usability, satisfaction and usefulness of the smartphone application. For a few studies, completing tasks using the application was also a method used for evaluation. Only one study which met the high user involvement criteria had explicit reference to empower patient decision and/or creating empathy by having a patient advisory board, whose role was to advise the scientific team of all aspects of the study and smartphone application. Some studies (Abdullah et al., 2016a;Matthews et al., 2016) had particularly low user numbers (ranging from 1-7 users) and there are questions as to whether such small samples can adequately capture users’ needs/wants and whether this constitutes a user-focused approach. The aims of these methods are to represent diversity in this voice and capture both an implicit and explicit understanding of users, tasks, environments and interactions so that technology can be designed better, and it can be argued that such small samples cannot provide the clinical, and industrial settings need to work collaboratively to establish mechanisms to enable technology development and the contribution of appropriate users to be accessible, inclusive, and representative.

From a more systemic perspective, this intersection between PPI and HCI may shed more light on the variability of user involvement both in this study and more generally. Both approaches combined describe the umbrella of the types of user involvement methods described in this review. However, both approaches individually have tensions that need to be considered, which are likely to impact the application of these methods. PPI approaches have limitations in that they are currently not versed in considering design within the tensions of cost and regulation, as PPI processes are often the result of existing government funding for research, rather than commercial and industrial funding, which is more typical of technology development. Tensions arise because of the differences in time management and resource allocation depending on the funding source, and there are questions regarding the adaptability of PPI practices for this. These practices tend to be focused on the clinical context and clinical task and are not versed in considering more subjective, hard-to-measure, and tacit aspects when designing technology such as user experience and everyday life practices, which are facets removed from direct clinical care. PPI approaches tend to traditionally linear, static approaches that do not evolve or iterate owing to new information, which is not suitable when designing technology where iteration is a requirement. This is because the origins of PPI do not stem from design or scientific disciplines, such as HCI, where the elicitation of need-finding is not just limited to the anticipated as it is in PPI but also unanticipated or implicit in nature [40]. For example, when reviewing user preferences on data visualization for remote monitoring technology, BD was touched upon, and it was found that the state of readiness and state of health as well as data literacy and familiarity with technology are all factors when considering user engagement with remote monitoring technology (Polhemus et al., 2020). The consideration of factors, such as state of readiness may not be readily captured by PPI.

For HCI approaches, unlike PPI processes, there is a lack of use of these formal methods in the context of mental health service and technology design. Therefore, there is little evidence of their effectiveness (Orlowski et al., 2016). PPI processes based on and in health care provision and improvement work are often conducted by professionals who have a skillset targeted toward working and engaging with those with mental health issues, whereas this cannot be said in the field of HCI where the training background is largely different with little or no experience in mental health. It is unclear whether this is a benefit of technology-based approaches as it has the potential, if planned and implemented well, to remove the power dynamics and hierarchy by not having a clinician taking lead which can sometimes negatively impact meaningful PPI contributions. Alternatively, it may be that if not designed with empathy and in conjunction with the advice from clinical persons or those with lived experience, HCI approaches could be a hindrance and a barrier to disclosure and engagement if the nonclinical professionals do not have the skillset to meaningfully engage those with mental health issues. This research highlights the need to upskill both communities to be better equipped, and future research should aim to explore this. Conclusively, user-focused approaches can provide a framework for PPI to embed participatory activities within the iterative, fast-paced development process of mental health technology development. Likewise, PPI has developed core standards around establishing an equal relationship between users and researchers, which can lend itself well in HCI approaches where this is not necessarily present.

2.7.6.2 Limitations of the review

The limitations in this paper may also contribute towards the lack of user-involvement mentioned in this paper. This paper reflects the current practices of user-involvement to the extent to which authors made this explicitly available in the chosen literature. It could be the case that not all authors disclosed the process of design in the paper for a variety of reasons. First of all, with strict word limits in the case of often complex papers in the field of mental health technology, authors may have decided to focus on other parts of the technology such as results and omit design and development of the technology. Likewise, the process of design could be described in other papers which may not have been picked up in this review. We did not contact authors to check if there was additional literature on user design nor interview authors about user design in the development of technology, but such practices might produce a more comprehensive review of user design practices in the future. It is also worth noting as a limitation that only research-led app development projects were assessed and there may be innovations in commercial and non-profit developments that have not been considered in this paper as only published literature was considered. Lastly, there may be limitations around the recommendations of best practice provided by the group of experts and future studies should consider more structured tools such as the Delphi method.

2.7.6.3 Conclusions

In conclusion, the current research provides evidence that despite recommendations on the involvement of user in the process of mental health technology design and evaluation, in this case specifically for bipolar disorder, there is large variability on whether the user is involved, how they are involved and the extent to which there is authentic empowerment of the voice of the user. This lack of user involvement in the development of technology echoes the findings from Ortiz et al (Ortiz et al., 2021) in which individual variability of experience of BD is not being considered in the development of such technology as adopting this approach is a core step in understanding this. The tensions between the design approaches used in PPI and HCI may shed some light on why there is variability of user involvement. Currently, both the design approaches work independently however future practices should aim to work together and encourage the awareness and mixing of methods. The findings of this research have been reviewed by an expert panel, including an individual with lived experience of bipolar disorder and recommendations were made for the design communities to establish better mechanisms for awareness, mixing of methods and increased user involvement.

2.8 User preferences for data visualisation for self-tracking technology in BD: a literature review

2.8.1 Introduction

To delve further into the account of individual variability in the design of self-technology in BD, we specifically investigated the role of user preferences in design particular for data visualisation in a further, systematic review of the literature. As defined by Matthews et al (Matthews, Murnane, & Snyder, 2017a) the experiences of the patterns of behaviours, symptoms and early warning signs that are revealed by self-monitoring in BD are highly individualistic, with the potential to promote preventive behaviours in some and ruminative unhealthy behaviours in others. The way this data is presented back the user is crucial in investigating, as it essentially determines whether technology plays a promoting role in self-management, or disruptive. Designs of high quality data visualisation of users require a sound understanding of individual variability, which can be achieved by understanding user perspectives and needs in design process, as indicated in the previous literature review which explored this (Majid et al., 2021). The review explores this in a systematic procedure, focusing on user preferences for data visualisation for those living with severe mental health conditions, with a focus on BD.

2.8.2 Methods

2.8.2.1 Identifying Relevant Studies

This protocol was registered on PROSPERO (International Prospective Register of Systematic Reviews, CRD42019139319) while the review was in its pilot phase [21]. We searched PubMed, IEEE Xplore, EMBASE, Web of Science, proceedings from the Association for Computing Machinery Conference on Human Factors in Computing Systems, and the Cochrane Library for original, peer-reviewed, or gray literature published in English between January 2007 and September 2021. Searches included combinations of terms such as mHealth, along with terms related to data visualization and neurological disease. Relevant papers were also identified from manual searches of included studies' reference lists. Studies were screened in 2 stages: abstract screening and full-text review. Eligibility criteria and screening forms were piloted on a set of 50 abstracts and 15 full-text reviews, and criteria were clarified or amended as needed. Two reviewers independently screened each abstract. In the case of disagreement, the abstract automatically proceeded to the full-text stage. Two reviewers then independently assessed each full-text paper for eligibility, and disagreements were resolved through discussion. If no consensus could be reached, a third member of the review team reviewed the paper and made a final determination. Agreement between reviewer pairs was determined through Cohen kappa. Deduplication, record management, and screening were conducted in CADIMA, an open-access systematic review software (Kohl et al., 2018).

2.8.2.2 Eligibility Criteria

We included studies if they met the following criteria:

- All or part of a study population was living with a mental health condition
- Users were ≥18 years of age
- Self-monitoring technology for laypeople to track, monitor, or manage their own health was investigated
- Results of any qualitative methods or integrated syntheses of mixed methods were reported
- Patient perspectives on visualizations of health or wellness data were reported

The following studies were ineligible for inclusion:

- Conditions that were not neurological in nature or not associated with mental health
- Perspectives on interface design, intervention design, or any component of RMT design unrelated to data visualization
- Visualizations limited to medication adherence or non–health-related data

- Perspectives of caregivers, health professionals, or others not living with a neurological or mental health condition

We purposefully remained broad in our definition of the term “data visualization” because it is understood differently by different people. Therefore, we included any format through which self-tracking technology displayed data to service users. We defined chronic mental health as depression, anxiety, or bipolar disorder.

2.8.2.3 Data Extraction, Critical Appraisal, and Qualitative Synthesis

Two authors independently reread each included study and extracted quotes related to data visualization preferences. When available, screenshots of data visualizations were also extracted. To ensure that analysis remained grounded in the context of the original studies, data extraction forms included a detailed description of each study’s objectives and methods, and annotated PDFs were preserved. Studies were critically appraised with the Mixed Methods Appraisal Tool (Pace et al., 2012). We then categorized studies as conceptually rich “key papers,” “satisfactory papers,” which are methodologically acceptable but provide only moderate value to the synthesis, and “fatally flawed papers,” which contain major methodological flaws (Dixon-Woods et al., 2007). We also noted “minimal impact papers,” which provided minimal contribution to the synthesis. Following a reading of the extracted text and its context, 2 authors independently coded data line by line, producing a draft coding frame. The coding frame was iteratively amended, refined, restructured and the data recoded until no additional codes or disagreements were identified, and categorized into “descriptive themes,” which described the structure and content of the codes. Analytical themes, which interpreted the coded data, were developed through iterative rereading and discussion of the codes and thematically organized data. Results

2.8.3 Results

2.8.3.1 Included studies

Searches returned 2928 unique records. Of these, 177 papers were included in full-text review and 35 were eligible for qualitative synthesis. Reviewer agreement was moderate during abstract screening (weighted κ=0.45) and substantial during full-text review (weighted κ=0.79) [22]. Our relatively low agreement during abstract screening was expected and mitigated by reviewing the full texts of all abstracts, which were judged eligible by at least one reviewer. Multiple papers were identified for 3 research projects, as indicated in the previous literature review: the SPIRIT study (n=3), the MoodRhythm app (n=2) [35,36], and the MONARCA (Monitoring, treatment and prediction of BD episodes) project (n=2) [37,38]. For these projects, all identified papers were analysed as 1 study. Thus, 31 unique studies were included. We identified 3 themes through content analysis: desire for data visualization, impact of visualizations on condition management, and visualization design considerations.

2.8.3.2 Desire for data visualisation

In 21 out of the 31 included studies, users expressed a desire for data visualisation including: mood and disease-specific symptom scores, physical activity, sleep patterns and quality and rhythm of daily activities.

2.8.3.3 Impact of visualisation on condition management

Generally, the results indicated a positive experience of the impact of visualisations using self-tracking tools on condition management. It was reported the data visualisations assisted in user recall of their experiences, especially considering the difficulty of reflecting on past experiences which has been indicated in other key studies such as (Matthews, Murnane, & Snyder, 2017a) (Murnane et al., 2016). It was reported that especially in the case of BD, and schizophrenia where the nature of the conditions can impact the perception of reality (Matthews et al., 2015), objective visualisation of retrospective symptoms presence can offer a sense of validation. Furthermore, visualisation could drive more engagement with self-tracking technology as especially in the case of BD and insomnia, data and insights could be perceived as rewards (Matthews et al., 2015). In a study with BD, it was indicated that the lack of data visualisation was thought to lead to study dropout (Bauer et al., 2018). Finally, users often described a desire of data visualisation in order to communicate with healthcare professionals, caregivers and others, irrespective of whether the technology was designed for this purpose.

2.8.3.4 Visualisation design preferences

Users often described expressed conflicting preferences for data visualisation (Bardram et al., 2013) (Murnane et al., 2016) (Buitenweg et al., 2019), indicating a considerable design challenge. Several studies reported the use of graphical preferences were preferred by most users but personal preferences varied highly (Bardram et al., 2013) again echoing the high level of individual variability when it comes to self-tracking technology for BD. To further support this, some users suggested that colour and images were powerful tools to increase engagement and meaning (Matthews et al., 2015), however some users described that this could trigger emotions related to self-image (Bauer et al., 2017) (Matthews et al., 2015) (Bardram et al., 2013). Users also discussed the value to adding contextual, qualitative information when interpreting visualised quantitative data (Bauer et al., 2017) (Bauer et al., 2018) (Matthews et al., 2015), such as the ability to annotate the data (e.g. via notes) as to provide individual context. In terms of timeliness, preferences for data visualisation varied from real-time, daily or even monthly feedback again reinforcing the high level of individual variability. Several studies concluded the importance for variability in designs, allowing users more control over the time and type of data they wish to see.

2.8.4.5 Conclusion

When designed appropriately, accounting for individual variability by employed user-centred approaches, data visualisations can potentially increase value and engagement and act as a highly appreciated component of self-monitoring technology for BD. The current research uncovered there is a desire for data visualisations, and they have important benefits for condition management such as assisting in user-recall and communications with healthcare professionals and caregivers. However, as demonstrated by design preferences for visualisation, a blanket approach when it comes to self-tracking technology may not be appropriate due to the contrasting and conflicting preferences of this user group. The ability to tailor and personalise technology to individuals' experiences, needs and preferences can potentially maximise the utility of technology. A route to this can be through the appropriate engagement of the user throughout technology development, design and deployment.

2.9 Chapter Summary

BD is leading cause of disability worldwide with significant impact on a person's quality of life and ability to function causing devasting social and economic consequences. The condition is characterised with marked and extreme mood disturbances from symptoms of mania, hypomania and depression ranging in severity, which define the major classification of the illness into BD1 and BD2. Research suggest that specifically looking at EWS, using aids such as EWS checklists, have proven utility in decrease rates of hospitalisation, time to episode and improved functioning for those with BD. However, the very nature of this condition poses challenges when it comes to self-tracking such as decreased self-awareness during period of being unwell (Lobban et al, 2011).

In the advent of pervasive technology, novel mobile technological solutions have been leveraged to aid in self-tracking in BD accounting for the challenges individuals may experience. Many of these technologies combine active (i.e., self-report) as well as automatically collected passive (i.e., activity, sleep patterns, heart rate etc) sensing techniques to provide insight into relapse for BD. There is some, albeit limited, proven utility of this type of technology in understanding relapse thus far. However, results are largely varied with a lack of accountability for the individual variability in the experience of this condition and how this may impact the development and design of technology. A method to explore individual variability and gain insight into users' needs and preferences are user-centred design approaches, as recommended in the literature. A literature review was conducted to investigate the current landscape of user-centred design approaches for current technology and results indicated that the adoption of these techniques is currently highly variable, with a significant proportion of studies reporting no-user involvement in the development and design of technology. Demonstrating that currently, individual variability is not being accounted for in this domain. This literature review also consisted of an expert review, who listed best practice principles to guide the design of self-tracking technology for BD. A second literature review was conducted on design preferences for self-tracking technology in the context of data visualisation for BD, where it was echoed that there is no “one size fits all” for the design of technology. It was concluded that technology should be tailored according to users’ needs and preferences, in order to maximize the potential that technology has in order to improve the lives of those living with this condition.

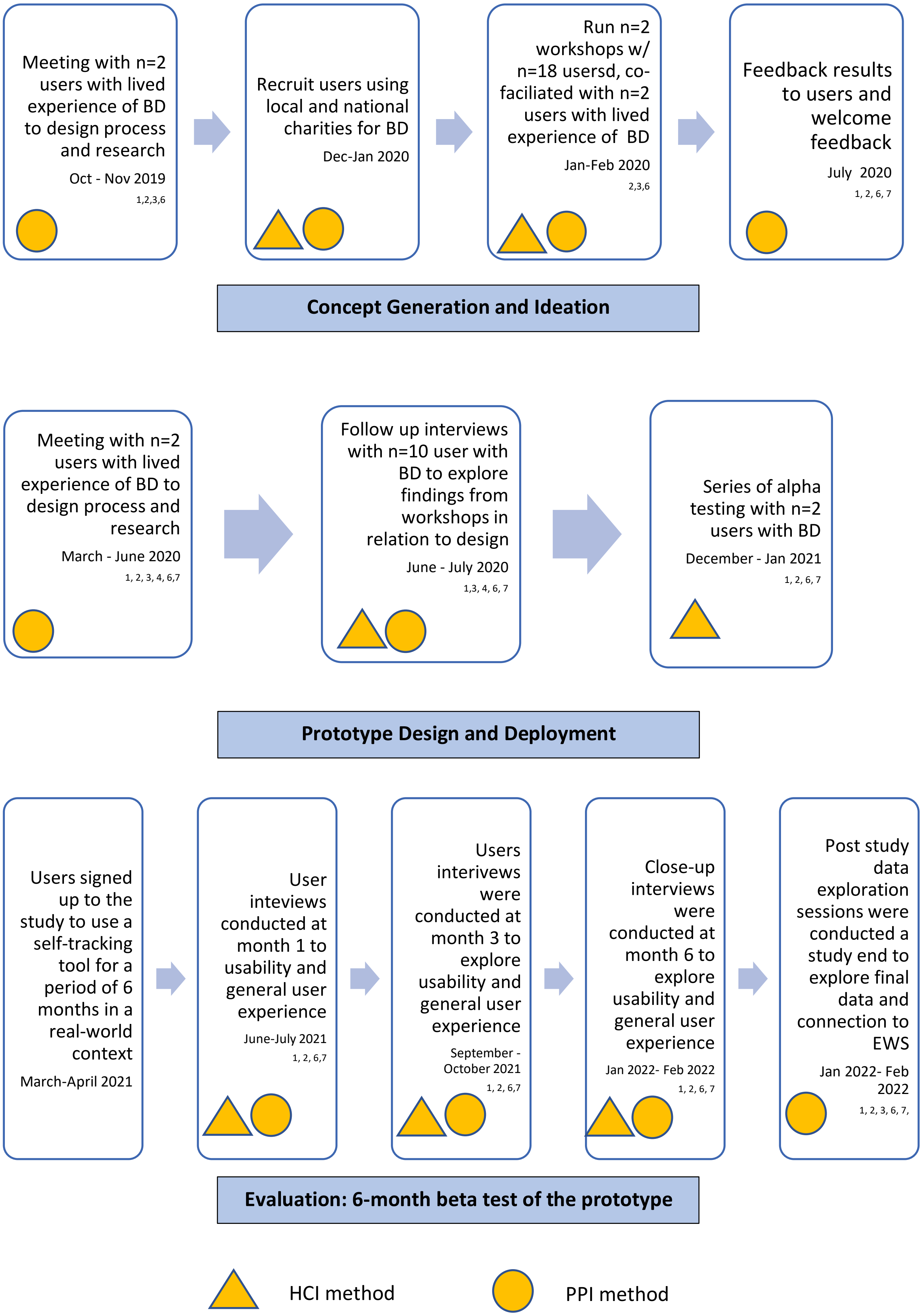
Chapter 3: Methodology

Chapter summary

- Provides a brief overview of the study design that was employed for the respective studies which will be further defined in respective study chapters: Chapter 4: Concept Generation and Ideation, Chapter 5: Prototype Design and Deployment and Chapter 6: Evaluation: 6 month beta test of the Prototype
- Provides an overview of the self-tracking tool that was employed in the studies.
- Provides an overview of qualitative and quantitative analytical methods used in this research.

3.1 Study Design

This research employed a mixed method approach to bring user's personal and lived experience of BD to the forefront in order to design and assess to assess a mobile self-tracking tool which uses passive and active sensing techniques to understand EWS for relapse for those with BD. In order to account for the variability in illness experience and to understand users' needs and preferences throughout the course of development. A full description of the study design is provided in Figure 2, indicating where the elements of the study were drawn from (e.g., HCI and/or PPI) and how they relate to the seven recommended principles.


Meeting with n=2 users with lived experience of BD to design process and research
Oct - Nov 2019
1,2,3,6
Recruit users using local and national charities for BD
Dec-Jan 2020
Run n=2 workshops w/ n=18 usersd, co-faciliated with n=2 users with lived experience of BD
Jan-Feb 2020
2,3,6
Feedback results to users and welcome feedback
July 2020
1, 2, 6, 7
Concept Generation and Ideation
Meeting with n=2 users with lived experience of BD to design process and research
March - June 2020
1, 2, 3, 4, 6,7
Follow up interviews with n=10 user with BD to explore findings from workshops in relation to design
June - July 2020
1,3, 4, 6, 7
Series of alpha testing with n=2 users with BD
December - Jan 2021
1, 2, 6, 7
Prototype Design and Deployment
Users signed up to the study to use a self-tracking tool for a period of 6 months in a real-world context
March-April 2021
User inteviews conducted at month 1 to usability and general user experience
June-July 2021
1, 2, 6,7
Users interviews were conducted at month 3 to explore usability and general user experience
September - October 2021
1, 2, 6,7
Close-up interviews were conducted at month 6 to explore usability and general user experience
Jan 2022- Feb 2022
1, 2, 6, 7
Post study data exploration sessions were conducted a study end to explore final data and connection to EWS
Jan 2022- Feb 2022
1, 2, 3, 6, 7,
Evaluation: 6-month beta test of the prototype
HCI method
PPI method

[1,2,3,4,5,6,7] Seven recommendations from Majid et al (2021)

Figure 2 Full study methodology

3.2 Self-tracking tool

The technology employed in this study is titled RADAR-Base and has been developed by the industrial sponsors of this research (National Institute of Health Research - Nottingham Biomedical Research Centre).

RADAR-Base stands for Remote Assessment of Disease and Relapse, and is an open-source mHealth platform able to leverage data from wearable and mobile technologies (https://radar-base.org/) which is outlined in Figure 5, as has proven applications in major depressive disorder, epilepsy and multiple sclerosis (Ranjan et al., 2019)

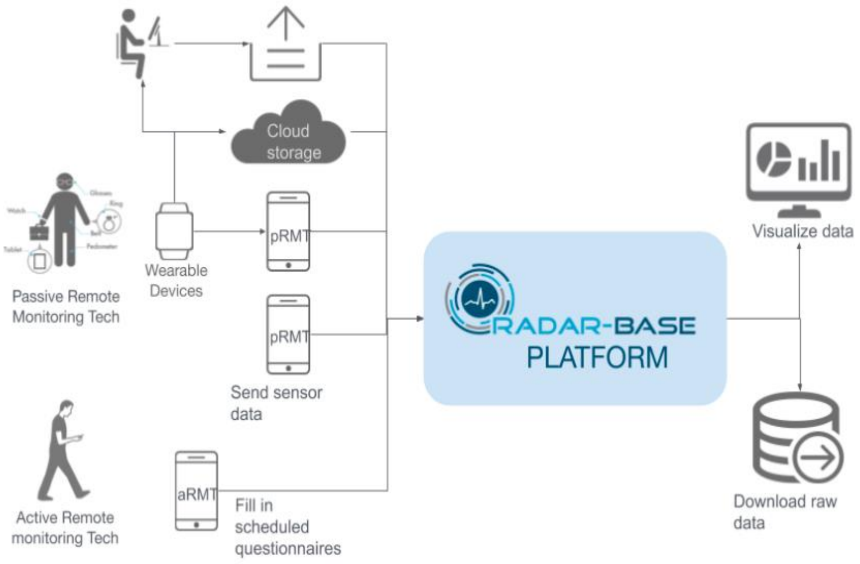


Figure 3 Outline of RADAR-Base Platform

Like other types of mobile self-tracking technology for BD (Ortiz et al., 2021), RADAR-Base employs both active and passive sensing methods. It consists of two applications, one for passive data collection *pRMT* and one for active data collection *aRMT*. The aRMT app allows for self-reported information to be inputted. The *pRMT* application allows for the passive collection of the following data sources:

- Phone usage
- Local weather
- Step count
- GPS location
- Which apps are being used and for how long
- Battery level
- Other Bluetooth devices in the vicinity
- Sleep information and heart rate (provided by user Fitbit connection if possible)
- Number of contacts on phone
- Ambient light

Though RADAR-Base has a proven use case in other mental health conditions such as depression (C. L. Stewart et al., 2018), it has not yet been explored in the context of BD. By adopting the close participatory design approach in this research, we aim to adapt RADAR-Base to be closely aligned to user's needs, preferences and individual variability. The process of which will be outlined in Chapter 5.

3.3 Methodology overview

An in-depth description of each research phase and associated methodology will be reported upon in later chapters in this thesis. However, in the interests of reader orientation, a brief description of overall methodology will be reported upon in this chapter.

3.3.1 Concept Generation and Ideation

Overview: This phase of the research employed two workshops to explore reasons, methods and challenges for self-tracking with n=18 user across two locations in the UK. Both workshops were three hours in total, with a one-hour lunch break in between. Workshops started with an icebreaker exercise followed by a three-part structure involving a series of interactive exercises.

Rationale: By establishing close relationships with two specific BD charities (Bipolar UK and Bipolar Lift CiC) we were able to explore methods that users would be comfortable with for meeting and sharing experiences. We tried to ensure our methods reflected current practices employed by these organizations. Both organizations run frequent workshops and group events which was the reason behind our choice of data collection.

3.3.2 Prototype Design and Deployment

Overview: This phase of research employed virtual follow up interviews (n=10) to explore findings from the workshops more specifically in relation to design, in order to create a working prototype ready for testing. When the prototype was ready, this was subject to a series of alpha testing to explore usability, user experience and functionality of the self-tracking tool.

Rationale: As this part of the research was conducted during the outbreak of COVID-19, there was a limited pool of methodological approaches that we could apply. The follow up sessions were initially meant to conducted in-person, in a subsequent workshop however they were conducted as virtual follow up interviews instead. The option to run a virtual workshop was considered however wgiven that the user number fell on the higher side (n=10) (Brown, 2022) as well as the novelty of online workshops at the given time, it was decided that interviews were the stronger methodological choice. As exploring usability was a core goal for this phase of this research, an alpha testing approach was employed with n=2 users. Alpha testing is a type of usability testing which is conducted at the early stages of development, to uncover early stage usability issues prior to full development (*Alpha Test*, n.d.) within test subjects who are internal to the development, and is also described as being within the process of development of mental health applications

3.3.3 Evaluation: 6-month Beta Test of the Prototype

Overview: The final prototype was enrolled for beta testing in a real world setting with n=8 users for a period of 6 months. The users engaged with both the passive and active sensing methods that made up the final prototype which is outlined in Chapter 5. The aims of this research was to explore the self-tracking tool in a real-world context; therefore, we provided the tool to be used in everyday life, aiming for minimal research interference unless during these specified time points, which users were aware of from the point of study sign up. Specifically at month 1, month 3, month 6 as well as after study completion to explore the self-tracking tool in an interview format. The interviews were aimed towards understanding the experience of using the self-tracking tool, paying particular focus towards usability and user preferences. The interviews conducted at the 3-month and post-study points explored the above but had an additional focus on data visualisation.

Rationale: Beta testing is a similar process to alpha testing, however users test a developed protype in a real world context, and is being more increasingly used in the general context of mental health technology (Shore et al., 2014) (Lewis et al., 2021) (Rickard et al., 2016) (Birrell et al., 2022). This underpinning our methodological choice at this phase of the research. Interviews were selected as an appropriate method to follow up as the aims of the research were to uncover individual variability, therefore required a method of data collection that facilitated this.

3.4 User recruitment

For the Concept Generation and Ideation phase, user recruitment was achieved through the assistance of both national and local charity organizations involved in this study who sent bulk messages to mailing list members who may be interested in the research, as well as other snowballing techniques via an existing PPI team based at the Institute of Mental Health at the University of Nottingham.

For the Prototype Design and Deployment phase of this research, the same users who were involved in the subsequent phase of the research were involved in the interviews as well as n=2 users who were involved in the series of alpha testing.

For the Evaluation phase of this research, a number of users who were involved in subsequent parts of this research were also involved in this phase which included testing the prototype in a real-world setting. A cohort-style user approach arose from this phase of research which will be reported upon this in this thesis. Further users were recruited through associated charities such as Bipolar UK.

3.5 Study Analysis

For qualitative analysis that was employed in this study, Braun and Clarkes thematic analysis approach (Braun & Clarke, 2006) was used. A reflexive thematic analysis approach was taken as it allowed the researcher to determine the outcome and focus of the findings, as opposed to being bound to a specific theoretical framework – as classified as Big Q thematic analysis. An inductive theoretical approach was taken to the analysis so that themes were derived directly from the data themselves. In terms of the themes, they were identified at a semantic level in order to identify and summarise the content of the data and capture the surface meaning (i.e., reflects what was explicitly said) - this was to keep as close as possible to the users account which is the primary purpose of this thesis. The thematic approach consisted of a six-step approach of: familiarizing yourself with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report.

In some cases, affinity mapping was used as a tool to visualise the thematic analysis results for those interview questions that had briefer, summary responses. Affinity mapping is prevalent tool in user experience research (Kara, 2019) used to discern patterns in notes and observations, thereby creating cohesive patterns or themes. Affinity mapping techniques is a quicker method compared to thematic analysis, and was used in areas where responses were brief and didn’t require a full thematic analysis, in case where a summary of the data was needed.

For quantitative analysis, we employed data analytic and statistic tools offered by Python and R programming libraries to understand the connections, if any, between the active and passive data. Using statistical methods to explore the connections between the active and passive is typical for analysis for this technology as outlined by Ortiz et al (2021).

Chapter 4: Concept Generation and Ideation

Chapter summary

- Limited research has investigated the everyday practices of self-tracking for BD, and it is unclear how the normative ontology that is seen in existing self-tracking technology discourses (e.g., the Quantified Self movement) is applicable to the domain of mental health
- To address this, this study explores reasons and methods ^{RQ1} for self-tracking, and challenges and opportunities for future mobile technology ^{RQ2}. To do this, and in line with PPI principles, we designed our study closely with individuals with lived experience of BD where a series of workshops and interviews were performed to explore how and why those with BD self-tracked.
- These results of this study contributed towards the knowledge for self-tracking in BD. The findings documented a high degree of complexity and individual variability in self-tracking, where over 50 methods of tracking were described. We suggest that future design in this domain should convert these complex experiences, accounting to this high level of variability.

3.1 Introduction

Self-tracking is a key method that many of those with BD employ to manage their everyday lives, extending into the use of digital devices like mobile phones to support a wide range of self-tracking practices (Murnane et al., 2016). While self-tracking has received some attention within HCI (I. Li, Dey, Forlizzi, et al., 2011; Luo, 2021)—particularly in the guise of the 'quantified self' movement (QS) (Choe et al., 2014; den Braber, 2016) —self-tracking's relevance to mental health, particularly in BD, has been examined only in limited ways (Matthews, Murnane, & Snyder, 2017a). Murnane et al. (Murnane et al., 2016) and Matthews et al. (Matthews, Murnane, & Snyder, 2017a) explored this in earlier studies and found the process of self-tracking in BD to be inherently complex as the sense of self and emotional state is in "flux, uncertain and unreliable". Both studies shed light on motivations and methods behind self-tracking however only provided a limited understanding on the internal process of self-tracking, especially on the process of turning a qualitative experience into something quantitative. Our research seeks to extend and deepen this understanding and to explore the nature of self-tracking practices for people with BD—as they are frequently designing their own self-tracking methods alongside clinical measures such as lists of 'early warning signs' (EWS) (Lobban et al., 2011) to both anticipate and manage conditions.

More specifically this research attempts to get to grips with three key issues related to work on BD and self-tracking. Firstly, current users are not being involved in the design of self-tracking technology for BD as much as they could be, and there are suggestions that HCI and PPI approaches should be hybridised to better empower users in the design and research of mental health technology (Majid et al., 2021), as indicated in former parts of this thesis. Secondly, there are tensions reported in the literature (Nicholas et al., 2017a) between what users want, what is theoretically sound, and the practical constraints of what can be measured with tracking-supporting technologies without infringing on confidentiality and privacy. Thirdly, many uses of self-tracking technology draw conceptually from the QS community, but in their nature adopt a specific normative ontology (Spiel et al., 2018). There are thus questions whether normative QS practices and ontologies are then applicable in the domain of mental health, especially in the context of BD where the sense of self is constantly in 'flux'. We will return to these issues in our discussion.

To address this, this study explores reasons and methods ^{RQ1} for self-tracking, and challenges and opportunities for future mobile technology ^{RQ2}. To do this, and in line with PPI principles, we designed our study closely with individuals with lived experience of BD where a series of workshops and interviews were performed to explore how and why those with BD self-tracked. As part of this we wanted to understand how this might also inform future technology, paying particular attention to current tensions in relation to some design practices and their limitations, i.e., we wanted to more broadly consider the working environment and researcher experience within mental health technology. If a self-tracking tool was designed which was based on users' needs, reflective of self-tracking practices as well as being embedded in proven clinical framework such as EWS (Lobban et al., 2011), it could have capacity to be a very useful method of self-management which is crucial for those with BD.

Specifically we unpack how those with BD convert their everyday experiences into something quantitative, and investigate the process of what constitutes a 'count' in the first place. In doing this, we demonstrate an example of ways to combine the principles of HCI and PPI—a design method of choice in mental healthcare research—in order to better ground research on BD and technology in ways that embed deeper participation. Combining principles of PPI and HCI to have users to assist in designing technology (as seen in participatory approaches in HCI) as well as in designing and running research (e.g., users co-facilitating design workshops) utilizing end-to-end user involvement.

4.2.1 BD and self-tracking

How people track in general is understood by some of the personal informatics models and tracking styles as seen in the HCI literature. For example, Li et al. (2011) describe a five stage process model of personal informatics (preparation, collection, integration, reflection and action) which was extended by Epstein et al. (D. Epstein et al., 2014) to account for other motivations of tracking beyond behaviour change. There is some, albeit limited literature in specifically understanding self-tracking in BD. Murnane et al. (Murnane et al., 2016) explored self-monitoring practices, attitudes and needs of individuals with BD using a survey with 552 users. They found that individuals reported that they primarily self-tracked items such as mood, sleep, finances, exercise and social interactions with an increasing trend towards the use of digital self-tracking methods. The study made suggestions towards the design of technology-based methods to be more condition-orientated, intuitive and proactive, including the need for sensing based technology methods. Matthews et al. (Matthews, Murnane, & Snyder, 2017a) takes this one step further and used user interviews (n=10) to explore how and why users engage in self-tracking, asking how the experience of BD influences self-tracking practices and what role technology plays in supporting this. They found users use a variety of methods to self-track to identify risky patterns that are indicative of episodes, as well as more positive trends that support recovery. They also found that users experience considerable challenges in self-tracking as their sense of self and emotional state is in 'flux, uncertain and unreliable'. This resonates with Rooksby et al. (Rooksby et al., 2014) who uncovered the chaotic and complex nature of tracking for activity tracker users.

Both studies report a positive perspective of the role of technology in self-tracking, if designed correctly. Poor usability and difficulty interpreting self-tracking data were the reported challenges in using self-tracking technology. Congruent with the aims of this research, the authors state that "currently, as designers we tend to focus on the quantitative, but as our users have shown, these numbers have deeply personal and qualitative associations". In an early paper, Martin and Lynch (A. Martin & Lynch, 2009) unpack this process of counting and how we interpret numbers. They argue that how the count is produced is largely dependent on who is doing the counting, what the count is for and the occupational and physical location of the counting event. Furthermore, the process of assigning numbers to "things" requires particular practices to render things accountable. The process of turning a everyday 'qualitative' experience into something quantitative has been considered within HCI research with key examples in the case of fitness tracker users (Rapp & Cena, 2016; Rooksby et al., 2014; Vanello et al., 2012). For example, the Tracker Goal Evolution Model described those motivations (which can derive from hedonic/eudanoimic needs) present themselves by qualitative goals. For example, the motivation to feel will translate into qualitative goals such as active lifestyle or lose weight. These qualitative goals, through trust and reflection, are then translated into quantitative goals that can be input into trackers (Niess & Woźniak, 2018). However, unlike the activities performed by fitness tracker users or other types of tracking in the QS community, it can be suggested that the process of tracking is distinct in the context of mental health. Patients and clinicians often describe the experience of bipolar disorder as "a rollercoaster" (Inder et al., 2008), where a person can flux between states of relapse, recovery, remission and recurrence. The comparison of the differences in interpretation compared to the QS community were touched upon by Matthews et al., (Matthews, Murnane, & Snyder, 2017a) who found that though practices are the same, the reasons and interpretation for QS practices were distinctly different compared to those for BD. For example, Martin (E. Martin, 2009) describes the concept of a "zero" day. A zero day is a day without change—an unremarkable day—which can be an indicator of improvement for those with BD, however this is unlikely to be treated in the same way for non-BD individuals in the QS community. This problem is of course part of a much broader class: for instance, a person's physical step in the QS community is represented by a set of algorithms that define and detect a particular, normative version of a step (Spiel et al., 2018), but there are questions as to whether such a step applies to 'everyone'.

In this research we draw on the idea of foregrounding normative ontologies for BD in and through our study of tracking practices. There are still unanswered questions such as how self-tracking helps people, how important EWS are when self-tracking, and how people feel specifically about mobile-based sensing methods of self-tracking. Although present research establishes that self-tracking is a deeply complex and personal process, the processes underpinning this have not received as much attention. Further, research to date has focused on those with BD "participating" in the research rather than being "involved" in the research such as developing research materials, undertaking interviews with research users, and identifying research priories, as seen in much of the PPI literature (Jennings et al., 2018). It remains unclear whether, by involving individuals with BD, research will produce the same or yield different results.

4.2.2 The current design of self-tracking technology

A full review of the current landscape of user involvement in the design of self-tracking technology has been explored in Chapter 2 of this doctoral research. To orientate the reader, a summary will be provided below in order to aid the understanding of the design of the current study.

Though the number of mobile applications for mental health is increasing, it is not clear that existing UX methods and techniques are appropriate or sufficient as identified in Chapter 1. Classical focus on usability and the delivery of the application may need to shift towards consideration of the ecosystem that surrounds everyday mental health experiences of the user and how this can be applied to design. Ethical concerns, heightened sensitivities and multiple stakeholder views (such as 'clinical' versus 'patient') likely require a fine balancing act when attempting design 'with' instead of design ‘for’. Goodwin et al. (Goodwin et al., 2016) states that there is a lack of parity of user involvement in the design of physical and mental health applications, where, for mental health, users are involved less frequently than for physical health apps. Furthermore, the design of such technology involves an intersection between healthcare and technology development, both of which traditionally have different approaches when involving the user, though there are exceptions to the rule. Patient and public involvement (PPI) dominates as a concept for ‘involvement and engagement’ within healthcare studies and interventions (Hoddinott et al., 2018). Key factors of this approach are outlined in Table 6. On the other hand, user-centred design approaches have—historically—been prevalent within HCI or service design more broadly. Participatory approaches that seek more direct user involvement in the design and delivery of digital technologies are also well-established. Thus, HCI design and healthcare-oriented approaches do have a number of overlaps: 1) they involve ‘users’ in research and development in some way; 2) they focus on understanding and empowering ‘users’; and 3) they make changes based on user responses. However, there are also key differences and limitations (Majid et al., 2021). Nevertheless, there is much work to do in blending these approaches, which is our focus here.

Key elements of healthcare-based approaches
• Evidence based in mental health research
• Traditionally a linear approach to involvement (from start to finish)
• A more standalone ‘event’ of participation rather than integrated into development
• Clinicians will often be involved (likely to be leading)
• Involvement is on a continuum (high-low) where high involvement is seen less frequently
• Users are often paid for contribution
• PPI organisers are often authors on works
• More commonly used in technology assessment

Table 6 Key elements of healthcare-based approaches

In prior parts of this thesis (Chapter 2), we found that only a small number of studies reported high user involvement, and that despite the presence of recommended standards for the involvement of the user in the process of design and evaluation, there is large variability in whether the user is involved, how they are involved and to what extent there genuine empowerment of the voice of the user, a purported aim for much healthcare technologies research. Our review developed seven principles for the mixing of methods present in PPI and HCI to ensure integration, which we detail below Table 1 as they informed our present study described in this chapter and provide key context for this.

1. Involve users in all stages of design and evaluation, including concept generation and ideation, prototype design and deployment, and evaluation stages with the goal of creating user empathy and empowerment. This process should have an adequate number of users to welcome diversity in thought. Equal representation is also a crucial consideration that needs to be considered when recruiting users.
2. Ensure early involvement as this will be cost-effective in the long run (avoid redesign and problems with use and implementation in the later stages).
3. Combine principles of PPI and HCI to not only have users to assist in designing technology but also in designing and running research (eg, users cofacilitating design workshops) and use end-to-end user involvement.
4. For academic and industry sectors to establish better mechanisms to access target user groups with lived experience of mental health issues, for example, by building relationships with existing patient-directed organizations such as charities and patient-led community groups.
5. Increase awareness of HCI and design communities in PPI principles and practices and increase awareness of PPI community in HCI and design methods or skills.
6. Encourage use and mixing of formal scientific or design methods with informal experiential and empathic practices to capture richness in understanding the dynamic requirements of technology users, which are cognizant of use in context.
7. Keep the user informed at all stages of the process, including final outcomes, future use, and next steps, which are often forgotten about

Table 7 Best practice recommendations for the design and evaluation of self-tracking technology for BD (Majid et al, 2021)

The current research aims to embed the recommended factors described above to understand self-tracking practices in those with BD and in future to inform the design of mobile self-tracking technology.

3.2 Study Approach

To ensure user involvement in the research, two individuals (P001 and P002) with lived experience of BD were closely involved in developing the study, via a series of workshops and follow up interviews. P001, P002 and the lead author met on a regular basis to organize the content of the workshops and user recruitment. P001 facilitated the workshop in one UK location and P002 managed the recruitment and facilitation of the workshop in another UK location, with the lead author being a joint facilitator. The lead author's role during the research was largely to be the 'middleman' between involvement with lived experience and involvement with research. The study had a consultative and collaborative approach throughout, where the lead author sought to empower those with lived experience to be co-designers and co-researchers in the study. The findings from the workshops and subsequent interviews were put together in a video presentation (in consultation with P001 and P002) and then shared with users of the broader study via email, as well as presented back to the users via Zoom. The presentation was aimed at gauging the accuracy of the results and eliciting responses.

3.2.1 Exploring methods for self-tracking

Table 10 indicates the types of tracking methods users reported during the workshops, along with examples of how this can be used to count experiences. Overall users indicated n=50 methods to self-track with multiple accounts of how these are used to count everyday experiences such as mood changes.

Types of tracking	*Examples*	*How it's used to count*
Pen and paper	Diaries, lists, calendar, post-it notes, self-made scales, keeping receipts, drawing/art, anniversary of life events	"Looking at number of activities on a calendar and see how they match to mood escalating"
Mental notes	Body scan, mental plan for day, tracking appetite, environment scan, personal hygiene, mindfulness, thought management	"Make a list of 20 things to do... if I get at least 5 done then that's good"
Social feedback	Share with friends/partner/carer	"Your carer or family friend can score your mood too, especially when you're unwell"
Technology based	Mobile apps, wearable technology, online mood scales, phone notes, bank balance, online forums, online research programs	"Use Bipolar UK mood scale…number 1-10 with qualitative statements…can relate to the qualitative statements and match the number"

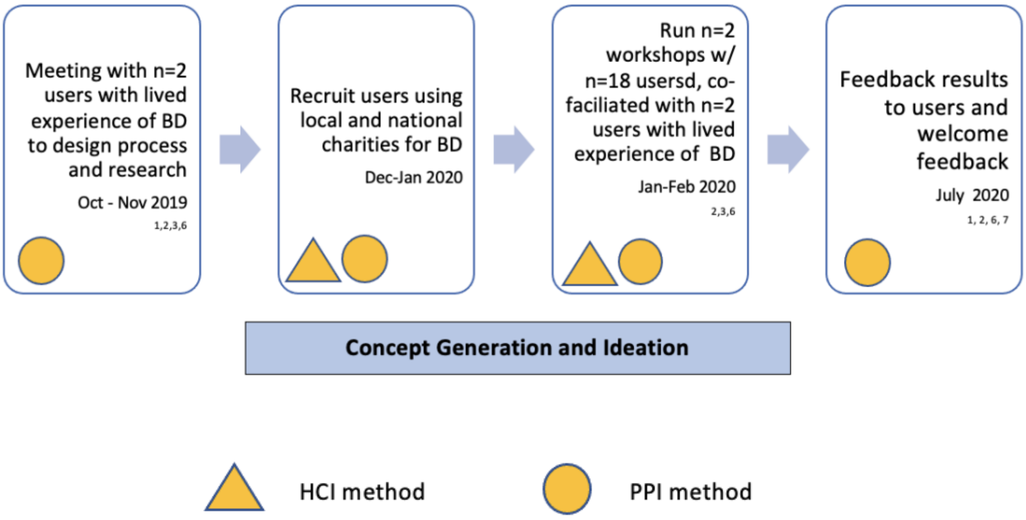
Table 8 Summary table of types of tracking

4.2.2 User Recruitment

User recruitment was achieved through the assistance of both national and local charity organizations involved in this study who sent bulk messages to mailing list members who may be interested in the research, as well as other snowballing techniques via an existing PPI team based at the Institute of Mental Health at the University of Nottingham.

4.2.3 Study Design

Figure 4 demonstrates the study design, indicating where the elements of the study were drawn from (e.g., HCI and/or PPI) and how they relate to the recommended seven principles. The study was conducted between October 2019 and July 2020. By establishing close relationships with two specific BD charities (Bipolar UK and Bipolar Lift CiC) we were able to explore methods that users would be comfortable with for meeting and sharing experiences. We tried to ensure our methods reflected current practices employed by these organizations. Both organizations run frequent workshops and group events which was the reason behind our choice of data collection. Our two workshops were held at public venues, for example one workshop was held at private meeting venue in a café and the other at a local community centre. Workshops were audio recorded followed by transcription. Subsequent follow-up interviews were conducted with those users who were willing to share further, to explore the findings specifically in relation to design.



1,2,3,4,5,6,7 Seven recommendations from Majid et al (2021)

Figure 4 Methodology for Concept Generation and Ideation

4.2.4 Workshop layout

Both workshops were three hours in total, with a one-hour lunch break in between. Workshops started with an icebreaker exercise followed by a three-part structure, which we detail below.

4.2.4.1 Exploring reasons for self-tracking

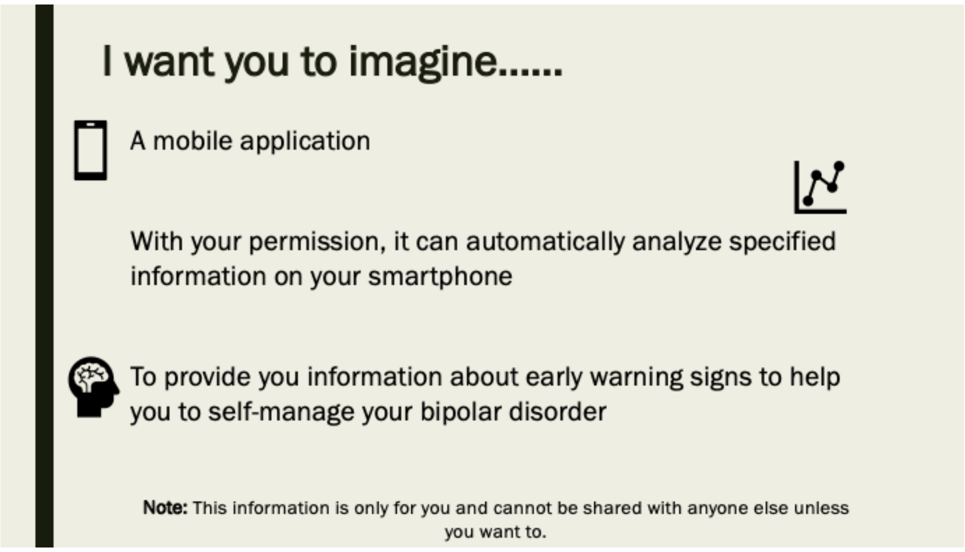
First, using an affinity mapping approach (Kara, n.d.), users were asked to share their thoughts on post-it notes to two questions: “how does self-tracking help you?” and “why do you self-track?”. All users were asked to answer the questions using post it notes and stick them on a large piece of flipchart paper corresponding to each question. Users were then split into two groups and given all the post it notes for each question and asked to discover and title emerging themes from the responses for the given questions. All users then had a chance to have a look at all themes and given three dot stickers to indicate which of the themes they found most important, for each question.

4.2.4.2 Exploring methods for self-tracking.

Next, users were given an A4 sheet of paper for use in a rapid sketching exercise (Levey, 2016), where users were asked to fold the paper until it had eight separate boxes, then were given eight minutes to draw, sketch and/or write the methods they currently use to self-track (spending one minute per box). Subsequently users were asked to form two separate groups to discuss the following question “how do you count or keep track of relevant factors relating to self-management?”.

4.2.4.3 Exploring challenges and opportunities.

Finally, users were asked to get into two groups and have an open discussion about the question “what challenges arise when self-tracking?”. The two workshop facilitators helped to guide the discussion in the two groups at each workshop. Users were provided an EWS checklist for mania and depression (Lobban et al., 2011) and given five dot stickers to indicate which EWS were relevant for them for depressive and manic episodes. Finally, a scenario was presented about mobile sensing technology which looked at smartphone data to sense EWS. Users were asked to share their views about this using post-it notes. The scenario was presented in PowerPoint format at the workshop (shown in Figure 5).



I want you to imagine……

A mobile application

With your permission, it can automatically analyze specified information on your smartphone

To provide you information about early warning signs to help you to self-manage your bipolar disorder

Note: This information is only for you and cannot be shared with anyone else unless you want to.

Figure 5 Scenario of mobile sensing technology that was presented to users at the workshops via PowerPoint presentation.

4.2.4.4 Workshop user sample

In total, we had n=18 users including the two facilitators with lived experience of BD. One third of users identified as women, two thirds as men. For age, 25.0% of users were between 25-34, 31.3% were between 35-44, 18.8% were between 45-54, 12.5% were between 55-64, and 12.5% were 65+. Of the users, 6.3% had more than three but less than five years of lived experience of BD, 18.8% had more than five but less than seven years and 75.0% had more than seven years. 50.0% had a diagnosis of Type I BD, 43.8% had a diagnosis of Type II BD and 6.3% preferred not to say. Two users preferred not to answer the demographic questionnaire. Ten users were willing to take part in follow-up interviews.

4.2.4.5 Analytic approach

For the first part of the workshop structure (Exploring reasons for self-tracking), users performed analysis themselves: as we mentioned, an affinity mapping was used with users organizing the answers to questions into themes and indicating importance via dot stickers. No further analysis was performed beyond this; hence results are presented as per the user’s discovery.

For the subsequent workshop sections, audio data was transcribed by the lead author and then analysed using Braun and Clarke’s thematic analysis phases (Braun & Clarke, 2006) using NVivo. The study employed a reflexive thematic analysis approach, enabling flexible exploration of findings without theoretical constraints. Inductive analysis derived themes directly from data. Themes were identified semantically to capture explicit content and align with users' accounts, the thesis's primary focus. The following process was undertaken: familiarizing yourself with the data, generating initial codes, searching for themes, reviewing themes, defining and name themes and producing the report. During the defining and naming themes stages, P001 and P002 were involved in analysis.

For the question “what challenges arise when self-tracking” under the final section of the workshop (Exploring challenges and opportunities), themes were ordered in hierarchical format which was dependent on the number of codes belonging to said theme (e.g., the theme with the highest number of codes were classified as most commonly described and so on). P001 and P002 also contributed to the understanding and summarizing of the interview data through a series of consultations with the doctoral researcher in order to understand opportunities for self-tracking technology as presented below.

4.2.4.6 Ethical Review

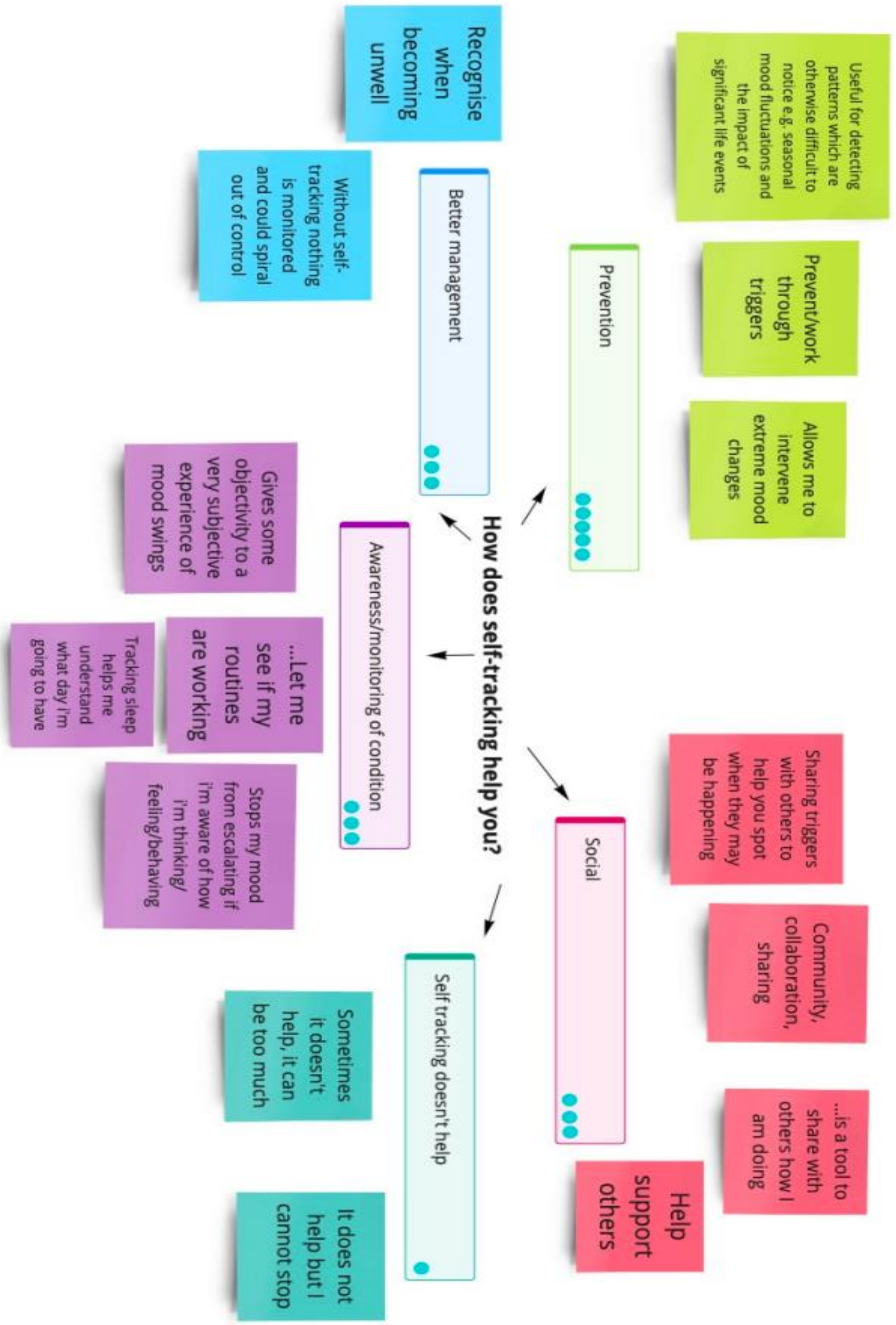
This study was subject to ethical review and approved by the School of Computer Science (reference: CS- 2019-R58) at the University of Nottingham, UK.

4.3 Results

In summary, workshops and interviews engaged users with three main topics: the reasons for self-tracking in BD, which methods were used, and finally a discussion around the challenges and opportunities in self-tracking practices. We discuss these in turn when describing our results.

4.3.1 Exploring reasons for self-tracking

First, we want to examine how users explored questions on “how does self-tracking help you?” and “why do you self-track?”. Figure 7 provides a reconstructed illustrative example of how the post-it notes were organized and using the dot sticker importance ratings for the specific question “how does self-tracking help you?” for all users. Actual images are not provided to protect the privacy of our users. Furthermore, user numbers are not provided here as these post-it notes were shared openly in the workshop, and users anonymously wrote answers. The full list of users discovered themes and level of importance is listed in Figure 6.



)

Figure 6 Scenario of mobile sensing technology that was presented to users at the workshops via PowerPoint presentation

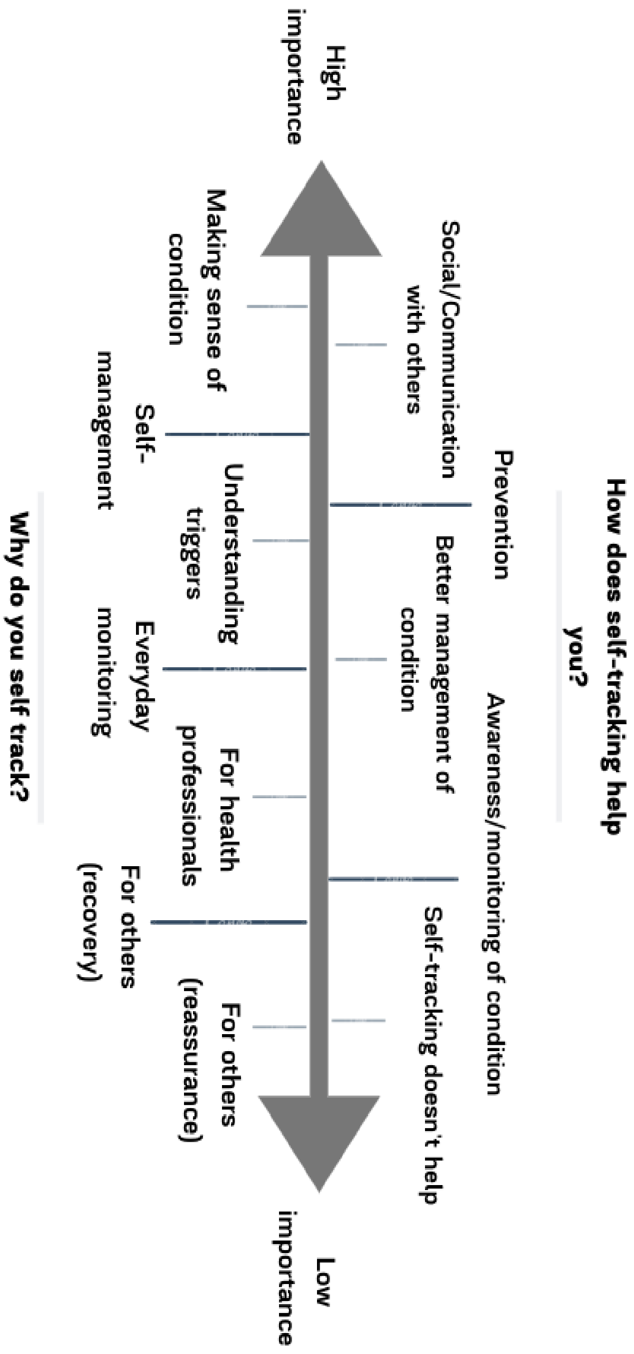


Figure 7 Themes for questions "how does self-tracking help you?" and "why do you self-track?" indicated by level of importance

4.3.2 User responses to how does self-tracking help you.

The most important theme that users discovered was titled SOCIAL/COMMUNICATION WITH OTHERS, where one user shared on a post-it note that "sharing triggers with others can help you spot when they may be happening" when describing periods of relapse. Another user shared that self-tracking could help them by “keeping balance with my family” and that it “is a tool to share with others to see how I’m doing”. The second most important theme as discovered by the users was PREVENTION; a user wrote that self-tracking can “prevent/work through triggers”. Further answers include that self-tracking can “help to intervene before extreme mood changes”, and that it is “useful for detecting patterns which are otherwise difficult to notice, for example seasonable mood fluctuations and the impact of significant life events”. The third most important theme discovered was BETTER MANAGEMENT OF CONDITION, an example from a user here saying that “knowledge is power” or another user answered that self-tracking “reduced overreliance on medication” and that it promoted “discipline”. The fourth most important theme was AWARENESS/MONITORING OF CONDITION, where a user noted that as a result of self-tracking “I understand the way I am and why” and that tracking certain factors such as sleep can “help me understand what day I’m going to have. Moods vary depending on sleep” and to “let me see if my routine is working”. The least important theme users discovered was SELF-TRACKING DOESN’T HELP, where one user indicated that self-tracking “doesn’t help, it can be too much” and lastly a user indicated that “it does not help but I can’t stop”.

4.3.3 User responses to why do you self-track

Using the same method as above, users categorized the answers to the question “why do you self-track?” into themes. The theme which was classified as being the most important was MAKING SENSE OF CONDITION, which had responses related to how self-tracking can “identify changes in mood” and to “to better understand my condition”. In line with this theme, another user indicated that they self-tracked “because it enables me to keep tabs on and consciously monitor my mood. Bipolar (especially Hypomania) has a habit of creeping up on me when I least expect it, so it pays to always be vigilant and not get lulled into a false sense of security. Self-tracking is a practical, straightforward way”. The second most important theme was titled SELF-MANAGEMENT, which had answers such as to “be more efficient”, “maintain balance” and to “stop spending all my money”. The next important theme was titled

UNDERSTANDING TRIGGERS; one user indicated that they self-track to “prevent/work through triggers” and that self-tracking through “writing feelings…helps me to put things into perspective when I calm down, it allows me to see how heavy the trigger was” and lastly to “to try and prevent triggers”. The fourth most important theme was titled EVERYDAY MONITORING/ACTIONS, where users provided answers such as they self-tracked “to not forget” and that it was part of “routine”. Another user indicated that “I don’t really self-track except when extremely low. Then it’s more putting it down on paper to be able to read back on it”. The fifth most important theme was titled FOR HEALTH PROFESSIONALS, where a user answered that they self-tracked “because my doctor told me so”, with another saying, “helps me and my psychiatrist to determine if my meds are working or need adjusting” and to share “with doc, shrink and mental health team”. Beyond health professionals, users split the last theme for others, into FOR OTHERS (RECOVERY) which had answers such as “help support others” and for others (reassurance) which included the following answers “show those I care about that I am trying to stay well”.

4.3.4 User responses to how do you count or keep track of relevant factors relating to self-management

Using numbers to self-track. Users indicated that they would use numerical scales such as the Bipolar UK scale [14] and other clinical and mood scales to understand how they are feeling. It terms of helpfulness, one user indicated that it is “really helpful to put a number to something that is extremely chaotic”. Users described that numbers can be emotional which can be demonstrated with the following quote “But the thing we have this negative image that between 1-5 your just below average, you’re a nothing you’re a nobody, and the self-esteem and all that kind of lovely stuff and that’s already kind of down at the bottom.. just ends up playing on it more..whereas I’m kind of like oooh I’m a 9… Actually I feel better, even though I feel shit.. It’s an interesting thing as the number affects me.. that number says something but I’m taking it to mean something different. …Because we are all wide in that way”. It was also described how the range of numbers employed differed for individuals: “1-10 I actually find more helpful… if it’s just small, like 1-3 I just don’t get it…it’s just not specific enough. 1-10 more options”. Some users stated that numbers are not for everyone, while others indicated they preferred a combination of quantitative and qualitative tracking: “it is just a number I think you can be like is it 3 and 4 but it’s the statement with the number then you can kind of relate... Relating to it, how to describe that stage.”

Using task completion to self-track. Users described the process of creating to-do lists and plans as a form of tracking, with the level at which this is completed being indicative of mood state. There were idiosyncratic views of what constitutes success in users account here. For example, one user stated, “make a list of 20 things to do... if I get at least 5 done then that’s good”, while another user said, “take three things of this to do list, turn the page over and write those down… this is more manageable and if you get one thing of this list today then that’s great”.

4.3.5 Exploring challenges and opportunities for self-tracking

Finally, we talk about how users responded to discussions around challenges and opportunities for self-tracking technologies. We will also touch on how users perceived the EWS checklists combined with their attitudes towards technological opportunities in self-tracking.

4.3.5.1 User responses to what challenges arise when self-tracking.

Table 9 indicates the key themes that arose from users when asked about challenges for self-tracking. These responses are presented in hierarchical order including subthemes and quotes. The themes are presented in table format to preserve the detail that was discovered from the thematic analysis, and to provide examples in relation to the subthemes. The most commonly occurring theme (a challenge) that arose from the analysis was about internal factors that prevent tracking. The least commonly occurring theme was related to self-tracking technology tracking more than current technology.

Themes	Sub-themes	Examples
Internal factors that prevent tracking *Personal factors that impact and prevent the process of self-tracking*	Difficulty tracking when unwell	“Self-tracking tends to go out of the window when we’re going on the down”
	Being honest with yourself and others when tracking	“I think sometimes I can also skew the results.”
	Poor memory and forgetting to track	“…but that pops up, with notifications and ask twice a day about how you’re feeling. I think that’s one of the biggest problems, because I forgot.”
	Lack of motivation	“It’s kind of motivation as well. You know, some days we wake up you think “I don’t want to do it anymore”
	Lack of discipline	“None of is any good if you’re not disciplining yourself to look at the information.”
Experience of tracking can be negative *The process of tracking can be a negative experience and a reminder of ill health*	Reminder of feeling low	“It will highlight that your mood is creeping down”
	Can become obsessive	“It can become obsessive for some people, which makes it a negative thing to self-track. So, you’ve got your 10,000 steps a day goal, but if you may not raise your head by a certain time, you might be like, well you’ve got a go for a run now for yourself”
	Tracking can be counterproductive	“But do I really want a reminder, I don’t know why I was feeling worthless, like a piece of shit. Is that going to trigger me off later on? By looking at it to remember that that’s the time and this happened and you’re going backwards”
	Too much insight is negative	“If you have too much insight, it can actually kill you because you know so much that there’s no end.”
	Tracking can be punishing yourself	“Tracking the moment after the moment and it’s quite like I said painful.”

Mental health stigma and shame *Avoiding tracking to hide mental health struggles because of external and societal pressures*	Not wanting to leave a record of your health	"In my case, I'm half Indian and half white and the way I live my life is actually more Indian Asian. So, what's happening Yeah, actually, things like I need to take note, you know, when they're saying do a journal do a diary, I wouldn't do any of that. I don't want people within my family to know that I'm unwell."
	Stigma around mental health	"So, the stigma that would then follow my family so my child would not be able to get married, because they'd be seen as being unwell because I'm actually the father and I was unwell, so if you see genetics and things like that, there's a lot of kind of that whole thing a stronghold in the community behind of it."
	Shame	"Because I think there's a lot of shame because well, I know I personally feel a lot of shame."
Lack of third-party support when self-tracking *No availability of support from clinicians and services when self-tracking, especially in situations when things are not going well*	Lack of support with the results of tracking	"when tracking they don't offer you any kind of support within the results or anything like that. That's all on you. So, whilst it's a good idea to do the self-tracking I think sometimes it's hard, particularly if you're not under a secondary mental health team or anything to follow through because you've got the information, you got the data, but you don't know really what to do with it."
	Clinicians are not open to the results of self-tracking	"Recently I had an incident where I wanted to share my psychiatrist my sleep tracking on my Fitbit and I showed her, but she wasn't even the slightest bit interested in it and thought why am I even bothering. Because I thought that was relevant and if I can evidence that my sleeping is improving then surely that's a good thing. Either she didn't have time for it or she wasn't bothered or you know it felt like it was a bit more wasted time and since then, I haven't done it"
Self-tracking technology should track more than current technology	No subthemes	"If you could chip us all, and monitor everything, spending, the whole lot"

Self-tracking technology should track as much as possible to monitor health and wellbeing		

Table 9 Themes for challenges during self-tracking

4.3.6 EWS and opportunities for sensing technology

For EWS, overall users engaged positively with the EWS checklists and indicated a number of common signs and symptoms that arose before an episode of mania and depression. A summary of the common EWS reported are described in Table 10. Users were asked how they felt about a mobile sensing technology to detect EWS via a scenario and asked to share their views via post-it notes. Users indicated openness towards this technology as indicated by the following answers: “interested to find out how useful it could be”, “would give it a go” and “it would make what I struggle to do manually, automatic”. Furthermore, users shared answers which described types of personalization and usability features such as it needs “simple language”, “personalized color layouts/colour choice and dark at night mode”, “easy to use analytics, archive of stats, records”. Other things that were uncovered in users’ answers were in relation to the types of data to sense by the following answers: “time spent of social media”, “measures how many times opening and closing apps” and “an app to measure frequency and speech of speech”. Lastly, users did describe a series of cautions in their answers such as its “not for those with paranoia” or “what if it got hacked?” and “is it safe?”.

EWS	Common symptoms
Depression	Low motivation/can’t get started Difficulty concentrating Low in energy Feeling tired/listless Loss of interest in activities
Mania	Ideas flowing too fast Stronger interest in sex Spending money more freely Racing thoughts More talkative

Table 10 Common EWS

4.4 Discussion

4.4.1 Exploring reasons for self-tracking

For RQ1, we found that the social, communicative aspect of self-tracking was of the highest importance for users (as per their voting); this was reinforced by responses to how self-tracking helped them, i.e., by making sense of the condition, which was also highly rated in importance. This finding reflects existing research on the social nature of tracking in HCI; for instance, Epstein., (D. A. Epstein et al., 2015) suggests we reconsider personal tracking as social tracking. This also reinforces findings of a similar study (Murnane et al., 2016) that asked similar questions on sociality in self-tracking, where users described the following themes: made health management more manageable, promoted self-awareness, reflection and empowerment and supported their interactions with clinicians. However, there are some key differences. Firstly, in relation to the detail of the themes that were discovered by our users around the reasons behind self-tracking compared to the aforementioned three themes in a similar study. Our users described 5 key themes for the question "how does self-tracking help you?" and 7 key themes for the question "why do you self-track?" with further personal insights via answers to this question. The level of detail in the reasons for self-tracking presented here are all per user discovery. Given the first-hand experience of our users who undertake the self-tracking on a daily basis, we felt as though the opportunity to be the experts here with little professional research or design influence will uncover better details for these crucial questions around self-tracking. Conversely, Murnane et al. (2016) employed a more traditional approach of a survey method with a large group of users (n=552) where the researchers themselves uncovered the themes from the answers to these questions.

Another notable difference is the emphasis on how self-tracking can help support interactions with clinicians, also as per Murnane et al. (2016). About two-thirds of the survey respondents in Murnane et al. (2016) reported using self-tracking data with discussions with healthcare professionals. However, in our findings, though the social communicative nature of self-tracking was a major driving factor, this was more in the context with sharing with family, friends and significant others rather than a clinical interaction. Sharing with health professionals was identified as a theme for “why do you self-track?”, however this was at the lower end of the spectrum of importance for our users. This comparison may shed light on the importance of the design of research and how it may yield different results, particularly as our study attempted to combine PPI with HCI approaches and thus push more towards users as co-designers of that research, such as for our workshops and workshop facilitation, where P001 and P002 (with lived experienced of BD) co-designed and led alongside the first author. This may have meant that other users felt more relaxed to share their experiences, away from probes connected to a more clinical environment. As mentioned, the Murnane et al. (2016) study used a survey approach where users may have associated this type of research with a clinical setting, which could have influenced the clinical focus in the results to these questions. In contrast, our workshops—sited in a café and a meeting house—avoided a university setting as this can be associated with clinical settings, thus providing a more neutral space for users. Given our sample was small, future studies should aim to replicate this finding in order to understand the true nature of the impact of the environment on the research findings.

We think research on BD and self-tracking, and mental health technology in general, can be more participatory using approaches akin to those described in this paper. The benefits to using this are as follows:

- Uncover rich details in findings that are led by the first-hand lived experiences of the users which can appropriately guide mental health technology research and design
- Create a more neutral, open research and design environment which allows for users to feel encouraged to share removed from expectations, fear or judgment
- Empower users to guide the research and design by experience

4.4.2 Exploring methods for self-tracking

In terms of self-monitoring practices for those with BD, our findings for RQ1 are congruent with the current research in terms of complexity (Matthews, Murnane, & Snyder, 2017a; Murnane et al., 2016). From the 18 user who attended the workshop, 50 methods of self-tracking were indicated which fell into the categories as follows: pen and paper, mental notes, social feedback and technology based. This too reflects the work of Murnane et al. [29]. Future work should consider expanding these categories beyond these four categories as seen in the work of Rooksby et al (2014) which considered type of technology such as physical devices, apps, exergames and web apps and also considered type of activities tracked such as walking, physical exercise, food and drink, weight and size and sleep.

Our novel findings are in relation to how these are used to count, turning a 'qualitative' experience into something quantitative. A characteristic example we saw was one user stating, "make a list of 20 things to do... if I get at least 5 done then that's good". This shows how situated and particular the perception of a successful day is for individuals, in this instance a 25% completion rate indicated a "good" day. Whereas, for current online to-do lists the model is to set a number of items with a particular deadline, where 100% completion equals ideal productivity. This is similar to the discourse of fitness trackers in the QS community and their particular normativity (Spiel et al., 2018), where an invisible ideal is set without accounting for individual differences; Spiel et al. state that "technologies do not facilitate a better life: they define it, without oversight, without transparency, using emotional design tricks to engage in a progressive redefinition of what it means to be human". Being able to set your own level of completion or productivity—an experience of a count—may be more suitable in the case of self-tracking technology for BD, as one's practices of turning experience into quantitative data is likely to differ from person to person. This is further echoed in the work of Rapp and Tirassa (2017) (Rapp & Tirassa, 2017) on the Theory of the Self for Personal Informatics who suggest that such personal informatics tools with a focus on behaviour change aim to modify specific behaviours (e.g. sedentariness) based on standards that may not be chosen by the user themselves. They theorise that by placing the user at the heart of the process of change, is more likely to increase wellbeing as the user can find their own goals and ways to "happiness" based on their own peculiarities. Further, our prior research (Chapter 2) looking at data visualization preferences concurs with this, specifically for those with central nervous system disorders, with BD forming a large part of sampled papers in a recent study: Polhemus et al. (Polhemus et al., 2020) found absolutely no consensus, where preferences ranged between graphical formats as well as non-graphical and textual descriptions. Polhemus et al. (2020) also found there was a large focus on how users valued the ability to provide contextual information when interpreting visualized data, such as annotating to provide "internal context" alongside a numeric score. In addition, Polhemus et al. (2020) described that the ability to provide the context behind the numbers connected the numeric with the qualitative and was seen as valuable when communicating with healthcare providers, as it enhanced the users' memories of past experiences. We find this is also echoed in personal informatics research where it is urged to recognise that data can be meaningful in the context it is produced but may lose meaning when it is removed from that context [9]. Future design consideration of self-tracking technology should consider the complex nature of this 'qualitative-to-quantitative' phenomenon, and accordingly we recommend the following design considerations:

- Provide the user with choice as to whether the self-tracking technology (e.g., tasks and feedback) is quantitative- or qualitative-focused
- Be able to define the quantitative if needed e.g., being able to pick numeric scales (e.g., 1-3 or 1-10), and define what numeric success constitutes (e.g., set a unique and adaptive percentage of success which moves away from a 100% completion model)
- The ability to augment quantitative data with qualitative experience (e.g., being able to annotate quantitative data with 'internal' context).

4.4.3 Challenges and opportunities for mobile technology

In line with Matthews et al., 2017 and Murnane et al., 2016, we found considerable challenges for self-tracking BD, with the most common being personal ‘internal’ factors that prevented tracking for RQ2. Users described that the sense of self is in ‘flux’, meaning that it is often difficult to self-track during periods of feeling unwell, issues around memory, motivation, or self-discipline. When presented with a scenario around mobile technology that could passively sense clinical factors, such as EWS, users displayed a degree of openness towards this technology; to sum up, one user stated such a technology would “make what I struggle to do manually, automatic”. Relatedly, there is a growing body of research on the utility of sensing various types of digital data as a form of self-tracking for BD severity which responds to this challenge, including sensing factors such as number of text messages (Beiwinkel et al., 2016; Faurholt-Jepsen et al., 2015; Puiatti et al., 2011), phone calls (Faurholt-Jepsen et al., 2015; Gruenerbl et al., 2014), emails (Puiatti et al., 2011), GPS data (Gruenerbl et al., 2014; Osmani et al., 2015a), voice features (Gideon et al., 2016; Karam et al., 2014a; Maxhuni et al., 2016; Muaremi et al., 2014; Vanello et al., 2012), accelerometer data (Bardram et al., n.d.-a; Garcia-Ceja et al., 2018; Vanello et al., 2012), and app usage (Alvarez-lozano et al., 2014a). In our study, users responded positively to Early Warning Sign checklists and described a series of common signs and symptoms, as well as specific symptoms, that arise before a period of relapse. This is in contrast with the ways in which current self-tracking mobile apps for BD tend to be developed without reference to clinical guidelines like EWS (Lobban et al., 2011). Beyond self-tracking factors suggested by current research, our users indicated the utility of many factors, outlined in our recommendations below. In our study, users also expressed caution. The experience of self-tracking can be negative, in that the process of self-tracking can be a reminder of feeling low and that too much insight can be negative, or counterproductive. As one user stated with reference to the mobile sensing technology scenario: “it is not for those with paranoia”. Self-tracking may be detrimental in the context of relapse where it serves as a reminder of deteriorating health.

Future designs of mobile self-tracking technology should consider these factors to support the self in self-tracking. In sum we recommend the following design considerations:

- Future designers should consider the use of EWS checklists as a clinically validated tool as an underpinning of self-tracking technology for BD
- Having passive sensing methods in mobile self-tracking technology for BD responds well to a commonly described challenge and designers and researchers should consider the following factors to sense (presented in hierarchical order): sleep levels, levels of social interaction, medication usage, wearable data, financial data, number of diary entries, food and calorie intake, to-do lists completion and employment information (e.g., sick days)
- The experience of self-tracking can sometimes be a negative one for those with BD. The level of user-control over self-tracking tools should be high; sensing streams should be manipulable in a meaningful way, what is collected must be controlled, and disengagement for periods should be easy and available without interactional costs.

4.4.4 Limitations

The data presented here is limited, has a small sample size, and is unlikely to represent the views of the larger population of those with lived experience of bipolar disorder. A large percentage of our user group were in points of remission, and it is likely that the account of those experiencing relapse in not represented in this research. The users in this research had existing practices of self-tracking and were likely to contribute a more detailed viewpoint compared to those who had just started their self-tracking journey. Future work should consider a larger sample with individuals across the spectrum of bipolar disorder severity and self-tracking practices, to account for the full self-tracking experience.

4.4.5 Future work

To explore how findings and recommendations from this research can be applied in practice, our future work aims to utilize this research for designing and deploying a prototype self-tracking mobile tool. This tool will incorporate both active and passive sensing to comprehend Early Warning Signs (EWS) of relapse for Bipolar Disorder on a broader scale. The findings and recommendations will be transformed into design suggestions and procedures, guiding upcoming technologies in the field of Bipolar Disorder self-tracking. These suggestions will also undergo additional review with individuals affected by Bipolar Disorder. Moreover, we believe our findings and recommendations could have broader applicability in health and well-being technology. Future work should consider this. Equally, while we focused on mobile self-tracking, future work should consider beyond this, e.g., other interaction modalities and scenarios, such as those enabled by sensing infrastructures developed for ubiquitous computing or Internet of Things. Though this work shed light on the complex experience of translating qualitative experiences into quantitative, future work should work out how to situate this into models seen in prior HCI research such as the Tracker Goal Evolution Model which describes this process in the case of activity trackers (Niess & Woźniak, 2018). The categories of and motivations around self-tracking seen in this work (such as pen and paper, mental notes, social feedback and technology-based) could be understood in more depth by relating to the categories seen in the work of Rooksby et al (Rooksby et al., 2014). Furthermore, future work should extend the understanding of reasons behind self-tracking to consider the kinds of questions that people ask about their information such as status, history, goals, discrepancies, context and factors (I. Li, Dey, & Forlizzi, 2011) to gain insights about their personal experiences.

4.6 Chapter Summary

We explored reasons, methods and challenges for self-tracking for those with lived experience of BD. Our findings for this differed from current literature, as previous studies indicated a more clinical basis to tracking compared to the more social, communicative basis to tracking we located. Differences of methods in investigating this may explain this, as the current study employed a more participatory method of research combining principles of PPI and HCI to create a peer-to-peer environment with the goal of empowering users. Our findings documented a large level of complexity in self-tracking, where 50 methods were described. We suggest that future designers should consider this to create technology with choices to convert these experiences, compared to the blanket normative ontology that is currently seen in the self-tracking space. The findings of this chapter will be built upon in subsequent chapter (Chapter 5: Prototype Design and Deployment) in order to guide the design of a self-tracking tool for BD, considering the complexity and individual variability of self-tracking that was uncovered in this research.

Chapter 5: Prototype design and deployment

Chapter summary

- The current study built on the findings in relation to self-tracking and BD from Chapter 4 in a series of follow up interview with n=10 users
- Specifically we looked at how the findings from the previous phase can guide the redesign of the mobile self-tracking app for BD $^{RQ4\ RQ7}$, how the app can be used to explore EWS ^{RQ5} and what is the general user experience of using the prototype ^{RQ6}
- Specifically, the findings were explored to guide the redesign of an existing self-tracking tool to be more closely aligned with user's needs, preferences and individual variability in BD.
- The findings led to a series of changes within the tool such as financial tracking, personalised medication tracking and a daily visual mood diary.
- A working protype was enrolled for an alpha testing series with a small sample of users of BD where findings are reported in relation to bugs, usability and general user experience.
- Lastly, the final prototype to be enrolled for a 6-month period of beta testing is described.

5.1 Introduction

The previous chapter (Chapter 4) explored the qualitative experiences of self-tracking for those with BD exploring reasons, methods and challenges for self-tracking as well as opportunities for self-tracking technology. In that study, it was found that self-tracking is a highly complex process, where over 50 methods of self-tracking were described. The previous study also demonstrated the value of combining methods of PPI and HCI to explore the complex nature of user needs, preferences and individual variability and found that by using these methods, observations were revealed that were different from that seen in the current literature. For example, our results revelated an emphasis on the social/communicative elements of self-tracking with family, friends and significant others when exploring reasons for self-tracking, whereas in the literature there is apparent reported emphasis on the clinically communicative aspects of self-tracking (Matthews, Murnane, & Snyder, 2017a; Murnane et al., 2016). The current chapter aims to explore and extend the previous work to guide the design of a novel self-tracking tool for those with BD paying particular attention to the following previous findings:

- Combining principles of PPI and HCI and adhering to best practice principles for the design and development of self-tracking technology for BD reveals findings that are led by the first-hand lived experience, and should be continued throughout the lifecycle of the project
- Provide the user with control and offer a level of tailoring with technology to account for individual variability, needs and preferences
- Embed EWS checklists into the design and development as not only do these have clinical validity but demonstrated user acceptance
- Embed passive sensing methods as these account for the commonly described challenges users face during the process of self-tracking

This research study explored the design of a self-tracking tool in three stages. Firstly, the findings of the workshops were put into a multimedia file (.MOV) and shared to all users who attended the workshop and subsequent follow up session were run to invite feedback and explore the accuracy of the findings. Secondly, a series of interviews were run with workshop particpants to explore the finding more in relation to design of the tool. Thirdly, a phase of alpha testing was ran with a small number of workshop particpants to explore bugs and acceptance of the prototype prior to final release which will be explored in subsequent chapters of this research (Chapter 6). The aims of this phase of research are to guide the design of the self-tracking tool to be closely aligned to the individual variability, needs and preferences of users which has limited exploration within the research thus far (Ortiz et al., 2021). Finally, this research aims to continue to contribute towards best practice of involvement of the user in the design of self-tracking technology for BD.

5.1.1 Design methods for self-tracking tracking in BD

The design of self-tracking tools for BD has been explored in a previous chapter of this research. However, in this section we will specifically recap on the methods of design to provide context on the methodology applied in this research. When exploring the methods for prototype design and deployment in our literature review (Majid et al, 2021), the following methods were commonly described: focus groups and workshops, sharing of wireframes and interviews. This guiding our choice in interviews at this stage which are commonly used in PPI practices (Abelson et al., 2016), as well as HCI. However, to draw on methods used in HCI we explored types of usability testing to guide the development of the product. Usability testing explores the user experience of interreacting with the application with the goal to identify any usability problems, collect qualitative and quantitative data and determine the user satisfaction of the product. Usability testing uncovers problems in design, discovers opportunities to improve the design and also learns about users' behaviours and preferences. This is usually done in an iterative format, during multiple stages of the development of the product as to ensure consistent improvement (Moran, 2019). Usability issues include anything that prevent task completion (such as the ability to complete self-report measures), takes someone off course, causes frustration or create confusion. Poor usability is one of the primary causes of failed adoption of health technology (Price et al., 2014), which is the driving factor of why it is imperative to be explored. Though usability testing is not a *research methology* in a traditional sense, it is a useful mechanism which can be used to understand the user experience of the tool in order guide the research findings.

Alpha testing is a type of usability testing which is conducted at the early stages of development, to uncover early stage usability issues prior to full development (*Alpha Test*, n.d.) within test subjects who are internal to the development, and described as being within the process of development of mental health applications (Price et al., 2014). Beta testing is a similar process however user test a developed protype in a real world context, and is being more increasingly used in the general context of mental health technology (Shore et al., 2014) (Lewis et al., 2021) (Rickard et al., 2016) (Birrell et al., 2022) and will be conducted to support the general understanding of the tool in a later chapter of this research. The level of usability testing, which can be seen on a spectrum from high to low, differs currently in the literature. For example, in our own review of user involvement in the design of self-tracking for BD, it was seen that low involvement was common and usability tended to be 1-dimenstionaly, typically involving the use of standardized and unstandardized questionnaire to measure usability, satisfaction and usefulness and explored in the later stages of development (e.g. evaluation phase) (Majid et al., 2021). The aim of this research is to guide the design of the self-tracking tool to be as closely align to users' needs, preferences and individual variability and therefore will use a multidimensional approach to explore this in the form of interviews and multiple phases of alpha testing over 3 weeks with a group of target users to explore usability, user experience, bugs and crashes.

5.2 Methodology

5.2.1 Study approach and design

To ensure the involvement of users in this research, two users with lived experience of BD (P001 and P002) were closely involved in the design of the study. Regular online meetings were conducted to design the research, draft the interview schedule and as well as outline the alpha testing plan. Both P001 and P002 were also users in the research, both of which partook in the follow up interviews, and P002 in the series of alpha testing. Principles of best practice for the design of mental health technology by (Majid et al., 2021) were referred to during the design of this study. Figure 8 visualizes the study design, indicating where the elements of the study were drawn from (e.g., HCI and/or PPI) and they relate to the seven recommended principles.

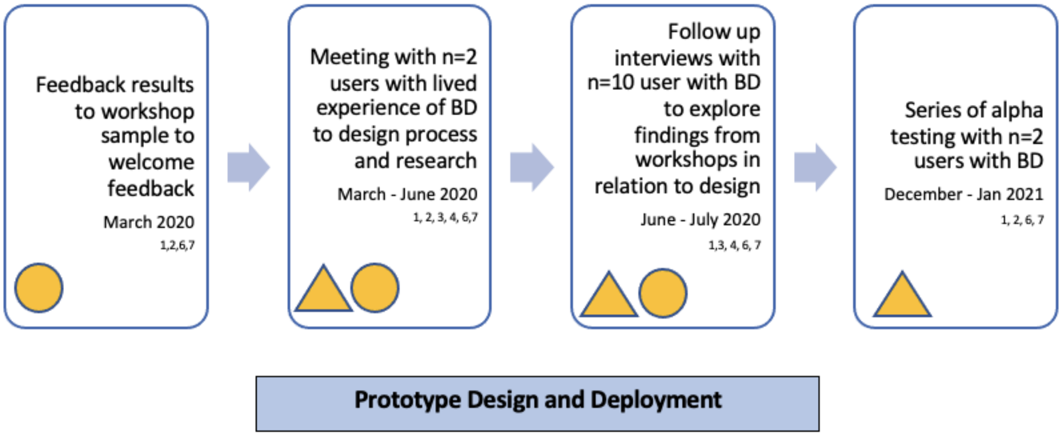


Figure 8 Study methodology for Prototype Design and Deployment

5.2.2 User recruitment

Users in this phase of research were existing users from the subsequent phases of research that are outlined in Chapter 4, all of which attended the self-tracking workshops. All users were reimbursed with a £20 Amazon voucher as a token of appreciation for their time in this research.

5.2.3 Self-tracking tool

The technology employed in this study is titled RADAR-Base (https://radar-base.org/) and has been developed by the industrial sponsors of this research (National Institute of Health Research - Nottingham Biomedical Research Centre). The details of this technology have been outlined in the methodology chapter of this doctoral research. However, in the interest of reader orientation, a brief summary of the technology will be provided in this chapter as a guide.

Like other types of mobile self-tracking technology for BD (Ortiz et al., 2021), RADAR-Base employs both active and passive sensing methods. It consists of two applications, one for passive data collection *pRMT* and one for active data collection *aRMT*. The aRMT app allows for self-reported information to be inputted. The *pRMT* application allows for the passive collection of the following data sources:

- Phone usage
- Local weather
- Step count
- GPS location
- Which apps are being used and for how long
- Battery level
- Other Bluetooth devices in the vicinity
- Sleep information and heart rate (provided by user Fitbit connection if possible)
- Number of contacts on phone

- Ambient light.

Though RADAR-Base has a proven use case in other mental health conditions such as depression, it has not yet been explored in the context of BD. By adopting the close participatory design approach in this research, we aim to adapt RADAR-Base to be closely aligned to user's needs, preferences and individual variability.

5.2.4 Interview schedule

This study used interviews to explore previous findings more in relation to design. The interviews were on an 1-1 basis, ranging between 60 to 90 minutes and were conducted remotely using Microsoft Teams. As the interviews were conducted at the height of the COVID-19 pandemic, there was no scope to conduct the interview on a face-to-face basis. Specifically, the interviews explored the following:

5.2.5 The importance of specific tracking factors

When exploring opportunities for self-tracking technology in Chapter 4, several desired active and passive tracking factors were uncovered. The first part of the interviews explored this in an interactive exercise with users where they were asked to rank the factors according to their preferences. The findings of which will reveal the average ranked preference of these specific factors to guide the subsequent re-design of RADAR-Base

5.2.6 Connecting RADAR-base to EWS

The second part of the interview aims to establish relationships between how RADAR-Base can be used to understand EWS, using the EWS checklist. Users were provided with a list of the specific data sources that RADAR-base could passively track (via the pRMT app) and asked to connect them to EWS that were relevant to them, encouraging further discussion on how they could prove useful. Users were also encouraged to speak about other useful tracking factors that have not been previously mentioned.

5.2.7 Design changes

The findings from above, as well as those that arose in Chapter 4, were discussed internally within the supervisory team of this research and then shared with the development team of the industrial sponsors of this study, to create design recommendations. The development team were responsible for implementation, maintenance and development of RADAR-Base. Design recommendations were put together according to user importance, as indicated in the subsequent interviews, which was balanced against development effort and project timelines. Due to the nature of doctoral work, only a finite amount of time was allocated to development.

5.2.8 Alpha testing phases

Once the design recommendation was implemented, an initial prototype was created for alpha testing. Two users who were involved in the series of research thus far, as well as a user who had been involved in the development in the research (P001) tested the prototype in a series of alpha testing lasting 5 weeks, with two rounds of prototype iteration preparing the prototype for testing in a real-world context with a larger group of target users. The process of alpha testing is outlined in Figure 9 below. This testing segment period was augmented with two check-in interviews, one after the first period of 1-week of testing and one was after the second period of 2-week testing. The interviews explored the usability, user experience and functionality of the self-tracking tool. Though usability testing is not a research methodology

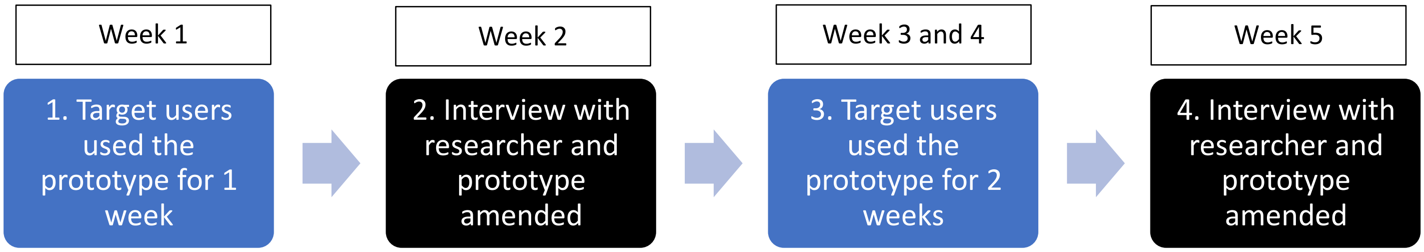


Figure 9 Alpha testing process

5.2.9 Analysis

5.2.9.1 Interviews

To establish hierarchy of preferences for specific tracking factors, a feature rank analysis was conducted (Finch, 2022). This is a process whereby users were asked to organize the specific tracking factors in order of preference and then an *average ranking analysis* was applied. Ranking questions calculate the average ranking for each answer choice so it can be determined which answer choice was most preferred overall. The answer choice with the largest average ranking is the most preferred choice.

Other parts of the interviews will be summarized as is, in verbatim to show connections between the specific tracking factors, RADAR-base pRMT and the EWS checklist.

5.2.9.2 Alpha testing

The findings from the alpha testing phase were categorised using affinity mapping techniques. Affinity mapping is a popular method used in design thinking, which organizes observations or ideas from a research study into distinct clusters (Experience, n.d.-a). Affinity mapping is not a research methodology; rather, it is a tool utilized in the User Experience industry to efficiently organize observations from usability testing methods, providing a swift and agile approach for structuring findings. The number of bugs and crashes will be reported at each alpha testing phase.

5.2.10 User sample

In total, we had n=10 users with lived experience engaged in the interviews including P001 and P002 who were involved the design and development of the research. For the alpha testing, a total of n=2 users engaged in this, one of which being P001.

5.3 Results

5.3.1 User interviews

5.3.1.1 Specific tracking factors

Feature rank analysis was applied to the n=10 user interviews and revealed the average preference for the specific tracking factors that were outlined during the Concept Generation and Ideation phase. On average, measuring sleep was ranked as the most preferred tracking factor, whereas tracking employment information was ranked as being the least preferred. The ranking of all the tracking factors is indicated in Table 13.

Specific tracking factor	**Ranking**
Measuring sleep	1
Keeping tabs on social interaction	2
Track access to medication	3
Connecting to wearables	4
Track finances	5
Diary feature for reflection	6
Track personal hygiene levels	7
Track food intake	8
Scales/graphs	9
To-do-lists	10
Track employment information	11

Table 11 Specific tracing factors in ranked order.

When discussing the importance of these tracking factors, it was indicated that MEASURING SLEEP is a "*crucial part of BD*" and underpins one of the most salient EWS when it comes to relapse. For example, P001 described that "*sleeping is the first thing that goes out the window for me when I'm not well. Too much sleep when feeling depressed and too little sleep when feeling manic… it's so important to track this (P001)*". When considering KEEPING TABS ON SOCIAL INTERACTION, behaviours such as increased phone usage was important to keep measuring, users described behaviours such as "*over-sharing emotional content on social media (P007)*" and "*increased Facebook messaging* *people from the past to re-connect or over-apologise (P009)"* when experiencing manic symptoms, and oppositely "*not being interested in social media, messenger or calls (P009)"* when experiencing depressive symptoms. TRACKING ACCESS TO MEDICATION was deemed an important factor by users as non-adherence to medication indicated a core EWS. Likewise, the use of certain PRN (as needed) medication indicated *"that things may be bad… and that you maybe need some additional support (P005)".* CONNECTING TO WEARABLES was also deemed as being highly insightful, however ranked more in the middle as not all users owned a wearable device. For those that did, they described factors such as *"tracking heart rate for anxiety (P001)*", "*tracking steps (P006)*" also provided useful insights into their EWS. TRACKING FINANCE, though described as being important *"as money spending during manic phases is so common (P001)"* wasn't ranked as being higher, as users commonly described that this could seem more like "*monitoring*" for them and would need to be developed in a way where the user is always in control of this.

Using a DIARY FEATURE FOR REFLECTION was described as useful as it allows the user to look back retrospectively and identify factors which influence the way they currently feel. For example, a user described that if her mood was low one day she would "*look back at her diary and identify the triggers such as forgetting to take my Lithium and it can help you to work backwards (P005)".* The ability to TRACK PERSONAL HYGIENE LEVELS was not commonly described as being important to users, but for some an insightful factor especially for EWS for depression. For example, one user described that "*the best barometer of me becoming depression was my relationship with my morning shower (P002)".* The other factor which fell into the lower levels of hierarchy included the ability to TRACK FOOD INTAKE. Again, this factor was not commonly described as being important, though for user users tracking things such as "calories" or "caffeine" provided insights into their mood. When SCALES/GRAPHS were explored, certain useful graphs were demonstrated as being useful such as the Bipolar UK Mood Scale (*Bipolar UK Mood Scale*, n.d.), clinically validated charts, visual mood ratings scales (e.g. smiley faces) or simply using a numeric scale between 1-10 to rate their mood. Using scale/graphs provides insights as one user described that *"the thing that amazed me about it was how* *much my mood fluctuates normally, even when I'm not having an episode. (P004)".* TO-DO LISTS though ranked as being less preferred, the process of writing a list "*helps with managing anxiety (P003*)" and the number of items completed provides insights into mood. The final self-tracking factor was explored and ranked the least preferred was TRACKING EMPLOYMENT INFORMATION. The reason behind this is applicability, as many users were not employed and in vast majority of their cases this was not related to their experience of BD. For those users who were in employment, they found that the amount they accessed support, or reasonable adjustments (such as sick leave) at work indicated how much "*I am struggling at that time" (P001).*

5.3.1.2 Connecting RADAR-base with EWS

The second part of the interview explored the connections between EWS and passive sensing methods that can be achieved via RADAR-Base (pRMT). Table 14 and 15 below indicate the early warning sign for depression and mania and outlines, from the user's perspective, which of the sensing methods via pRMT can be used to provide insight into the respective EWS. User quotes are also provided for additional details on the connection. The final column outlines those tracking factors that are not currently collected by pRMT but were described as useful in providing insights into the respective EWS.

Early warning signs for depression	**RADAR-Base pRMT**	**User quotes**	**Non pRMT tracking factors**
Low motivation/can't get started	– Sleep information (via Fitbit) (P001, P002) – Phone usage (P003) – GPS (P003, P006) – Step count (P006) – Which apps are being used and for how long (P004, P005, P008)	"For phone usage, this would be more, and GPS this would be less" (P003) "Just social media usage will be sufficient" (P004, P005) "Gmail app usage will decrease" (P008)	– To-do-lists (P005)
Difficulty concentrating	– Phone usage (P003) – Which apps are being used and for how long (P003) – Step count (P005)	"Lots of things open and not spending lot of time on them" (P003)	– Driving patterns (P001)
Low in energy	– Heart rate information (via Fitbit) (P003)	"For example, phone usage	

	– Phone usage (P003)	increased, heart rate would be the same” (P003)	
Feeling tired/listless	– Sleep information (via Fitbit) (P001, P002, P003) – Heart rate information (via Fitbit) (P003, P005) – Heart rate information (via Fitbit) (P005)	“Sleep information would be increased” (P003)	
Loss of interest in activities	– Step count (via Fitbit or smartphone) (P001) – Phone usage (P003) – Which apps are being used and for how long (P003, P008)	“Both phone usage and app usage will be decreased” (P003) “Camera app will be decreased as I usually enjoy photography” (P008)	– To do lists (P005) – Employment information (P005)
Sleeping too much	– Sleep information (via Fitbit) (P001, P002, P003, P005, P006) – Heart rate information (via Fitbit) (P003) – Phone usage (P008) – Step count (P005)	“Changing alarm app to later” (P008)	
Thinking about suicide/death	– GPS location (P004) – Phone usage (P006, P008) – Which apps are being used and for how long (P001)	“GPS showing danger locations like bridges” (P004) “Calls to crisis team increases” (P008)	– Search history (P003, P008) – Diary entries (P005)
Want to be alone	– GPS location (P001) – Phone usage (P003, P008)	“Phone usage will be decreased” (P003)	– Social media usage (P009)
Less talkative	– Phone usage (P003, P008) – Which apps are being used and for how long (P005)		– Audio recordings of phone conversations (P003)
Neglecting hygiene or personal appearance	– Sleep information (via Fitbit) (P003) – Phone usage (P009) – GPS location (P009) – Ambient light (P009)		– Hygiene tracking (P003, P009)
Feeling anxious	– Heart rate (via Fitbit) (P003, P005) – Which apps are being used and for how long (P008)	“Heart rate will be increased” (P003) “More Youtube app searching for ASMR/Relaxing videos” (P008)	
Less interest in sex	– Which apps are being used and for how long (P003)	“Such as a decline in dating apps” (P003)	
Negative thoughts pop into your mind			
Afraid of going crazy			– Diary entries (P008)

Disinterest in food	– Which apps are being used and for how long (P003)	“Not using food delivery apps” (P003)	– Weight information (P001,P008)
Feeling sad or wanting to cry	– Phone usage (P001)		
Loss of interest in people	– Phone usage (P001, P003, P005, P008”) – GPS location (P003,) – Which apps are being used and for how long (P005)	“Phone usage being less” (P003) “Decreased WhatsApp usage” (P008)	
Not being able to get up in the morning	– Sleep information (via Fitbit) (P002, P003, P005) – GPS location (P003) – Step count (P003, P005) – Phone usage (P006, P009) – Ambient light (P009)	“Sleep info, GPS and step count being low” (P003)	– To-do-lists (P005)
Senses seem duller	– Phone usage (P001)		
Worrying a lot	– Which apps used and usage timing (P001)		
Cannot face normal tasks	– Step count (P001) – Phone usage (P003) – Heart rate information (via Fitbit) (P005) – Which apps used and usage timing	“Phone usage being more” (P003) “Not being able to open Gmail app” (P008)	– To-do-lists (P005) – Employment information (P005)
Cannot get off to sleep	– Sleep information (via Fitbit) (P001, P002, P003) – Ambient light (P003)		
Feeling irritable	– Sleep information (via Fitbit) (P001) – Heart rate information (via Fitbit) (P003) – Phone usage (P006) – Which apps used and usage timing (P006)	“Heart rate being high” (P003)	
Feeling very guilty	–		– Diary entries (P005)
Using sleeping tablets	– Sleep information (via Fitbit) (P003, P005) – Heart rate information (via Fitbit) (P005) – Step count (P005)	“More deep sleep” (P003)	
Using street drugs	– Heart rate information (via Fitbit) (P005) – GPS location (P005) – Step count (P005) – Phone usage (P005)		– Levels of spending (P005) – Driving patterns (P005)
Waking early in the morning with feelings of dread	– Heart rate information (via Fitbit) (P002, P003) – Sleep information (via Fitbit) (P003) – Phone usage (P008)	“Phone calls to crisis team will increase” (P008)	

Table 12 Table with EWS for depression, and RADAR-Base pRMT can be used to provide insights into the respective EWS

Early warning signs for mania $^+$	**RADAR-Base pRMT**	**User quotes**	**Other tracking factors**
Ideas flowing too fast	– Phone usage (P003, P008) – Which apps used and usage timing (P003, P004, P006, P008) – Battery level (P003)	*"Social media usage"* (P004) *"Notes app on my phone"* (P008)	– Speech rate (P002) – Search history (P003)
Stronger interest in sex	– Which apps used and usage timing (P003) – Heart rate information (via Fitbit) (P003, P005)		– Emergency contraceptive usage (P008)
Spending money more freely	– Which apps used and usage timing (P003) – Step count (P003)		– Levels of spending (P002,P004, P005, P006, P008, P009, P010)
Racing thoughts	– Phone usage (P003, P008) – Which apps used and usage timing (P003, P008) – Sleep information (via Fitbit) (P003, P005) – Heart rate information (via Fitbit) (P009)	*"Increased use of notes app"* (P008) *"Increased heart rate* (P009)	– Driving patterns – Search history (P003) – Speech rate (P002)
More talkative	– Phone usage (P002, P003, P005, P008)	*"Increased volume of texts"* (P002)	
Being disinhibited or outrageous	– Heart rate (via Fitbit) – Which apps used and usage timing (P003)	*"Social media posting more"* (P003) *"Content of text message would be insightful"* (P002)	– Social media usage (P002)
Can't get off to sleep	– Heart rate (via Fitbit) (P002) – Sleep information (via Fitbit) (P003, P006, P010) – Ambient light (P003)		
Feeling irritable	– Heart rate (via Fitbit) (P003)	*"Heart rate increased"* (P003)	
Energetic/very active	– Heart rate (via Fitbit) (P002, P003, P005, P008) – Sleep information (via Fitbit) (P003, P008) – Step count (P003, P005) – GPS location (P003, P006)		– Driving patterns (P006) – To-do lists (P008)
Feeling creative	– Heart rate (via Fitbit) (P003) – Which apps used and usage timing (P003, P008)	*"Buying more creative products, using creative apps"* (P003) *"Notes app, photo app and camera increased usage"* (P008)	– Artistic output (e.g poems) (P002) – To-do lists (P008) – Drawings on paper (P008)
Feeling emotionally high	– Heart rate (via Fitbit) (P003)		
Not needing much sleep	– Sleep information (via Fitbit) (P002, P003, P008)		

	– Step count (P005) – Heart rate (via Fitbit) (P005)		
Involved in many projects	– Step count (P003) – Phone usage (P003, P008) – Which apps used and usage timing (P008)	*“Use of email apps and phone calls will be increased”* (P008)	– To-do lists (P003)
Difficulty concentrating	– Phone usage (P003) – Which apps used and usage timing (P003, P006)	*“Lots of apps open and discarded”*	
Feeling strong or powerful			– Levels of spending (P006) – Search history (P003)
Reckless pleasure seeking	– Sleep information (via Fitbit) (P007) – Which apps used and usage timing (P003) – Heart rate (via Fitbit) (P003, P007) – Phone usage (P004, P007) – GPS location (P006) – Local weather (P007) – Step count (P007) – Number of contacts (P007)	*“Apps such as social media”* (P004) *“All of these data streams will show increases for me”* (P007)	– Levels of spending (P004)
Being uncooperative			
Wanting to party all night	– Sleep information (via Fitbit) (P005, P006) – Which apps used and usage timing (P003) – Heart rate (via Fitbit) (P005) – Step count (P005, P010) – Phone usage (P002)	*“Using lots of social apps”* (P003) *“Step count being quicker than normal in a small period of time”* (P010)	
Feeling very important	–		
Thinking my thoughts are being controlled	– Phone usage (P003) – Which apps used and usage timing (P003)		
Auditory hallucinations			
Anxious	– Heart rate (via Fitbit) (P003) – Which apps used and usage timing (P009)		
Having bizarre thoughts	– GPS location – Phone usage (P002)	*“Location and time are important* (P004) *“Text message content would be insightful”* (P002)	
Senses seem sharper			– Driving patterns (P003)
Feeling very religious	– Which apps used and usage timing (P003, P006)	*“Using religious apps – prayers and things”* (P003).	

+EWS for mania also include EWS for hypomania

Table 13 Table with EWS for mania, and RADAR-Base pRMT can be used to provide insights into the respective EWS

5.3.2 Design changes based on user interviews

The findings from the interviews were discussed within the internal team in order to develop a list of design and development changes, that were informed by the user's needs, preferences and individual variability and to tailor RADAR-Base towards this. Table 16 outlines the final changes that were decided upon by the internal team, as well as rational and the process that was adopted. Design changes were balanced according to user preferences and accounting for

Design and development change	**Rationale**	**Process**
To add a medication question to the aRMT app so that users can track their daily medication	Track access to medication was highly ranked in terms of being a preferred tracking factor	aRMT app will deliver a daily questionnaire to the user to input whether they have taken their AM dose of medication, and repeat this for the PM dose of medication
The add the ability to track spending via pRMT	The utility of spending information was indicated by the specific tracking factors, as it was ranked as being moderately preferred. It was also described as useful non-pRMT tracking factor forEWS, especially for mania.	TrueLayer was integrated into RADAR-Base pRMT (www.truelayer.com) which allows for the collection of financial information from mobile banking apps on users, through secure autentication. phone, through users' consent. Further details about Truelayer are provided in Appendix 1
To add a diary feature to the aRMT app to allow for reflection	The utility of the diary feature was described as moderately preferred by users, and also indicated as an additional useful non pRMT tracking factor for EWS, especially for depression	
To add user preferred scales/graphs to the aRMT app to increase engagement	The following scales/graphs were described as commonly used by users: -Visual mood diary using smiley faces (e.g. Daylio app (*Daylio - Journal, Diary and Mood Tracker*, n.d.) - Bipolar UK Scale – a 10-item mood scale with descriptions of mood (*Bipolar UK Mood Scale*, n.d.) -Quick Inventory of Depressive Symptomology (QIDS-SR) -Altman Self-Rating Mania Scale (ASRM) -Snaith-Hamilton Pleasure Scale (SHAPS) The full scale/graphs can be found in Appendix 2	These graphs were implemented into the aRMT app of RADAR-Base
The ability to turn certain tracking factors on/off according to user control	The theme around user control was acknowledged in the subsequent interviews, but also acknowledged in Chapter 4. The ability to pick and choose which tracking factors were on/off and when was imperative to the whole user experience, especially with more sensitive tracking described factors such as levels of spending	The pRMT app could be personalised according to users' preference where tracking factors could be turned on/off at any point during the study.

Table 14 Design and development change

5.3.3 Alpha Testing

A working prototype of RADAR-Base aRMT and pRMT was enrolled for a series of alpha testing with n=2 users. Both apps went through two phases of testing. At the end of phases, feedback was taken and suggestions for improvement and iteration were put into development before the next phase of testing started. The results of the affinity mapping exercise are outlined in Figure 11, with categories and corresponding quotes from users according to the alpha testing phase. The number of bug fixes and the types of changes that were done at each phase are outlined in the subsequent Figure 10

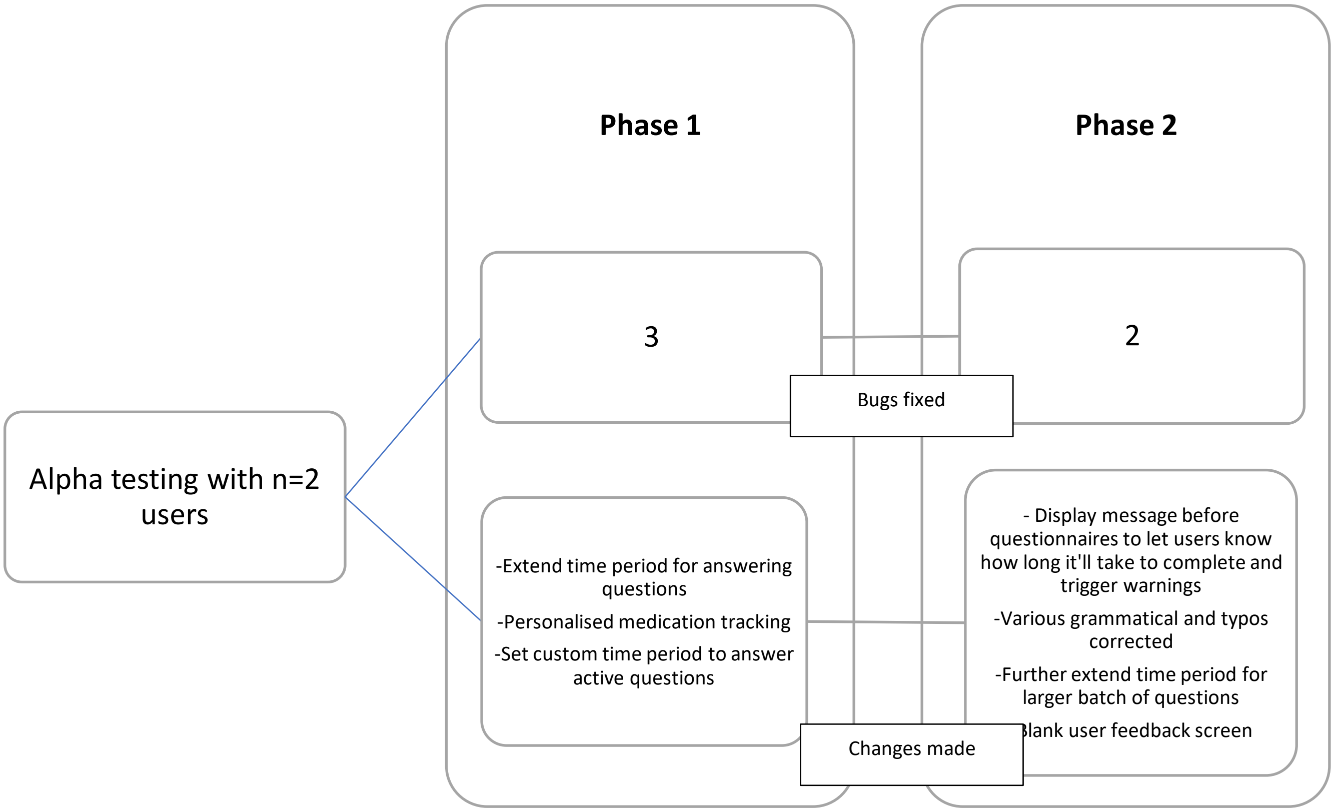
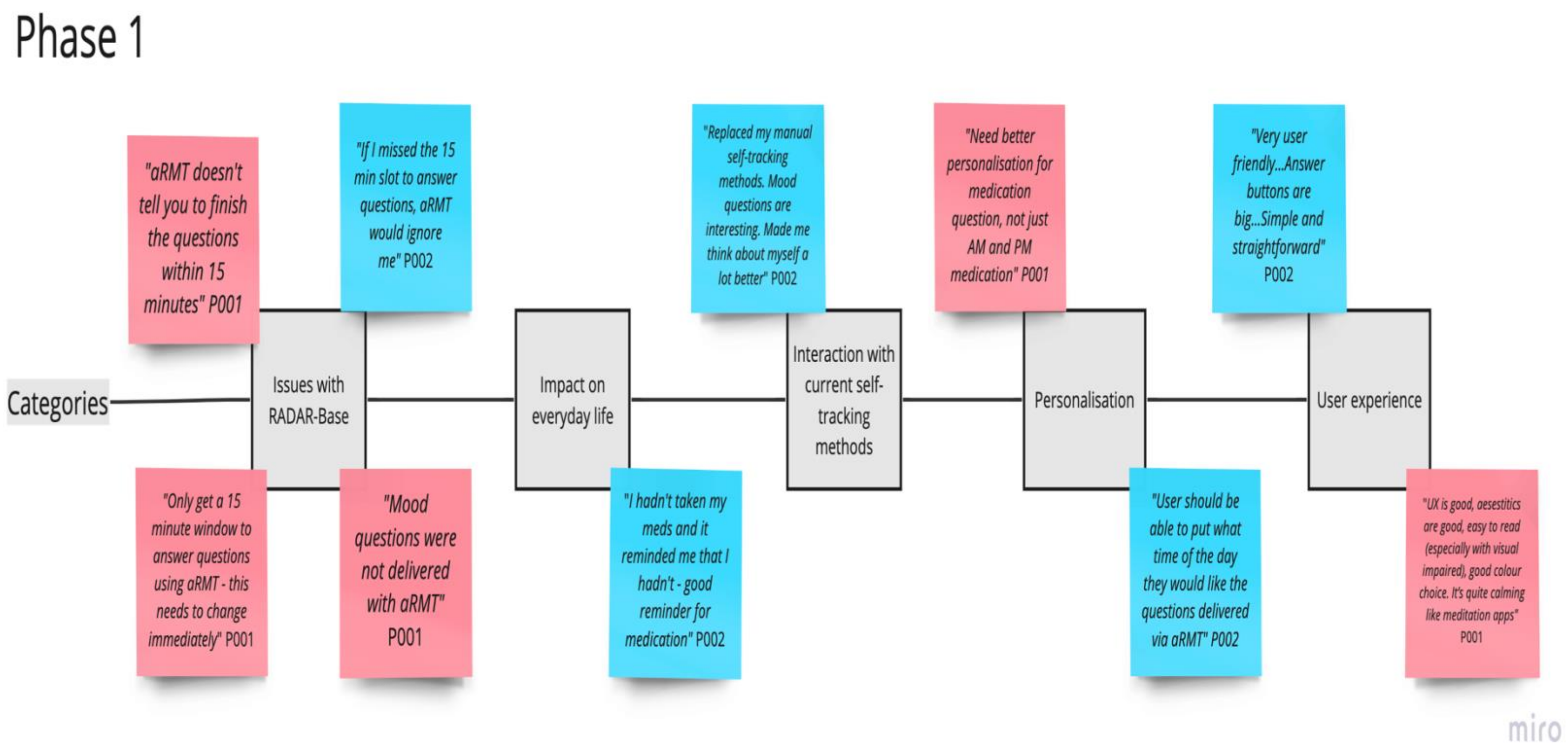


Figure 10 Bugs and changes made from Alpha Testing



"aRMT doesn't tell you to finish the questions within 15 minutes" P001

"If I missed the 15 min slot to answer questions, aRMT would ignore me" P002

"Only get a 15 minute window to answer questions using aRMT - this needs to change immediately" P001

"Mood questions were not delivered with aRMT" P001

"I hadn't taken my meds and it reminded me that I hadn't - good reminder for medication" P002

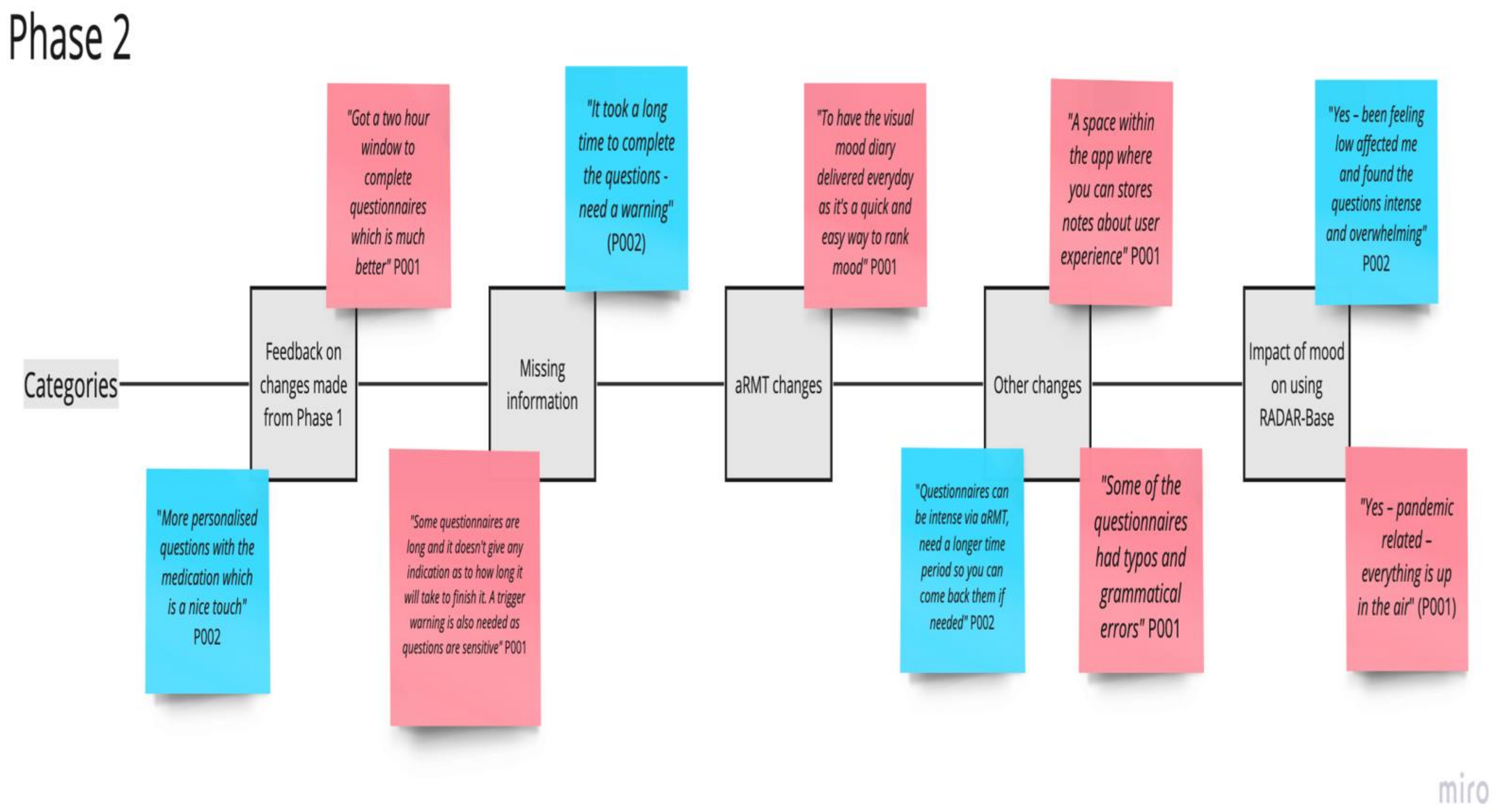
"Replaced my manual self-tracking methods. Mood questions are interesting. Made me think about myself a lot better" P002

"Need better personalisation for medication question, not just AM and PM medication" P001

"User should be able to put what time of the day they would like the questions delivered via aRMT" P002

"Very user friendly...Answer buttons are big...Simple and straightforward" P002

"UX is good, aesestitics are good, easy to read (especially with visual impaired), good colour choice. It's quite calming like meditation apps" P001



"Got a two hour window to complete questionnaires which is much better" P001

"More personalised questions with the medication which is a nice touch" P002

"It took a long time to complete the questions - need a warning" (P002)

"Some questionnaires are long and it doesn't give any indication as to how long it will take to finish it. A trigger warning is also needed as questions are sensitive" P001

"To have the visual mood diary delivered everyday as it's a quick and easy way to rank mood" P001

"A space within the app where you can stores notes about user experience" P001

"Questionnaires can be intense via aRMT, need a longer time period so you can come back them if needed" P002

"Some of the questionnaires had typos and grammatical errors" P001

"Yes - been feeling low affected me and found the questions intense and overwhelming" P002

"Yes - pandemic related - everything is up in the air" (P001)

Figure 11 Quotes from Alpha Testing phases

5.3.3 Final version of RADAR-Base

The final version of RADAR-base is outlined in Figure 12, which outlined the final features for pRMT, aRMT and other features that arose as a result of the above findings. The diagram identifies existing features of RADAR-Base, and those that arose as a result of the findings from the current study and in Chapter 4. Figure 13 provides the screenshots of the final product to give a visual insight into the final product which will be evaluated in the next chapter.

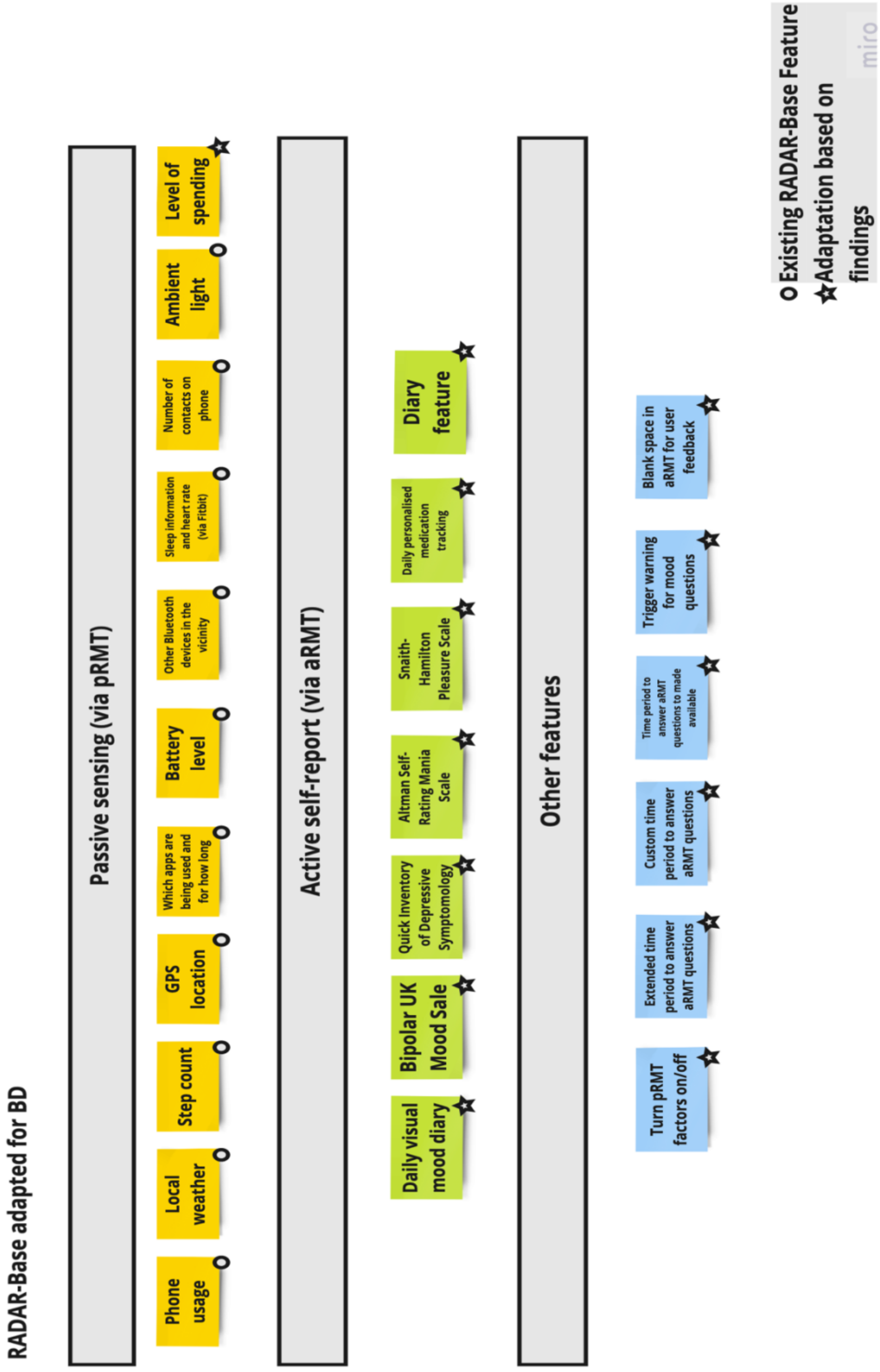
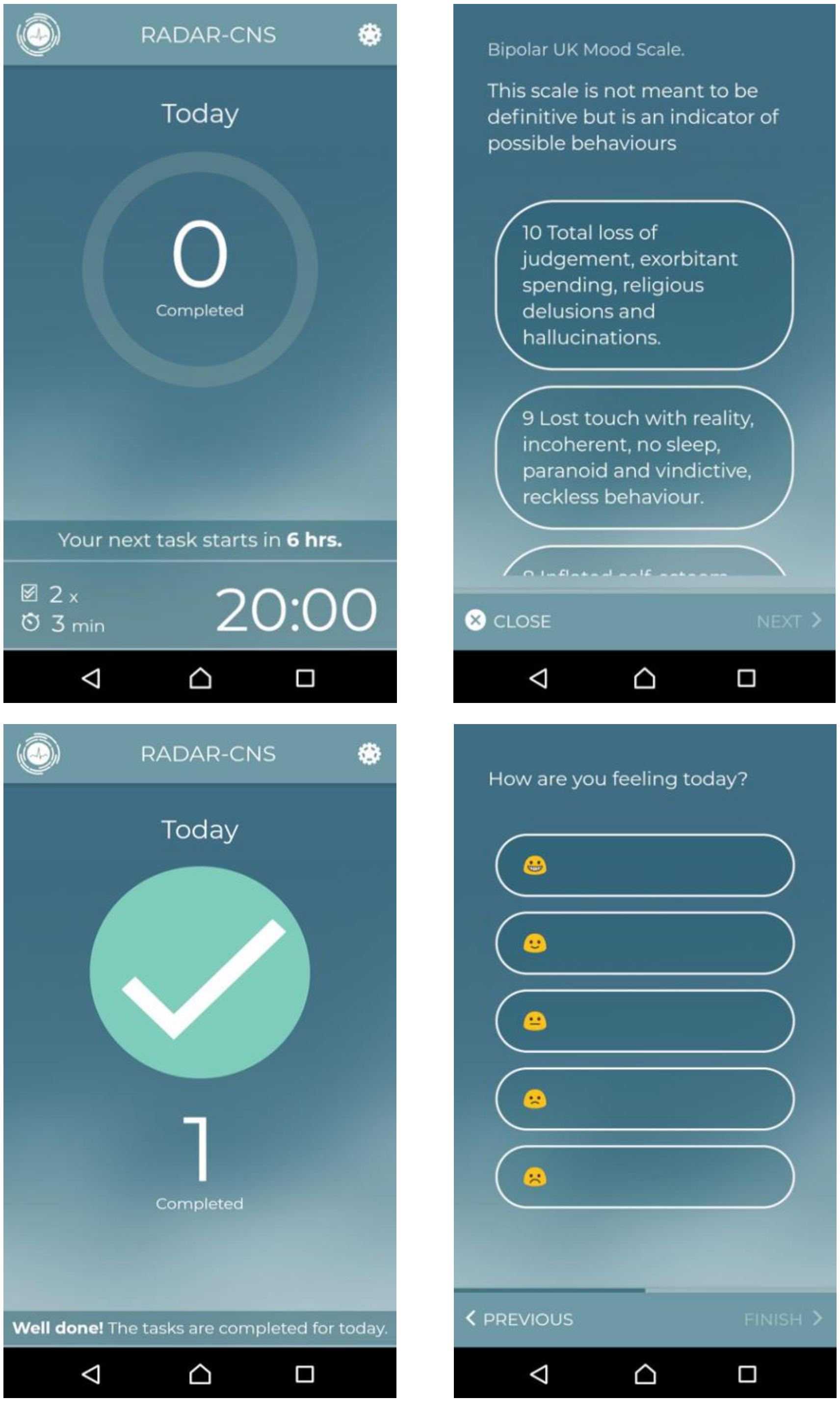


Figure 12 Final version of RADAR-Base



Your next task starts in **6 hrs.**

Bipolar UK Mood Scale.

This scale is not meant to be definitive but is an indicator of possible behaviours

10 Total loss of judgement, exorbitant spending, religious delusions and hallucinations.

9 Lost touch with reality, incoherent, no sleep, paranoid and vindictive, reckless behaviour.

Well done! The tasks are completed for today.

How are you feeling today?

Figure 13 Screenshots of RADAR-Base

5.4 Discussion

5.4.1 Principal results

One of the primary aims throughout this body of work is to combine principles of PPI and HCI and adhere to best practice principles for the development of self-tracking technology for BD (Majid et al, 2021). The processes outlined during this research aim to contribute towards a working example of best practice to ensure the technology is developed as closely aligned to user's needs, preferences and individual variability. The level of user involvement that has been present in this study, as well as Chapter 4 haas notable differences compared to the user-involvement that is seen in similar studies which has been indicated in the literature review conducted in Chapter 2 (Majid et al, 2021). Our findings in relation this reveal the complex nature of self-tracking in terms of user preference for tracking factors, the utility of tracking technology and the connection to individual EWS for mania and depression. This further contributing to RQ3.

The findings in relation to specific tracking features for BD ordered by user preference are novel findings and have not been discussed in other key papers (Ortiz et al., 2021) (Matthews, Murnane, & Snyder, 2017a) (Murnane et al., 2016) (Matthews, Murnane, Snyder, et al., 2017a). Though specific tracking factor in relation to EWS have been described in key paper, to our knowledge there has been a lack of exploration around user preference, and we hope that our findings can guide future researchers/designers in this domain. In the next section, we will focus on some of the core findings around self-tracking factors and EWS.

The findings in relation to technology usage during BD episodes has been explored in a core paper by (Matthews, Murnane, Snyder, et al., 2017a), who found explore specific indicators such as technology mediated communication, streaming behaviour (e.g. Netflix), e-commerce, gambling and gaming etc and the manifestations of this indicators dependant on depression or mania. Many of these findings revealed in this research echo the findings from this study, such as the decrease or increase of phone usage being indicative of depression and mania respectively. As well as validate some the tracking indicators that were indicated that were not previously mentioned such as social media usage and search history having utility in understanding depression and mania.

Measuring sleep was universally described as the most preferred tracking factor by users, and highly referenced in terms of increases/decreases in relation to various EWS for both mania and depression. This is echoed in a wealth of literature surround sleep disruptions and BD (Gold & Sylvia, 2016) (Gruber et al., 2011) and also a core passive sensing factor currently employed in self-tracking technology for BD (Ortiz et al, 2021). As users described, during mania or hypomania sleep disruptions typically presented as a decreased need for sleep with studies reporting findings that 69%-99% of BD reported a lessened need for sleep during manic episodes or difficulties in falling and/or staying asleep (Robillard et al., 2013). During depression, sleep disruptions typically presented as hypersomnia, or excessive sleepiness. Hypersomnia is largely prevalent in BD with rates between 38%-78% (Kaplan et al., 2015). This core tracking factor will be available to track using the pRMT app on RADAR-Base for users, using connection to a Fitbit if available

Another novel finding is in relation to the utility of spending levels, and their connection to EWS. Though this was moderately preferred by the group of users and described as something that would have to be "carefully designed, with full user control", it was prescribed as an insightful tracking factor in relation to key EWS for mania including: spending money more freely, feeling strong or powerful and reckless pleasure seeking. This was usually described in the context of increasing for EWS for mania. The relationship between BD and financial difficulties is a well-established one, where it was described individuals with increased manic symptoms are 70% more likely to report spending large amounts of money. Patients with BD may have reduced ability to manage personal finances, which may arise as a result of increased levels of impulsivity (Cheema et al., 2015). Given the user described utility of this tracking factor, which is supported by available research, it has been incorporated as a passive sensing method (via pRMT) in the final design of RADAR-Base for BD.

The three other tracking factors which arose from this research which have received little attention in the self-tracking literature thus far are the following: track personal hygiene levels, track employment information and driving patterns. Though the utility of these tracking factors were not universally mentioned by users, but they still provide insight into the individual variability of self-tracking in BD which should be considered by future researchers and designers in this domain. In relation to personal hygiene levels, a decrease in this has been described during depressive phases where one user described that the *"the best barometer of me becoming depression was my relationship with my morning shower" (P002).* When connecting this to the literature for depressive symptoms, in a very recent study it described that though issues concerning personal hygiene and grooming has been described n personal narratives for person, they have been rarely included in symptoms lists or outcomes in research studies (V. Stewart et al., 2022). This study explored this further and found that for some users, like in the current research, personal hygiene and grooming standards were an integral aspect of their depression, with could negatively impact recovery. Unfortunately, there was no scope within the current research to consider this factor further in the context of self-tracking technology for BD, but future research should consider this.

The other two tracking factors that arose that were of interest were employment information and driving patterns. In respect to employment information, it only applied to a handful of users and for those it did, accessing things such as "sick leave or reasonable adjustments at work" were indicative of deteriorating mood. When examining the available literature, in a systematic review it was found that BD appears to lead to workplace underperformance and up to 40-50% are impacted with a downward slide in occupational status over time (Marwaha et al., 2013). In relation to driving patterns, this was a new tracking factor that arose from the EWS exercise and was described as being useful particularly for many EWS for mania from the user's perspective. Driving fast was described as one of the three types of disinhibited behaviours when hypomanic in a survey with a larger sample of users with BD where 62.5% of the sample of n=88 described this behaviour. Though both the utility of driving behaviour, as well as employment information is clear, the limitations on the development resources of this research meant they were not able to explore in the context of self-tracking technology. Again, future research would benefit from exploring this.

Another core aim of this research was to understand individual variability within the experience of BD and its role with technology development. The complex nature of self-tracking in BD has been extensively discussed in Chapter 4, however this complexity juxtaposed on EWS, and sensing methods has not yet explored. The findings described in Table 11 and 12 provide a key insight into this complex nature. Not only do users describe a combination of different EWS that are relevant to them, but they also describe a combination of different sensing methods which could have utility for the EWS. For example, for the depression EWS *cannot face normal tasks*, one user describes step count as being useful, another describes heart rate as being useful whereas another describes app usage as being useful. When asked further information, one user described that their *“phone usage would be more*” (P003) (i.e., increased) for this EWS, whereas for another user it was that app usage would be low (i.e., decreased) such as not being able to “*open the Gmail app*” (P008). As it can be seen, there is no consensus and one’s persons experience, even of this single EWS, does not necessarily reflect that of another. Within related literature, a dichotomy exists within technology indicators with increases for mania and decreased for depression (Matthews, Murnane, Snyder, et al., 2017b). However, even for this specific EWS this is not the same as for one user increased usage relates to this EWS, and for another decreased usage is an indicator.

This echoes the findings from Ortiz et al (2021) about individual variability and its role in technology development, as well as findings from Majid et al. As there is no consensus in illness experiences, designer and researcher should strive to provide a level of tailoring so that technology can be adapted towards needs and preferences. In our attempt to do this via the exploration of RQ4, we allow users to turn tracking factors on/off (via pRMT), personalise their medication tracking (via aRMT), customise the time period they wish to answer questions about mood and wellbeing (via aRMT). Many of these design changes arose from the alpha testing series, demonstrating the valuable nature of early involvement in the development of technology to be more aligned to users' needs and preferences, The impact this has on usability and general user experience will be explored in the next chapter of this doctoral research.

5.4.2 Limitations

Due to constraints of doctoral research in relation to time and resources, not all self-tracking factors that arose from this research were able to be explored in the context of our developed tool. These factors are important to the individual variability that exists within the experience of BD and future research should explore their utility in the design of self-tracking tools for BD. To expand further on these constraints, in an ideal scenario, the exploration of user's needs, preferences and individual variability should have guided the development of a self-tracking tool "from scratch". However, given the constraints on time and resources, an existing tool (RADAR-base) was adapted. How this may have primed the users' responses in the interviews has not been considered. This research used the same users that have been involved in the lifecycle of this research from Chapter 4, both in the interviews and alpha testing. The impact this has had on the findings has not yet been explored and will be considered in the next stages of this research.

Also, the constraints on time and resources did not allow the exploration of patterns of different EWS together. Each EWS of depression or mania can have non-health related reasons for change and EWS are therefore only predictive of becoming unwell when several occur together. Potentially users might prefer to draw their own conclusions from the information given to them by the app or they might want the app itself to analyse the pattern of changes in EWS to give a warning to users. They might want to get the analysis from RADAR-base and have the raw output to decide themselves. It is possible that this is a feature that the user would choose to select on an individual preference basis. We were not able to explore this, but it may be a future topic for research and app development.

There are also some limitations to report upon due to the focus on individual variability in the current study. Firstly, as the solution aims to be individually tailorable, it can be argued that the tool is not generalisable and cannot be applied to a wider sample. Secondly, beyond looking and findings from an individual level, the findings in this chapter did not look at the role of outcomes. The goal of evidence-based medicine is to improve medical outcomes based on the highest quality evidence available, however the current research did not look at how outcomes were improved based on the implementation of the tool. Future research should consider this, and to investigate the impact of the tool of relevant mental health related outcomes.

5.5 Chapter Summary

The current study combines principles of HCI and PPI to explore self-tracking factors and their connection to EWS to guide the re-design of an existing self-tracking tool (RADAR-Base) to sense EWS for BD. The primary findings for this chapter were in relation to design which are reported below. RADAR-Base has both active and passive sensing capabilities and has proven utility for other conditions such as depression, epilepsy and multiple sclerosis. A series of follow-up interviews with n=10 user was run to explore this, which guided the redesign of RADAR-base as per RQ4. Financial tracking was adapted as a capability for passive sensing and a range of user-led, active, self-report capabilities were added such as a daily visual mood diary, Bipolar UK mood scale, personalised medication tracking, diary for reflection, and a range of clinically validated tools to measure depression, mania and anhedonia. To account for individual variability in the experience of BD, a level of tailoring was also developed where users can turn tracking factors on/off as they wish. As per RQ6, a working prototype was enrolled for an alpha testing series with n=2 users with BD, which aimed to explore bug, usability and general user experience. The results of this further guided the design of RADAR-base to be more user-orientated, such as extending time periods to answer questionnaires, trigger warnings for sensitive questionnaires, making the medication tracking more personalised. The final version of RADAR-Base for BD was readied for testing in a real-world context, with a larger group of users which will form the basis of the next chapter (Chapter 6) of this doctoral work.

Chapter 6: Evaluation: 6-month Beta test of the prototype

Chapter summary

- This chapter reports on the 6-month beta test of the prototype, in a real-world context, with n=8 users with BD
- Specifically this section of the research explores how much of the data collected by the app relates to EWS ^{RQ8}, whether the app is usable and acceptable ^{RQ9} and what is the researchers experience of applying the overall methodology in the research ^{RQ10}
- To ensure the high level of involvement of users by combining principles of HCI and PPI, users were consulted at various points of the study via a series of interviews the 1-month, 3-month, 6-month and post study periods.
- Data produced from the study was provided in dashboard format to the users for them to establish connection between their data and their early warning signs of mania and depression.
- Our findings highlight interesting subjective interpretations of the data and their connections to EWS of mania and relapse e.g., users uncovered connections between frequency of their data and respective EWS. These findings contribute towards the knowledge for self-tracking in BD.
- Key findings between frequency of certain data streams and mood changes were also uncovered in the statistical analysis such as a decrease in app usage during onset on manic symptoms. This findings are specific to the app deployed in this research.

6.1 Introduction

The experiences of self-tracking in BD have been explored in Chapter 4 in this research, which explored reasons, methods and challenges for self-tracking using methods which combined principles of HCI and PPI to highly involve the user in these explorations. In a continuation of using these methods, the subsequent Chapter 5 aimed to explore the findings from Chapter 4 more in relation to the design of self-tracking technology and the findings were used to adapt an existing self-tracking tool which uses passive and active sensing method to be more closely aligned with the needs, preferences and individual variability of those with BD. A series of user-led changes were implemented as outlined in Chapter 5. A final version of RADAR-Base was developed, ready for testing in a real-world setting with a larger group of users with lived experience of BD. The overall aims of this research are to enrol a working prototype for a 6-month period of beta testing with those with BD to explore usability, feasibility and acceptability and to explore the relationship of the data to EWS and mood changes. Specifically, we will explore the following research questions not yet explored:

RQ8: How much of the data collected by the sensing system (both passive and active data) relates to EWS?

RQ9: Is this type of technology usable and acceptable for people living with this condition?

RQ10: What are the researchers experience of applying the overall methodology employed in research thus far?

The current chapter will outline the process of evaluating this tool with a group of users over a period of 6-months, using a beta-testing method. One-to-one Interviews were conducted with users at 1-month, 3-month, and 6-month points in the study, as well as a post study data exploration interview. Quantitative data was collected by both the passive (via pRMT) and active sensing (via self-report aRMT) capabilities of RADAR-Base, and the findings of these sensing methods in relation to gaining insights in EWS of mania and depression will be reported upon this chapter.

For RQ9, usability is defined as *"extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* (1SO 9241-11:2018). In terms of measuring usability, the System Usability Scale is a widely used scale as a quick method for exploring usability and will be employed in this study. Acceptability is defined by how well "an intervention will be received by the target population" (Ayala and Elder, 2011) and can be assessed by quantitatively through measures of acceptability or satisfaction or qualitatively through probing questions which are focused on understanding how they are interacted with a “intervention” (Moore et al, 2015). The latter method will be used in this study to explore usability through a series of interviews with users.

6.2 Methodology

6.2.1 Study approach and design

To ensure the involvement of the users in the evaluation of the prototype and following the best practice principles outlined by Majid et al (2021) users were consulted on an ongoing basis throughout the course of the evaluation phase. Specifically at month 1, month 3, month 6 as well as after study completion to explore the self-tracking tool. The aim of this research was to explore the self-tracking tool in a real-world context; therefore, we provided the tool to be used in everyday life, aiming for minimal research interference unless during these specified time points, which users were aware of from the point of study sign up. Figure 14 visualizes the study design, indicating where the elements of the study were drawn from (e.g., HCI and/or PPI) and how they relate to the seven recommend principles.

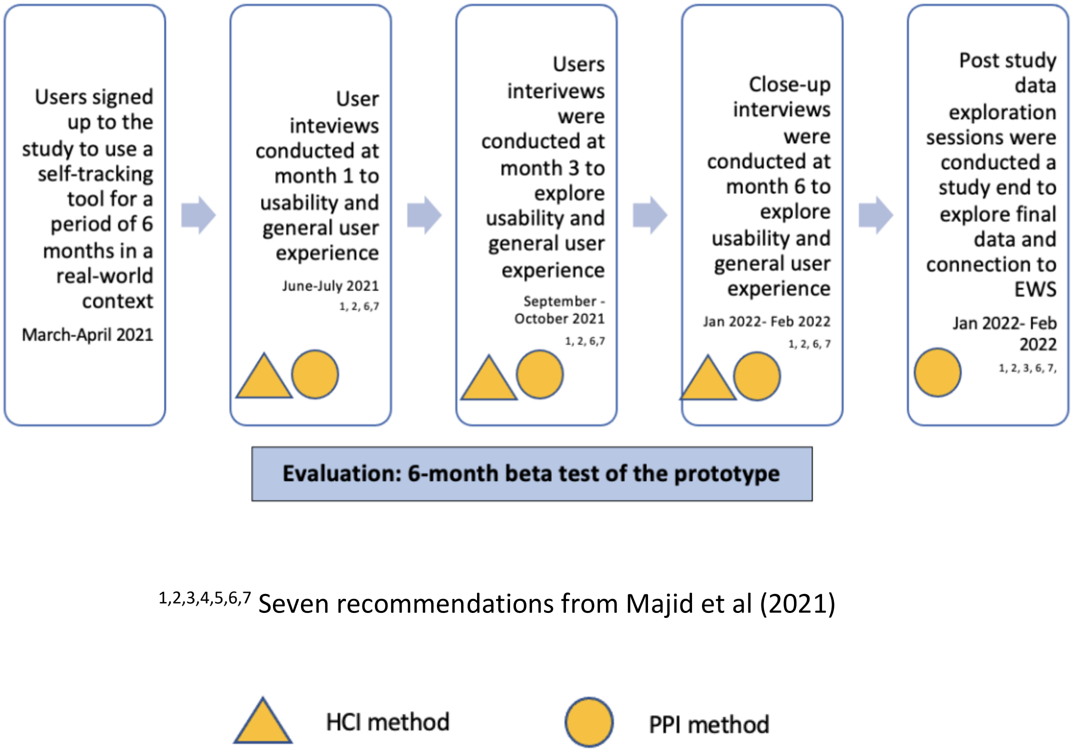


Figure 14 Evaluation methodology

6.2.2 Study recruitment

A number of users who partook in previous parts of this research such as the workshops in Chapter 4 and interviews and alpha testing in Chapter 5 also partook in this phase of evaluation. This ongoing *cohort-style* user recruitment style arose from this series of research as many users expressed an interest in testing the tool, given their contributions towards its development thus far. As the research spanned over a period of 2 years and involved a series of consultation in the form of interviews and workshops, the author of this work formed ongoing relationships with the users. As empowering the users in the development of the tool is one of the primary aim of this study, the author worked with jointly with some users on research, development and process of analysis. Though this approach is not alwayd seen in the development of self-tracking tools in BD as outlined by the findings of the user involvement landscape by Majid et al., (2021), it was seen in one study which formed part of the review. In the study by Bauer et al., (2017) employed an advisory group approach which consisted of eight “consumers” and “consumer advocates” who met monthly to advise the SPIRIT scientific team on all aspects of trial design and conduct, which was resulted in changes to the app and study. The Spirit App is smartphone self-monitoring application that allows patients to undertake modules and complete questionnaires for mental health assessment, similar the tool being developed in this study. Though not commonly seen in technology development for BD, the development of this cohort user recruitment will be reported upon this chapter, providing a reflexive account of the experiences of applying this group from the researcher’s perspective. The impact this style of recruitment on the results of this research will also be explored in the discussion chapter. This is not necessarily defined as a research finding but rather an important insight gained from the experience of conducting the research.

New users were also recruited to the study via local charities who have been involved in previous parts of the research such as Bipolar UK (national) and Bipolar Lift CiC (Nottingham) where a study leaflet was distributed through mailing lists. Doctoral researcher SM also attended some peer support meeting at Bipolar UK where the study was showcased in order to drive user recruitment. Additional recruitment was conducted via Nottingham Healthcare NHS Foundation Trust, where the study received ethical approval. Study leaflets were distributed to clinicians who worked with those with BD, who were able to forward this to those who may be interested.

6.2.3 Study process

6.2.3.1 Recruitment and sign up

Prospective users emailed doctoral researcher SM to express interest in the study. A user information sheet (Appendix 3) was then emailed to the user, which outlined the full process of the research as well as a consent form (Appendix 4). For those users who were still interested in the meeting, a virtual meeting was conducted via Microsoft Teams were SM outlined the process of the study and filled out core study paperwork such as demographic information and completion of the EWS checklist (Lobban et al, 2011). In order to tailor RADAR-Base towards the user, they were asked to complete a questionnaire about their current medication (Appendix 5) and asked which passive data streams they would like ON/OFF during the data collection process (Appendix 6). These could be changed at any point during the study. In addition to this, users were asked when they would receive to aRMT notifications to complete self-report measures, and this included day of the week and time. During the meeting, users were assisted in downloading the active (aRMT) and passive (pRMT) apps from Google Play and received some brief training on how to answer questions about their mood and medication over the course of data collection. They were also made aware of the interviews at the 1-month, 3-month, 6-month and post study points and that the researcher would be in touch closer to the time to arrange these.

6.2.3.2 RADAR-Base adapted to BD

The final version of RADAR-Base which has been enrolled for this period of beta testing is outlined in Figure 12.

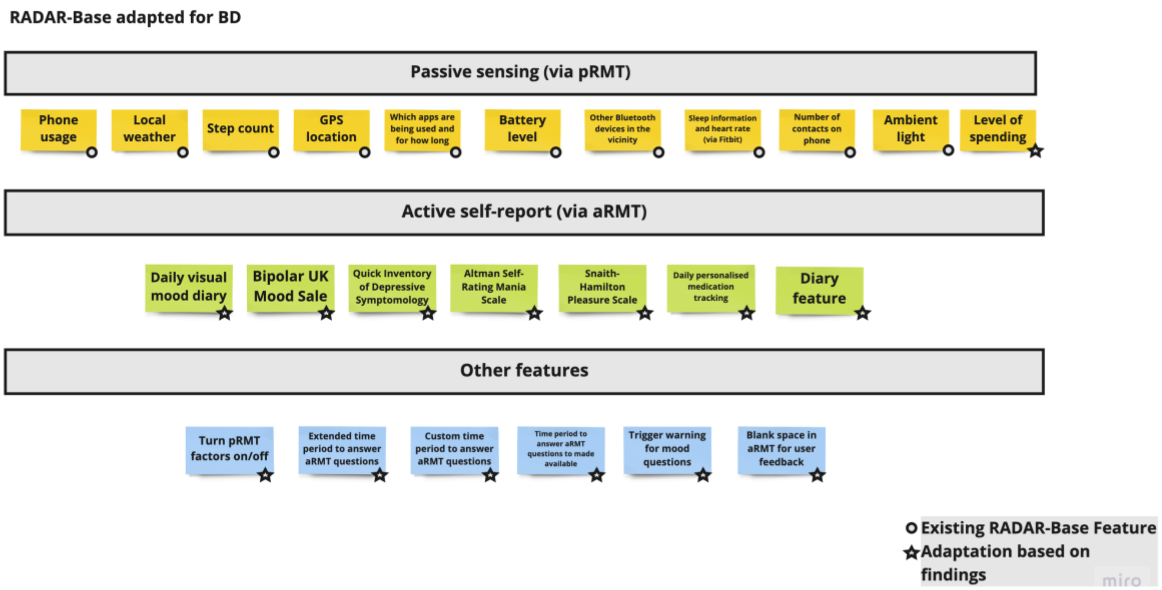


Figure 15 Final version of RADAR-Base

Passive sensing (via pRMT)

Passive data collection via pRMT was conducted automatically for the duration of the study (6-months) and the configuration of data collection will be according to user choice, depending on which data streams they choose to have ON/OFF. Data was passively collected dependant on whether the user's smartphone is:

- Turned on
- Connected to the internet
- pRMT app remained downloaded on the user's phone

If the pRMT app was deleted from the users' phone, this could cease data collection. However, the user was free to delete the app at any point during data collection without providing any reason as part of their right to withdraw from the study.

Active sensing (via aRMT)

Active self-report data collection via aRMT was conducted throughout the duration of the study (6-months) and will consist of daily and weekly self-reporting.

The following self-report measures will be delivered daily:

- Visual mood diary (Appendix 2.1)
- Personalised medication tracking (Appendix 5)

The following self-report measured will be delivered weekly:

- Bipolar UK mood scale (Appendix 2.2)
- Quick Inventory of Depressive Symptomology (QIDS-SR) (Appendix 2.3)
- Altman Self-Rating Scale (ASRM) (Appendix 2.4)
- Snaith-Hamilton Pleasure Scale (SHAPS) (Appendix 2.5)

6.2.4 Experience sampling (ESM)

For the purposes of this research, we will use ESM to deliver the questionnaires to users. ESM is described as a structured diary methods where users are asked to report their thought, feelings, behaviours and/or environment on multiple occasions over time. It consists of asking users to provide these self-reports at random occasions during the waking hours of a normal week. (Larson & Csikszentmihalyi, 2014). This methodology offers a number of advantages when compared to traditional assessments of mental health experiences, which mainly rely on using retrospective data, and may be open to recall bias. With ESM, there is a short space of time between the occurrence of an event and the report of details which decreases the likelihood of memory bias. As this happens on multiple occasions, ESM can provide a rich and descriptive data set which details the users' daily experiences which can provide key assessments into the temporal relation between numerous variables; in this case being actively collected and passively collected data in understanding EWS in BD. Also, as ESM is a method used in real time in everyday contexts, it has high levels of ecological validity (Lenaert et al., 2019).

The field of ESM is still in its early days but systematic reviews have reported high acceptance and feasibility with those with BD (aan het Rot et al., 2012; Naslund et al., 2015) with some studies demonstrating effectiveness in gaining a deeper understanding of how symptoms unfold over time. For example, when examining the perceptions of daily "hassles" and "uplifts" in BD, using ESM it was found that individuals who had more previous depressive episodes and current depressive symptoms experienced negative events as more stressful. Another study explored coping mechanism to both. Another study in bipolar disorder examined coping mechanisms or response styles to both positive and negative mood and found that low mood predicted increased rumination, which reduced mood further. High positive affect predicted high-risk taking, which resulted in even higher positive affect, demonstrating a continual cycle of escalating mood levels (Pavlickova et al., 2013). Finally, another key study found that on average BD patients reported less positive affect and more negative affect, when compared to controls (Havermans et al, 2010) and spent more time at home alone, and/or engaged in passive activities (Havermans et al, 2007)

The literature was reviewed when considering how to the design of ESM for the current research. A key recent paper indicated some guidelines for the application of ESM for mental health (Myin-Germeys et al., 2018). Firstly, it has been suggested that completing a questionnaire shouldn't take longer than 2 minutes and on average should contain 30-

60 items as including more items is likely it minimises reactivity to this method. For our study, the visual mood diary, personalised medication tracking, and Bipolar UK mood scale are considered 1-item each and the subsequent items (QIDS-SR, ASRM and SHAPS) have 16, 5 and 14 items respectively, totalling at 38 items which is within this guideline.

When looking towards the number of notifications per day, a systematic review for ESM in BD indicates this can range from 2-10 observations per day (aan het Rot et al., 2012). In terms of compliance and response predictors, another recent study conducted with those with psychosis found acceptable compliance with an average response rate of 78%. Compliance also declined across days, reaching a low on the 5th day with 73% (95% CI Compliance also varied significantly across assessments depending on the time within a day ($p < .001$), with highest compliance between 12 p.m. and 1.30 p.m. and lowest compliance between 7.30 a.m. and 9 a.m. Persons with psychosis were less compliant than healthy users (70% vs. 83%. Also, females ($p = .002$) and older users ($p < .001$) were slightly more compliant. In this study, it was suggested that methodological studies investigating compliance in ESM protocols using more than six days are needed to clarify the reducing compliance. Our research will contribute partly towards the current gap as the observation period is six months. In terms of observations per day, in line with the ethos of this study, we asked users when they would like to receive the questionnaires in terms of days and time in order to maintain compliance due to the longitudinal nature of the study. Finally, and most importantly, we want to maintain the key notion of the user being at the heart of the research and do so by providing choice and control of when they would provide data.

6.2.5 Interview schedule

User interviews were conducted at the 1 month, 3 month and 6 month and post-study points in the study. As this research was being conducted at the height of COVID-19, all interviews were conducted remotely using Microsoft Teams. The interviews were aimed towards understanding the experience of using the self-tracking tool, paying particular focus towards usability and user preferences. The interviews conducted at the 3-month and post-study points explored the above but had an additional focus on data visualisation, with the aim of drawing connections between the aRMT and pRMT collected data and specified early warning signs of mania and depression. Interview outlines are described below.

6.2.5.1 1- month and 3-month point interview

These two interviews employed the same structure to explore experience of using the app in real-world context. Specifically, the following questions were asked during the interviews.

1. How are you finding taking part in the study?
2. Do you have any thoughts or feeling about using the app?
3. What do you like most about using the app?
4. What do you like the least about using the app?
5. How can we improve your experience of using the app

In addition, at the 3-moint point interview a snapshot of the processed aRMT and pRMT data was also provided as a probe to explore data utility and visualisation. The following additional questions were asked:

Data based questions:

1. What stands out to you from this data and why?
2. For those parts of the data that stand out, what do you think the data means?
3. Looking at your early warning sign checklist, how do you think your data connects to what you indicated here?
4. How can be improve data the way we present the data to you?
5. How often would you like to receive this type of data?

6.2.5.2 6-month point interview

This interview was focused on drawing on the collective data collection experience, with a focus on usability, preference and utility of the self-tracking tool in a real-world context. Specifically, the following questions were asked during the interviews:

6.2.5.3 Experience based questions [A]

1. How did you find using the app[+] over the period of 6 months?

2. What were the best parts of using the app?
3. What were the worst parts of using the app?
4. What problems does this app solve for you? What doesn't it solve?
5. Do you think if this app was available commercially that you would use this?

A. *These questions were repeated for both aRMT and pRMT apps of RADAR-Base*

6.2.5.4 Post study data exploration interview

Once data from aRMT and pRMT was processed, a final interview was held with users to explore the data in the form of a user dashboard and its connections to EWS of mania and depression. A description of the user dashboard can be found in the next section. Usability was also explored in a quantitative format, by the administration of the System Usability Scale (SUS). The SUS scale is a 10-item questionnaire with statements of system usability with five response options: from strongly agree to strongly disagree. It is a highly reliable tool which is commonly used in measuring usability for digital products and services (Affairs, 2013). Using the tool can provide confidence in the execution of the tool, in relation to the impact of usability, and thus make the findings of more interest which is the underpinning of its use in the post study data exploration interview. Specifically, the following additional questions were asked during the interviews to explore the data and EWS:

1. What stands out to you from this data and why?
2. For those parts of the data that stand out, what do you think the data means?
3. Looking at your early warning sign checklist, how do you think your data connects to what you indicated here as an early warning sign?
4. What parts of the data are considered normal and what parts are considered out of the ordinary for you?
5. How can be improve data the way we present the data to you?
6. How often would you like to receive this type of data?

6.2.6 Data analysis

6.2.6.1 Quantitative analysis

User data was initially processed in MS Excel to carry out descriptive analysis in order to summarise the data over the 6-month course of data collection and present this to the user in the form of a user dashboard.

For the different tracking factors of pRMT, the following method of descriptive analysis was applied:

- **Which apps being used and for how long**: total number of interactions of all apps per week
- **Phone usage:** total time of active phone usage per day
- **Battery level:** mean level of battery level per day
- **Other Bluetooth devices in the vicinity:** total number of connected Bluetooth devices per day
- **Number of contacts on phone**: total number of phone contacts per week
- **Phone light:** mean level of phone light per day
- **Step count**: total number of steps per day
- **GPS location:** total amount of movements per day
- **Sleep information and heart rate (via Fitbit***)*: no analysis was conducted as no users had a Fitbit connection
- **Ambient light**: average light levels per day
- **Levels of spending**: average amount of spending per day

Given the granularity of app usage data, where the use of individual apps was shown daily, this data was not incorporated in the user dashboard on Tableau but sent as a separate interactive MS Excel document to the users via email. An example of this is shown in Appendix 7

For the different aRMT tracking factors, the following descriptive analysis was conducted:

- **Daily visual mood diary:** the number of times an item on the visual diary (i.e., emoji) was chosen

- **Bipolar UK mood scale**: the numeric values were converted to mood categories as per the BPUK scale (0-1 = severe depression, 2-3 mild to moderate depression, 4-6 = balanced mood, 7-8 = hypomania, 9-10= mania)
- **Quick Inventory of Depressive Symptomology (QIDS-SR):** The total QIDS-SR weekly score was totalled and scored according to instruction provided with the instrument (Rush et al., 2003). Total scores range from 0-27 where the lower scores indicate little, or no depression and higher scores indicate very severe depression.
- **Altman Self-Rating Mania scale (ASRM):** The total ASRM score was totalled and scored according to instructions provided with the instrument (E. G. Altman et al., 1997). Total ASRM scores range from 0–20, where scores of 6 or greater indicate significant manic or hypomanic symptoms
- **Snaith-Hamilton Pleasure scale (SHAPS):** The total SHAPS score was computed by summing scores across four response categories where higher scores indicate less hedonic tone, and hence more anhedonia symptoms according to Snaith et al., 1995.
- **Daily personalised medication tracking**: Total yes/no responses per day were calculated

The findings from the aRMT and pRMT descriptive analysis were transferred to Tableau Software, which is a data visualisation tool. From this, individual's user dashboards were created which indicated all user data in one interactive visualisation which they could access from their own computers. Instructions were provided on how to access their dashboard via email will follow up calls if necessary.

Statistical analysis

In order to analyse factors that trigger a change in behaviour of the users, we employed data analytic and statistic tools offered by Python and R programming libraries to understand the connections, if any, between the active and passive data. Using statistical methods to explore the connections between the active and passive is typical for analysis for this technology as outlined by Ortiz et al (2021).

6.2.6.2 Qualitative analysis

In this study, the qualitative analysis employed Braun and Clarke's thematic analysis approach (Braun & Clarke, 2006). A reflexive thematic analysis approach was chosen, allowing the researcher to ascertain findings' outcomes and focuses, unbound by a specific theoretical framework such as the classified Big Q thematic analysis. The analysis adopted an inductive theoretical approach, deriving themes directly from the data itself. These themes were identified at a semantic level to summarize content and capture surface meaning, which closely aligns with the users' accounts—central to this thesis' purpose. The thematic analysis followed a six-step process: becoming familiar with the data, generating initial codes, identifying themes, reviewing them, defining and naming them, and producing the final report.

In some cases, affinity mapping was used as a tool to visualise the thematic analysis results for those interview questions that had more brief responses. Affinity mapping is prevalent tool in user experience research (Kara, 2019) used to discern patterns in notes and observations, thereby creating cohesive patterns or themes.

NVivo 12 will used for thematic data analysis and Miro will be used for the affinity mapping exercises.

6.2.7 Ethical approval

This research received ethical approval from the Department of Computer Science at the University of Nottingham reference: CS- 2019-R58. This research also received full ethical approval from Nottingham Healthcare NHS Foundation Trust IRAS 295693.

6.2.8 User sample

Out of the n=14 prospective users that expressed interest, n=8 users were enrolled onto the study. The sample was an equal split between gender with 50% identifying as male and 50% identifying as female. In terms of age, a range was seen where 37.5% were aged between 25-34, 37.5% were aged between 35-44, 12.5% were aged between 45-54 and finally were 12.5% were aged between 55-64. In terms of lived experience, a large part of the sample (87.5%) had more than 7 years lived experience, whereas 12.5% had more that 5 years but less than 7 years lived experience. Majority of the sample had a diagnosis of BD2 (87. %) and a smaller proportion (12.5%) (n=1) had a diagnosis of BD1. 37.% of the sample had already been involved in other parts of the study such as the interviews and workshops in previous chapters. Out of the n=8 users who were enrolled onto the study, n=6 completed the study. Completion of the study was defined as partaking in the full 6 months of data collection. The full details of the sample are outlined in the Table. 15 below.

User number	Gender	Age	Lived Experience	Diagnosis Category	Already involved in the study	Study completion
P001	Female	35-44	More than 7 years	BD2	No	Yes
P002	Male	25-34	More than 7 years	BD2	No	Yes
P003	Female	25-34	More than 5 years but less than 7 years	BD2	Yes	Yes
P004	Female	35-44	More than 7 years	BD2	No	Yes

P005	Female	35-44	More than 7 years	BD2	No	No
P006	Male	55-64	More than 7 years	BD2	Yes	Yes
P007	Male	25-34	More than 7 years	BD2	Yes	No
P008	Male	45-54	More than 7 years	BD1	No	Yes

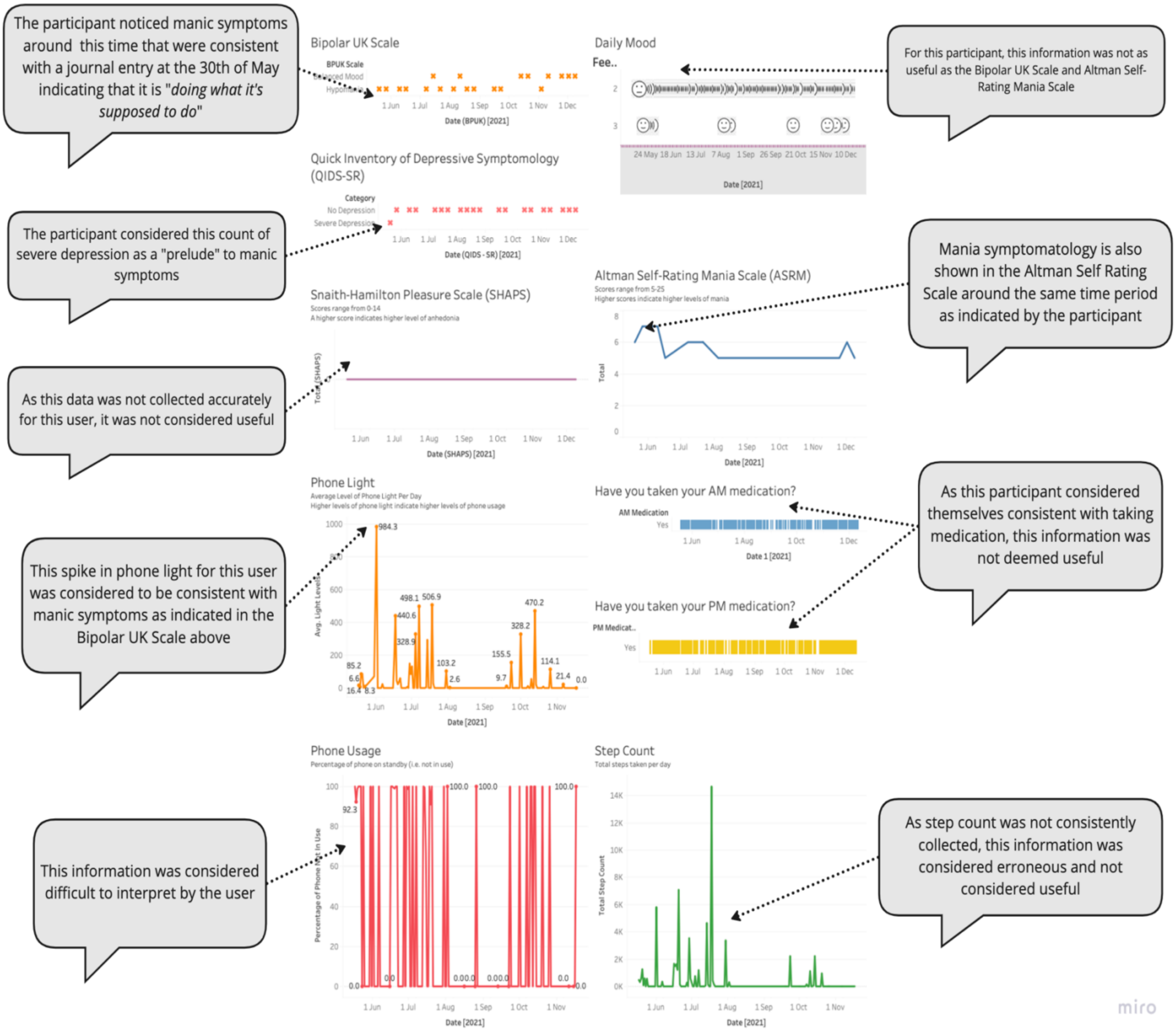
Table 15 Characteristics of sample

6.3 Results

6.3.1 Quantitative results

6.3.1.1. Descriptive analysis

Descriptive analysis was conducted in MS Excel and used to create individual dashboards in Tableau Software, which is a data visualisation tool. These dashboards were then shared with users to explore the data in connection with EWS. The n=6 user dashboards can be found in Appendix 8.1, 8.2, 8.3, 8.4, 8.5, 8.6. In the post-study data exploration interview, the user dashboard was presented to users and EWS were explored with n=4 users. N=2 users did not partake in these as the researcher lost contact with them. A full example of how this dashboard was used in the interview is exemplified in Figure 16 with findings from P001. Annotations are made on the dashboard to exemplify the type of feedback that was received on the different elements of the dashboard and what, for this user, was considered useful and not useful and most importantly how this connected to EWS of mania and depression.



The participant noticed manic symptoms around this time that were consistent with a journal entry at the 30th of May indicating that it is "*doing what it's supposed to do*"

The participant considered this count of severe depression as a "prelude" to manic symptoms

As this data was not collected accurately for this user, it was not considered useful

This spike in phone light for this user was considered to be consistent with manic symptoms as indicated in the Bipolar UK Scale above

This information was considered difficult to interpret by the user

For this participant, this information was not as useful as the Bipolar UK Scale and Altman Self-Rating Mania Scale

Mania symptomatology is also shown in the Altman Self Rating Scale around the same time period as indicated by the participant

As this participant considered themselves consistent with taking medication, this information was not deemed useful

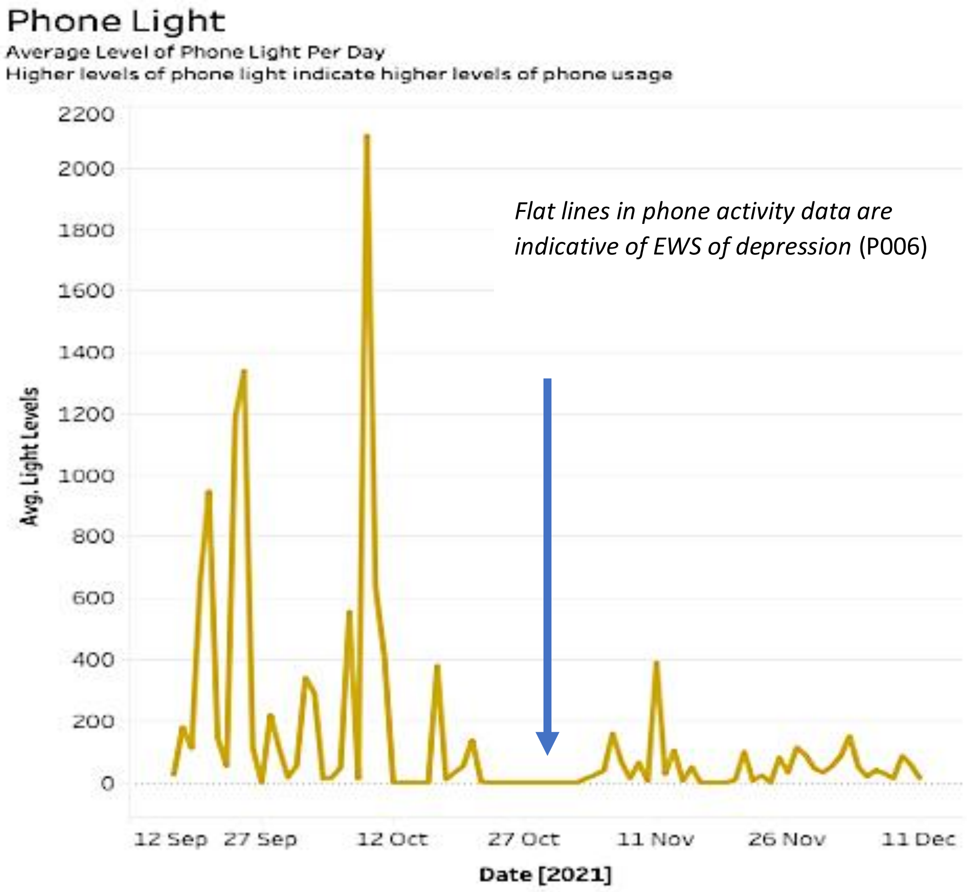
As step count was not consistently collected, this information was considered erroneous and not considered useful

Figure 16 Example of dashboard used in the post-study exploration interview for P001

6.3.1.2 EWS of depression

Phone usage data

Lack of phone usage was described as an EWS of depression, as P006 described that *"using phone less equals depressive symptoms… So, some indication of being low and depressed is generally less activity in the data" (P006).* Specifically, the following flat line in the data was referred to as a marker for depressed mood for this user is demonstrated in Figure 17. The user further described that there are "*occasional little flats, but that's that a particularly bulk of inactivity in my part" (P006).* This was indicated by other users who indicated that "*using phone less indicated depressive symptoms*" (P001).

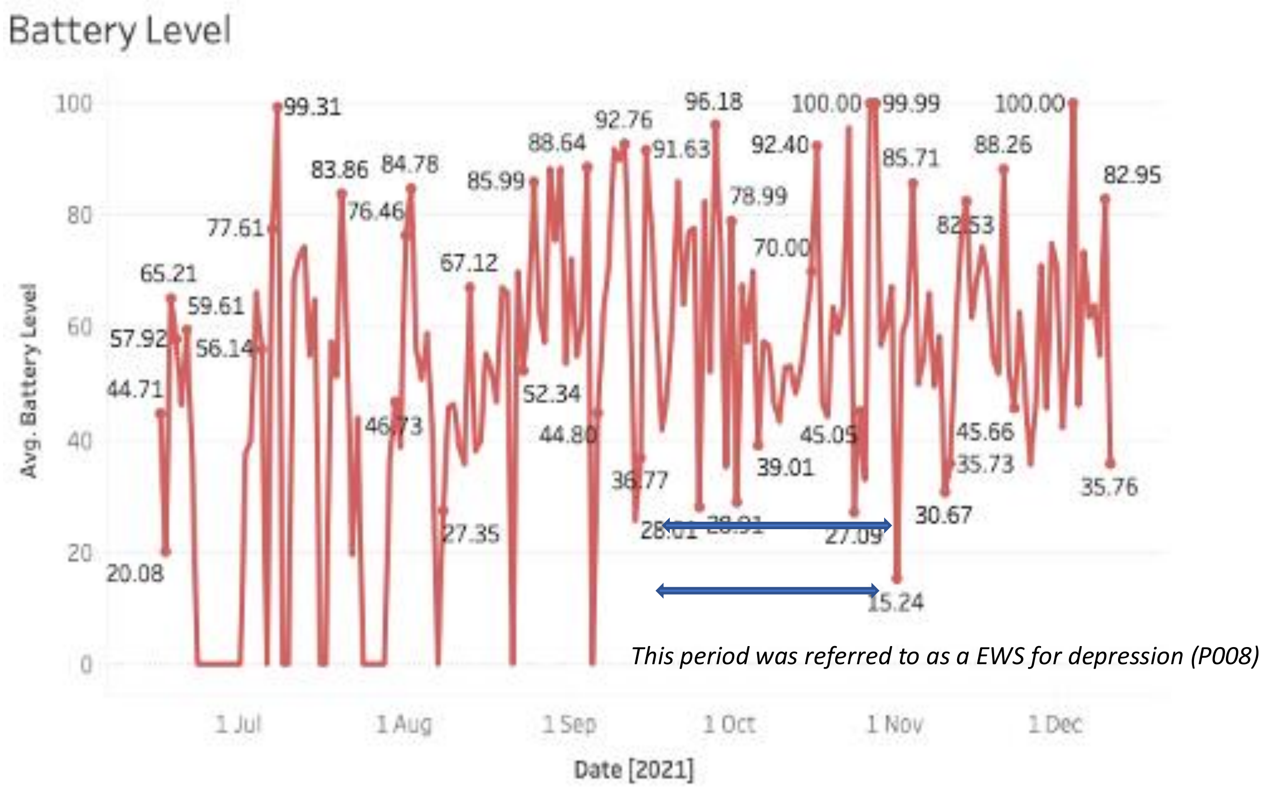


Flat lines in phone activity data are indicative of EWS of depression (P006)

Figure 17 Phone light levels for P006

Battery level data

Battery level data was also indicated as being useful in providing insight into EWS for depression, as the following user described *"the battery level one, its useful as it shows that I haven't been using my phone that much" (P008).* It was indicated that periods of low data for battery level were indicative of EWS, where the following dip indicated in Figure 18, is where the user described "*I would imagine I was depressed or something like that, as I'm not charging my phone, you know*?" (P008)



This period was referred to as a EWS for depression (P008)

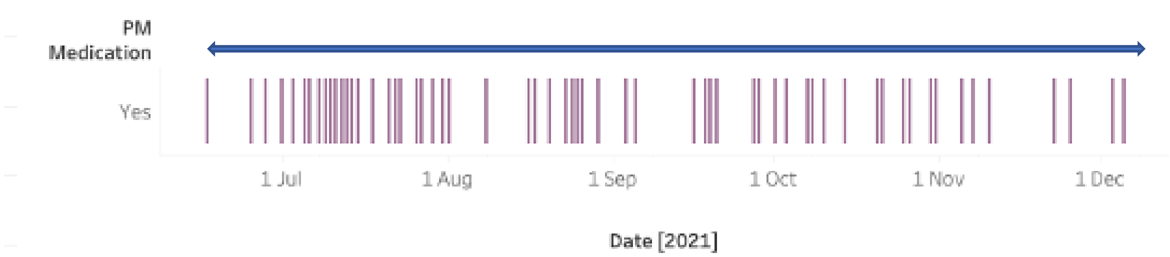
Figure 18 Battery levels for P008

App usage data

App usage data was also referred to as having utility for EWS of depression, where one user described that during periods of depression *"I'll kind of just leave the apps, you know I can just be in bed"* (P008).

Active self-report data

The daily personalised medication question was particularly described as being useful for some users where it was described that "I *was filling it out every night, but then as things* *progressed, I think that was a period where I was either a bit depressed or a bit manic and not noticing I should fill it in".* This in indicated in the data in Figure 19.



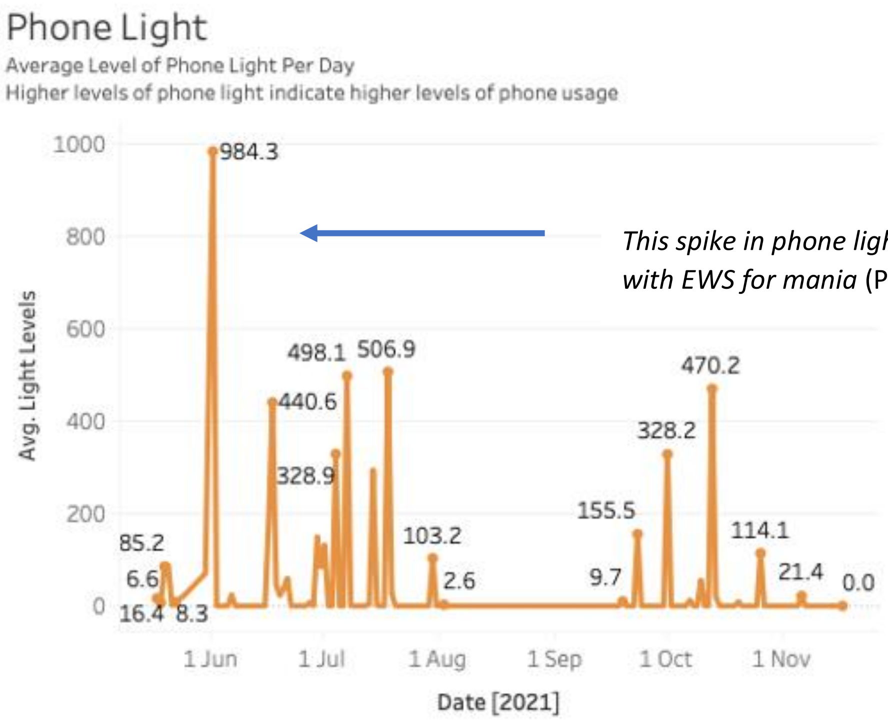
These blanks in the active data were indicative of periods of mania and depression (P008)

Figure 19 Daily medication tracking for P008

6.3.1.3 EWS for Mania

Phone light levels

For the following user, increased phone light was indicative of signs of hypomania. In particular, the following EWS for mania were connected to increased phone light (P001): Can't get off to sleep, racing thoughts, involved in many projects (Figure 20). For this user, not only was this consistent with the active (self-report) data which indicated hypomania (Figure 21) around the same time period, but also corresponded to a journal entry from the 30th of May indicating hypomania showing that self-tracking tool "*is doing what it's supposed to do*" (P001).



This spike in phone light levels is consistent with EWS for mania (P001)

Figure 20 Phone light levels for P001

On this occasion, the pRMT and aRMT collected data consistently described hypomanic mood (P001)

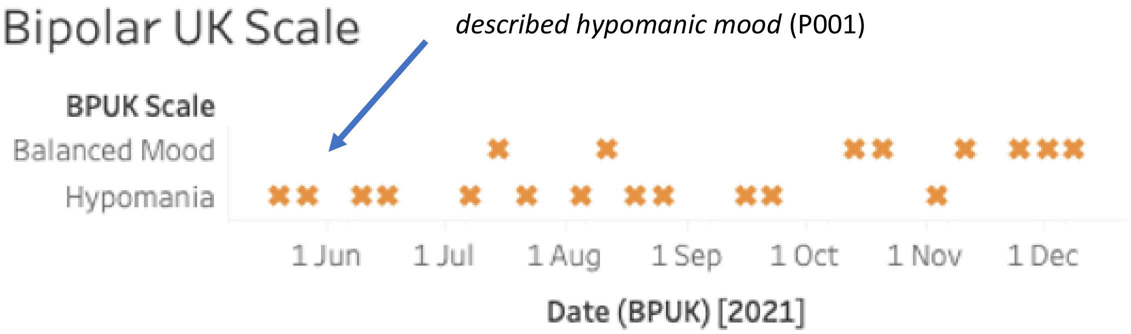


Figure 21 Bipolar UK Scale for P001

App usage

For some users, increased app usage was indicative of manic mood. For example, it was described that "*I'm flicking through all these apps, it's where my mood is elevated*" (P008). Generally, it was described that "*using the phone more was indicative of mania*" (P001).

6.3.1.4 Statistical analysis

The subsequent section analysed EWS and connection to the descriptive data from the user's subjective perspective. To explore this objectively, we employed data analytic and statistic tools offered by Python and R programming libraries to understand the connections, if any, between the active and passive data. Due to the small sample size and issues with data collection, only a summary of the analysis will be provided below.

Pre-processing:

The individual observations gathered for each user were combined into a single file. Columns that played no significant role in the analysis, like the user ID and source ID, were dropped or removed from the file. To simplify the analysis, we converted the epoch time to the standard date and time format.

Main findings

There are some parameters that correlate with one another while integrating the results of different parameters for each user individually.

- P003 and P004 use their phones less (indicated by a decrease in app usage) as their mood changes to a higher level. Simultaneously, their phone light intensity also decreases with a higher mood level. These two results can be correlated as the phone light intensity is dependent on how long the phone is being used (while accessing applications)
- The battery life for P002 is on the lower side with an increase in mood level. Hence, we can understand the reason for a decrease in phone light intensity with a higher mood level.
- P003 charges their phone less often with a higher mood level. This corresponds to the lower battery life as well as a decrease in app usage and phone light.

Other parameters

Due to data inconsistencies, indicators such as step count, Bluetooth devices, and relative location could not yield valuable insights. Data from all individuals were not included since their active and passive datasets did not overlap at different observations. The analysis of these datasets is based on daily observations rather than weekly observations such as the following aRMT questionnaires (Bipolar UK Scale, QIDS-SR, SHAPS and ASRM) that were included in the study design. This was driven, once again, by the fact that relatively few data points coincided in reference to a weekly timeframe as many users did not fill out consistently, or the aRMT app did not deliver this consistently due to technical issues.

6.3.1.4 Qualitative results

1 month point interviews

All users (n=6) took part in month 1 interviews. Three overall themes were found: positive experiences of using the app, negative experiences of using the app and issues with notifications which are further defined and exemplified in the Table 16 below. Overall, users described that the app was user friendly, and that there was utility in the passive data collection. Users also described some negative experiences of using the application, such as a consciousness of being tracked. Some users also describe that the daily mood scale was too simplistic, which may not have sufficiently captured the complexity of mood. It was also described that some users would forget to fill in the daily medication questionnaire delivered by the aRMT app. The final theme that was found was system issues which impacted the user experience. Prior to the first round of user interviews, the app had some technical issues impacting aRMT notifications.

Theme	Sub-theme	Quote
Positive experiences *This describes the positive aspects of using the aRMT and pRMT apps of RADAR-Base adapted for BD*	The app is user friendly	*"You know it's very quick, very easy, no long-winded bits to it"* (P006)

	Passive collection is useful	*"So yeah, it is easy to use… And I guess because it's more of a passive thing that it asked me to do something rather than me actively choosing to do something"* (P002)
	Serves as a reminder for key self-management acts (such as taking my medication)	*"You know it's been okay because I think it helps to remind me to check to see that I have taken my medication even though yes I do" (P001)*
Negative experiences of using the app *This describes the negative aspects of using the aRMT and pRMT apps of RADAR-Base adapted for BD*	Being conscious of being tracked	*"What can others determine based upon my activity? … It's a weird time as well. With the pandemic and stuff. And with this whole issue that we've got going on with trustworthiness around governments and what's going on in terms of, you know, this kind of tin hat, Big Brother type stuff. So that's what I'm thinking to myself"* (P004)
	Daily mood scale is too simplistic	"*I think in terms of the mood tracking, when it's um, it does scale of like 5 different smiley faces. Yeah, things that might not be sufficient for me*" (P002)
	Forgetting to fill out medication questions	*"It's just occasionally I forget. So, I might most days I do both the 12:00 o'clock and the 8:00 o'clock. But sometimes I might miss one, especially if I'm doing something."* (P006)
System issues *This describes the nature of glitches, bugs and issues that were experienced by users at the 1-month point*	Notification issues	*"No problem at all, apart from I think we've got the schedule incorrect with the prompts. I think they need to be changed just in order for me to engage better"* (P004)

	Glitchy and buggy	*"Fix the glitches, make sure I can use my phone without interruption."* (P001)

Table 16 Themes for 1-month point interviews

3-month point interviews

A total of n=5 interviews were conducted. Compared to 1-month point interviews, users were also provided with a snapshot of their data during this interview. This was in the form of a MATLAB Figure for each pRMT data streams that the user consented to being sensed, overlayed with their active data, which could be filtered by dates (Figure 24). An example of this for P008 is provided for their Bluetooth devices data, as well as aRMT data at the top. This snapshot was used as probe to explore data utility. A full interactive version in the form of a dashboard was providing at study close.

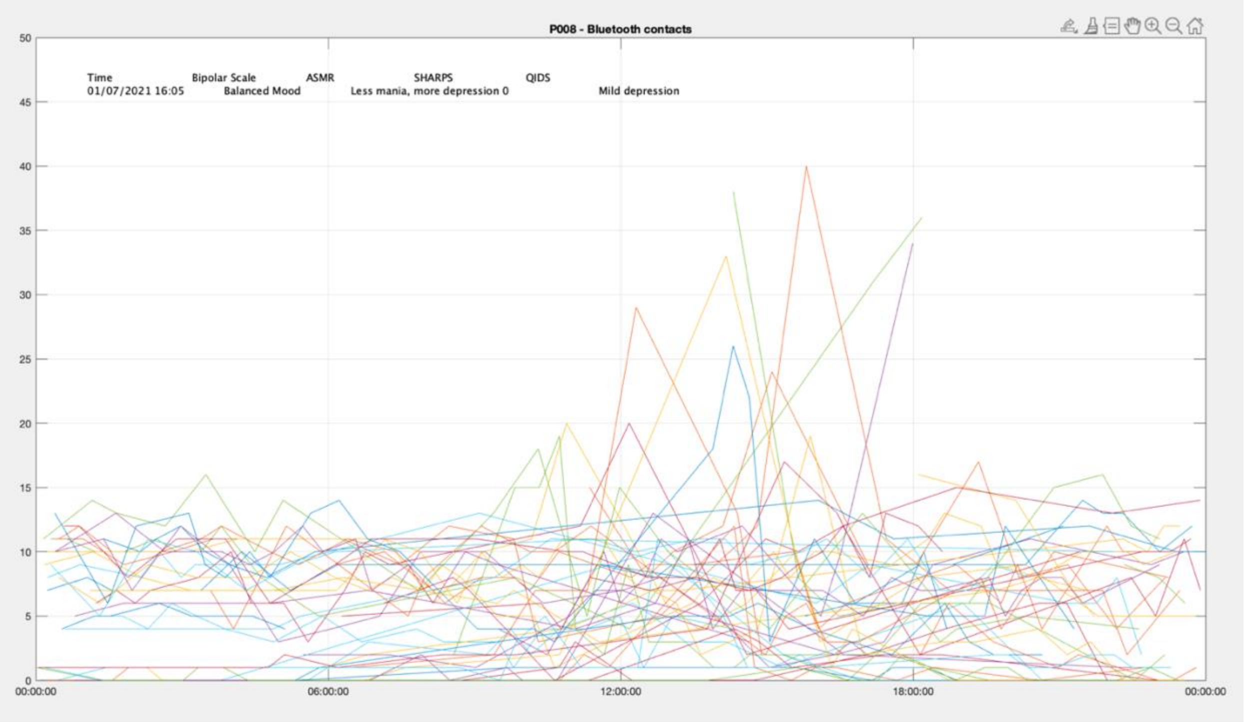


Figure 22 Data "snapshot" for Bluetooth devices in the vicinity (collected via pRMT) and aRMT data

Themes, subthemes and quotes from the user interviews are outlined in Table 17. Similarly, to month 1, positive experiences and negative experiences of using the app emerged as a theme. For positive experiences, it emerged that the app was user-friendly, integrated into everyday life and that the questions are useful. In relation to negative experiences of using the app, users described that they would forget to fill in the questions delivered by the aRMT app and that too much data can be a negative experience. System issues were also captured in the interviews, where some users described the app as being glitchy and with bugs. Finally, the utility of the snapshot of data was explored where users described that the data gave insights into BD triggers and that the data is consistent with actual behaviour. However, there were also concerns that too much data is not beneficial to the management of BD.

Theme	Sub-theme	Quote
Positive experiences *This describes the positive aspects of using the aRMT and pRMT apps of RADAR-Base adapted for BD*	The app is user friendly	*"I was going to say it's user friendly. When it works it works. It's simple and straight forward." (P002)*
	The app integrates into everyday life	*"It's just become routine to do it so… It's almost subconscious" (P006)*
	The questions are useful	*"I think the questions are really well, they're very pertinent." (P003)*
Negative experiences of using the app *This describes the negative aspects of using the aRMT and pRMT apps of RADAR-Base adapted for BD*	Would forget to tell in the questions	*"I just spent three weeks in Portugal and then a week in London and stuff. And when I'm on holidays I kind of forget about my phone. So sometimes I think I'll shoot. I'm not doing this stuff for the study" (P003)*
	Notification issues	*"The window of time to complete the questions needs to be longer" (P008)*
Utility of data *This describes the perceived usefulness of the data snapshot*	Data gives insight into triggers	*"It's there like you can find out certain triggers that you didn't know about, you can actually either avoid or better manage those triggers" (P003)*
	Data is consistent with behaviour	*"It's this consistent with how I've been using it, so most of the time in the daylight hours I would be using it and it tapers off as night falls, and then I turn my phone off at night, then turn it back on when I'm moving around to do whatever in the morning" (P001)*

	Too much data is not good	*“The graph is unnecessary for me as a user. As someone who is trying to monitor BD, maybe too much data would be coming away from that.” (P003)*

Table 17 3-month interview themes

6-month point interviews

An affinity mapping exercise was conducted on the interview for the aRMT and pRMT apps which form RADAR-Base for BD. The full findings of this can be found in Figure 25

Active data collection app

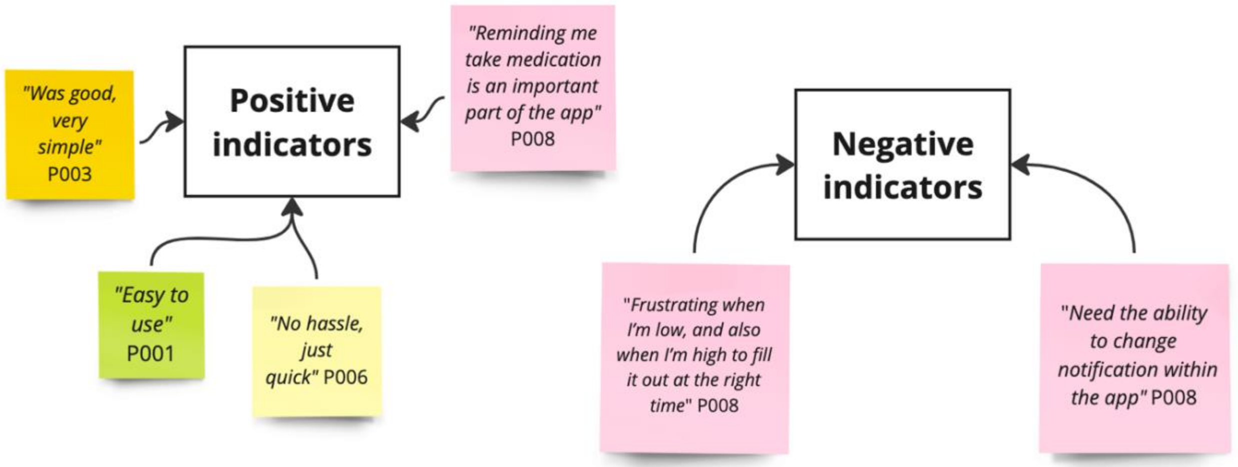
When considering the collective experience of use over the period of 6 months, there were several positive and negative indicators of the aRMT app. For positive indicators, users described an intuitive user experience, with ease of use and simplicity. For negative indicators, difficulties arose during mood shifts, and the lack of ability to change notifications within the app was seen as a frustration. The best parts of the app were the aRMT content (such as questionnaires etc) as well as the ability to tailor the app, such as getting prompts at specified times. However, the notifications were also deemed the worst parts of the app due to the technical issues that were experienced during the study. Most importantly, when exploring the utility of this app, it was described that it facilitated self-awareness, especially for those who are still learning about their EWS. However, the app didn't negate the reality of BD as when mood shifts arise, "*they just come*" (P008).

Passive data collection app

The collective experience of using the pRMT over the period of 6 month, the app was largely not noticed by majority of users as it was automatic and collecting data in the background. This was the very thing appreciated by users when asked what the best parts of the app were, as it had a lack of interference in their daily lives. Specific PRMT tracking factors and output were appreciated by users including step count and data feedback, where the latter provided useful insights. Similar to aRMT, technical issues were considered the worst parts of the app such as the data not being collected properly, as well as notification issues. The pRMT solved a core problem for users through automatic data collection, as mood shifts can impact self-report data input as a user described that "*when I'm depressed, I don't remember*" (P008).

aRMT

How did you find using the app over the period of 6 months?



Positive indicators

"Was good, very simple" P003

"Easy to use" P001

"No hassle, just quick" P006

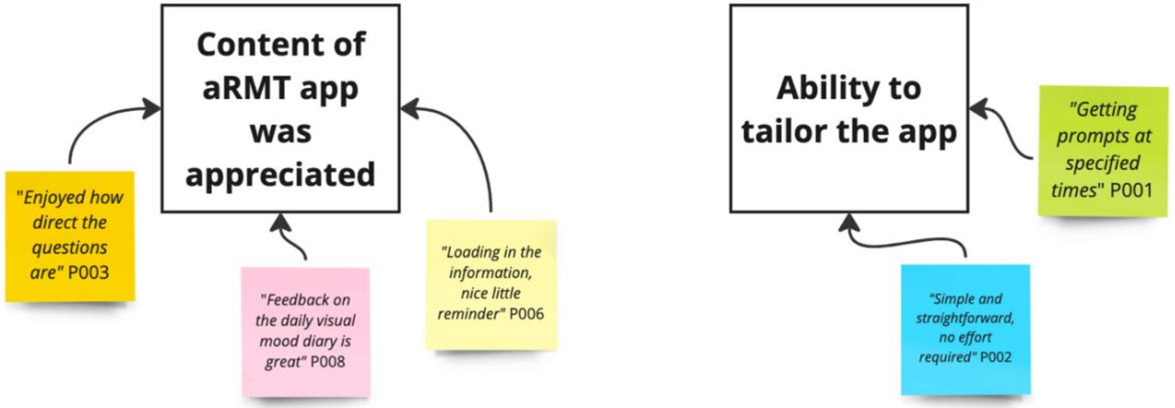
"Reminding me take medication is an important part of the app" P008

Negative indicators

"Frustrating when I'm low, and also when I'm high to fill it out at the right time" P008

"Need the ability to change notification within the app" P008

What were the best parts of using the app?



Content of aRMT app was appreciated

"Enjoyed how direct the questions are" P003

"Feedback on the daily visual mood diary is great" P008

"Loading in the information, nice little reminder" P006

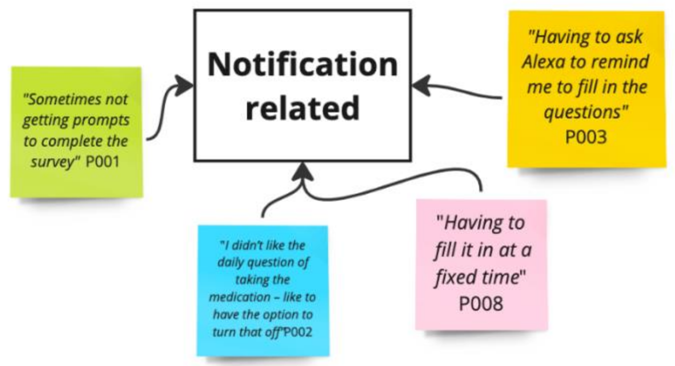
Ability to tailor the app

"Simple and straightforward, no effort required" P002

"Getting prompts at specified times" P001

miro

What were the worst parts of using the app?



Notification related

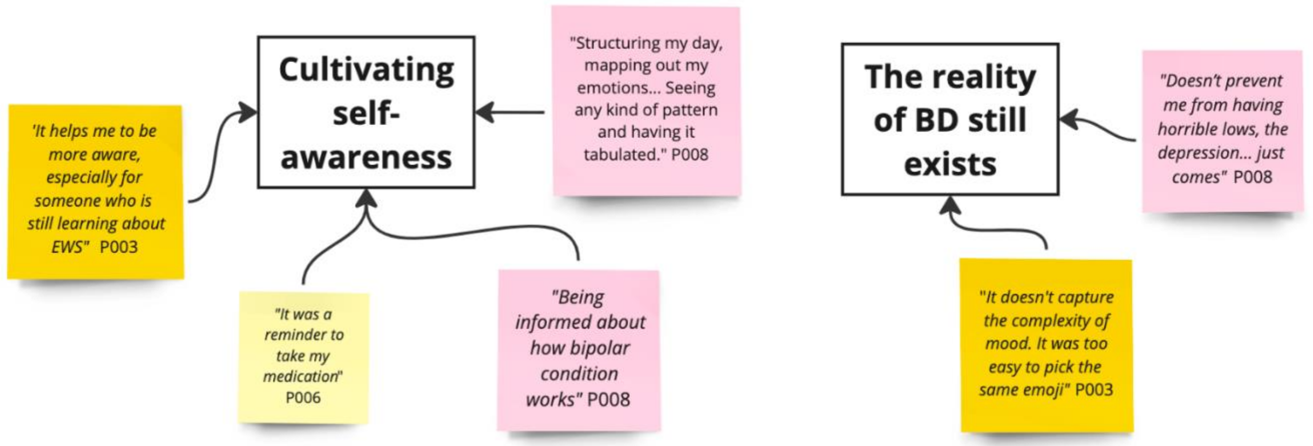
"Sometimes not getting prompts to complete the survey" P001

"Having to ask Alexa to remind me to fill in the questions" P003

"I didn't like the daily question of taking the medication – like to have the option to turn that off" P002

"Having to fill it in at a fixed time" P008

What problems does this app solve for you? What doesn't it solve?



Cultivating self-awareness

'It helps me to be more aware, especially for someone who is still learning about EWS" P003

"Structuring my day, mapping out my emotions... Seeing any kind of pattern and having it tabulated." P008

"It was a reminder to take my medication" P006

"Being informed about how bipolar condition works" P008

The reality of BD still exists

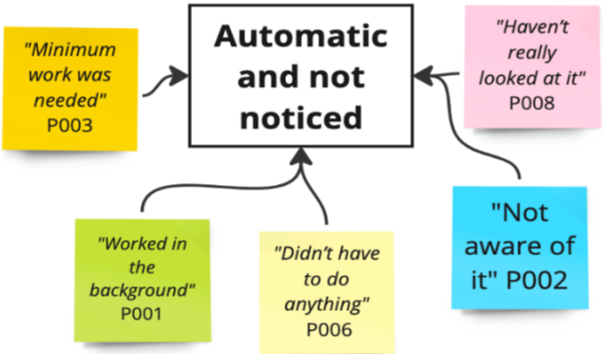
"Doesn't prevent me from having horrible lows, the depression... just comes" P008

"It doesn't capture the complexity of mood. It was too easy to pick the same emoji" P003

miro

pRMT

How did you find using the app over the period of 6 months?



Automatic and not noticed

"Minimum work was needed" P003

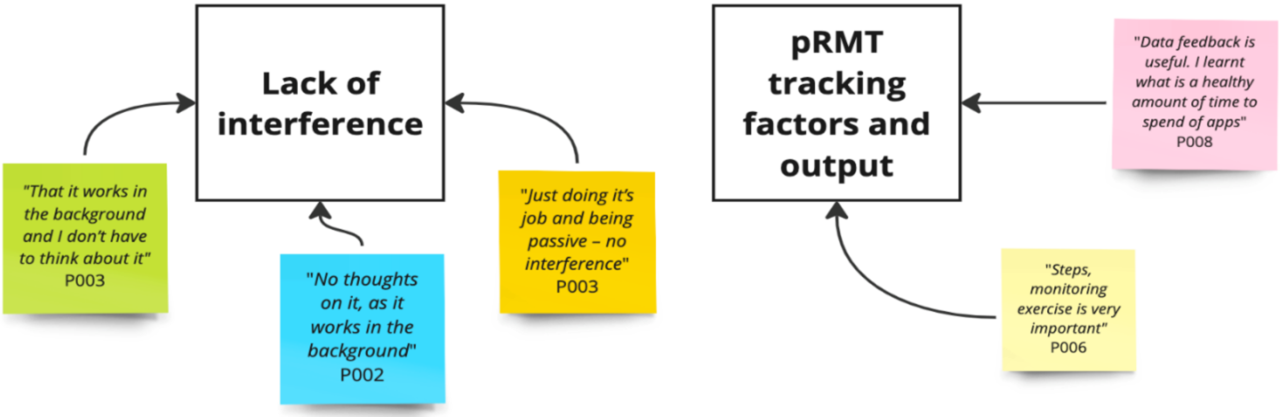
"Worked in the background" P001

"Didn't have to do anything" P006

"Haven't really looked at it" P008

"Not aware of it" P002

What were the best parts of using the app?



Lack of interference

"That it works in the background and I don't have to think about it" P003

"No thoughts on it, as it works in the background" P002

"Just doing it's job and being passive – no interference" P003

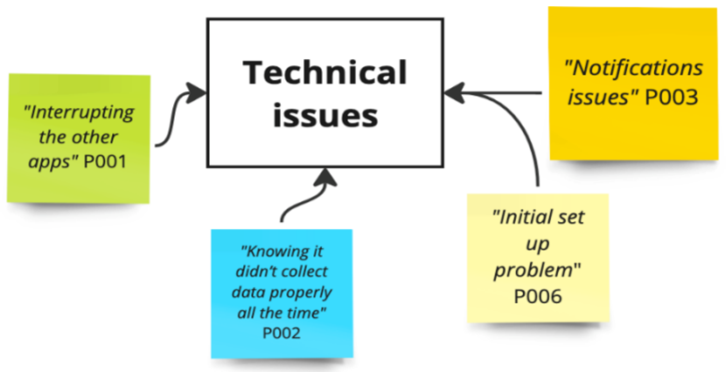
pRMT tracking factors and output

"Data feedback is useful. I learnt what is a healthy amount of time to spend of apps" P008

"Steps, monitoring exercise is very important" P006

miro

What were the worst parts of using the app?



Technical issues

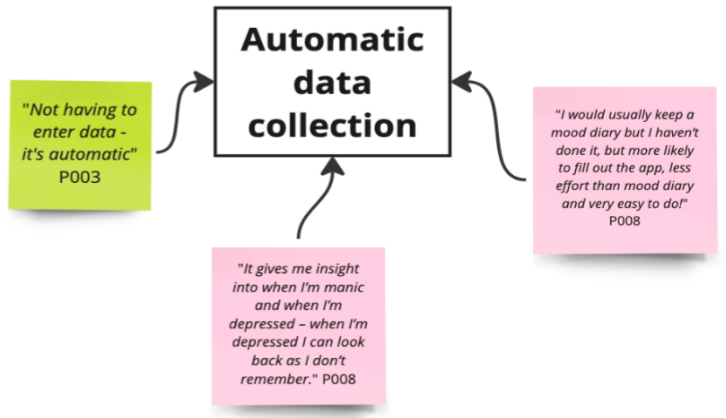
"Interrupting the other apps" P001

"Knowing it didn't collect data properly all the time" P002

"Initial set up problem" P006

"Notifications issues" P003

What problems does this app solve for you? What doesn't it solve?



Automatic data collection

"Not having to enter data - it's automatic" P003

"It gives me insight into when I'm manic and when I'm depressed – when I'm depressed I can look back as I don't remember." P008

"I would usually keep a mood diary but I haven't done it, but more likely to fill out the app, less effort than mood diary and very easy to do!" P008

miro

Figure 23 Full affinity map for aRMT and pRMT app

6.3.1.5 *Suggestions for improvement*

Throughout the course of the interviews, suggestions of improvement were made by users. A full list of user-led recommendations is provided below.

- Reward-based messaging for engagement with the application such as:
    - Confetti splatter for inputting data
    - Smiley face for engagement
- To have a mood diary function for reflection
- Personalized coaching based on insights from the app
- Adding a community element to the app
- Notification settings within the app (such as the ability to turn aRMT notifications off)
- The ability to choose aRMT self-report measures

Table 18 User-led recommendations for improvement of RADAR-Base for BD

6.3.1.6 User experience metrics

Total SUS scores, as well as average SUS scores (83.3) are plotted in Figure 26. According to the accepted interpretation of SUS score, any score above 80.3 is considered to reflect excellent usability, with the standard average being around 68. Further details are found in the Figure 26.

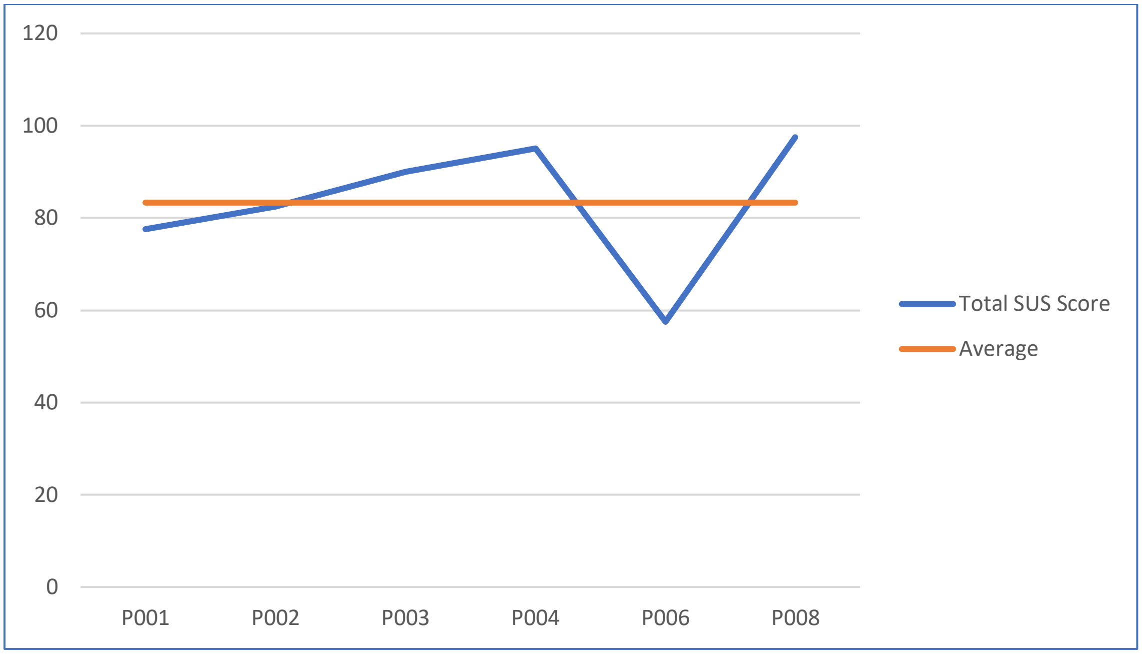


Figure 24 Line chart for SUS Scores

6.4 Discussion

This study investigated the utility, usability, feasibility and acceptability of a self-tracking tool which uses passive and active sensing methods to understand EWS for those with BD. The tool, RADAR-Base, was tailored towards BD through a series of user-centred design research which combined principles of user-involvement in HCI and PPI to understand user needs, preferences and individual variability and to ensure the tool aligned as closely as possible to these. Specifically, we explore the following research questions:

RQ8: How much of the data collected by the sensing system (both passive and active data) relates to EWS?

RQ9: Is this type of technology usable and acceptable for people living with this condition?

RQ10: What are the researchers experience of applying the overall methodology employed in research thus far?

The findings from this study will be structured around these research questions in this discussion to provide an overview of RADAR-Base, it's connection to EWS as well as report on user-experience.

6.4.1 pRMT, aRMT, connections and self-management

The answer to these particular questions was explored from two perspective, first the user's subjective perspective when viewing their data and secondly from an objective, statistical perspective. Given that the user's perspective is considered the most important throughout the dialogue of this research, we will consider this first and then make comparisons to the objective analysis that was carried out.

For EWS for depression, it was described that a reduction in phone usage data was indicative of these early warning signs where one user described that "*using phone less equals depressive symptoms*" (P006). This direction of low frequency of data streams also emerged in the battery level data. Where a lack of battery charge indicated a lack of phone usage which was indicative of EWS for depression. This low frequency was also reflected in app usage data, again where low frequency (i.e., usage) was connected to EWS for depression, as when feeling low it was described as "*I'll just kind of leave the apps…just be in bed*" (P008). There we no particular mentions of specific EWS when exploring depression, where most users described that these patterns in the data can be applied to their early warnings for depression generally, rather than single EWS. Though this could be the case, there are several limitations around the comprehension of the data and questions as to whether the dashboard was too complex for users, or whether the applied methodology was appropriate to explore the data which could have impact user answers to these questions. This will be explored in the later limitations section.

For EWS for mania, phone light was described as a useful tracking factor for the following EWS: can't get off to sleep, racing thoughts, involved in many projects (P001). Oppositely to depression, the increase of phone light was connected to these EWS. For one user, this increased in phone light (via pRMT) was also reflected in the aRMT data as the Bipolar UK scale was also showing hypomania during the same time period (P001). This user also verified this by her own journal entry at the same time point where she described she was displaying hypomanic symptoms, indicating it's doing "what it's supposed to do". This increase of usage was also reflected in app usage data by another user to described that during EWS for mania "*I'm flicking through all these apps" (P008).*

Though this exploration from the user's perspective had limitations, it provides useful insights into how users make sense of their data, especially around pattern recognition and its connection to EWS where it was universally confirmed by users that low frequency in the data is typically indicative of EWS for depression, whereas high frequency in the data is typically indicative of EWS for mania. This reflects findings in a similar study by (Matthews, Murnane, Snyder, et al., 2017a) which described this pattern as an indicator of mania and depression of technology-mediated behaviours. This exploration of the data from the user's perspective was not seen in similar studies for self-tracking technology for BD (Beiwinkel et al., 2016; Abdullah et al., 2016a; Palmius et al., 2017; C. H. Cho et al., 2019; C.-H. Cho et al., 2020), and is unique to this research. When systematically reviewing the current self-tracking technology for BD, it was found that individual variability was not considered as a potential predictor of mood shift for BD (Ortiz et al, 2021). There was a call for the tailored monitoring of symptoms of BD, where this study has made the first attempt in trying to understand this individual variability by capturing the subjective interpretation of the data from the user's perspective. The next section will consider this from an objective understanding, through the application of statistical analysis.

Our brief descriptive statistical analysis found that for app usage data, an increase of daily mood data (more manic symptoms) was associated with a decrease of app usage data, except for 1 user who showed an increased in daily mood data and more app usage. For battery level data, there was a cluster of users who showed increase in battery life when daily mood increased (n=2), and another cluster who showed decreased in battery life when daily mood increased (n=2). For phone light levels, similar findings were found where most users (80%) showed a decrease in phone light levels as daily mood increased, and n=1 users displayed higher phone light levels as daily mood increases.

When comparing the subjective interpretations of the data, compared to the objective statistical understandings, there are the following observations. Firstly, for some users it was indeed the case that when their mood increased (more manic symptoms) that the frequency of the data for specific tracking factors increased such as app usage, battery level and phone light. However, for some users, the opposite was true where an increase of their mood (more manic symptoms) was associated with a decrease of app usage, battery levels and phone light levels, and this was the cluster where most of our users fell into. This again drawing back to the individual variability experienced in BD, where the presence of a straightforward indicator of EWS for mania and depression does not exist and users fell into different distinct clusters of increase and decrease in self-tracking factor data frequency and its relationship with mood. This again is in conflict with findings from a similar study (Matthews, Murnane, Snyder, et al., 2017a)

When connecting this to the existing literature, the same was found by a study by Muaremi et al (2014) when exploring the utility of call duration, number of speakers turns and speech features in relation to BD symptomology, it was revealed that no specific pattern was identified and that all users showed different predictors. Other studies found more uniformed findings such as higher depressive symptoms relating to a decrease in social communication and physical activity (Beiwinkal et al, 2016), increases in manic symptoms relation to higher conversation time and movement (non-sedentary duration) (Abdullah et al, 2016), higher depressive symptoms relating to few changes in geolocation data (Palmius et al, 2017), higher manic symptoms and decreased in the number of hours of sleep the next night (Li et al, 2019). As can be seen, these findings are highly variable, all refer to different tracking factors and cannot be connected to each other, nor the current research.

Some studies report of mixed findings with no distinction between manic and depressive symptoms. For example, in a study which explored the utility of accelerometer activity and symptoms, increased accelerometer activity was correlated with both mania and depression scores (Zulueta et al, 2019). This was theorised to be related to the presence of mixed episodes in BD, where an individual can experience both depressive and manic symptomology in one episode, which has not been accounted, nor mentioned by our users in this study. However future studies need to consider the possibility of symptoms that overlap and how this is reflected in self-tracking technology for BD.

In conclusion, we gained insights into the subjective interpretation of passive and active sensed self-tracking data from the users' perspective where our user uncovered patterns of data frequency and connected this to EWS of mania and depression. An example being an increase in phone light levels (pRMT) being related to a hypomanic episode (P001), which was also confirmed in the active reported data (Bipolar UK scale) as well as the users own personal journal which indicated hypomanic symptoms around the time same time period. Generally, users collectively assumed that an increase in sensing data frequency was related to more manic EWS, whereas a decrease was related to more depressive EWS. When investigating this from an objective lens using descriptive statistics, though we did indeed find similar findings in a small group of users, the vast majority of users showed contrasting findings. For battery level data, it was described that a decrease was related to more manic mood, and this was similarly found for phone light levels and app usage data. These findings highlight the variability that exists in the data from both a subjective, as well as objective level and future research should aim in bridging this gap to explore the full utility of self-tracking technology for BD.

6.4.2 Examining the usability and acceptability of our self-tracking tool for BD

Throughout the interview points in this study, RADAR-Base for BD was universally well received, with a SUS score indicating excellent usability. The usability of RADAR-Base was explored, and it was described as "*simple*", "*easy to use*" and "*just quick*". Other useful aspects of the self-tracking tool were that it served as a reminder for key self-management behaviours, such as taking medication. Another key finding from the interview was how the apps integrated into everyday life, where it was described that "*it's become almost routine to do so…it's almost subconscious*" (P006). More negative experiences of the app centred around notification issues where an issue with RADAR-base throughout the study meant that users did not receive their notifications for the aRMT app when they were meant to. The time period for the questions was described as too short (3 hours) by users, which were not able to change due to the technical limitations of RADAR-Base. Future studies should consider a longer time period for data input, as this is likely to increase engagement with any self-report data. Also, due to the current limitations within the app, users were not able to customise their notifications within the app, but rather had to manually get them changed by asking the researcher. This was seen as a large inconvenience for users, and in future the ability to change notifications within the app should be explored.

When looking specifically at the positive aspects of the aRMT, the content was well appreciated by user such as the nature of the questions as well as the data feedback that was generated. The other aspects that were appreciated was the ability to tailor the app, such as "*getting prompts at specified times*" P006. In terms of problems the app solved, users described that the aRMT app helped them to cultivate self-awareness, especially for those who were *“still learning about EWS”* (P003). This is a pertinent finding to some of the challenges that were described in other sections of this doctoral work in relation to the sense of self being in flux in BD, due to the extreme shifts in mood. From our perspective, if app does indeed contribute to self-awareness and contributes towards am individual understanding the nature of BD, this is a huge step in the right direction for the future of self-tracking in BD. The reality of BD was touched upon in the feedback where it described that though the app does cultivated self-awareness, it doesn’t negate the reality of BD. From one’s users’ perspective it “*doesn’t prevent me from having horrible lows, the depression… just comes*” (P008). In future, the identification of such lows using this source of data could in future be linked specific advice on how to cope or manage it. There are a growing number of psychoeducation interventions, some of which are digitalised (Gadelrab et al., 2022), that might be linked, or a personalised message could be devised by the user with steps to be taken or considered. Taking this one step further, beyond psychoeducation, this form of personalised solution embedded into everyday life has been called Ecological Momentary Intervention (EMI) (Balaskas, 2021) which can be delivered via digital methods.

For pRMT, from the user’s perspective this was largely unnoticed as it was automatic, as one user described *“minimum work was needed”* (P003). This lack of interference was described as one of the valued parts of the app. The worst part of the pRMT app was technical in nature, where the issues with data collection and notification issues served as obvious problems for the users. Though the notification issues have been covered in other sections, the issue around data collection will be touched upon briefly here. RADAR-base pRMT had intermittent issues with data collection meaning that for some streams of data such as step count, GPS location and levels of spending, there was little or no data collection. A full description of this will be provided in the limitations. As can be seen in the subsequent chapter there was large utility reported from the user’s perspective for all three of these data streams. Not only was this disappointing to the users, but also served as a large limitation to the study, especially given that the ability to sense levels of spending had been developed bespoke for the RADAR-Base for BD. To round off on the experience of the pRMT app, users described that automatic data collection solves the inherent challenge of self-tracking when unwell, as a user described “*it gives me insight into when I’m manic and when I’m depressed – when I’m depressed, I can look back as I don’t remember.*" (P008). This challenge of tracking when unwell is well described in the literature (Matthews et al., 2016; Murnane et al., 2016), as well as covered in previous chapters of this research, If, indeed, the self-tracking tool can assist with the well described challenge, this is another positive step in the right direction for the future of self-tracking in BD.

6.4.3 Personal reflections of a design researcher

In this final section in our discussion, we turn to consider the nature of the investigation into BD we conducted, and the ways in which the first author featured within that process. Thus, we offer a reflexive account of this, presented in the first person by the first author which will be built upon in the final Discussion (Chapter 7) chapter:

This paper formed part of my doctoral research, where I was part of a multidisciplinary doctoral training course in Computer Science. I have a background in pure psychology (BSc in Psychology and MSc in Psychological Research Methods) with experience of conducting research in the mental health sector. It was a long learning exercise of how to apply the methods and skills I learnt in Psychology and working in the mental health sector, to then develop technology which is something you had to learn “on the job”, as they say. There are huge benefits from having the background in Psychology for the task at hand, as empathy is something that that I was able to cultivate in previous roles and could enthusiastically apply during the process of design to ensure that the users were feeling empowered to shape technology in meaningful ways. My background helped me to build relationships with users and enable roles such as lived experience co-facilitators and co-designers, both at the heart of the methods developed as part of this research. It also pushed me to notice the current limitations in design of mental health technology and reflect upon how user involvement is variable and sometimes tokenistic, and that we as *designers and researchers need to be better at involving users in mental health technology, even when that can be difficult. The ability to put myself in their shoes, by having spent much time in the mental health sector, meant I became an advocate for user needs and felt dedicated to them being heard. I was invited to peer-to-peer events, formed key relationships with BD charities and also had many users comment on how seen they felt "seen" during talks about the research, even one woman saying "it had brought tears to my eyes" because it is so rarely, genuinely, asked what those with BD want in relation to treatment and management, which unfortunately is the sad reality of severe mental illness.*

However, it was the same experience that limited me at times. Often, I felt I was a "jack of all trades and master of none" during the course of this research. There were times where I was a psychologist, other times where I was a UX researcher, some other times wearing the hat of an HCI academic and even playing the role of a data scientist. Often, I felt like I wasn't doing any of them well, I was just doing them because that is what the field requires. Mental health technology is such a multifaceted discipline, albeit in its infancy where the requirements of what is needed to create good technology is still being explored. As it's in its infancy, it lacks precedence for what constitutes good practice as it is the job for us, as current researchers in the field, to create this. The absence of a framework can sometimes be confusing and other times, troubling. For example, working with those with severe mental health issues carries associated risks. When this is being conducted in a more traditional clinical setting, those risks are known and there are associated training and operating procedures to ensure the safety of the researcher. However, when doing the same work in a non-clinical setting, a different approach is needed: operating procedures are not as regimented as those in a clinical setting, for obvious reasons. In the process of empowering those users with severe mental health issues to be co-designers or co-researchers, the therapeutic line which you are trained on in formal clinical training (during my experiences as a mental health researcher), became a lot more blurred. For example, accepting users' connection requests on social media or giving them access to personal phone numbers, conducting meetings alone, etc., are all *things that I would not do in previous research roles prior to my PhD, in line with particular risk procedures in place; however, I willingly did during my current doctoral research. Whether it’s right or wrong, risky or not risky is something that is up for debate. But what I do know is that there is huge amount of up-skilling needed in the mental health technology space where contributing disciplines need to be better at sharing practices and expertise to ensure appropriate development, not only to improve the design and deployment of technology, but also to support the development of researchers who are trying to be the “jack of all trades”. More help is needed for them to be the “master of none”. Appropriate supervision is also needed from a multi-disciplinary team, where it’s important to highlight knowledge gaps and work (humbly) towards filling those. Sharing best practice around researcher development, including researcher safety (such as regular debrief and counselling) and wellbeing, is needed to design this new breed of researcher which is not a psychologist and not even a computer scientist but something along the lines of a computer psychologist.*

6.4.4 Limitations

The study had several limitations which impacted the research. Firstly, this phase of the research was conducted during the height of COVID-19, which had a detrimental impact on mental health globally (Talevi et al., 2020). This would have significantly impacted our users, who are already on the journey managing their mental health daily. The added pressure of COVID-19 could have impacted the level of engagement and participation in the research, such as the post study interviews. For example, though the study was opened for recruitment via local NHS sites, which itself was an intensive process considering the nature of the study, not a single user signed up for the study though some expressed interests in the study. When compared to the research conducted pre-COVID-19 such as Concept Generation and Ideation phase, this was an issue not seen. Exploring the literature on the impact of COVID-19 on BD it was found the following were likely to be factors of impact on willingness to engage restrictive measures, biological rhythm changes, sleep problems, social isolation etc. The impact on this on the willingness and capabilities of our users, and prospective users in this research should be considered when interpretating our recruitment and engagement levels (Mind, 2021). More specifically, the impact of COVID-19 on mental illness was witnessed during the study by the researcher during interviews where many of the user expressed worry, anxiety and depression during periods of lock down. This was not just limited to the user themselves, but also the researcher who was experiencing similar feelings however with the added pressure of having to feel "strong" in front of users, bearing in mind her role to empower and advocate. This was, at times, extremely difficult for the researcher. The collective researcher-user mental health during these times is critical to consider when interpreting this research.

During the deployment of the prototype during the evaluation phase, there were a whole host of technical issues. Firstly, data from the data streams levels of spending, GPS and step count were not accurately collected from the pRMT app which meant there it was a large amount of missing data. This was a large disappointment for the research as well as the users, as these were three data streams that were predicted to have significant` utility into understanding EWS. Especially with levels of spending which had been designed and developed into RADAR-base as a result of user feedback. This missing data significantly impacted the month 3 interviews, where it was planned for the data to be feedback to the user. The large amounts missing data impacted the data visualisation as users expected to see this data, which in turn impacted engagement. Also, as RADAR-base was managed by the industrial sponsors of this doctoral research, there was a disconnect between the researcher and RADAR-base. The research only received the data once it had been collected at distinct time points, where it was often difficult to decipher issues until the data was processed. In other words, there was no active monitoring of the data. The researcher learnt about the data streams not being collected close to the 3-month interview point which also impacted the researcher's interpretation of the data and subsequent visualisation. In hindsight, presented the data in a MATLAB figure was deemed confusing by the users and the research should have considered more user-friendly methods of presenting this data. But given the issues at the time, this was overlooked. A better approach would have been to consult users more about their preferences on data visualisation and strive towards this.

Another technical issues with RADAR-Base for BD was the delivery of the aRMT self-report measures. Largely the daily delivered questionnaires such as the visual mood diary, personalised medication tracking were delivered correctly. However, the other self-report measures such as the Bipolar UK scale, QIDS-SR, ASRM, SHAPS were not delivered correctly which resulted in a huge amount of data loss. As even when they were delivered correctly, the three-hour slot to answer the questions was often missed by our users. This was because the users considered this to be too much of a short time frame to input the necessary data. This was a technical limitation of RADAR-Base that the time period to answer any self-report measures via aRMT could only be extended to three hours. The impact this had on the study was large, and it led to the redundancy of these measures from a statistical perspective as data was largely missing, infrequent and couldn’t be used to draw insights. Again, this was hugely disappointing to our users as these self-report questions were user-led and the insights gained from them did not live up to expectations of the predicted impacted

Lastly as the tool was highly personalised to the specific user, it can be argued that this may impact the overall generalisability of the tool, making it less generalisable to a larger sample. Future research should consider the scalability of this level of personalisation as well as the impact on generalisability. It can also be argued that the non-alignment of the qualitative and quantitative data illustrates the shortcomings of relying on self-stories, as the non-alignment could be attributed to erroneous recall.

6.5 Chapter Summary

A final prototype was developed and enrolled for a 6-month period of beta testing with n=8 users. Out of the n=8 that were enrolled in the study, n=6 completed the study which was defined by the full 6-month of data collection. Users were consulted during the various points of the 6 months via a series of interviews which were conducted at the 1-month, 3-month, 6-month and post study periods. The interviews focused on gaining insights into the user experience of RADAR-base for BD, to ensure we were empowering users to be involved in the design and research of the self-tracking tool. In the post study data exploration interview, users were given an interactive dashboard of their descriptive data and asked to interpret this and connect this to their EWS. A series of interesting findings emerged where users collectively described a pattern between the frequency of their data and the connection to EWS; where high frequency in the data was related to EWS for mania and low frequency in the data was connected to depression. As an example, P001 described a spike in phone light levels to be connected to increased manic symptoms, which was also demonstrated in her Bipolar UK scale (via aRMT) which displayed hypomanic state around the same time period. This user also confirmed that her own personal journey entry also exhibited hypomanic symptoms at the same time period demonstrating that RADAR-Base is doing what “it’s meant to do”. Though our statistical findings do not necessarily align with the pattern that the users perceived in the data, it nonetheless sheds light on the complexity and individual variability in the experience of BD symptoms. Future research should aim in bridging this gap, as there were universal acceptance and utility of using RADAR-Base for BD, where users described that that it helped to cultivate self-awareness and ameliorate some of the challenges experienced during self-tracking, such as the ability to self-track when unwell.

Chapter 7: Discussion

Chapter summary

- The key contribution of this doctoral thesis is towards the knowledge of user involvement in the design and research of self-tracking in BD, serving as a first of its kind in combining methods seen in both HCI and PPI
- Using this approach was advantageous in getting closer to understanding the complex individual variability in self-tracking, as our findings revealed intimate and intricate details of the experience of self-tracking and how technology can be tailored towards this.
- The concept of individual variability was further expanded upon through the findings, such as subjective experience of quantitative data and the importance of tailoring and customisability.
- How the self-tracking tool positively responded to the inherent challenges of self-tracking via the ability to passively sense data and specifically track clinically validated EWS is also explored.
- Using a *cohort-style* user group was beneficial to the attrition rate of this study as the dropout rate was low (25%), compared to other studies in BD.
- Finally, the implications of applying this methodology are reported upon, as well the as the researchers personal account of this.

7.1 Implications for methodology of user involvement; drawing from PPI and HCI

One of the most salient contributions of this doctoral work are towards user involvement in the design of self-tracking technology in BD. In an earlier chapter (Chapter 2), there was an exploration around the methods employed to design and research in this unique, yet emergent, domain of mental health technology. Mental health technology is an intersection point between healthcare and technology; therefore, we explored the approaches of user involvement within these two similar, yet distinct domains. Approaches in HCI and technology generally exhibited an extended "vocabulary" when it came to involving the user from user-centred design (UCD) to more participatory democratic forms, considerations of non-use, design fictions, critical engagements and various other approaches (Baumer, 2015). These approaches typically employ an iterative process, where the user is involved in the cyclical development of the product from concept development to deployment (McCurdie et al., 2012). Whereas in healthcare approach, PPI is a more dominant form of involvement (Jennings et al., 2018). This process of involvement is more linear, where the patient is involved throughout discrete parts of research development. PPI practices are on a spectrum stretching from low involvement to high involvement. The lower ends of this spectrum as referred to as tokenistic involvement, whereas the higher levels are when the patient is fully embedded. Aiming to establish an equal relationship between the researchers and PPI user, with shared decision making (often referred to as "co-production") is at the high end of this participation ladder and described as the pinnacle by many researchers and clinicians, if not PPI users themselves. Co-design is often seen at the top of this hierarchy. Though the approaches seen in HCI and PPI have distinct mechanism and processes, it was concluded that they have three common goals when it comes to involving the user. This includes the following: 1) they aim to involve the user in research and development 2) they focus on understanding and empowering the user 3) they aim to make changes based on user feedback.

These approaches of involving the user were explored within a scoping, academic literature review of current self-tracking of technology for BD and it was found that there was a large variability in how and when users are involved (Majid et al, 2021). Out of the 11 novel smartphone apps included in this review, 4 (36%) self-monitoring apps were classified as having no mention of user involvement in design, 1 (9%) self-monitoring app was classified as having low user involvement, 3 (36%) self-monitoring apps were classified as having medium user involvement, and 2 (18%) self-monitoring apps were classified as having high user involvement. Despite the presence of extant approaches for the involvement of the user in the process of design and evaluation, there is large variability in whether the user is involved, how they are involved, and to what extent there is a reported emphasis on the voice of the user, which is the aim of such design approaches. Demonstrating that currently, individual variability is not being accounted for in this domain as highlighted by Ortiz et al., 2021. It was concluded that technology should be tailored according to users' needs and preferences, in order to maximize the potential that technology has in order to improve the lives of those living with this condition. In a bid to make changes in this domain, an expert review was carried out within this review to develop best practice ingredients of user involvement, which resulted in seven guiding principles. Essentially the best practice ingredients described that users should be involved early in the design process, and this should not just be limited to the design itself, but also to associated research ensuring end-to-end involvement. Communities in health care–based design and HCI design need to work together to increase awareness of the different methods available and to encourage the use and mixing of the methods as well as establish better mechanisms to reach the target user group.

The motivation to empower users to become *co-designer*s and/or *co-researchers* became the driving factor for this research and heavily influenced the methodology of the redesign of RADAR-Base for BD. The re-design of the self-tracking tool took part in three phases which are outlined in Chapter 4, Chapter 5 and Chapter 6 which are Concept Generation and Ideation, Prototype Design and Deployment and Evaluation, respectively. Throughout these sections, we strived to closely embed the best practice ingredients of user-involvement as described in our review.

For the Concept Generation and Ideation phase of this research, not only did we involve users in the design of the self-tracking tool (n=18), but we also involved them (n=2) in them in the design of the research to explore reasons, methods and challenges for self-tracking for BD, as well as opportunities for technology for self-tracking. We ran a series of workshops in which n=18 users partook in, across two different locations in the UK (Nottingham and Northampton). Two users with lived experience of BD co-facilitated the workshops in order to create a more peer-to-peer environment. Using these methods, we uncovered novel findings in the understanding of self-tracking beyond other key studies which explored this using more traditional research methods such as surveys and interviews (Matthews, Murnane, & Snyder, 2017b; Murnane et al., 2016). For example, the social nature of self-tracking was deemed to be the most important for our users, whereas in similar literature the clinical basis of self-tracking had more emphasis. This also sheds light on the differences that arise dependant who the app is being designed for. In clinical research, clinical researchers may design what they think is best for the patient and make this more palatable for the patient, so the patients and clinicians could collectively then judge the success of this in clinical terms. Whereas in our research as the focus was on the user in more holistic sense, and in relation to their own self-management, where differing findings were revealed. Our findings also shed light on the complexity of self-tracking, where over 50 methods of self-tracking method were described in the workshops. Overall, it was concluded self-tracking technology should be designed to account for personal struggles and individual variability for those with BD experience and provide users with the ability to have passive sensing methods coupled with increased control, constitution of 'successes', and ability to avoid tracking easily. All of which we implemented in the next phase of research.

The next phase of the research was the Prototype Design and Deployment phase where the findings from the workshop were further investigated in relation to the design of RADAR-Base for BD, focusing on EWS for mania and depression. We explored the findings with n=10 users who partook in the workshops to investigate this. Our findings uncovered the utility of a number of novel self-tracking tracking factors which have received little reference in the literature thus far such as levels of spending, personal hygiene levels, employment information and driving patterns. These self-tracking factors, as well as already established sensing capabilities of RADAR-Base, were connected to specific EWS for mania and depression. A description on how these factors would change according to the specific EWS was also provided from the users' perspective. For example, for the presence of the EWS *reckless pleasure seeking* for mania, users connected the increase of pRMT collected sensing data such as app usage, phone usage, GPS location as well as levels of spending. The findings from the interview were used to create design recommendation for RADAR-Base, and to tailor this towards the needs, preferences and individual variability of those with BD. Several changes were made such as the ability to sense levels of spending, implementation of user-led self-report measures (via aRMT app), as well as increased tailoring such as the ability to turn the sensing factors ON/OFF. A working prototype was established and enrolled for a series of alpha testing with n=2 users, where more changes were commissioned to closely align the self-tracking tool to the experience of BD. As most of our users during this series of work were repeatedly involved and consulted during the research, we established a type of *cohort-style* group of users. Most of which expressed interest in the full testing of the app, as they had been so closely involved in the development and design thus far.

A final prototype was developed and enrolled for a 6-month period of beta testing with n=8 users. Out of the n=8 that were enrolled in the study, n=6 completed the study which was defined by the full 6-month of data collection. Users were consulted during the various points of the 6 months via a series of interviews which were conducted at the 1-month, 3-month, 6-month and post study periods. The interviews focused on gaining insights into the user experience of RADAR-base for BD, to ensure we were empowering users to be involved in the design and research of the self-tracking tool. In the post study data exploration interview, users were given an interactive dashboard of their descriptive data and asked to interpret this and connect this to their EWS. A series of interesting findings emerged where users collectively described a pattern between the frequency of their data and the connection to EWS; where high frequency in the data was related to EWS for mania and low frequency in the data was connected to depression. As an example, P001 described a spike in phone light levels to be connected to increased manic symptoms, which was also demonstrated in her Bipolar UK scale (via aRMT) which displayed hypomanic state around the same time period. This user also confirmed that her own personal journey entry also exhibited hypomanic symptoms at the same time period demonstrating that RADAR-Base is doing what “it’s meant to do”. Though our exploratory descriptive statistical do not necessarily align with the pattern that the users perceived in the data, it nonetheless sheds light on the complexity and individual variability in the experience of BD symptoms. Future research should aim in bridging this gap, as there were universal acceptance and utility of using RADAR-Base for BD, where users described that that it helped to cultivate self-awareness and ameliorate some of the challenges experienced during self-tracking, such as the ability to self-track when unwell. Finally, RADAR-base did not achieve much of the promised functionality so there were key aspects of its potential acceptability and usability that could not be tested. If it did achieve better functionality, the complexity of the data and its relationships could be explored in a granular way.

Key takeaways

- This series of research is an example of high-level involvement of the user in the design and research for self-tracking for BD, serve as a first of its kind it combining methods seen in both PPI and HCI.
- The intimate and intricate details of self-tracking, as well as the recommendations to develop a tool that was more closely aligned to the experience of BD are key advantages of using this approach.
- The tool was universally demonstrated good levels of usability and acceptability, with further recommendations for improvement.
- Using a cohort-style user group served large benefits to the attrition of the study which had a low drop out (25%), compared to what is seen in literature reporting dropout rates between 32%-38%.
- However, the impact of employing this approach, as well as the personal experience of applying this approach and extended methodology will be reported upon in the next section of the discussion.

7.2 Reflections on the methodology and impact on the research

A reflexive personal account of conducting research at this intersection of healthcare and technology was provided in Chapter 6 of this research and will be elaborated upon here.

7.2.1 Our methodology: help or hindrance?

Though the application of this type of approach served benefits for the research in terms of motivations as demonstrated in the low dropout rates in the study, one must consider the collective experience of the methodology and how this impacts the results of the study. This will be reflected in first person account from the researcher's experience.

This doctoral research aimed to cultivate relationships with users to empower them to become co-researchers and co-designers for self-tracking technology for BD. I already had an existing background in mental health, with several years of working in the assessment and treatment of various mental health conditions. This served as a great tool to cultivate the empathy and understanding needed for the research motivations, and to build relationships with users. It also served as a great motivating factor of knowing the harsh realities of those living with severe mental health and being motivated to improve this. Having this unique background, which was applied to the development of technology is not commonplace and how this impacted the results and motivations of users must be considered. Firstly, to work with those with mental health, especially in the context of research was a fine-tuned task. Essentially you are going in with the attitude "I know things are not great for you and I want to try and help". This requires a level of confidence, and the humility to step down and understand you are not the expert here. For many researchers, having control over the research is a large prerogative. Especially within doctoral studies where you are essentially trying to become the expert. However, in the case I was not the expert and had to understand that I would never be. The control was rightly given to the experts to guide the research as per their needs. Looking back in reflection, this requires a high level of confidence and motivation. This motivation never ceased and the goal of making sure the users felt as though they were the expert at every point during the study was always the desired outcome. But the process of trying to understand the perspective from multiple users with multiple experiences was in itself a complex task, it required active listening and looking for meaning in ambiguous contexts. I was able to draw on my clinical experience such the delivery of group and individual therapy in assisting with workshops and interviews respectively. The ability to understand safeguarding, how to keep users safe during the research, was also extremely crucial to the tasks. For example, what does one do when a user is visibility ill, or a harm to themselves and others. Due to my background in mental health, I had a clear idea on how *to deal with such situations. But this is not common amongst my peers in the Department of Computer Science who may be working on similar types of projects. This is one insight into the level of upskilling needed within this community.*

However, is this closeness helpful or a hindrance in the context of this research must be considered. Questions such as did our users have a positive experience of using self-tracking tool because of the tool? Or was it because of the relationships that arose as a result? For example, in one excerpt of feedback from a user she was describing how she kept forgetting to fill out the questionnaires via the aRMT via the following quote "and then I remembered I've got to fill out the questionnaires for Shazmin" (P003). Not the research study, but Shazmin. It must be considered that using this type of methodology meant that users were emotionally connected to the relationship within the research, and maybe not the tool itself. For example, during interviews, it was often the case that the conversation would deviate away from the schedule and the simple question of "how are you?" would open a conversation just between two human beings, sharing and talking about life itself. As a goal to empower users, these conversations were not shied away from but welcomed in the attempt to make users feel confident, feel welcomed and most importantly feel heard. As the latter parts of the research were conducted during the heights of COVID-19, where there was an additional pressure of mental health and loneliness, these conversations increased with some users simply having nobody else to speak to during this time. Again, how that impacted the results of the study is something to consider. Only future studies which attempt to replicate the results of this study would shed light on these implications. For example, it would be of interest to understand how the results of a future study would be impacted without the type of researcher-user relationship cultivated in this study, or with future researcher who had a different background (e.g., non mental health related).

The researcher-user relationship that we established cultivated a level of support and assistance in the usage of the self-tracking for BD. For example, users were assisted by the *researcher to download the app, they had the researchers contact number if anything went wrong, they had regular check ins about the apps and reimbursed for their time (via Amazon vouchers). The point of contact was always me whom they had worked with to develop the app and then use the app over the course of almost 2 years. The level of personalised support and assistance, especially with a single researcher, in a study that was bigger with a larger user pool just would not be feasible, nor sustainable and how that would impact the outcomes are yet to be explored. Also, the regular check ins with a single researcher could also had created a sense of accountability, which may have influenced the study outcomes. For example, did the user members continue with the study because of their relationship with the researcher to support her motivations for doctoral studies? These are questions which must be explored within the design of similar research studies*

7.2.2 Personal difficulties of employing this methodology

Employing this methodology was not without its challenges from a personal perspective. The norms of what apply in traditional research did not apply when using this methodology Coming from a mental health background, there are a numerous policies and guidance in your conduct with the "patient". For example, there is lone working policy where if you are seeing a patient by yourself, it was mandatory that another researcher knows the name, location, meeting start time, meeting end time and a check in at the end of the meeting. Due to this study sitting in the intersection between healthcare and technology, these polices were not tailored to this and simply not available. It was often the case where meetings were conducted in coffee shops, phone calls on personal numbers and a number of users were sending requests on social media platforms such as LinkedIn and Instagram. There were sometimes difficulties in knowing where the line when you are trying to empower someone essentially be your colleague, yet they are also the subject in the study. This was particularly difficult for me where I experienced dissonance between being an advocate and an also a responsible scientist.

Furthermore, it was particularly challenging working with mental health during the peaks of COVID-19. The impact of COVID-19 on general mental health were a huge cause of concern, where the pandemic had huge psychological impact on individuals (Talevi et al., 2020). The impact that COVID-19 on those with existing mental health issues was hugely significant, with 1/3 of young people and adults experienced worsening mental health the pandemic (What Has the Impact of the Pandemic Been on Mental Health?, n.d.). Having to motivate myself, to work with a user group who have the additional pressure of COVID-19 on their existing mental health issues was a hard task. It was taxing mentally, as I also had a role to perform which was to empower users and make them feel heard. But as the saying goes "it was like trying to pour from an empty cup" during this collectively testing time for all. The individual struggles of COVID-19 hugely impacted the latter parts of this research, namely the evaluation phase where both the user group as well as the researcher were trying to get through the social restrictions, anxiety and shock of COVID-19. Prior to COVID-19, our meetings were more in person, in workshop- style settings where the aim was to get everyone involved and excited about the future of self-tracking. There was a tangible feeling of community where users came together to discuss their experiences for the development of science. However, after COVID-19, these in person methodologies were no longer feasible and we had to move onto more 1-1 methods using MS Teams. This felt almost 2 dimensional compared to our methods pre-COVID and significantly impacted our engagement. Meetings with users got shorter and shorter and it was a struggle to sometimes get in contact with some users. Upkeeping the momentum of advocacy and empowerment was difficult during these latter phases of the study, impacting the quality of our results in this phase.

Key takeaways

- The background of the researcher could be contributing factor to success when trying to embed such a methodology which aims to empower patients as it requires a high level of empathy. Such as having a background in mental health and clinical practice (e.g., group, individual therapy) as the researcher did in this doctoral work.

- Future initiatives should look to upskill researchers who work in this space in emotional intelligence such as empathy who do not come from a mental health background in order to create more meaningful relationships with this unique set of users.
- Future research should be considering the researcher-user relationship in the process of applying this methodology and investigate the impact of this on the adoption of technology.
- There is an overwhelming need for the development of policies and procedures that safeguard researchers conducting research in severe mental health (e.g., lone working policy) which don't fall under traditional healthcare departments in universities and research organisation (i.e., NHS). Such as the case in this research which was conducted in the School of Computer Science.
- Specifically, there should be a policy guiding how to conduct such advocatory research in the context of severe mental health focusing on where to draw boundaries to protect both the user-researcher such as the Professional Boundaries and Relationship Policy as seen in the NHS.

7.3 Implications for self-tracking technology

The current research has numerous implications for self-tracking technology for BD. Implications apply to the design of self-tracking technology which have largely been covered in the previous section. However, this stretched beyond this to understanding the individual variability of BD, needs and preferences specifically for this technology which will be discussed in this section. Lastly, we will explore the role of validation in the design of self-tracking for BD focusing on EWS and passive sensing.

7.3.1 Individual variability and self-tracking

The lack of user involvement in the design and development may explain why individual variability of illness expereince is not considered in current self-tracking technology for

BD (Ortiz et al, 2021). Beyond our review which explored this (Majid et al, 2021), it was also reflected in a large review of the past decade on HCI and affective health (including BD) which found that out of 76 studies involved, only 16 studies reported clinical evaluations which was descried as those users who use mental health services or who meet the formal criteria for a specific mental health problem (Sanches et al., 2019). In response to this, this series of research has highlighted many findings in the individual differences of the experience of BD and connected this to the design and development of BD. This was firstly explored in the Concept Generation and Ideation phase when exploring reasons, methods and challenges self-tracking. Specifically, when explorations methods of self-tracking it was revealed at how truly complex this is for those with BD. N=18 users talked about over 50 methods that they use to track their mental health with even more complex descriptions of how they apply and use these methods. There was some exploration on how numbers (i.e., qualitative information) are used to make sense of a qualitative experience such as mood changes, or EWS. For some users, they found counting items done on a to-do-list as a useful mechanism of understanding their mood. stated, "*make a list of 20 things to do... if I get at least 5 done then that's good*", while another user said, "*take three things of this to do list, turn the page over and write those down… this is more manageable and if you get one thing of this list today then that's great".* The completion of 5 things done of 20 constituted successes for one user, whereas for another user getting 1 out of 3 things done was success. This is just one insight into how variable the individual experience of a good day in BD can be.

This individual variability doesn't just apply to self-tracking to BD, but also the self-tracking discourse in HCI more generally and the concept of "normative ontology"; where an ideal is set within self-tracking technology without accounting for individual differences. In this paper, Spiel describes the mechanism of producing a step using self-tracking technology and questions whether the steps that are counted are produced in the same way from human to human. Indeed, she argues that currently a human step is "whatever a set of algorithms running on a portable computational device can detect and process". Martin and Lynch (A. Martin & Lynch, 2009) unpicks the process of counting and aptly summarize that "*how a count is produced depends very much on how is doing the counting, what the count if for, and the occupational and physical location of the counting event*". To relate this back to BD, Emily Martin (E. Martin, 2009) in her book Bipolar Expeditions described the concept of a "zero" day for her. A zero day is a day without change—an unremarkable day—which can actually be an indicator of improvement for those with BD, however this is unlikely to be treated in the same way for non-BD individuals who engage in self-tracking who interpret this as an "off day". Understanding this individual variability in our experiences of life, especially in the context of mental health, is hugely important consideration for future designers and researchers in the space as they have a responsibility to design technological experiences with meaning, but not experiences that define what is means to be "*well"* as this is inherently subjective, with high levels of variability dependent on the context.

In the Prototype Design and Deployment phase of this research, there further insights into the individual experience of BD and it's relation to self-tracking technology. An extensive list of useful self-tracking factors was provided in this research and some of what had received little attention in the self-tracking research thus far such as levels of spending, driving patterns and employment information. It was found that the utility of these tracking factors varied from user to user, and it was imperative for users to have control of which of these tracking factors they were engaging with in the use of technology, as they can significantly impact their willingness and trust of the system. Specially for more sensitive tracking factors described such as levels of spending. Being able to tailor the system, based on individual variability was a large output from this research which was not just limited to the tracking factors (via pRMT) but also the timing of self-report outcomes (via aRMT). The lack of tailoring was self-tracking technology for BD was described as a paint point by Ortiz et al., 2021, and by offering this to our users in the current research we hope to made first step at improving this.

In the evaluation phase, individual variability and its connection to self-tracking technology was largely accounted for in exploration of the data produced in the study. The subjective interpretation of the data was compared to the statistical exploration of the day, where interpretative discrepancies arose. The subjective experience, or self-expereince, of interpreting self-tracking data is of huge benefits, as these individuals are the experts of their own experience. Statistical methods operate at the different level, change is looked at through the lens of numbers, and mathematical significance is set at a fixed level. But as we have seen, and explored in previous sections, there may need to a level of tailoring and personalization to how we objectively measure in this case. Statistical models need to be tailored to be sensitive the definition of change set by the users, as this will differ from user to user. This was also echoed in the findings by Ortiz et al, where the review highlighted the difficulty in developing an overarching prediction algorithm for a variable group of patient and advocated for the development of personalized prediction models alongside other key authors such as Palmius et al., 2017; Pratap et al., 2019. Future research should aim to bridge the gap between subjective interpretation and objective analysis in self-tracking data by exploring the utility of personal predictive models.

7.3.2 The role of validation: clinical tools and passive monitoring

Another current pain point in the design of self-tracking tool is the lack of validity in terms of research data and clinical tools used (Nicholas et al., 2017a). This, combined with a lack of user involvement, suggested there are inherent issues in the current app marketplace for BD. The current research aimed to ameliorate some of these pain points but employing the use of validated clinical tools such as the EWS checklist (Lobban et al., 2011) and also collected data for a period of 6 months using the development self-tracking tool. The use of the early warning signs was largely welcomed by users, where in all tasks where these were employed, users identified a list of EWS that were relevant to them. In latter parts of the research, users also were able to use the EWS and connect this to self-tracking factors available in RADAR-Base and beyond. In the feedback during the evaluation tool, users are described the self-tracking tool as being helpful for cultivating awareness "especially for someone who is new to learning about EWS" (P003). In the context of this research, the EWS checklists were used for exploration both from the researcher and users' perspective. When trying to connect this to the data, this came with difficulties as only one user was able to draw connections between the EWS and the data. Whereas for other users, it was much more difficult to draw connections between specific EWS and the data, as for many of them there is an overlap. Unlike other studies in this space, there was no specific focus on the predictive quality of the self-tracking data and EWS. Therefore, we are unable to draw conclusions between EWS and self-tracking data beyond that from the users' subjective experience, as the EWS data wasn't used in the statistical analysis. The reason being as it's a static measure of symptoms, not subject to change unlike the other aRMT data, which was used in the analysis, as the primary aim of the analysis was to understand if there were changes in the data according to experienced mood. EWS data would be useful in drawing predictions in passive monitoring (via pRMT) and active monitoring (via aRMT) data. In other words, how much of the change in the data is relevant to specific EWS. But as mentioned in previous sections of this study, in order to draw predictions, algorithms first must be personalized to understand the individual variability that is present in EWS. From the findings of this research, one method of achieving this would be by the involvements of users and the use of validated clinical tools such as the EWS checklist in the development and tailoring of predictive algorithms. Though this was beyond the scope of the current work, future research should consider this in order to move away from the issues that are seen in this domain of self-tracking tech for BD.

To account for the inherent challenges that are experienced during self-tracking for those with BD, there was a call to passively monitor and intervene (Murnane et al., 2016) in order to reduce users efforts and burdens, as opposed to more actively inputted self-report data, and to also support the continuous collection of data. To account for this challenge, we enrolled a self-tracking tool that has both passive monitoring and active monitoring capability. The utility of this was explored and the results proved that the passive collection of data does indeed assist in the challenges that those with BD experience in self-tracking as exemplified by this key quote by a user when examining their collective experience of the pRMT app "*it gives me insight into when I'm manic and when I'm depressed – when I'm depressed, I can look back as I don't remember" (*P008). The utility of this was further demonstrated when exploring the data with users, where patterns of the data was described to provide insight into EWS for mania and depression. The findings of this research demonstrate the utility of passively collection data in terms of user acceptance and for the exploration of EWS. Future studies to aim to explore this further.

Key takeaways

- The subjective experience of converting a qualitative expereince into something quantitative (i.e., what numbers mean) has high degree of individual variability within BD
- The ability to customize technology are useful features in aligning technology to users' needs and preferences
- When looking at quantitative data, it is important to give value to the users' subjective interpretation of the data, rather than just relying on statistical methods which have sensitivity thresholds set a fixed level rather than being personalized
- From the users' subjective experience, employing the EWS checklist as a clinically valid tool to explore relapse was useful in this study. However future studies should consider the predictive quality of this tool in the context of active and passive sensing methods.
- Users showed acceptance of passive sensing methods in responding to some of the inherent challenges users experience when self-tracking

7.4 Implications for bipolar disorder

Cultivating self-awareness plays a key role in psychological treatment for BD, which when in adjunct to pharmacological treatment, has the largest evidence base in the treatment and management of BD including reducing onset of new episodes, length of hospital stay and adherence to drug treatment (Rabelo et al., 2021). The current research provides evidence on the usability and acceptance of novel self-tracking tool, which has been designed closely to users' needs and preferences in order to aid self-awareness in BD. Cultivating self-awareness was indicated by our users as a positive outcome of RADAR-base usage and users understood more about their EWS through the data.

This study also tackles the lack of parity of the involvement of users with physical and mental health issues in the development of technology, where those with the latter are less likely to be involved than the former (Goodwin et al., 2016). Why this is the case has been touched upon in Chapter 2 and it has been described that there are inherent difficulties and systemic issues that make it difficult to involve users with mental health issues. However, the current study proves that high level involvement is possible. But one thing that hasn't been considered, and that has been learnt though the course of this research, is the real-life gravity of stigma in BD. Stigma is defined as "negative or discriminatory attitudes that others have about mental illness" (Borenstein, 2020). When considering the research, studies has shown that high levels of self-perceived stigma are associated the lower levels of functioning in BD (Vázquez et al., 2011). When investigating this further in a large study (n=80), it was found that the three predictors of functioning were severity of depression, perceived social support and internalised stigmatisation (Cerit et al., 2012). When describing functioning in this study, it was defined as "social function" including household relationships, participation in social activities, daily activities and hobbies, taking initiative and self-sufficiency, occupation and others. The role of this perceived stigma on users' ability and willingness to be involved in research and design should be considered, as it could play a significant role. This perceived stigma is something I witnessed during this research, when on one occasion I gave a talk to boost study recruitment at Bipolar UK, I talked about the aims of the research which were to empower users to be parts of self-tracking for BD. I was inundated with emails after this with people thanking me, with one lady saying that she had been brought to tears because it's not very often that they get asked "*what they want*" when it comes to BD. The perceived stigma and feeling of disenfranchisement are common in BD and as researchers and designers must become mindful of when thinking about the barriers of involvement. It is our role as researchers and designers to tailor our practices to true advocacy, moving away from disenfranchisement, if we truly want to get close to user's needs, preferences and experiences.

Key takeaways

- Stigma in BD could play a role in user involvement in research and design of self-tracking technology.
- It is the role of future research and designers to consider this as a barrier to involvement, and strive to tailor their practices towards advocacy with a sensitivity towards the high levels of marginalisation in BD

7.5 Responsible Research and Innovation Practices

The current research aligns with the framework for Responsible Research and Innovation as outlined by EPSRC, who have partly funded this research (EPSRC, 2023). This framework outlines the process that seeks to encourage the cultivation of creativity and the creation of opportunities for scientific and innovative endeavours that align with societal values and serve the greater public good. Responsible research and innovation recognize that innovation can give rise to inquiries and ethical dilemmas, frequently carries ambiguity in its objectives and motivations, and remains unpredictable in its potential outcomes, whether positive or negative. The framework outlines a process titled the AREA approach defined by the following steps: anticipate, reflect, engage and act.

The current research aligns with the first step framework by anticipating the role of the user in the development of self-tracking technology, considering the impact of low user involvement. Secondly, throughout the central thesis this was reflected upon often with involvement of users with BD to ideate what constitutes better research and development in the space. In terms of engagement, those with BD closely collaborated with the researcher to define research, conduct data collection as well as analysis, ultimately defining the trajectory the research would take as per the final “act” stage.

7.6 Overall Limitations

There were inherent limitations of this research which had significant impact on the scope of this work. First it is important to consider the sample. Whether the *cohort style* user group that we employed in this study has been a help or a hindrance has been considered in other part of this discussion. The other aspect to consider is the presence of relapse within our sample, as a large percentage of our user group were in points of remission experiencing euthymic mood for large parts of the study. Much of our sample (75%) had more than 7 years lived expereince of BD in Concept Generation and Ideation phase as well as Prototype Design and Deployment phase. This was similar for the evaluation phase where an even larger proportion of the sample (87.5%) had over 7 years lived experience of BD. This sample may not represent the experiences, needs and preferences of those compared to those who earlier in their journey of BD and future research should consider this. This is highlighted by the number of users in the different packages of research who already had existing methods of self-tracking demonstrating some level of insight into their condition, as well a detailed viewpoint when it came to making suggestions towards RADAR-base. However, future studies should consider those who are earlier in their BD and self-tracking journey to compare findings as it could be the case that needs, and preferences will be distinct dependant on where you are on this trajectory of recovery.

There were a host of technical limitations of RADAR-base that impacted the study, as well as the quality of data in this study. For the Prototype Design and Deployment phase, there were limitations in terms of time and resources relevant to making changes that were user led, and a number had to be chosen according to the project timelines and feasibility. RADAR-base was used as a base for exploration for the users, which may primed their thinking of what self-tracking for BD should look like. It prevented the “blue-sky” thinking, which is a form of creative brainstorming with no limitations, and not guided by what you have done or capable of doing. However, there were constraints in thinking based on the capabilities of RADAR-Base in this research. Future studies should promote “blue-sky” thinking in future and allows users to think from scratch without any perceived limitation, to see whether there is difference in thinking in relation to design.

During the deployment of the prototype during the evaluation phase, there were a whole host of issues. Firstly, data from the following data streams levels of spending, GPS and step count were just not accurately collected from the pRMT app which meant there it was a large amount of missing data or in some cases, no data. This was a large disappointment for the research as well as the users, as these were three data streams that were predicted to have large utility into understanding EWS. Especially with levels of spending which had been designed and developed into RADAR-base as a result of user feedback. This missing data significantly impacted the month 3 interviews, where it was planned for the data to be feedback to the user. The large amounts missing data impacted the data visualisation as users expected to see this data, which in turn impacted engagement. Also, as RADAR-base was managed by the industrial sponsors of this doctoral research, there was a disconnect between the researcher and RADAR-base. The research only received the data once it had been collected at distinct time points, where it was often difficult to decipher issues until the data was processed. In other words, there was no capabilities of the active monitoring of the data. The researcher learnt about the data streams not being collected close to the 3-month interview point which also impacted the researcher's interpretation of the data and subsequent visualisation. In hindsight, presented the data in a MATLAB figure was deemed confusing by the users and the research should have considered more user-friendly methods of presenting this data. But given the issues at the time, this was overlooked.

Another technical issues with RADAR-Base for BD was the delivery of the aRMT self-report measures. Largely the daily delivered questionnaires such as the visual mood diary, personalised medication tracking were delivered correctly. However, the other self-report measures such as the Bipolar UK scale, QIDS-SR, ASRM, SHAPS were not delivered correctly which resulted in a huge amount of data loss. As even when they were delivered correctly, the three-hour slot to answer the questions was often missed by our users. This was a technical limitation of RADAR-Base that the time period to answer any self-report measures via aRMT could only be extended to three hours. The impact this had on the study was large, and it led to the redundancy of these measures from a statistical perspective as data was largely missing, infrequent and couldn't be used to draw insights from the passively collected data. Again, this was hugely disappointing to our users as these self-report questions were user-led and the insights gained from them didn't live up to expectations of what they could have been.

7.7 Overall Conclusions

The overall aim of this doctoral work was to explore the use of a mobile self-tracking tool which uses passive and active sensing techniques to understand EWS for relapse for those with BD, which is closely aligned to the personal and lived experience of its users through high levels of user-involvement throughout the course of design and evaluation. To ensure this high level of involvement, the research drew upon participatory involvement methods from both HCI and PPI. From the related work section of this thesis (Chapter 2), it was concluded that technology should be tailored according to users' needs and preferences, in order to maximize the potential that technology has in improving the lives of those living with this condition. In a subsequent chapter of this doctoral research (Chapter 4), we explored reasons, methods and challenges for self-tracking for those with lived experience of BD via two user-led workshops with 18 users. Compared to previous studies on this topic, our findings revealed that the social, communicative basis to tracking was highly important, compared to the more clinical basis of tracking that was seen in other studies. Our findings also documented a large level of complexity in self-tracking, where 50 methods were described. Chapter 5 leveraged the findings on the intimate and intricate nature of self-tracking to guide the re-design of an existing self-tracking tool (RADAR-Base) via a series of interviews (n=10). Financial tracking was adapted as a capability for passive sensing and a range of user-led, active, self-report capabilities were added such as a daily visual mood diary, Bipolar UK mood scale, personalised medication tracking, diary for reflection, and a range of clinically validated tools to measure depression, mania and anhedonia. To account for individual variability in the experience of BD, a level of tailoring was also developed where users can turn tracking factors on/off as they wish. A working prototype was subject to a series of alpha testing with two users which led to the prototype being user-orientated, such as extending time periods to answer questionnaires, trigger warnings for sensitive questionnaires, making the medication tracking more personalised. In the Evaluation Chapter (Chapter 6), a final prototype was developed and enrolled for a 6-month period of beta testing with eight users. A series of findings emerged where users collectively described a pattern between the frequency of their data and the connection to EWS; where high frequency in the data was related to EWS for mania and low frequency in the data was connected to depression. Though our statistical findings did not align with the pattern that the users perceived in the data, it nonetheless sheds light on the complexity and individual variability in the experience of BD symptoms, which was a core deliverable of this research. Users demonstrated acceptability of the tool, with indication that the tool helped to ameliorate some of the challenges that one experiences during self-tracking. However, there were also several improvements which should be considered by future research. In the final chapter of the thesis (Chapter 7), the core limitations of this doctoral work are discussed such as the technical difficulties of RADAR-Base and COVID-19 and how this impacted the research, especially during the evaluation phase of this research. Finally, a personal reflexive account of applying this methodology was provided with core insights on the researchers' role in the process. Suggestions for the community of mental health technology research were also put forward such as initiatives to upskill in empathy and safeguarding policies for researchers working in advocatory research in mental health.

References

aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*, *32*(6), 510–523. https://doi.org/10.1016/j.cpr.2012.05.007

Abdullah, S., Matthews, M., Frank, E., Doherty, G., Gay, G., & Choudhury, T. (2016a). Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association*, *23*(3), 538–543. https://doi.org/10.1093/jamia/ocv200

Abdullah, S., Matthews, M., Frank, E., Doherty, G., Gay, G., & Choudhury, T. (2016b). Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association*, *23*(3), 538–543. https://doi.org/10.1093/jamia/ocv200

Abelson, J., Wagner, F., DeJean, D., Boesveld, S., Gauvin, F.-P., Bean, S., Axler, R., Petersen, S., Baidoobonso, S., Pron, G., Giacomini, M., & Lavis, J. (2016). PUBLIC AND PATIENT INVOLVEMENT IN HEALTH TECHNOLOGY ASSESSMENT: A FRAMEWORK FOR ACTION. *International Journal of Technology Assessment in Health Care*, *32*(4), 256–264. https://doi.org/10.1017/S0266462316000362

Affairs, A. S. for P. (2013, September 6). *System Usability Scale (SUS)*. Department of Health and Human Services. https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

Alpha Test. (n.d.). Retrieved January 26, 2023, from https://www.productplan.com/glossary/alpha-test/

Altman, E. G., Hedeker, D., Peterson, J. L., & Davis, J. M. (1997). The Altman Self-Rating Mania Scale. *Biological Psychiatry*, *42*(10), 948–955. https://doi.org/10.1016/S0006-3223(96)00548-3

Altman, M., Huang, T. T. K., & Breland, J. Y. (2018). Design Thinking in Health Care. *Preventing Chronic Disease*, *15*. https://doi.org/10.5888/pcd15.180128

Alvarez-lozano, J., Frost, M., Osmani, V., Bardram, J., Kessing, L. V., & Faurholt-jepsen, M. (2014a). Tell Me Your Apps and I Will Tell You Your Mood: Correlation of Apps Usage with Bipolar Disorder State. *In ACM Proceedings 7th International Conference on Pervasive Technologies Related to Assistive Environments, Rhodes Island*.

Alvarez-lozano, J., Frost, M., Osmani, V., Bardram, J., Kessing, L. V., & Faurholt-jepsen, M. (2014b). Tell Me Your Apps and I Will Tell You Your Mood: Correlation of Apps Usage with Bipolar Disorder State. *In ACM Proceedings 7th International Conference on Pervasive Technologies Related to Assistive Environments, Rhodes Island*.

Angst, J. (2002). Historical aspects of the dichotomy between manic–depressive disorders and schizophrenia. *Schizophrenia Research*, *57*(1), 5–13. https://doi.org/10.1016/S0920-9964(02)00328-6

Angst, J., & Marneros, A. (2001). Bipolarity from ancient to modern times: Conception, birth and rebirth. *Journal of Affective Disorders*.

Anthes, E. (2016). Mental health: There's an app for that. *Nature*, *532*(7597), 20–23. https://doi.org/10.1038/532020a

Arnstein, S. R. (1969). A Ladder of Citizen Participation. *Journal of the American Institute of Planners*, *35*(4), 216–224. https://doi.org/10.1080/01944366908977225

Arnstein's Ladder of Citizen Participation. (n.d.). Retrieved September 14, 2020, from https://www.citizenshandbook.org/arnsteinsladder.html

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Association, T. A. P. (1987). Diagnostic and statistical manual for mental disorders. *DSM-IIIR* 3–24.

Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado-Terriza, J. A., Lee, S., Pomares, H., & Rojas, I. (2015). Design, implementation and validation of a novel open framework for agile development of mobile health applications. *BioMedical Engineering OnLine*, *14*(2), S6. https://doi.org/10.1186/1475-925X-14-S2-S6

Bardram, J. E., Frost, M., Szanto, K., Faurholt-Jepsen, M., Vinberg, M., & Kessing, L. V. (n.d.-a). Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the MONARCA System. *Mental Health*, 10.

Bardram, J. E., Frost, M., Szanto, K., Faurholt-Jepsen, M., Vinberg, M., & Kessing, L. V. (n.d.-b). Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the MONARCA System. *Mental Health*, 10.

Bardram, J. E., Frost, M., Szántó, K., Faurholt-Jepsen, M., Vinberg, M., & Kessing, L. V. (2013). Designing mobile health technology for bipolar disorder: A field trial of the monarca system. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2627–2636. https://doi.org/10.1145/2470654.2481364

Batra, S., Baker, R. A., Wang, T., Forma, F., DiBiasi, F., & Peters-Strickland, T. (2017). Digital health technology for use in patients with serious mental illness: A systematic review of the literature. *Medical Devices (Auckland, N.Z.)*, *10*, 237–251. https://doi.org/10.2147/MDER.S144158

Bauer, A. M., Hodsdon, S., Hunter, S., Choi, Y., Bechtel, J., & Fortney, J. C. (2017). Lessons from the deployment of the SPIRIT app to support collaborative care for rural patients with complex psychiatric conditions. *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 772–780. https://doi.org/10.1145/3123024.3125610

Bauer, A. M., Iles-Shih, M., Ghomi, R. H., Rue, T., Grover, T., Kincler, N., Miller, M., & Katon, W. J. (2018). Acceptability of mHealth augmentation of Collaborative Care: A mixed methods pilot study. *General Hospital Psychiatry*, *51*, 22–29. https://doi.org/10.1016/j.genhosppsych.2017.11.010

Baumer, E. P. S. (2015). Usees. *Proceedings of CHI*.

Beiwinkel, T., Kindermann, S., Maier, A., Kerl, C., Moock, J., Barbian, G., & Rössler, W. (2016). Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study. *JMIR Mental Health*, *3*(1), e2. https://doi.org/10.2196/mental.4560

Bipolar UK Mood Scale. (n.d.). Bipolar UK. Retrieved February 8, 2022, from https://www.bipolaruk.org/faqs/mood-scale

Bipolar UK, U. (n.d.). *Bipolar UK Mood Scale*. Bipolar UK. Retrieved February 8, 2022, from https://www.bipolaruk.org/faqs/mood-scale

Birrell, L., Furneaux-Bate, A., Debenham, J., Spallek, S., Newton, N., & Chapman, C. (2022). Development of a Peer Support Mobile App and Web-Based Lesson for Adolescent Mental Health (Mind Your Mate): User-Centered Design Approach. *JMIR Formative Research*, *6*(5), e36068. https://doi.org/10.2196/36068

Blanco, C., Grant, J., Petry, N. M., Simpson, H. B., Alegria, A., Liu, S.-M., & Hasin, D. (2008). Prevalence and Correlates of Shoplifting in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *American Journal of Psychiatry*, *165*(7), 905–913. https://doi.org/10.1176/appi.ajp.2008.07101660

Bos, F., Schreuder, M., Doornbos, B., Snippe, E., Bruggeman, R., Krieke, L. V. D., Haarman, B., Wichers, M., & George, S. (2021). Prospective early warning signals to detect transitions to manic and depressive episodes in bipolar disorder. *European Psychiatry*, *64*(S1), S78–S79. https://doi.org/10.1192/j.eurpsy.2021.237

Bowden, C. L. (1998). New Concepts in Mood Stabilization: Evidence for the Effectiveness of Valproate and Lamotrigine. *Neuropsychopharmacology*, *19*(3), Article 3. https://doi.org/10.1016/S0893-133X(98)00025-6

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, Matt., (2020). *How to Build a Participant List for UX Workshops*. Nielsen Norman Group. Retrieved March 23, 2023, from https://www.nngroup.com/articles/workshop-participant-lists/

Buitenweg, D. C., Bongers, I. L., Mheen, D. van de, Oers, H. A. M. van, & Nieuwenhuizen, C. van. (2019). Cocreative Development of the QoL-ME: A Visual and Personalized Quality of Life Assessment App for People With Severe Mental Health Problems. *JMIR Mental Health*, *6*(3), e12378. https://doi.org/10.2196/12378

Bush, N. E., Dobscha, S. K., Crumpton, R., Denneson, L. M., Hoffman, J. E., Crain, A., Cromer, R., & Kinn, J. T. (2015). A Virtual Hope Box Smartphone App as an Accessory to Therapy: Proof-of-Concept in a Clinical Sample of Veterans. *Suicide and Life-Threatening Behavior*, *45*(1), 1–9. https://doi.org/10.1111/sltb.12103

Busk, J., Faurholt-Jepsen, M., Frost, M., Bardram, J. E., Kessing, L. V., & Winther, O. (2020). Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach. *JMIR MHealth and UHealth*, *8*(4), e15028. https://doi.org/10.2196/15028

Cade, J. F. J. (1949). Lithium Salts in the Treatment of Psychotic Excitement. *Medical Journal of Australia*, *2*(10), 349–352. https://doi.org/10.5694/j.1326-5377.1949.tb36912.x

Cao, L., Mohan, K., Xu, P., Ramesh, B., Ramesh, B., & Of, B. (2009). *A framework for adapting agile development*.

Carta, M. G., Aguglia, E., Balestrieri, M., Calabrese, J. R., Caraci, F., Dell'Osso, L., Di Sciascio, G., Drago, F., Faravelli, C., Lecca, M. E., Moro, M. F., Nardini, M., Palumbo, G., & Hardoy, M. C. (2012). The lifetime prevalence of bipolar disorders and the use of antidepressant drugs in bipolar depression in Italy. *Journal of Affective Disorders*, *136*(3), 775–780. https://doi.org/10.1016/j.jad.2011.09.041

Cerit, C., Filizer, A., Tural, Ü., & Tufan, A. E. (2012). Stigma: A core factor on predicting functionality in bipolar disorder. *Comprehensive Psychiatry*, *53*(5), 484–489. https://doi.org/10.1016/j.comppsych.2011.08.010

Cheema, M. K., MacQueen, G. M., & Hassel, S. (2015). Assessing personal financial management in patients with bipolar disorder and its relation to impulsivity and response inhibition. *Cognitive Neuropsychiatry*, *20*(5), 424–437. https://doi.org/10.1080/13546805.2015.1076722

Cho, C. H., Lee, T., Kim, M. G., In, H. P., Kim, L., & Lee, H. J. (2019). Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: Prospective observational cohort study. *Journal of Medical Internet Research*, *21*(4). https://doi.org/10.2196/11029

Cho, C.-H., Lee, T., Lee, J.-B., Seo, J. Y., Jee, H.-J., Son, S., An, H., Kim, L., & Lee, H.-J. (2020). Effectiveness of a Smartphone App With a Wearable Activity Tracker in Preventing the Recurrence of Mood Disorders: Prospective Case-Control Study. *JMIR Mental Health*, *7*(8), e21283. https://doi.org/10.2196/21283

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1143–1152. https://doi.org/10.1145/2556288.2557372

Cipriani, A., Pretty, H., Hawton, K., & Geddes, J. R. (2005). Lithium in the prevention of suicidal behavior and all-cause mortality in patients with mood disorders: A systematic review of randomized trials. *The American Journal of Psychiatry*, *162*(10), 1805–1819. https://doi.org/10.1176/appi.ajp.162.10.1805

Clemente, A. S., Diniz, B. S., Nicolato, R., Kapczinski, F. P., Soares, J. C., Firmo, J. O., & Castro-Costa, É. (2015). Bipolar disorder prevalence: A systematic review and meta-analysis of the literature. *Brazilian Journal of Psychiatry*, *37*, 155–161. https://doi.org/10.1590/1516-4446-2012-1693

Colombo, C., Fossati, A., & Colom, F. (2012). Bipolar Disorder. *Depression Research and Treatment*, *2012*, 525837. https://doi.org/10.1155/2012/525837

Coryell, W., Leon, A. C., Turvey, C., Akiskal, H. S., Mueller, T., & Endicott, J. (2001). The significance of psychotic features in manic episodes: A report from the NIMH collaborative study. *Journal of Affective Disorders*, *67*(1), 79–88. https://doi.org/10.1016/S0165-0327(99)00024-5

Daylio—Journal, Diary and Mood Tracker. (n.d.). Daylio. Retrieved January 26, 2023, from https://daylio.net/

Dell'Osso, B., Camuri, G., Cremaschi, L., Dobrea, C., Buoli, M., Ketter, T. A., & Altamura, A. C. (2017). Lifetime presence of psychotic symptoms in bipolar disorder is associated with less favorable socio-demographic and certain clinical features. *Comprehensive Psychiatry*, *76*, 169–176. https://doi.org/10.1016/j.comppsych.2017.04.005

Demissie, M., Hanlon, C., Birhane, R., Ng, L., Medhin, G., & Fekadu, A. (2018). Psychological interventions for bipolar disorder in low- and middle-income countries: Systematic review. *BJPsych Open*, *4*(5), 375–384. https://doi.org/10.1192/bjo.2018.46

den Braber, M. (2016). The Emergence of Quantified Self as a Data-driven Movement to Promote Health and Wellness. *Proceedings of the First Workshop on Lifelogging Tools and Applications*, 1. https://doi.org/10.1145/2983576.2983584

Dixon-Woods, M., Booth, A., & Sutton, A. J. (2007). Synthesizing qualitative research: A review of published reports. *Qualitative Research*, *7*(3), 375–422. https://doi.org/10.1177/1468794107078517

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015). A lived informatics model of personal informatics. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 731–742. https://doi.org/10.1145/2750858.2804250

Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., & Munson, S. (2014). Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data. *Proceedings of the 2014 Conference on Designing Interactive Systems*, 667–676. https://doi.org/10.1145/2598510.2598558

Ertugrul, A., & Meltzer, H. Y. (2003). Antipsychotic drugs in bipolar disorder. *The International Journal of Neuropsychopharmacology*, *6*(3), 277–284. https://doi.org/10.1017/S1461145703003560

Evans, S., Banerjee, S., Leese, M., & Huxley, P. (2007). The impact of mental illness on quality of life: A comparison of severe mental illness, common mental disorder and healthy population samples. *Quality of Life Research*, *16*(1), 17–29. https://doi.org/10.1007/s11136-006-9002-6

Faurholt-Jepsen, M., Bauer, M., & Kessing, L. V. (2018). Smartphone-based objective monitoring in bipolar disorder: Status and considerations. *International Journal of Bipolar Disorders*, *6*(1), 6. https://doi.org/10.1186/s40345-017-0110-8

Faurholt-Jepsen, M., Vinberg, M., Frost, M., Christensen, E. M., Bardram, J. E., & Kessing, L. V. (2015). Smartphone data as an electronic biomarker of illness activity in bipolar disorder. *Bipolar Disorders*, *17*(7), 715–728. https://doi.org/10.1111/bdi.12332

Fearon, P., Kirkbride, J. B., Morgan, C., Dazzan, P., Morgan, K., Lloyd, T., Hutchinson, G., Tarrant, J., Fung, W. L. A., Holloway, J., Mallett, R., Harrison, G., Leff, J., Jones, P. B., & Murray, R. M. (2006). Incidence of schizophrenia and other psychoses in ethnic minority groups: Results from the MRC AESOP Study. *Psychological Medicine*, *36*(11), 1541–1550. https://doi.org/10.1017/S0033291706008774

Fekadu, A., Hanlon, C., Thornicroft, G., Lund, C., Kaaya, S., Alem, A., Gureje, O., Cleare, A. J., Patel, V., & Prince, M. J. (2015). Care for bipolar disorder in LMICs needs evidence from local settings. *The Lancet Psychiatry*, *2*(9), 772–773. https://doi.org/10.1016/S2215-0366(15)00222-9

Ferrari, A. J., Stockings, E., Khoo, J.-P., Erskine, H. E., Degenhardt, L., Vos, T., & Whiteford, H. A. (2016). The prevalence and burden of bipolar disorder: Findings from the Global Burden of Disease Study 2013. *Bipolar Disorders*, *18*(5), 440–450. https://doi.org/10.1111/bdi.12423

Finch, H. (2022). An introduction to the analysis of ranked response data. *Practical Assessment, Research, and Evaluation*, *27*(1). https://doi.org/10.7275/tgkh-qk47

Ford, G. (2020, November 6). Prevalence vs. Incidence: What is the difference? *Students 4 Best Evidence*. https://s4be.cochrane.org/blog/2020/11/06/prevalence-vs-incidence-what-is-the-difference/

Fortuna, K. L., DiMilia, P. R., Lohman, M. C., Bruce, M. L., Zubritsky, C. D., Halaby, M. R., Walker, R. M., Brooks, J. M., & Bartels, S. J. (2018). Feasibility, Acceptability, and Preliminary Effectiveness of a Peer-Delivered and Technology Supported Self-Management Intervention for Older Adults with Serious Mental Illness. *Psychiatric Quarterly*, *89*(2), 293–305. https://doi.org/10.1007/s11126-017-9534-7

Gadelrab, R., Simblett, S., Hook, J., Rickwood, S., Martinez, J., Johnstone, M., Flower, C., Bourne, S., Young, A., & Macritchie, K. (2022). Creating a Digital Psychoeducation Programme for bipolar disorder in the COVID-19 pandemic. *European Psychiatry*, *65*(S1), S569–S569. https://doi.org/10.1192/j.eurpsy.2022.1458

Garcia-Ceja, E., Riegler, M., Jakobsen, P., Torresen, J., Nordgreen, T., Oedegaard, K. J., & Fasmer, O. B. (2018). Motor Activity Based Classification of Depression in Unipolar and Bipolar Patients. *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 316–321. https://doi.org/10.1109/CBMS.2018.00062

Geddes, J. R., Burgess, S., Hawton, K., Jamison, K., & Goodwin, G. M. (2004). Long-term lithium therapy for bipolar disorder: Systematic review and meta-analysis of randomized controlled trials. *The American Journal of Psychiatry*, *161*(2), 217–222. https://doi.org/10.1176/appi.ajp.161.2.217

Geddes, J. R., Calabrese, J. R., & Goodwin, G. M. (2009). Lamotrigine for treatment of bipolar depression: Independent meta-analysis and meta-regression of individual patient data from five randomised trials. *The British Journal of Psychiatry*, *194*(1), 4–9. https://doi.org/10.1192/bjp.bp.107.048504

Geddes, J. R., & Miklowitz, D. J. (2013). Treatment of bipolar disorder. *The Lancet*, *381*(9878), 1672–1682. https://doi.org/10.1016/S0140-6736(13)60857-0

Gideon, J., Provost, E. M., & McInnis, M. (2016). Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2359–2363. https://doi.org/10.1109/ICASSP.2016.7472099

Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., ... & Brugha, T. S. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The lancet*, *386*(9995), 743-800.

Gold, A. K., & Sylvia, L. G. (2016). The role of sleep in bipolar disorder. *Nature and Science of Sleep*, *8*, 207–214. https://doi.org/10.2147/NSS.S85754

Goodwin, F. K., & Jamison, K. R. (2007). *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression*. Oxford University Press, USA.

Goodwin, J., Cummins, J., Behan, L., & O'Brien, S. (2016). *Development of a mental health smartphone app: Perspectives of mental health service users: Journal of Mental Health: Vol 25, No 5*. https://www.tandfonline.com/doi/full/10.3109/09638237.2015.1124392?casa_token=41ZFiXxQv-QAAAAA:fI5wZ6tdICAIayIJQYC3DmsS1oZzr6Im9NvpTTkZ9IDnZgURc79TgVd992hE4f-xQT14avA3dX7zhQ

Griswold, K. S., & Pessar, L. F. (2000). Management of Bipolar Disorder. *American Family Physician*, *62*(6), 1343–1353.

Gruber, J., Miklowitz, D. J., Harvey, A. G., Frank, E., Kupfer, D., Thase, M. E., Sachs, G. S., & Ketter, T. A. (2011). Sleep matters: Sleep functioning and course of illness in bipolar disorder. *Journal of Affective Disorders*, *134*(1), 416–420. https://doi.org/10.1016/j.jad.2011.05.016

Gruenerbl, A., Osmani, V., Bahle, G., Carrasco, J. C., Oehler, S., Mayora, O., Haring, C., & Lukowicz, P. (2014). Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. *Proceedings of the 5th Augmented Human International Conference on - AH ’14*, 1–8. https://doi.org/10.1145/2582051.2582089

Hall, C. L., Brown, S., James, M., Martin, J. L., Brown, N., Selby, K., Clarke, J., Williams, L., Sayal, K., Hollis, C., & Groom, M. J. (2019). Consensus workshops on the development of an ADHD medication management protocol using QbTest: Developing a clinical trial protocol with multidisciplinary stakeholders. *BMC Medical Research Methodology*, *19*(1), 126. https://doi.org/10.1186/s12874-019-0772-2

Hamilton, M. (1960). A RATING SCALE FOR DEPRESSION. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*(1), 56–62.

Hidalgo-Mazzei, D., Mateu, A., Reinares, M., Murru, A., del Mar Bonnín, C., Varo, C., Valentí, M., Undurraga, J., Strejilevich, S., Sánchez-Moreno, J., Vieta, E., & Colom, F. (2016). Psychoeducation in bipolar disorder with a SIMPLe smartphone application: Feasibility, acceptability and satisfaction. *Journal of Affective Disorders*, *200*, 58–66. https://doi.org/10.1016/j.jad.2016.04.042

Hidalgo-Mazzei, D., Reinares, M., Mateu, A., Juruena, M. F., Young, A. H., Pérez-Sola, V., Vieta, E., & Colom, F. (2017). Is a SIMPLe smartphone application capable of improving biological rhythms in bipolar disorder? *Journal of Affective Disorders*, *223*, 10–16. https://doi.org/10.1016/j.jad.2017.07.028

Hidalgo-Mazzei, D., Reinares, M., Mateu, A., Nikolova, V. L., Bonnín, C. del M., Samalin, L., García-Estela, A., Pérez-Solá, V., Young, A. H., Strejilevich, S., Vieta, E., & Colom, F.

(2018). OpenSIMPLe: A real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder. *Journal of Affective Disorders*, *241*, 436–445. https://doi.org/10.1016/j.jad.2018.08.048

Hlastala, S. A., Frank, E., Kowalski, J., Sherrill, J. T., Tu, X. M., Anderson, B., & Kupfer, D. J. (2000). Stressful life events, bipolar disorder, and the “kindling model.” *Journal of Abnormal Psychology*, *109*(4), 777–786. https://doi.org/10.1037/0021-843X.109.4.777

Hoddinott, P., Pollock, A., O’Cathain, A., Boyer, I., Taylor, J., MacDonald, C., Oliver, S., & Donovan, J. L. (2018). How to incorporate patient and public perspectives into the design and conduct of research. *F1000Research*, *7*, 752. https://doi.org/10.12688/f1000research.15162.1

Levey, Yael., (2016) *How to Run a Crazy Eights exercise | I Am Not My Pixels*. (n.d.). Retrieved February 18, 2022, from https://www.iamnotmypixels.com/how-to-use-crazy-8s-to-generate-design-ideas/

Inder, M. L., Crowe, M. T., Moor, S., Luty, S. E., Carter, J. D., & Joyce, P. R. (2008). “I actually don’t know who I am”: The impact of bipolar disorder on the development of self. *Psychiatry*, *71*(2), 123–133. https://doi.org/10.1521/psyc.2008.71.2.123

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., … Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Jauhar, S., & Young, A. H. (2019). Controversies in bipolar disorder; role of second-generation antipsychotic for maintenance therapy. *International Journal of Bipolar Disorders*, *7*(1), 10. https://doi.org/10.1186/s40345-019-0145-0

Jennings, H., Slade, M., Bates, P., Munday, E., & Toney, R. (2018). Best practice framework for Patient and Public Involvement (PPI) in collaborative data analysis of qualitative mental health research: Methodology development and refinement. *BMC Psychiatry*, *18*(1), 213. https://doi.org/10.1186/s12888-018-1794-8

J Bauer, A. M., Hodsdon, S., Bechtel, J. M., & Fortney, J. C. (2018). Applying the principles for digital development: case study of a smartphone app to support collaborative care for rural patients with posttraumatic stress disorder or bipolar disorder. *Journal of Medical Internet Research*, *20*(6), e10048.

Johansson, C., & Werbart, A. (2020). Am I Really Bipolar? Personal Accounts of the Experience of Being Diagnosed With Bipolar II Disorder. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.482715

Johnson, S. L., & Roberts, J. E. (1995). Life events and bipolar disorder: Implications from biological theories. *Psychological Bulletin*, *117*, 434–449. https://doi.org/10.1037/0033-2909.117.3.434

Judd, L. L., & Akiskal, H. S. (2003). Depressive episodes and symptoms dominate the longitudinal course of bipolar disorder. *Current Psychiatry Reports*, *5*(6), 417–418. https://doi.org/10.1007/s11920-003-0077-2

Judd, L. L., Schettler, P. J., Akiskal, H. S., Maser, J., Coryell, W., Solomon, D., Endicott, J., & Keller, M. (2003). Long-term symptomatic status of bipolar I vs. Bipolar II disorders. *The International Journal of Neuropsychopharmacology*, *6*(2), 127–137. https://doi.org/10.1017/S1461145703003341

Kahlbaum, K. (2013). *Die Katatonie*. BoD – Books on Demand.

Kaplan, K. A., McGlinchey, E. L., Soehner, A., Gershon, A., Talbot, L. S., Eidelman, P., Gruber, J., & Harvey, A. G. (2015). Hypersomnia Subtypes, Sleep and Relapse in Bipolar Disorder. *Psychological Medicine*, *45*(8), 1751. https://doi.org/10.1017/S0033291714002918

Kara, P. (n.d.). *Affinity Diagramming: Collaboratively Sort UX Findings & Design Ideas*. Nielsen Norman Group. Retrieved February 15, 2022, from https://www.nngroup.com/articles/affinity-diagram/

Karam, Z. N., Provost, E. M., Singh, S., Montgomery, J., Archer, C., Harrington, G., & Mcinnis, M. G. (2014a). ECOLOGICALLY VALID LONG-TERM MOOD MONITORING OF INDIVIDUALS WITH BIPOLAR DISORDER USING SPEECH. *Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing / Sponsored by the Institute of Electrical and Electronics Engineers Signal Processing Society. ICASSP (Conference)*, *2014*, 4858–4862. https://doi.org/10.1109/ICASSP.2014.6854525

Karam, Z. N., Provost, E. M., Singh, S., Montgomery, J., Archer, C., Harrington, G., & Mcinnis, M. G. (2014b). ECOLOGICALLY VALID LONG-TERM MOOD MONITORING OF INDIVIDUALS WITH BIPOLAR DISORDER USING SPEECH. *Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing / Sponsored by the Institute of Electrical and Electronics Engineers Signal Processing Society. ICASSP (Conference)*, *2014*, 4858–4862. https://doi.org/10.1109/ICASSP.2014.6854525

Karanti, A., Kardell, M., Joas, E., Runeson, B., Pålsson, E., & Landén, M. (2020). Characteristics of bipolar I and II disorder: A study of 8766 individuals. *Bipolar Disorders*, *22*(4), 392–400. https://doi.org/10.1111/bdi.12867

Kebede, D., Alem, A., Shibire, T., Deyassa, N., Negash, A., Beyero, T., Medhin, G., & Fekadu, A. (2006). Symptomatic and functional outcome of bipolar disorder in Butajira, Ethiopia. *Journal of Affective Disorders*, *90*(2), 239–249. https://doi.org/10.1016/j.jad.2005.11.009

Keck, P. E., McElroy, S. L., Havens, J. R., Altshuler, L. L., Nolen, W. A., Frye, M. A., Suppes, T., Denicoff, K. D., Kupka, R., Leverich, G. S., Rush, A. J., & Post, R. M. (2003). Psychosis in bipolar disorder: Phenomenology and impact on morbidity and course of illness. *Comprehensive Psychiatry*, *44*(4), 263–269. https://doi.org/10.1016/S0010-440X(03)00089-0

Kendall, T., Morriss, R., Mayo-Wilson, E., Marcus, E., & on behalf of the Guideline Development Group. (2014). Assessment and management of bipolar disorder: Summary of updated NICE guidance. *BMJ*, *349*(sep25 5), g5673–g5673. https://doi.org/10.1136/bmj.g5673

Ketter, T. A. (2010). Diagnostic Features, Prevalence, and Impact of Bipolar Disorder. *The Journal of Clinical Psychiatry*, *71*(6), 27652. https://doi.org/10.4088/JCP.8125tx11c

Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, *7*(1), 8. https://doi.org/10.1186/s13750-018-0115-5

Kroon, J. S., Wohlfarth, T. D., Dieleman, J., Sutterland, A. L., Storosum, J. G., Denys, D., de Haan, L., & Sturkenboom, M. C. (2013). Incidence rates and risk factors of bipolar disorder in the general population: A population-based cohort study. *Bipolar Disorders*, *15*(3), 306–313. https://doi.org/10.1111/bdi.12058

Lagan, S., Ramakrishnan, A., Lamont, E., Ramakrishnan, A., Frye, M., & Torous, J. (2020). Digital health developments and drawbacks: A review and analysis of top-returned apps for bipolar disorder. *International Journal of Bipolar Disorders*, *8*(1), 39. https://doi.org/10.1186/s40345-020-00202-4

Larson, R., & Csikszentmihalyi, M. (2014). The Experience Sampling Method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2

Lenaert, B., Colombi, M., van Heugten, C., Rasquin, S., Kasanova, Z., & 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. https://doi.org/10.1080/09602011.2017.1330214

Lewis, M., Palmer, V. J., Kotevski, A., Densley, K., O’Donnell, M. L., Johnson, C., Wohlgezogen, F., Gray, K., Robins-Browne, K., & Burchill, L. (2021). Rapid Design and Delivery of an Experience-Based Co-designed Mobile App to Support the Mental Health Needs of Health Care Workers Affected by the COVID-19 Pandemic: Impact Evaluation Protocol. *JMIR Research Protocols*, *10*(3), e26168. https://doi.org/10.2196/26168

Li, H., Mukherjee, D., Krishnamurthy, V. B., Millett, C., Ryan, K. A., Zhang, L., Saunders, E. F. H., & Wang, M. (2019). Use of ecological momentary assessment to detect variability in mood, sleep and stress in bipolar disorder. *BMC Research Notes*, *12*(1), 791. https://doi.org/10.1186/s13104-019-4834-7

Li, I., Dey, A., Forlizzi, J., Höök, K., & Medynskiy, Y. (2011). Personal informatics and HCI: Design, theory, and social implications. *CHI ’11 Extended Abstracts on Human Factors in Computing Systems*, 2417–2420. https://doi.org/10.1145/1979742.1979573

Li, I., Dey, A. K., & Forlizzi, J. (2011). Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 405–414. https://doi.org/10.1145/2030112.2030166

British National Formulary *Lithium carbonate | Drugs | BNF content published by NICE*. (n.d.). Retrieved February 18, 2023, from https://bnf.nice.org.uk/drugs/lithium-carbonate/

Liverpool, S., Webber, H., Matthews, R., Wolpert, M., & Edbrooke-Childs, J. (2019). A Mobile App to Support Parents Making Child Mental Health Decisions: Protocol for a Feasibility Cluster Randomized Controlled Trial. *JMIR Research Protocols*, *8*(8), e14571. https://doi.org/10.2196/14571

Lloyd, T., Kennedy, N., Fearon, P., Kirkbride, J., Mallett, R., Leff, J., Holloway, J., Harrison, G., Dazzan, P., Morgan, K., Murray, R. M., & Jones, P. B. (2005). Incidence of bipolar affective disorder in three UK cities: Results from the ÆSOP study. *British Journal of Psychiatry*, *186*(2), 126–131. https://doi.org/10.1192/bjp.186.2.126

Lobban, F., Solis-Trapala, I., Symes, W., Morriss, R., & ERP Group, University of Liverpool. (2011). Early warning signs checklists for relapse in bipolar depression and mania: Utility, reliability and validity. *Journal of Affective Disorders*, *133*(3), 413–422. https://doi.org/10.1016/j.jad.2011.04.026

López Zurbano, S., González-Pinto Arrillaga, A., & López-Peña, P. (2015). Gender differences in bipolar disorder. *Psychopathology in Women: Incorporating Gender Perspective into Descriptive Psychopathology, 2015, ISBN 978-3-319-05869-6, Págs. 641-659*, 641–659. https://dialnet.unirioja.es/servlet/articulo?codigo=5355943

Luo, Y. (2021). Designing Multimodal Self-Tracking Technologies to Promote Data Capture and Self-Reflection. *Companion Publication of the 2021 ACM Designing Interactive Systems Conference*, 11–15. https://doi.org/10.1145/3468002.3468232

Majid, S., Reeves, S., Figueredo, G., Brown, S., Lang, A., Moore, M., & Morriss, R. (2021). The Extent of User Involvement in the Design of Self-tracking Technology for Bipolar Disorder: Literature Review. *JMIR Mental Health*, *8*(12), e27991. https://doi.org/10.2196/27991

Mantere, O., Suominen, K., Leppämäki, S., Valtonen, H., Arvilommi, P., & Isometsä, E. (2004). The clinical characteristics of DSM-IV bipolar I and II disorders: Baseline findings from the Jorvi Bipolar Study (JoBS). *Bipolar Disorders*, *6*(5), 395–405. https://doi.org/10.1111/j.1399-5618.2004.00140.x

Martin, A., & Lynch, M. (2009). Counting Things and People: The Practices and Politics of Counting. *Social Problems*, *56*(2), 243–266. https://doi.org/10.1525/sp.2009.56.2.243

Martin, E. (2009). *Bipolar Expeditions: Mania and Depression in American Culture*. Princeton University Press.

Marwaha, S., Durrani, A., & Singh, S. (2013). Employment outcomes in people with bipolar disorder: A systematic review. *Acta Psychiatrica Scandinavica*, *128*(3), 179–193. https://doi.org/10.1111/acps.12087

Mason, B. L., Brown, E. S., & Croarkin, P. E. (2016). Historical Underpinnings of Bipolar Disorder Diagnostic Criteria. *Behavioral Sciences*, *6*(3). https://doi.org/10.3390/bs6030014

Matthews, M., Abdullah, S., Murnane, E., Voida, S., Choudhury, T., Gay, G., & Frank, E. (2016). Development and Evaluation of a Smartphone-Based Measure of Social Rhythms for Bipolar Disorder. *Assessment*, *23*(4), 472–483. https://doi.org/10.1177/1073191116656794

Matthews, M., Murnane, E., & Snyder, J. (2017a). Quantifying the Changeable Self: The Role of Self-Tracking in Coming to Terms With and Managing Bipolar Disorder. *Human–*

Computer Interaction, *32*(5–6), 413–446. https://doi.org/10.1080/07370024.2017.1294983

Matthews, M., Murnane, E., & Snyder, J. (2017b). Quantifying the Changeable Self: The Role of Self-Tracking in Coming to Terms With and Managing Bipolar Disorder. *Human–Computer Interaction*, *32*(5–6), 413–446. https://doi.org/10.1080/07370024.2017.1294983

Matthews, M., Murnane, E., Snyder, J., Guha, S., Chang, P., Doherty, G., & Gay, G. (2017a). The double-edged sword: A mixed methods study of the interplay between bipolar disorder and technology use. *Computers in Human Behavior*, *75*, 288–300. https://doi.org/10.1016/j.chb.2017.05.009

Matthews, M., Voida, S., Abdullah, S., Doherty, G., Choudhury, T., Im, S., & Gay, G. (2015). In situ design for mental illness: 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2015. *MobileHCI 2015 - Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 86–97. https://doi.org/10.1145/2785830.2785866

Maxhuni, A., Muñoz-Meléndez, A., Osmani, V., Perez, H., Mayora, O., & Morales, E. F. (2016). Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive and Mobile Computing*, *31*, 50–66. https://doi.org/10.1016/j.pmcj.2016.01.008

McCurdie, T., Taneva, S., Casselman, M., Yeung, M., McDaniel, C., Ho, W., & Cafazzo, J. (2012). mHealth Consumer Apps: The Case for User-Centered Design. *Biomedical Instrumentation & Technology*, *46*(s2), 49–56. https://doi.org/10.2345/0899-8205-46.s2.49

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M. A., Petukhova, M., & Kessler, R. C. (2007). Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *64*(5), 543–552. https://doi.org/10.1001/archpsyc.64.5.543

Miklowitz, D. J., Goodwin, G. M., Bauer, M. S., & Geddes, J. R. (2008). Common and Specific Elements of Psychosocial Treatments for Bipolar Disorder: A Survey of Clinicians Participating in Randomized Trials. *Journal of Psychiatric Practice®*, *14*(2), 77–85. https://doi.org/10.1097/01.pra.0000314314.94791.c9

Moran., Kate (2019) *Usability Testing 101*. Nielsen Norman Group. Retrieved January 26, 2023, from https://www.nngroup.com/articles/usability-testing-101/

Morgan, A. J., Reavley, N. J., Ross, A., Too, L. S., & Jorm, A. F. (2018). Interventions to reduce stigma towards people with severe mental illness: Systematic review and meta-analysis. *Journal of Psychiatric Research*, *103*, 120–133. https://doi.org/10.1016/j.jpsychires.2018.05.017

Morriss, R. (2004). The early warning symptom intervention for patients with bipolar affective disorder. *Advances in Psychiatric Treatment*, *10*(1), 18–26. https://doi.org/10.1192/apt.10.1.18

Morriss, R., Faizal, M. A., Jones, A. P., Williamson, P. R., Bolton, C. A., & McCarthy, J. P. (2007). Interventions for helping people recognise early signs of recurrence in bipolar disorder. *Cochrane Database of Systematic Reviews*, *1*. https://doi.org/10.1002/14651858.CD004854.pub2

Morriss, R., Lobban, F., Riste, L., Davies, L., Holland, F., Long, R., Lykomitrou, G., Peters, S., Roberts, C., Robinson, H., & Jones, S. (2016). Clinical effectiveness and acceptability of structured group psychoeducation versus optimised unstructured peer support for patients with remitted bipolar disorder (PARADES): A pragmatic, multicentre, observer-blind, randomised controlled superiority trial. *The Lancet Psychiatry*, *3*(11), 1029–1038. https://doi.org/10.1016/S2215-0366(16)30302-9

Muaremi, A., Gravenhorst, F., Grunerbl, A., Arnrich, B., & Troester, G. (2014). Assessing Bipolar Episodes Using Speech Cues Derived from Phone Calls. *Pervasive Computing Paradigms for Mental Health*. https://doi.org/10.1007/978-3-319-11564-1

Murnane, E. L., Cosley, D., Chang, P., Guha, S., Frank, E., Gay, G., & Matthews, M. (2016). Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: Implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association*, *23*(3), 477–484. https://doi.org/10.1093/jamia/ocv165

Murray, E., Hekler, E. B., Andersson, G., Collins, L. M., Doherty, A., Hollis, C., Rivera, D. E., West, R., & Wyatt, J. C. (2016). Evaluating Digital Health Interventions. *American Journal of Preventive Medicine*, *51*(5), 843–851. https://doi.org/10.1016/j.amepre.2016.06.008

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. https://doi.org/10.1002/wps.20513

N, P., Kea, S., O, C., Jr, G., Gm, G., & M, D. V. (2018). Group-Personalized Regression Models for Predicting Mental Health Scores From Objective Mobile Phone Data Streams: Observational Study. *Journal of Medical Internet Research*, *20*(10). https://doi.org/10.2196/10194

Naslund, J. A., Marsch, L. A., McHugo, G. J., & Bartels, S. J. (2015). Emerging mHealth and eHealth interventions for serious mental illness: A review of the literature. *Journal of Mental Health*, *24*(5), 321–332. https://doi.org/10.3109/09638237.2015.1019054

National Institute of Health Research. (2012). *Briefing notes for researchers: Public involvement in NHS, public health and social care research*.

National Institute of Health Research. (2014, September 24). *Overview | Bipolar disorder: Assessment and management | Guidance | NICE*. NICE. https://www.nice.org.uk/guidance/cg185

Nicholas, J., Fogarty, A. S., Boydell, K., & Christensen, H. (2017a). The Reviews Are in: A Qualitative Content Analysis of Consumer Perspectives on Apps for Bipolar Disorder. *Journal of Medical Internet Research*, *19*(4), e105. https://doi.org/10.2196/jmir.7273

Nicholas, J., Fogarty, A. S., Boydell, K., & Christensen, H. (2017b). The Reviews Are in: A Qualitative Content Analysis of Consumer Perspectives on Apps for Bipolar Disorder. *Journal of Medical Internet Research*, *19*(4), e105. https://doi.org/10.2196/jmir.7273

Niendam, T. A., Tully, L. M., Iosif, A.-M., Kumar, D., Nye, K. E., Denton, J. C., Zakskorn, L. N., Fedechko, T. L., & Pierce, K. M. (2018). Enhancing early psychosis treatment using smartphone technology: A longitudinal feasibility and validity study. *Journal of Psychiatric Research*, *96*, 239–246. https://doi.org/10.1016/j.jpsychires.2017.10.017

Niess, J., & Woźniak, P. W. (2018). Supporting Meaningful Personal Fitness: The Tracker Goal Evolution Model. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3173745

Orlowski, S., Matthews, B., Bidargaddi, N., Jones, G., Lawn, S., Venning, A., & Collin, P. (2016). Mental Health Technologies: Designing With Consumers. *JMIR Human Factors*, *3*(1), e4. https://doi.org/10.2196/humanfactors.4336

Ortiz, A., Maslej, M. M., Husain, M. I., Daskalakis, Z. J., & Mulsant, B. H. (2021). Apps and gaps in bipolar disorder: A systematic review on electronic monitoring for episode prediction. *Journal of Affective Disorders*, *295*, 1190–1200. https://doi.org/10.1016/j.jad.2021.08.140

Osmani, V., Gruenerbl, A., Bahle, G., Haring, C., Lukowicz, P., & Mayora, O. (2015a). Smartphones in Mental Health: Detecting Depressive and Manic Episodes. *IEEE Pervasive Computing*, *14*(3), 10–13. https://doi.org/10.1109/MPRV.2015.54

Osmani, V., Gruenerbl, A., Bahle, G., Haring, C., Lukowicz, P., & Mayora, O. (2015b). Smartphones in Mental Health: Detecting Depressive and Manic Episodes. *IEEE Pervasive Computing*, *14*(3), 10–13. https://doi.org/10.1109/MPRV.2015.54

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, *49*(1), 47–53. https://doi.org/10.1016/j.ijnurstu.2011.07.002

Palmius, N., Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Geddes, J. R., Goodwin, G. M., & De Vos, M. (2017). Detecting Bipolar Depression From Geographic Location Data. *IEEE Transactions on Biomedical Engineering*, *64*(8), 1761–1771. https://doi.org/10.1109/TBME.2016.2611862

Pavlickova, H., Varese, F., Smith, A., Myin-Germeys, I., Turnbull, O. H., Emsley, R., & Bentall, R. P. (2013). The Dynamics of Mood and Coping in Bipolar Disorder: Longitudinal Investigations of the Inter-Relationship between Affect, Self-Esteem and Response Styles. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0062514

Polhemus, A., Novak, J., Majid, S., Simblett, S., Bruce, S., Burke, P., Dockendorf, M. F., Temesi, G., & Wykes, T. (2020). *Data visualization in chronic neurological and mental health* *condition self-management: A systematic review of user perspectives (Preprint)*. https://doi.org/10.2196/preprints.25249

Post, R. M., Denicoff, K. D., Leverich, G. S., Altshuler, L. L., Frye, M. A., Suppes, T. M., Rush, A. J., Jr, P. E. K., McElroy, S. L., Luckenbaugh, D. A., Pollio, C., Kupka, R., & Nolen, W. A. (2003). Morbidity in 258 Bipolar Outpatients Followed for 1 Year With Daily Prospective Ratings on the NIMH Life Chart Method. *The Journal of Clinical Psychiatry*, *64*(6), 12282.

Pratap, A., Atkins, D. C., Renn, B. N., Tanana, M. J., Mooney, S. D., Anguera, J. A., & Areán, P. A. (2019). The accuracy of passive phone sensors in predicting daily mood. *Depression and Anxiety*, *36*(1), 72–81. https://doi.org/10.1002/da.22822

Office of Health and Improvement Disparities (2022) Premature mortality in adults with severe mental illness (SMI). (n.d.). GOV.UK. Retrieved January 23, 2023, from https://www.gov.uk/government/publications/premature-mortality-in-adults-with-severe-mental-illness/premature-mortality-in-adults-with-severe-mental-illness-smi

Price, M., Yuen, E. K., Goetter, E. M., Herbert, J. D., Forman, E. M., Acierno, R., & Ruggiero, K. J. (2014). mHealth: A Mechanism to Deliver More Accessible, More Effective Mental Health Care. *Clinical Psychology & Psychotherapy*, *21*(5), 427. https://doi.org/10.1002/cpp.1855

Miklowitz, D. J., & Scott, J. (2009). Psychosocial treatments for bipolar disorder: Cost-effectiveness, mediating mechanisms, and future directions. *Bipolar disorders*, *11*, 110-122.

Puiatti, A., Mudda, S., Giordano, S., & Mayora, O. (2011). Smartphone-centred wearable sensors network for monitoring patients with bipolar disorder. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3644–3647. https://doi.org/10.1109/IEMBS.2011.6090613

Rabelo, J. L., Cruz, B. F., Ferreira, J. D. R., Viana, B. de M., & Barbosa, I. G. (2021). Psychoeducation in bipolar disorder: A systematic review. *World Journal of Psychiatry*, *11*(12), 1407–1424. https://doi.org/10.5498/wjp.v11.i12.1407

Ranjan, Y., Rashid, Z., Stewart, C., Conde, P., Begale, M., Verbeeck, D., Boettcher, S., The Hyve, Dobson, R., Folarin, A., & The RADAR-CNS Consortium. (2019). RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices. *JMIR MHealth and UHealth*, *7*(8), e11734. https://doi.org/10.2196/11734

Rapp, A., & Cena, F. (2016). Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, *94*, 1–17. https://doi.org/10.1016/j.ijhcs.2016.05.006

Rapp, A., & Tirassa, M. (2017). Know Thyself: A Theory of the Self for Personal Informatics. *Human–Computer Interaction*, *32*(5–6), 335–380. https://doi.org/10.1080/07370024.2017.1285704

Rickard, N., Arjmand, H.-A., Bakker, D., & Seabrook, E. (2016). Development of a Mobile Phone App to Support Self-Monitoring of Emotional Well-Being: A Mental Health Digital Innovation. *JMIR Mental Health*, *3*(4), e6202. https://doi.org/10.2196/mental.6202

Robillard, R., Naismith, S. L., & Hickie, I. B. (2013). Recent Advances in Sleep-Wake Cycle and Biological Rhythms in Bipolar Disorder. *Current Psychiatry Reports*, *15*(10), 402. https://doi.org/10.1007/s11920-013-0402-3

Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. (2014). Personal tracking as lived informatics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1163–1172. https://doi.org/10.1145/2556288.2557039

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Sanches, P., Höök, K., Sas, C., Janson, A., Karpashevich, P., Nadal, C., Qu, C., Daudén Roquet, C., Umair, M., Windlin, C., & Doherty, G. (2019). HCI and Affective Health: Taking stock of a decade of studies and charting future research directions. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–17. https://doi.org/10.1145/3290605.3300475

Schärer, L. O., Krienke, U. J., Graf, S.-M., Meltzer, K., & Langosch, J. M. (2015). Validation of life-charts documented with the personal life-chart app – a self-monitoring tool for bipolar disorder. *BMC Psychiatry*, *15*(1), 49. https://doi.org/10.1186/s12888-015-0414-0

Schuurmans, J., van der Linden, J. L., van Ballegooijen, W., Ruwaard, J., Stek, M. L., Smit, J. H., & Riper, H. (2016). Tablet-based support for older adults with severe mood disorders treated in an ambulatory geriatric psychiatry setting: Protocol of a feasibility study of the eCare@Home platform. *Internet Interventions*, *6*, 22–28. https://doi.org/10.1016/j.invent.2016.09.001

Scott, D., & Happell, B. (2011). The High Prevalence of Poor Physical Health and Unhealthy Lifestyle Behaviours in Individuals with Severe Mental Illness. *Issues in Mental Health Nursing*, *32*, 589–597. https://doi.org/10.3109/01612840.2011.569846

Shore, J. H., Aldag, M., McVeigh, F. L., Hoover, R. L., Ciulla, R., & Fisher, A. (2014). Review of Mobile Health Technology for Military Mental Health. *Military Medicine*, *179*(8), 865–878. https://doi.org/10.7205/MILMED-D-13-00429

Simon, J., Pari, A. A. A., Wolstenholme, J., Berger, M., Goodwin, G. M., & Geddes, J. R. (2021). The costs of bipolar disorder in the United Kingdom. *Brain and Behavior*, *11*(10), e2351. https://doi.org/10.1002/brb3.2351

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A Scale for the Assessment of Hedonic Tone the Snaith–Hamilton Pleasure Scale. *The British Journal of Psychiatry*, *167*(1), 99–103. https://doi.org/10.1192/bjp.167.1.99

Spiel, K., Kayali, F., Harrer, S., Horvath, L., Sicart, M., Penkler, M., & Hammer, J. (2018). *Fitter, Happier, More Productive?The Normative Ontology of Fitness Trackers*. 10.

Spinuzzi, C. (2005). The Methodology of Participatory Design. *Technical Communication*, *52*(2), 163–174.

Stange, J. P., Zulueta, J., Langenecker, S. A., Ryan, K. A., Piscitello, A., Duffecy, J., McInnis, M. G., Nelson, P., Ajilore, O., & Leow, A. (2018). Let Your Fingers Do the Talking: Passive Typing Instability Predicts Future Mood Outcomes. *Bipolar Disorders*, *20*(3), 285. https://doi.org/10.1111/bdi.12637

Steele, A., Maruyama, N., & Galynker, I. (2009). Psychiatric symptoms in caregivers of patients with bipolar disorder: A review. *Journal of Affective Disorders*, *121*, 10–21. https://doi.org/10.1016/j.jad.2009.04.020

Stewart, C. L., Rashid, Z., Ranjan, Y., Sun, S., Dobson, R. J. B., & Folarin, A. A. (2018). RADAR-base: Major Depressive Disorder and Epilepsy Case Studies. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and*

Ubiquitous Computing and Wearable Computers - UbiComp '18, 1735–1743. https://doi.org/10.1145/3267305.3267540

Stewart, V., Judd, C., & Wheeler, A. J. (2022). Practitioners' experiences of deteriorating personal hygiene standards in people living with depression in Australia: A qualitative study. *Health & Social Care in the Community*, *30*(4), 1589–1598. https://doi.org/10.1111/hsc.13491

Talevi, D., Socci, V., Carai, M., Carnaghi, G., Faleri, S., Trebbi, E., Bernardo, A. di, Capelli, F., & Pacitti, F. (2020). Mental health outcomes of the CoViD-19 pandemic. *Rivista di Psichiatria*, *55*(3), 137–144.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606-613.

Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Palmius, N., Osipov, M., Clifford, G. D., Goodwin, G. M., & De Vos, M. (2016a). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders*, *205*, 225–233. https://doi.org/10.1016/j.jad.2016.06.065

Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Palmius, N., Osipov, M., Clifford, G. D., Goodwin, G. M., & De Vos, M. (2016b). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders*, *205*, 225–233. https://doi.org/10.1016/j.jad.2016.06.065

Vanello, N., Guidi, A., Gentili, C., Werner, S., Bertschy, G., Valenza, G., Lanatá, A., & Scilingo, E. P. (2012). Speech analysis for mood state characterization in bipolar patients. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2104–2107. https://doi.org/10.1109/EMBC.2012.6346375

Vázquez, G. H., Kapczinski, F., Magalhaes, P. V., Córdoba, R., Lopez Jaramillo, C., Rosa, A. R., Sanchez de Carmona, M., & Tohen, M. (2011). Stigma and functioning in patients with bipolar disorder. *Journal of Affective Disorders*, *130*(1), 323–327. https://doi.org/10.1016/j.jad.2010.10.012

Vazquez-Montes, M. D. L. A., Stevens, R., Perera, R., Saunders, K., & Geddes, J. R. (2018). Control charts for monitoring mood stability as a predictor of severe episodes in patients with bipolar disorder. *International Journal of Bipolar Disorders*, *6*(1), 7. https://doi.org/10.1186/s40345-017-0116-2

Faurholt-Jepsen, M., Busk, J., Frost, M., Vinberg, M., Christensen, E. M., Winther, O., ... & Kessing, L. V. (2016). Voice analysis as an objective state marker in bipolar disorder. *Translational psychiatry*, *6*(7), e856-e856.

Mind (2021) What has the impact of the pandemic been on mental health? (n.d.). Retrieved January 31, 2023, from https://www.mind.org.uk/coronavirus-we-are-here-for-you/coronavirus-research/

Wykes, T., Lipshitz, J., & Schueller, S. M. (2019). Towards the Design of Ethical Standards Related to Digital Mental Health and all Its Applications. *Current Treatment Options in Psychiatry*, *6*(3), 232–242. https://doi.org/10.1007/s40501-019-00180-0

Yee, C. S., Vázquez, G. H., Hawken, E. R., Biorac, A., Tondo, L., & Baldessarini, R. J. (2021). Long-Term Treatment of Bipolar Disorder with Valproate: Updated Systematic Review and Meta-analyses. *Harvard Review of Psychiatry*, *29*(3), 188–195. https://doi.org/10.1097/HRP.0000000000000292

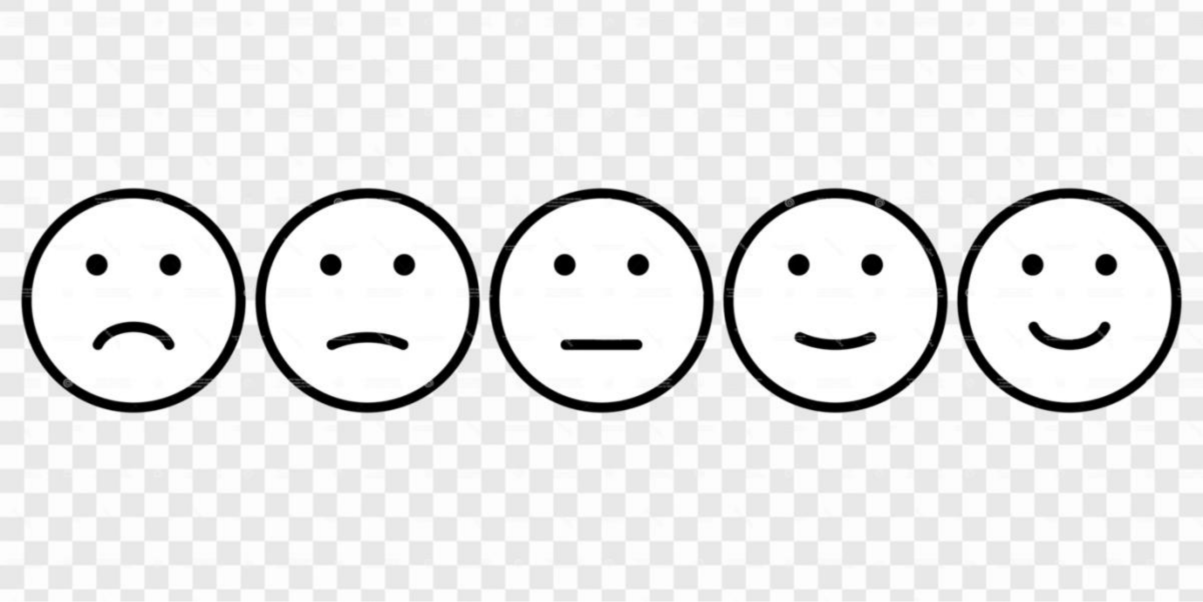
Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (2000). Young mania rating scale. *Journal of Affective Disorders*.

Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S. A., McInnis, M., Ajilore, O., Nelson, P. C., Ryan, K., & Leow, A. (2018). Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study. *Journal of Medical Internet Research*, *20*(7), e241. https://doi.org/10.2196/jmir.9775

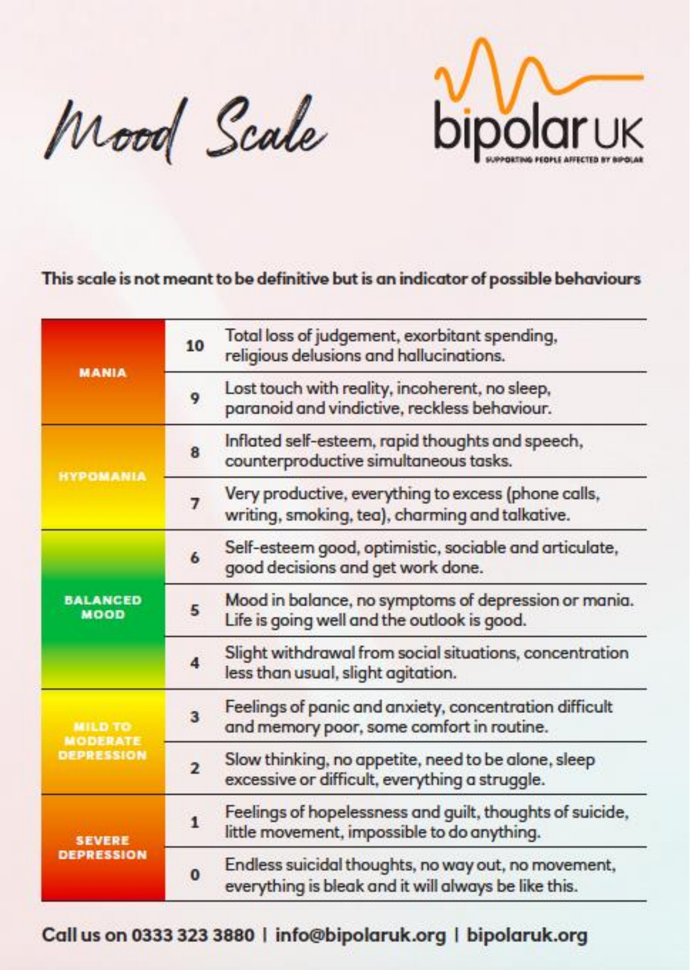
Appendices

1. Full scales/graphs used in the study

1.1 Visual mood diary using smiley faces (e.g. Daylio app (*Daylio - Journal, Diary and Mood Tracker*, n.d.)



1.2. Bipolar UK Scale – a 10-item mood scale with descriptions of mood (Bipolar UK, n.d.)



Mood Scale

bipolar UK

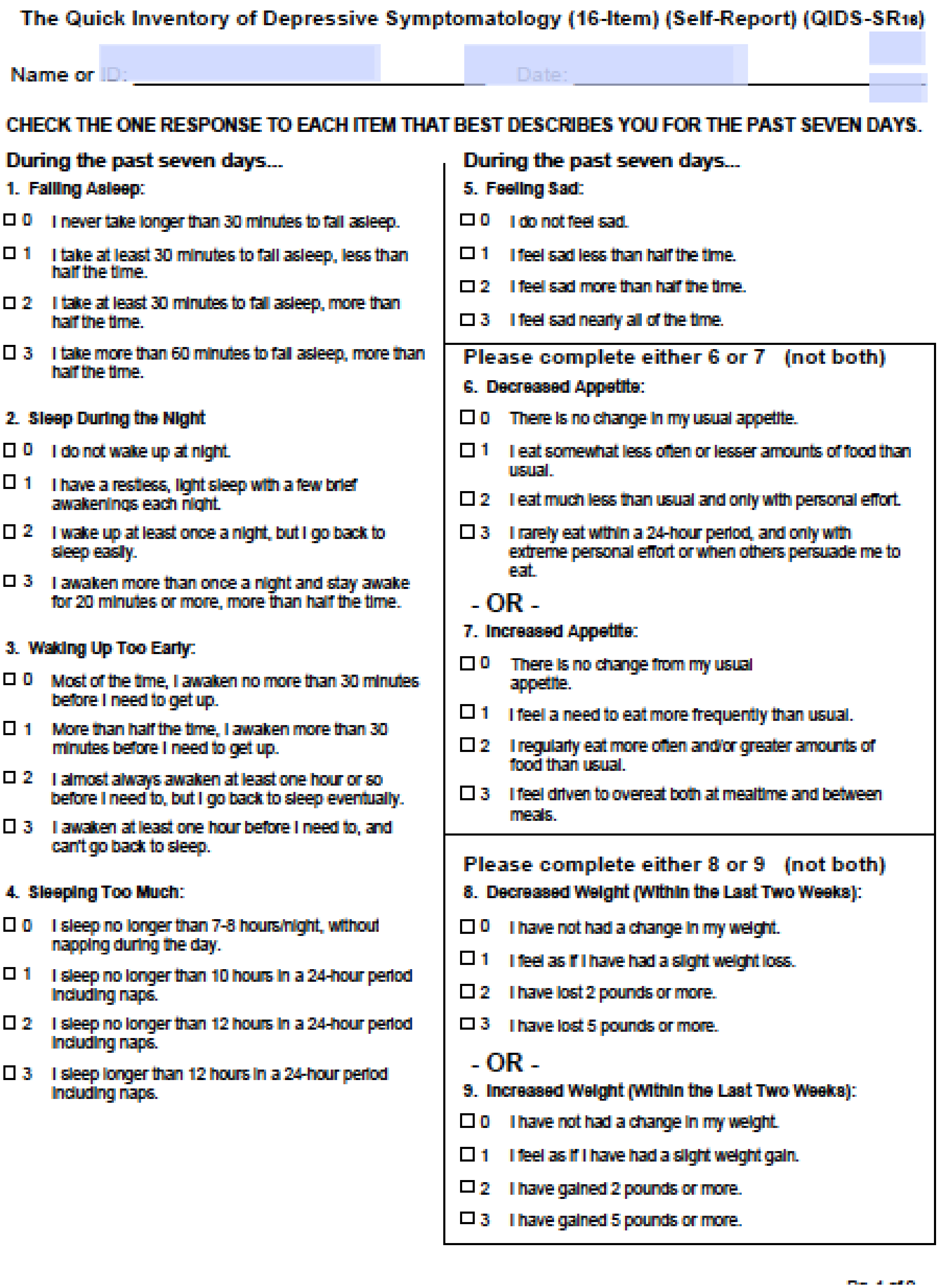
SUPPORTING PEOPLE AFFECTED BY BIPOLAR

This scale is not meant to be definitive but is an indicator of possible behaviours

MANIA	**10**	Total loss of judgement, exorbitant spending, religious delusions and hallucinations.
	9	Lost touch with reality, incoherent, no sleep, paranoid and vindictive, reckless behaviour.
HYPOMANIA	**8**	Inflated self-esteem, rapid thoughts and speech, counterproductive simultaneous tasks.
	7	Very productive, everything to excess (phone calls, writing, smoking, tea), charming and talkative.
BALANCED MOOD	**6**	Self-esteem good, optimistic, sociable and articulate, good decisions and get work done.
	5	Mood in balance, no symptoms of depression or mania. Life is going well and the outlook is good.
	4	Slight withdrawal from social situations, concentration less than usual, slight agitation.
MILD TO MODERATE DEPRESSION	**3**	Feelings of panic and anxiety, concentration difficult and memory poor, some comfort in routine.
	2	Slow thinking, no appetite, need to be alone, sleep excessive or difficult, everything a struggle.
SEVERE DEPRESSION	**1**	Feelings of hopelessness and guilt, thoughts of suicide, little movement, impossible to do anything.
	0	Endless suicidal thoughts, no way out, no movement, everything is bleak and it will always be like this.

Call us on 0333 323 3880 | info@bipolaruk.org | bipolaruk.org

1.3. Quick Inventory of Depressive Symptomology (QIDS-SR)



The Quick Inventory of Depressive Symptomatology (16-Item) (Self-Report) (QIDS-SR$_{16}$)

Name or ID: ______________________ Date: ______________________

CHECK THE ONE RESPONSE TO EACH ITEM THAT BEST DESCRIBES YOU FOR THE PAST SEVEN DAYS.

During the past seven days…

1. Falling Asleep:

☐ 0 I never take longer than 30 minutes to fall asleep.
☐ 1 I take at least 30 minutes to fall asleep, less than half the time.
☐ 2 I take at least 30 minutes to fall asleep, more than half the time.
☐ 3 I take more than 60 minutes to fall asleep, more than half the time.

2. Sleep During the Night

☐ 0 I do not wake up at night.
☐ 1 I have a restless, light sleep with a few brief awakenings each night.
☐ 2 I wake up at least once a night, but I go back to sleep easily.
☐ 3 I awaken more than once a night and stay awake for 20 minutes or more, more than half the time.

3. Waking Up Too Early:

☐ 0 Most of the time, I awaken no more than 30 minutes before I need to get up.
☐ 1 More than half the time, I awaken more than 30 minutes before I need to get up.
☐ 2 I almost always awaken at least one hour or so before I need to, but I go back to sleep eventually.
☐ 3 I awaken at least one hour before I need to, and can't go back to sleep.

4. Sleeping Too Much:

☐ 0 I sleep no longer than 7-8 hours/night, without napping during the day.
☐ 1 I sleep no longer than 10 hours in a 24-hour period including naps.
☐ 2 I sleep no longer than 12 hours in a 24-hour period including naps.
☐ 3 I sleep longer than 12 hours in a 24-hour period including naps.

During the past seven days…

5. Feeling Sad:

☐ 0 I do not feel sad.
☐ 1 I feel sad less than half the time.
☐ 2 I feel sad more than half the time.
☐ 3 I feel sad nearly all of the time.

Please complete either 6 or 7 (not both)

6. Decreased Appetite:

☐ 0 There is no change in my usual appetite.
☐ 1 I eat somewhat less often or lesser amounts of food than usual.
☐ 2 I eat much less than usual and only with personal effort.
☐ 3 I rarely eat within a 24-hour period, and only with extreme personal effort or when others persuade me to eat.

\- OR -

7. Increased Appetite:

☐ 0 There is no change from my usual appetite.
☐ 1 I feel a need to eat more frequently than usual.
☐ 2 I regularly eat more often and/or greater amounts of food than usual.
☐ 3 I feel driven to overeat both at mealtime and between meals.

Please complete either 8 or 9 (not both)

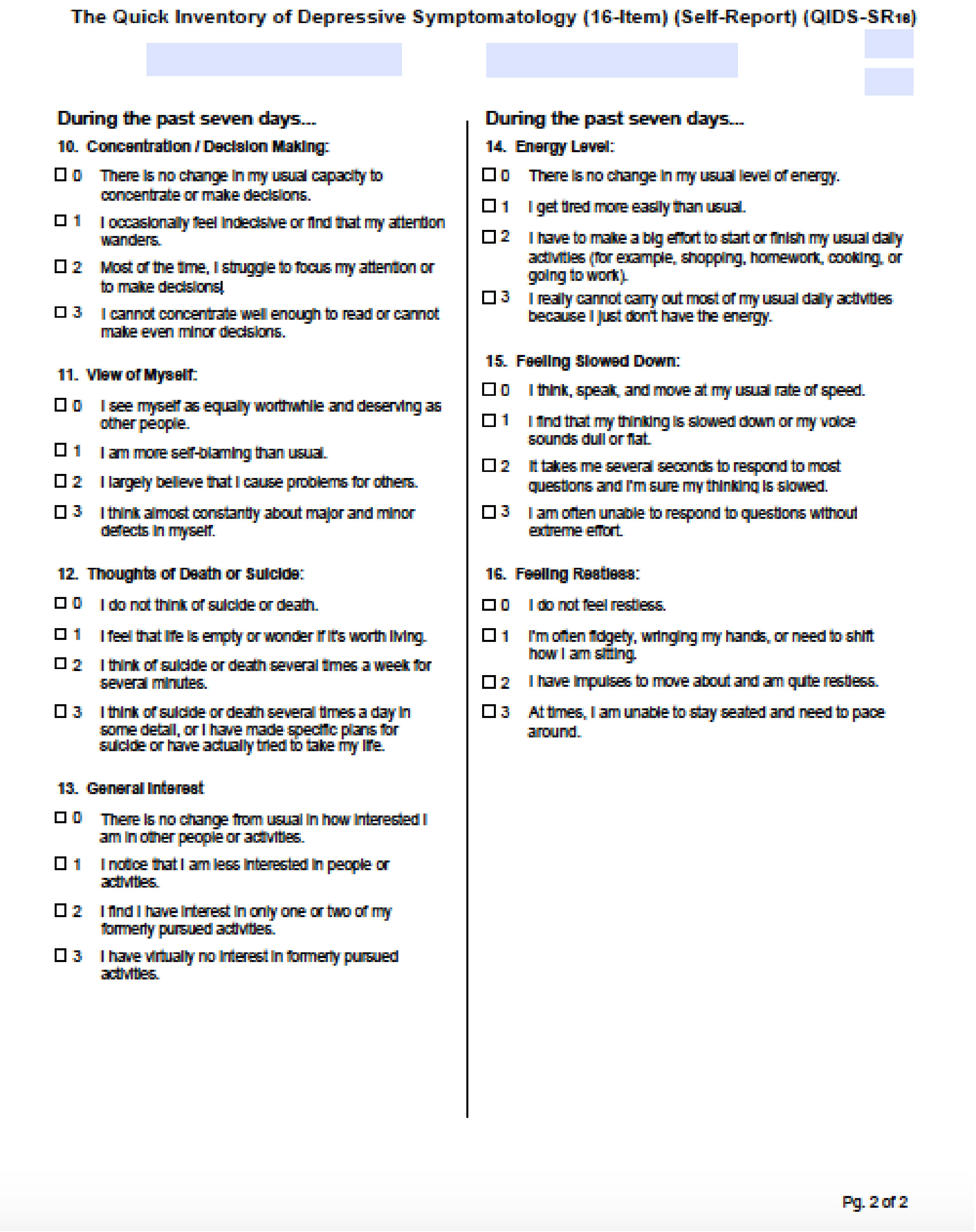
8. Decreased Weight (Within the Last Two Weeks):

☐ 0 I have not had a change in my weight.
☐ 1 I feel as if I have had a slight weight loss.
☐ 2 I have lost 2 pounds or more.
☐ 3 I have lost 5 pounds or more.

\- OR -

9. Increased Weight (Within the Last Two Weeks):

☐ 0 I have not had a change in my weight.
☐ 1 I feel as if I have had a slight weight gain.
☐ 2 I have gained 2 pounds or more.
☐ 3 I have gained 5 pounds or more.



The Quick Inventory of Depressive Symptomatology (16-Item) (Self-Report) (QIDS-SR16)

During the past seven days…

10. Concentration / Decision Making:

☐ 0 There is no change in my usual capacity to concentrate or make decisions.

☐ 1 I occasionally feel indecisive or find that my attention wanders.

☐ 2 Most of the time, I struggle to focus my attention or to make decisions.

☐ 3 I cannot concentrate well enough to read or cannot make even minor decisions.

11. View of Myself:

☐ 0 I see myself as equally worthwhile and deserving as other people.

☐ 1 I am more self-blaming than usual.

☐ 2 I largely believe that I cause problems for others.

☐ 3 I think almost constantly about major and minor defects in myself.

12. Thoughts of Death or Suicide:

☐ 0 I do not think of suicide or death.

☐ 1 I feel that life is empty or wonder if it's worth living.

☐ 2 I think of suicide or death several times a week for several minutes.

☐ 3 I think of suicide or death several times a day in some detail, or I have made specific plans for suicide or have actually tried to take my life.

13. General Interest

☐ 0 There is no change from usual in how interested I am in other people or activities.

☐ 1 I notice that I am less interested in people or activities.

☐ 2 I find I have interest in only one or two of my formerly pursued activities.

☐ 3 I have virtually no interest in formerly pursued activities.

During the past seven days…

14. Energy Level:

☐ 0 There is no change in my usual level of energy.

☐ 1 I get tired more easily than usual.

☐ 2 I have to make a big effort to start or finish my usual daily activities (for example, shopping, homework, cooking, or going to work).

☐ 3 I really cannot carry out most of my usual daily activities because I just don't have the energy.

15. Feeling Slowed Down:

☐ 0 I think, speak, and move at my usual rate of speed.

☐ 1 I find that my thinking is slowed down or my voice sounds dull or flat.

☐ 2 It takes me several seconds to respond to most questions and I'm sure my thinking is slowed.

☐ 3 I am often unable to respond to questions without extreme effort.

16. Feeling Restless:

☐ 0 I do not feel restless.

☐ 1 I'm often fidgety, wringing my hands, or need to shift how I am sitting.

☐ 2 I have impulses to move about and am quite restless.

☐ 3 At times, I am unable to stay seated and need to pace around.

Pg. 2 of 2

1.4. Altman Self-Rating Mania Scale (ASRM)

ASRM

Instructions

1. On this questionnaire are groups of five statements; read each group of statements carefully.

2. Choose the one statement in each group that best describes the way you have been feeling for the past week.

3. Circle the number next to the statement you picked. 4. Please note: The word "occasionally" when used here means once or twice; "often" means several times or more; "frequently" means most of the time.

1. Positive Mood

- ○ I do not feel happier or more cheerful than usual.
- ○ I occasionally feel happier or more cheerful than usual.
- ○ I often feel happier or more cheerful than usual.
- ○ I feel happier or more cheerful than usual most of the time.
- ○ I feel happier or more cheerful than usual all of the time.

2. Self-Confidence

- ○ I do not feel more self-confident than usual.
- ○ I occasionally feel more self-confident than usual.
- ○ I often feel more self-confident than usual.
- ○ I feel more self-confident than usual.
- ○ I feel extremely self-confident all of the time.

3. Sleep Patterns

- ○ I do not need less sleep than usual.
- ○ I occasionally need less sleep than usual.
- ○ I often need less sleep than usual.
- ○ I frequently need less sleep than usual.
- ○ I can go all day and night without any sleep and still not feel tired.

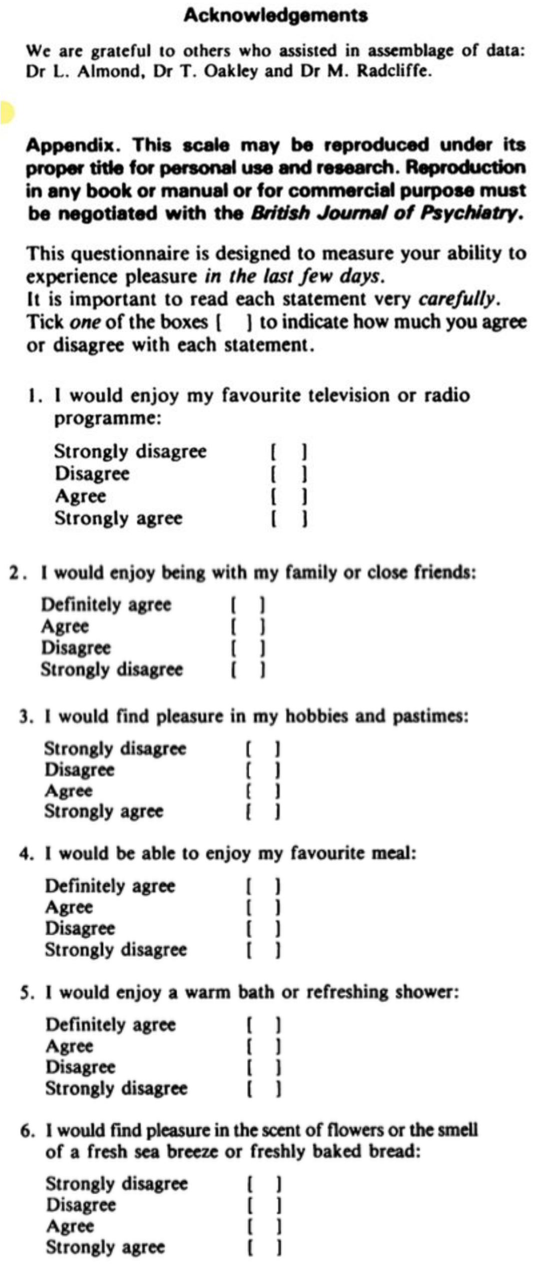
4. Speech

○ I do not talk more than usual.
○ I occasionally talk more than usual.
○ I often talk more than usual.
○ I frequently talk more than usual.
○ I talk constantly and cannot be interrupted.

5. Activity Level

○ I have not been more active (either socially, sexually, at work, home or school) than usual.
○ I have occasionally been more active than usual.
○ I have often been more active than usual.
○ I have frequently been more active than usual.
○ I am constantly active or on the go all the time.

1.5. Snaith-Hamilton Pleasure Scale (SHAPS)



Acknowledgements

We are grateful to others who assisted in assemblage of data: Dr L. Almond, Dr T. Oakley and Dr M. Radcliffe.

Appendix. This scale may be reproduced under its proper title for personal use and research. Reproduction in any book or manual or for commercial purpose must be negotiated with the *British Journal of Psychiatry*.

This questionnaire is designed to measure your ability to experience pleasure *in the last few days*.
It is important to read each statement very *carefully*.
Tick *one* of the boxes [ ] to indicate how much you agree or disagree with each statement.

1. I would enjoy my favourite television or radio programme:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]

2. I would enjoy being with my family or close friends:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

3. I would find pleasure in my hobbies and pastimes:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]

4. I would be able to enjoy my favourite meal:

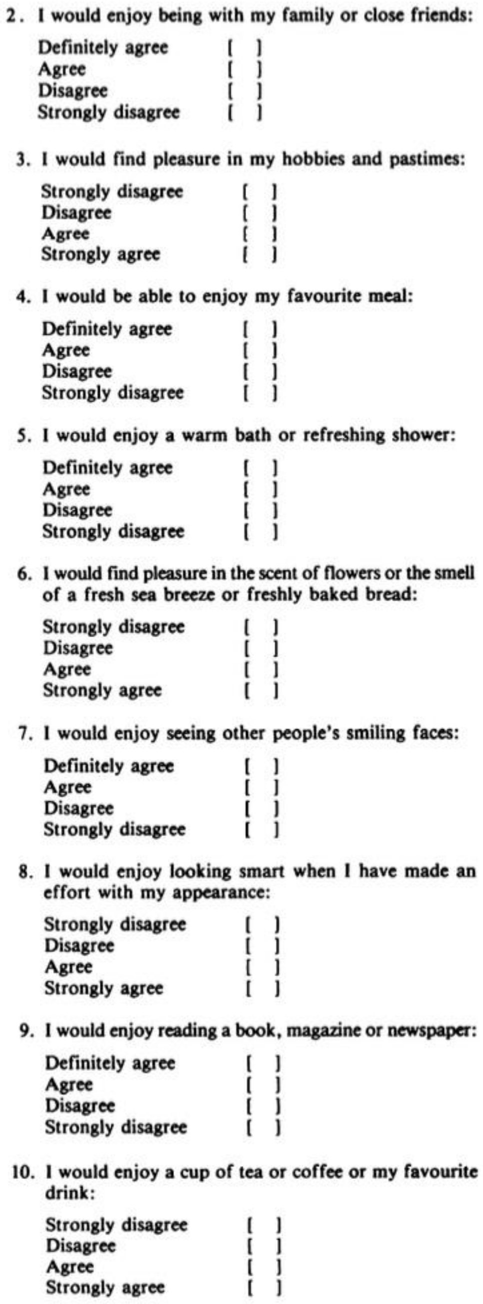
Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

5. I would enjoy a warm bath or refreshing shower:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

6. I would find pleasure in the scent of flowers or the smell of a fresh sea breeze or freshly baked bread:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]



2. I would enjoy being with my family or close friends:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

3. I would find pleasure in my hobbies and pastimes:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]

4. I would be able to enjoy my favourite meal:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

5. I would enjoy a warm bath or refreshing shower:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

6. I would find pleasure in the scent of flowers or the smell of a fresh sea breeze or freshly baked bread:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]

7. I would enjoy seeing other people's smiling faces:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

8. I would enjoy looking smart when I have made an effort with my appearance:

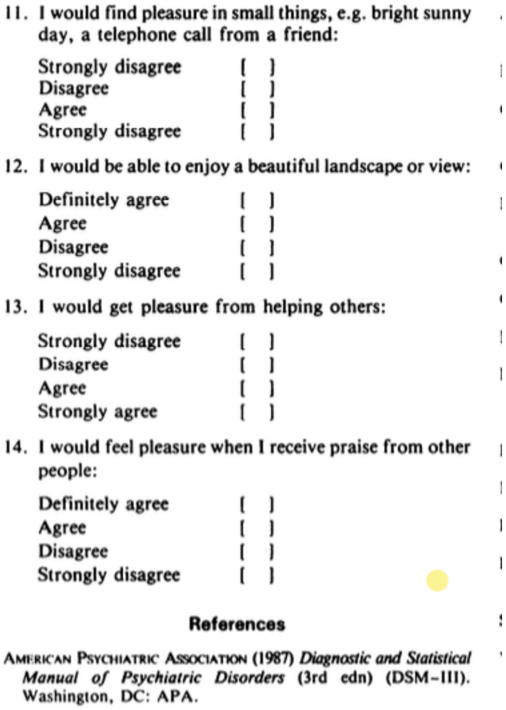
Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]

9. I would enjoy reading a book, magazine or newspaper:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

10. I would enjoy a cup of tea or coffee or my favourite drink:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]



11. I would find pleasure in small things, e.g. bright sunny day, a telephone call from a friend:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly disagree [ ]

12. I would be able to enjoy a beautiful landscape or view:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

13. I would get pleasure from helping others:

Strongly disagree [ ]
Disagree [ ]
Agree [ ]
Strongly agree [ ]

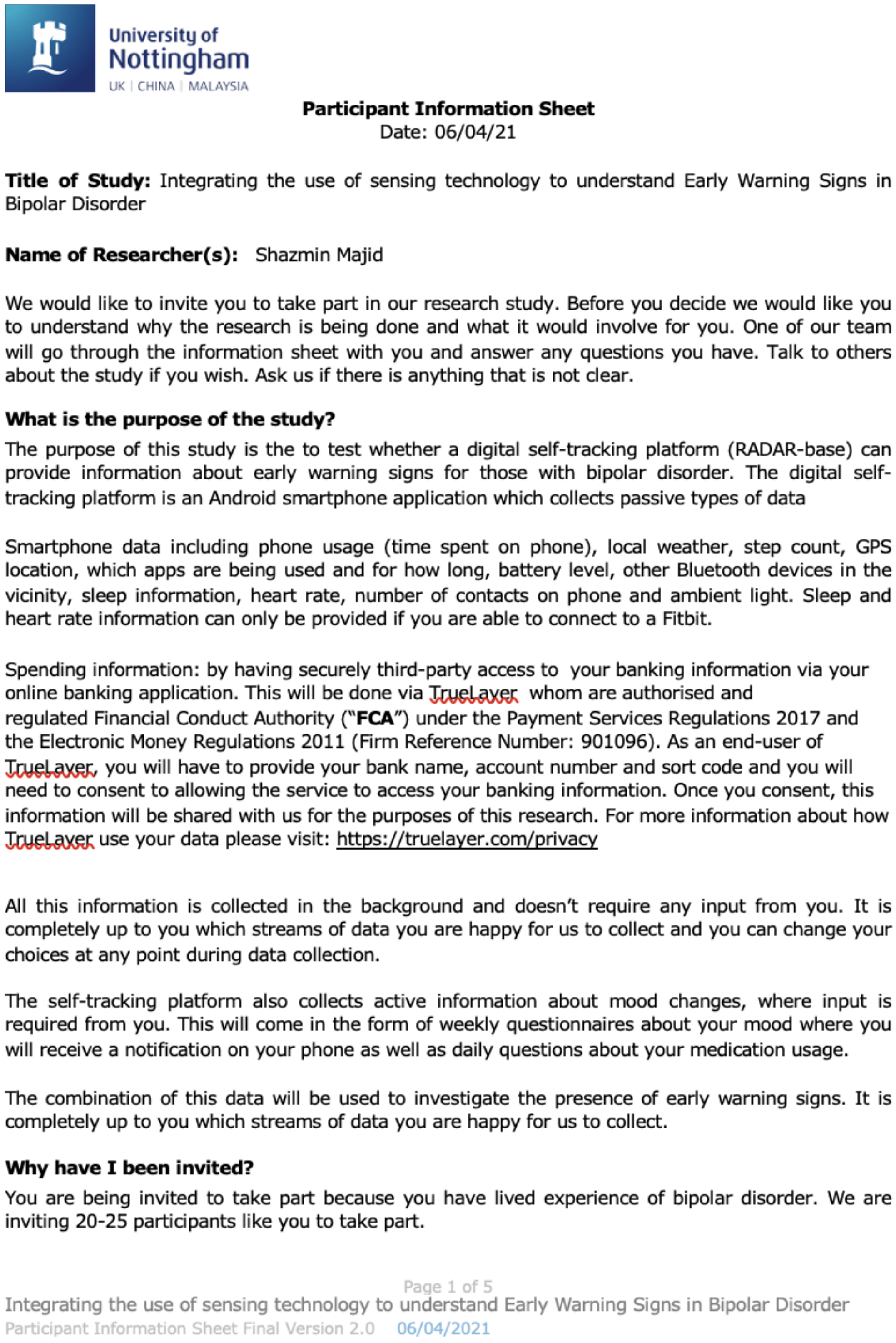
14. I would feel pleasure when I receive praise from other people:

Definitely agree [ ]
Agree [ ]
Disagree [ ]
Strongly disagree [ ]

References

AMERICAN PSYCHIATRIC ASSOCIATION (1987) *Diagnostic and Statistical Manual of Psychiatric Disorders* (3rd edn) (DSM-III). Washington, DC: APA.

2. Participant Information Sheet



University of Nottingham
UK | CHINA | MALAYSIA

Participant Information Sheet
Date: 06/04/21

Title of Study: Integrating the use of sensing technology to understand Early Warning Signs in Bipolar Disorder

Name of Researcher(s): Shazmin Majid

We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. One of our team will go through the information sheet with you and answer any questions you have. Talk to others about the study if you wish. Ask us if there is anything that is not clear.

What is the purpose of the study?

The purpose of this study is the to test whether a digital self-tracking platform (RADAR-base) can provide information about early warning signs for those with bipolar disorder. The digital self-tracking platform is an Android smartphone application which collects passive types of data

Smartphone data including phone usage (time spent on phone), local weather, step count, GPS location, which apps are being used and for how long, battery level, other Bluetooth devices in the vicinity, sleep information, heart rate, number of contacts on phone and ambient light. Sleep and heart rate information can only be provided if you are able to connect to a Fitbit.

Spending information: by having securely third-party access to your banking information via your online banking application. This will be done via TrueLayer whom are authorised and regulated Financial Conduct Authority ("**FCA**") under the Payment Services Regulations 2017 and the Electronic Money Regulations 2011 (Firm Reference Number: 901096). As an end-user of TrueLayer, you will have to provide your bank name, account number and sort code and you will need to consent to allowing the service to access your banking information. Once you consent, this information will be shared with us for the purposes of this research. For more information about how TrueLayer use your data please visit: https://truelayer.com/privacy

All this information is collected in the background and doesn't require any input from you. It is completely up to you which streams of data you are happy for us to collect and you can change your choices at any point during data collection.

The self-tracking platform also collects active information about mood changes, where input is required from you. This will come in the form of weekly questionnaires about your mood where you will receive a notification on your phone as well as daily questions about your medication usage.

The combination of this data will be used to investigate the presence of early warning signs. It is completely up to you which streams of data you are happy for us to collect.

Why have I been invited?

You are being invited to take part because you have lived experience of bipolar disorder. We are inviting 20-25 participants like you to take part.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form via email. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

If you decide to take part, please contact the researcher, using the details provided at the bottom of these sheet to organise a virtual meeting.

At the first meeting, you will be asked to fill out a questionnaire with information about you, your medication and your early warning signs. You will be supported to download an Android app called RADAR-base from Google Play onto your phone and you will receive training on how to use the app to answer the questionnaire about mood changes, as well as instructions on how to delete the app in the case you want to withdraw. You will be provided with a form with a list of the data streams that can be collected and will be asked to indicate which ones you are comfortable with being turned "on" or "off". You are able to change your choices at any point during data collection by letting the researcher know. You will also be given the phone number so you can call the researcher about any issues you may be having with app. You will receive a £20 voucher at this point compensation for your time and effort.

The app will stay installed on your phone for a period of 6 months, unless you decide to withdraw from the study. During these 6 months, you will receive a notification once a week to complete a series of short questionnaire about your mood and mental halth that week. You will also receive daily questions about your medication usage and mood. You will also have meetings with the researcher on months 1, 3 and 6 where you will be asked to talk about your experience with using the application and any changes you think are needed. The researcher will also discuss your data with you at this point. This will happen in graph form, where the researcher will show you changes in frequency of the data streams that you have chosen for us to collect. You will be asked what you think the changes mean. To provide an example, if you have chosen for us to collect data on "phone usage", the researcher will show you a summary of how much your phone usage has changed over the course of a month and ask you what you think about these changes. These monthly meeting will be held virtually on Microsoft Teams where you will be recorded. It will be completely up to you whether this is video recording or audio only. At the end of the six months, you will have a final "close up" meeting with the researcher where you will be invited to share your reflections and also be assisted in deleting the application from your phone. You will receive a £20 voucher at the "close up" meeting for your time and effort.

Expenses and payments

Participants will be provided with a £20 voucher for their participation at the first meeting and also receive a £20 voucher at the "close up" meeting. This will either be a Love2Shop or Amazon voucher. Travel expenses will be offered for any visits incurred as a result of participation.

Will my taking part in the study be kept confidential?

We will follow ethical and legal practice and all information about you will be handled in confidence. If you join the study, the data collected for the study will be looked at by authorised persons from the University of Nottingham who are organising the research. They may also be looked at by authorised people to check that the study is being carried out correctly. All will have a duty of confidentiality to you as a research participant and we will do our best to meet this duty. All information which is collected about you during the research will be kept **strictly confidential**, stored in a secure and locked office, and on a password-protected database. Any information about you which leaves the University will have your name and address removed (anonymised) and a unique code will be used so that you cannot be recognised from it. Anonymised data may also be stored in data archives for future researchers interested in this area.

With the informed consent from the participants, we will collect the following types of passively collected data: phone usage, local weather, step count, GPS location, which apps are being used and for how long, battery level, other Bluetooth devices in the vicinity, sleep information, heart rate, number of contacts on phone, ambient light, spending information and car usage data. It is completely up to the participants which types of passively data they are happy for us to collect, which they can change at any point by informing the researcher. We will also collect weekly mood change data and daily medicate usage data, which will be provided by you via the self-tracking platform. All smartphone data will be non-identifiable as you will be assigned a Unique Identification Number which can only be accessed by the lead researcher. All data will be stored on the secure University of Nottingham servers. Data transmission to and from the secure from the application will be via a secure Hypertext Transfer Protocol Secure Connection (HTTPS). Data collected via your smartphone will be encrypted and uploaded to secure servers by Wi-Fi or mobile data connection. Data will be temporarily cached on the smartphone until an appropriate connection is available and will then be automatically deleted from the phone memory.

We would also like to record the monthly interviews, which will be video or audio only as per your choice. When transcribing, we will give participants pseudonyms (e.g. participant 1, participant 2) to ensure anonymity. Once transcribed, we will destroy the raw video or audio files. The information produced from the workshop will be used to write academic papers and a PhD thesis.

The University may store your data for up to 25 years and for a period of no less than 10 years after the research project finishes. The researchers who gathered or processed the data may also store the data indefinitely and reuse it in future research. After this time your data will be disposed of securely. During this time all precautions will be taken by all those involved to maintain your confidentiality, only members of the research team will have access to your personal data.

Although what you say during this research study is confidential, should you disclose anything to us which we feel puts you or anyone else at any risk, we may feel it necessary to report this to the appropriate persons.

What are the possible benefits of taking part?

We are passionate about designing digital technology for people with bipolar, with people with bipolar disorder. Your involvement in this study helps us to understand what digital self-tracking means to individuals and how it can be helped in self-management, especially in understanding early warning signs. The findings of this research aim to shape the future of digital self-tracking for those with bipolar disorder, and your contribution to this will be invaluable. Finally, at the end of the study, you will also be provided with a summary document of your data over the period of 6 months of collection which you can use as you wish.

What are the possible disadvantages and risks of taking part?

For some participants, having data collected by smartphone might cause some distress. Talking about mood changes over time as well as early warning signs can also be distressing. Both of these can be perceived as a disadvantage of participation. You will be asked to complete daily questions about your mood and medication which should take a few minutes to complete. You will also be asked to complete a set of questionnaires about your mood and mental health which will take approximately 25 minutes to complete once a week. The ability to dedicate the time to complete these questionnaires is something to consider when signing up to the study.

Participation to this study is completely voluntary and you can withdraw at any time without giving reason.

Another potential risk in this study is one of identification (i.e. losing / leaking data, being identified as a participant, having accounts attributed to me) though every effort will be made to keep the data safe and anonymised as described in the section titled "Will my taking part in the study be kept confidential?" on Page 2.

What if there is a problem?

If you consent to the study, we will send you an email asking you if you want to provide a contact number and or email address of family/friend/member of a mental health team. This is to provide a route for a follow-up, if necessary, to ensure that the research has not impacted you in a negative way. If you do provide this information, please let this person know that you will be doing so. This information will be stored in the password-protected database at the university.

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. If you remain unhappy and wish to complain formally, you can do this by contacting the School Ethics Committee. Both of these contact details are listed at the bottom of the sheet.

What will happen if I don't want to carry on with the study?

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and without your legal rights being affected. You are able with withdraw by deleting the application and data collection will completely stop at this point. If you withdraw then the information collected so far will be removed from the research up until the point of publication, when it is no longer feasible to exclude data. Participants will not be identifiable in the data at this point.

What will happen to the results of the research study?

The results of this study will be used to write an article which will be submitted for scientific publication as well as form part of a thesis as part of a PhD qualification for Shazmin Majid. Direct quotes and may be used in the publication and you will not be identified in any of these materials.

Who is organising and funding the research?

This research is being organised by the University of Nottingham and is being funded by ESPRC Centre for Doctoral Training in "My Life in Data" in conjunction with Nottingham NIHR Biomedical Research Centre 'Mental Health Technology" theme.

Who has reviewed the study?

All research in the University of Nottingham is looked at by a group of people, called a Research Ethics Committee REF CS- 2019-R58 to protect your interests. This study has been reviewed and approved by the School of Computer Science Research Ethics Committee.

Further information and contact details

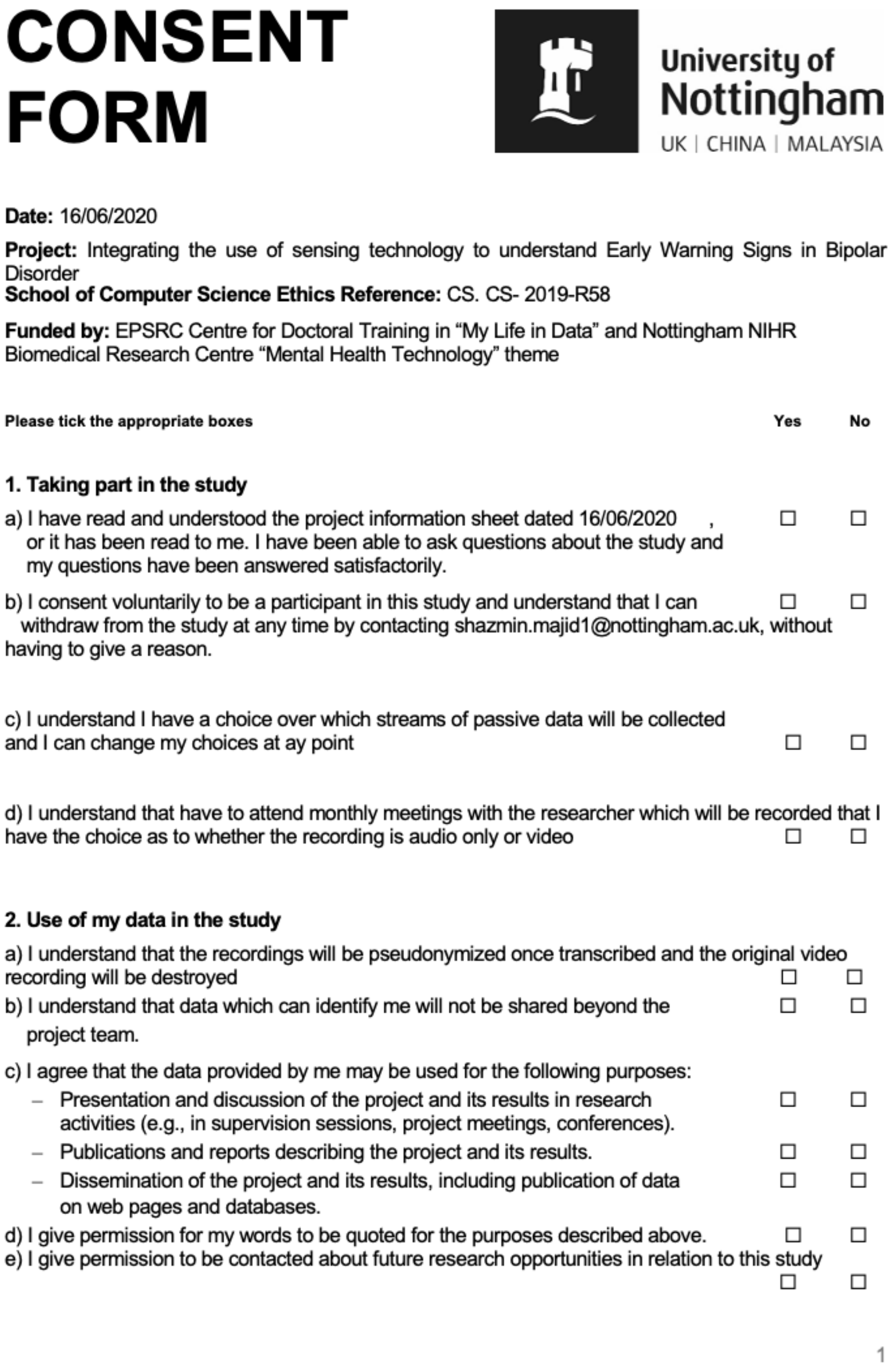
Lead Researcher:
Shazmin Majid
Room C07
Department of Computer Science
Wollaton Road
Jubilee Campus – The University of Nottingham
NG8 1BB
Shazmin.majid@nottingham.ac.uk

+447713508981

Supervisor/PI:
Dr Stuart Reeves
Room C15
Department of Computer Science
Wollaton Road
Jubilee Campus – The University of Nottingham
NG8 1BB
Stuart.reeves@nottingham.ac.uk

Research Ethics Committee:
Cs-ethicsadmin@cs.nott.ac.uk

3. Participant Consent Form



CONSENT FORM

University of Nottingham
UK | CHINA | MALAYSIA

Date: 16/06/2020

Project: Integrating the use of sensing technology to understand Early Warning Signs in Bipolar Disorder
School of Computer Science Ethics Reference: CS. CS- 2019-R58

Funded by: EPSRC Centre for Doctoral Training in “My Life in Data” and Nottingham NIHR Biomedical Research Centre “Mental Health Technology” theme

Please tick the appropriate boxes	**Yes**	**No**
1. Taking part in the study		
a) I have read and understood the project information sheet dated 16/06/2020 , or it has been read to me. I have been able to ask questions about the study and my questions have been answered satisfactorily.	☐	☐
b) I consent voluntarily to be a participant in this study and understand that I can withdraw from the study at any time by contacting shazmin.majid1@nottingham.ac.uk, without having to give a reason.	☐	☐
c) I understand I have a choice over which streams of passive data will be collected and I can change my choices at ay point	☐	☐
d) I understand that have to attend monthly meetings with the researcher which will be recorded that I have the choice as to whether the recording is audio only or video	☐	☐
2. Use of my data in the study		
a) I understand that the recordings will be pseudonymized once transcribed and the original video recording will be destroyed	☐	☐
b) I understand that data which can identify me will not be shared beyond the project team.	☐	☐
c) I agree that the data provided by me may be used for the following purposes:		
– Presentation and discussion of the project and its results in research activities (e.g., in supervision sessions, project meetings, conferences).	☐	☐
– Publications and reports describing the project and its results.	☐	☐
– Dissemination of the project and its results, including publication of data on web pages and databases.	☐	☐
d) I give permission for my words to be quoted for the purposes described above.	☐	☐
e) I give permission to be contacted about future research opportunities in relation to this study	☐	☐

Please tick the appropriate boxes	**Yes**	**No**

3. Reuse of my data

a) I give permission for the data that I provide to be reused for the sole purposes of future research and learning.	☐	☐
b) I understand and agree that this may involve depositing my anonymized processed data in a data repository, which may be accessed by other researchers.	☐	☐

4. Security of my data

a) I understand that safeguards will be put in place to protect my identity and my data during the research, and if my data is kept for future use.	☐	☐
b) I confirm that a written copy of these safeguards has been given to me in the University's privacy notice, and that they have been described to me and are acceptable to me.	☐	☐
c) I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data.	☐	☐

5. Copyright

a) I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented.	☐	☐
b) I wish to be publicly identified as the creator of the following works: audio/video recordings, drawings, and any other material I produced or co-produced during the study.	☐	☐

6. Signatures (sign as appropriate)

Click or tap here to enter text. ____ Click or tap here to enter text. ____ Click or tap here to enter text.

Name of participant (IN CAPITALS) Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

SHAZMIN MAJID ____ Signature ____ Date

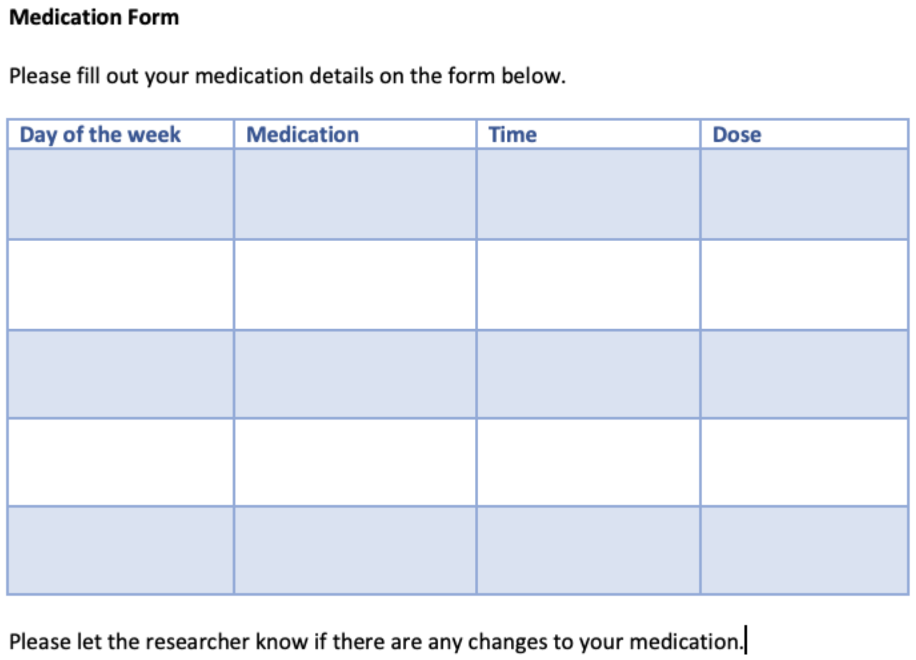
7. Researcher's contact details

Name: Shazmin Majid

Phone: 077135 08981

Email: shazmin.majid1@nottingham.ac.uk

4. Current Medication Form



Medication Form

Please fill out your medication details on the form below.

Day of the week	Medication	Time	Dose

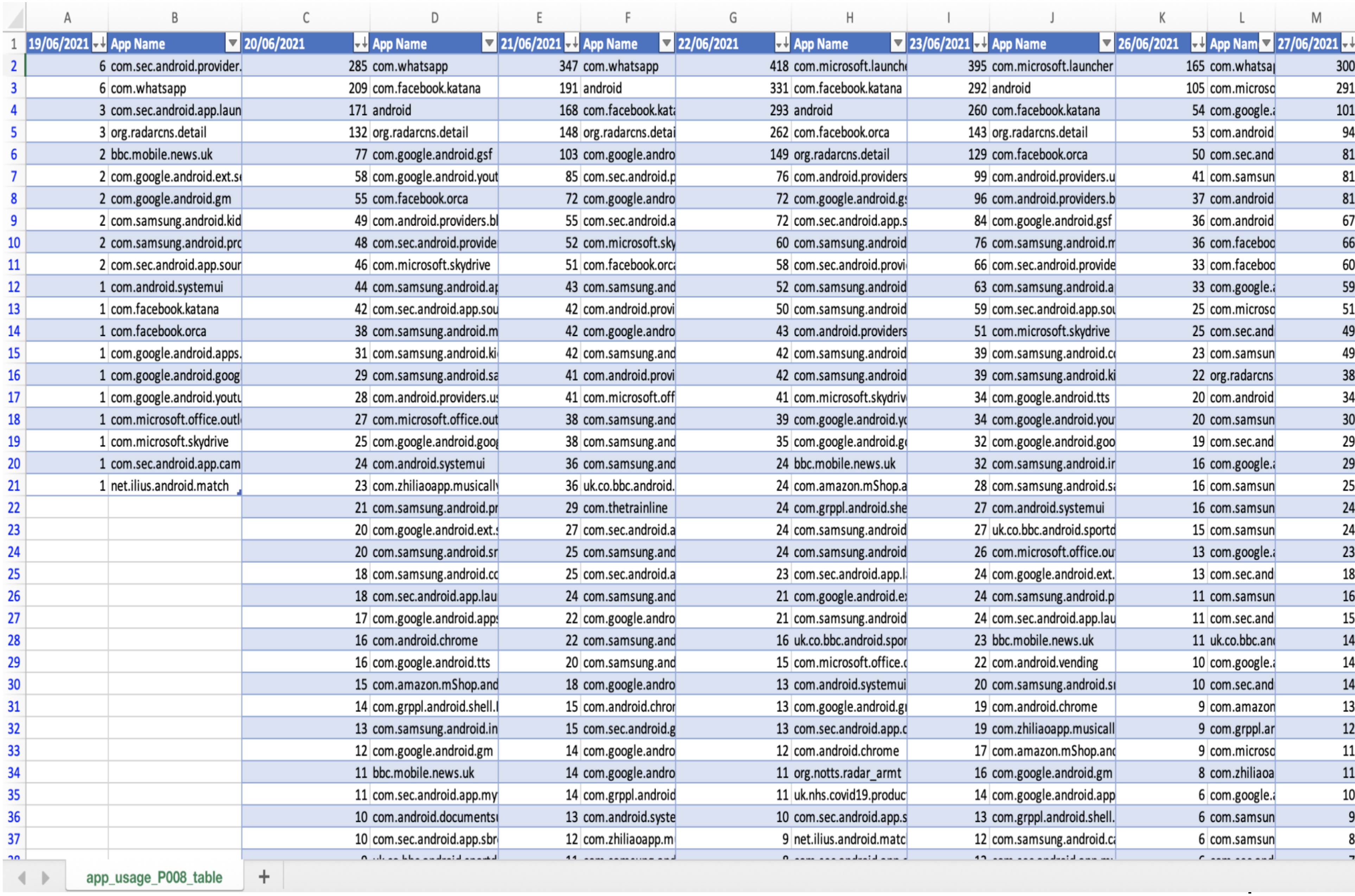
Please let the researcher know if there are any changes to your medication.

5. Data collection stream form

Please indicate which streams of data you would like “ON” or “OFF”. You are able to change this at any point during the study.

Data streams	
Phone usage (via smartphone)	
Local weather (via smartphone)	
Step count (via Fitbit or smartphone)	
GPS location (via Fitbit or smartphone)	
Which apps used and usage timing (via smartphone)	
Battery level (via smartphone)	
Other Bluetooth devices in the vicinity (via smartphone)	
Sleep information (via Fitbit)	
Heart rate (via Fitbit)	
Number of contacts on the phone (via smartphone)	
Ambient light (via smartphone)	
Levels of spending measured by mobile banking (via smartphone)	

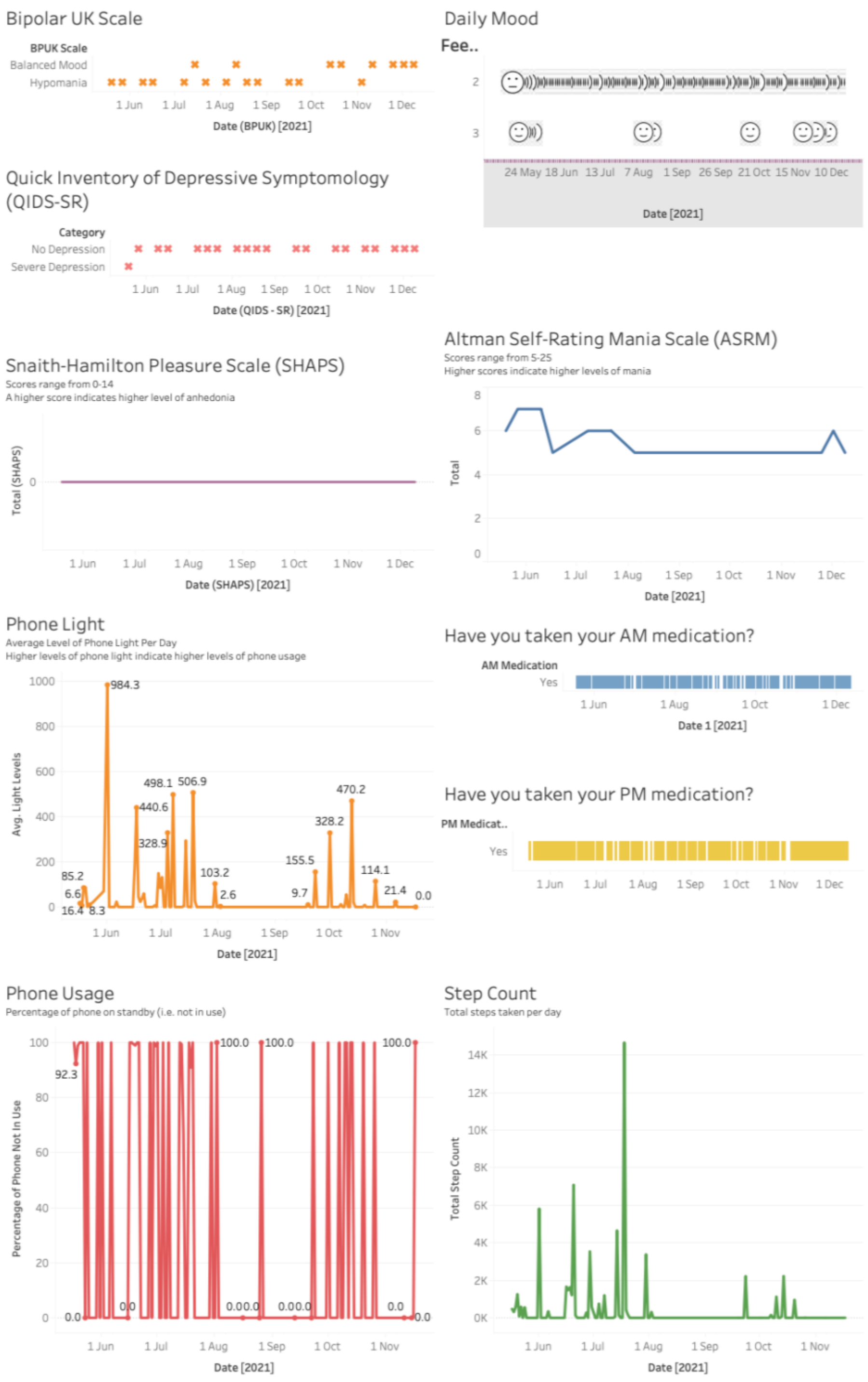
6. App usage data example



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	19/06/2021	App Name	20/06/2021	App Name	21/06/2021	App Name	22/06/2021	App Name	23/06/2021	App Name	26/06/2021	App Nam	27/06/2021
2	6	com.sec.android.provider.	285	com.whatsapp	347	com.whatsapp	418	com.microsoft.launch	395	com.microsoft.launcher	165	com.whatsa	300
3	6	com.whatsapp	209	com.facebook.katana	191	android	331	com.facebook.katana	292	android	105	com.micros	291
4	3	com.sec.android.app.laun	171	android	168	com.facebook.kat	293	android	260	com.facebook.katana	54	com.google.	101
5	3	org.radarcns.detail	132	org.radarcns.detail	148	org.radarcns.detai	262	com.facebook.orca	143	org.radarcns.detail	53	com.android	94
6	2	bbc.mobile.news.uk	77	com.google.android.gsf	103	com.google.andro	149	org.radarcns.detail	129	com.facebook.orca	50	com.sec.and	81
7	2	com.google.android.ext.s	58	com.google.android.yout	85	com.sec.android.p	76	com.android.providers	99	com.android.providers.u	41	com.samsun	81
8	2	com.google.android.gm	55	com.facebook.orca	72	com.google.andro	72	com.google.android.g	96	com.android.providers.b	37	com.android	81
9	2	com.samsung.android.kid	49	com.android.providers.bl	55	com.sec.android.a	72	com.sec.android.app.s	84	com.google.android.gsf	36	com.android	67
10	2	com.samsung.android.pr	48	com.sec.android.provide	52	com.microsoft.sky	60	com.samsung.android	76	com.samsung.android.r	36	com.facebo	66
11	2	com.sec.android.app.sour	46	com.microsoft.skydrive	51	com.facebook.orc	58	com.sec.android.provi	66	com.sec.android.provide	33	com.facebo	60
12	1	com.android.systemui	44	com.samsung.android.a	43	com.samsung.and	52	com.samsung.android	63	com.samsung.android.a	33	com.google.	59
13	1	com.facebook.katana	42	com.sec.android.app.sou	42	com.android.provi	50	com.samsung.android	59	com.sec.android.app.so	25	com.micros	51
14	1	com.facebook.orca	38	com.samsung.android.m	42	com.google.andro	43	com.android.providers	51	com.microsoft.skydrive	25	com.sec.and	49
15	1	com.google.android.apps.	31	com.samsung.android.ki	42	com.samsung.and	42	com.samsung.android	39	com.samsung.android.c	23	com.samsun	49
16	1	com.google.android.goog	29	com.samsung.android.sa	41	com.android.provi	42	com.samsung.android	39	com.samsung.android.ki	22	org.radarcns	38
17	1	com.google.android.yout	28	com.android.providers.u	41	com.microsoft.off	41	com.microsoft.skydriv	34	com.google.android.tts	20	com.android	34
18	1	com.microsoft.office.outl	27	com.microsoft.office.out	38	com.samsung.and	39	com.google.android.y	34	com.google.android.you	20	com.samsun	30
19	1	com.microsoft.skydrive	25	com.google.android.goo	38	com.samsung.and	35	com.google.android.g	32	com.google.android.goo	19	com.sec.and	29
20	1	com.sec.android.app.cam	24	com.android.systemui	36	com.samsung.and	24	bbc.mobile.news.uk	32	com.samsung.android.ir	16	com.google.	29
21	1	net.ilius.android.match	23	com.zhiliaoapp.musicall	36	uk.co.bbc.android.	24	com.amazon.mShop.a	28	com.samsung.android.s	16	com.samsun	25
22			21	com.samsung.android.pr	29	com.thetrainline	24	com.grppl.android.she	27	com.android.systemui	16	com.samsun	24
23			20	com.google.android.ext.	27	com.sec.android.a	24	com.samsung.android	27	uk.co.bbc.android.sportd	15	com.samsun	24
24			20	com.samsung.android.sr	25	com.samsung.and	24	com.samsung.android	26	com.microsoft.office.ou	13	com.google.	23
25			18	com.samsung.android.c	25	com.sec.android.a	23	com.sec.android.app.l	24	com.google.android.ext.	13	com.sec.and	18
26			18	com.sec.android.app.lau	24	com.samsung.and	21	com.google.android.e	24	com.samsung.android.p	11	com.samsun	16
27			17	com.google.android.app	22	com.google.andro	21	com.samsung.android	24	com.sec.android.app.lau	11	com.sec.and	15
28			16	com.android.chrome	22	com.samsung.and	16	uk.co.bbc.android.spo	23	bbc.mobile.news.uk	11	uk.co.bbc.an	14
29			16	com.google.android.tts	20	com.samsung.and	15	com.microsoft.office.	22	com.android.vending	10	com.google.	14
30			15	com.amazon.mShop.and	18	com.google.andro	13	com.android.systemui	20	com.samsung.android.s	10	com.sec.and	14
31			14	com.grppl.android.shell.	15	com.android.chro	13	com.google.android.g	19	com.android.chrome	9	com.amazon	13
32			13	com.samsung.android.in	15	com.sec.android.g	13	com.sec.android.app.	19	com.zhiliaoapp.musicall	9	com.grppl.a	12
33			12	com.google.android.gm	14	com.google.andro	12	com.android.chrome	17	com.amazon.mShop.an	9	com.micros	11
34			11	bbc.mobile.news.uk	14	com.google.andro	11	org.notts.radar_armt	16	com.google.android.gm	8	com.zhiliaoa	11
35			11	com.sec.android.app.my	14	com.grppl.android	11	uk.nhs.covid19.produc	14	com.google.android.app	6	com.google.	10
36			10	com.android.documents	13	com.android.syste	10	com.sec.android.app.s	13	com.grppl.android.shell.	6	com.samsun	9
37			10	com.sec.android.app.sbr	12	com.zhiliaoapp.m	9	net.ilius.android.matc	12	com.samsung.android.c	6	com.samsun	8

7. Participant dashboards

7.1 Dashboard for P001 (Presented in Tableau Reader 2021.4)



Bipolar UK Scale

Daily Mood

Quick Inventory of Depressive Symptomology (QIDS-SR)

Altman Self-Rating Mania Scale (ASRM)

Scores range from 5-25

Higher scores indicate higher levels of mania

Snaith-Hamilton Pleasure Scale (SHAPS)

Scores range from 0-14

A higher score indicates higher level of anhedonia

Phone Light

Average Level of Phone Light Per Day

Higher levels of phone light indicate higher levels of phone usage

Have you taken your AM medication?

Have you taken your PM medication?

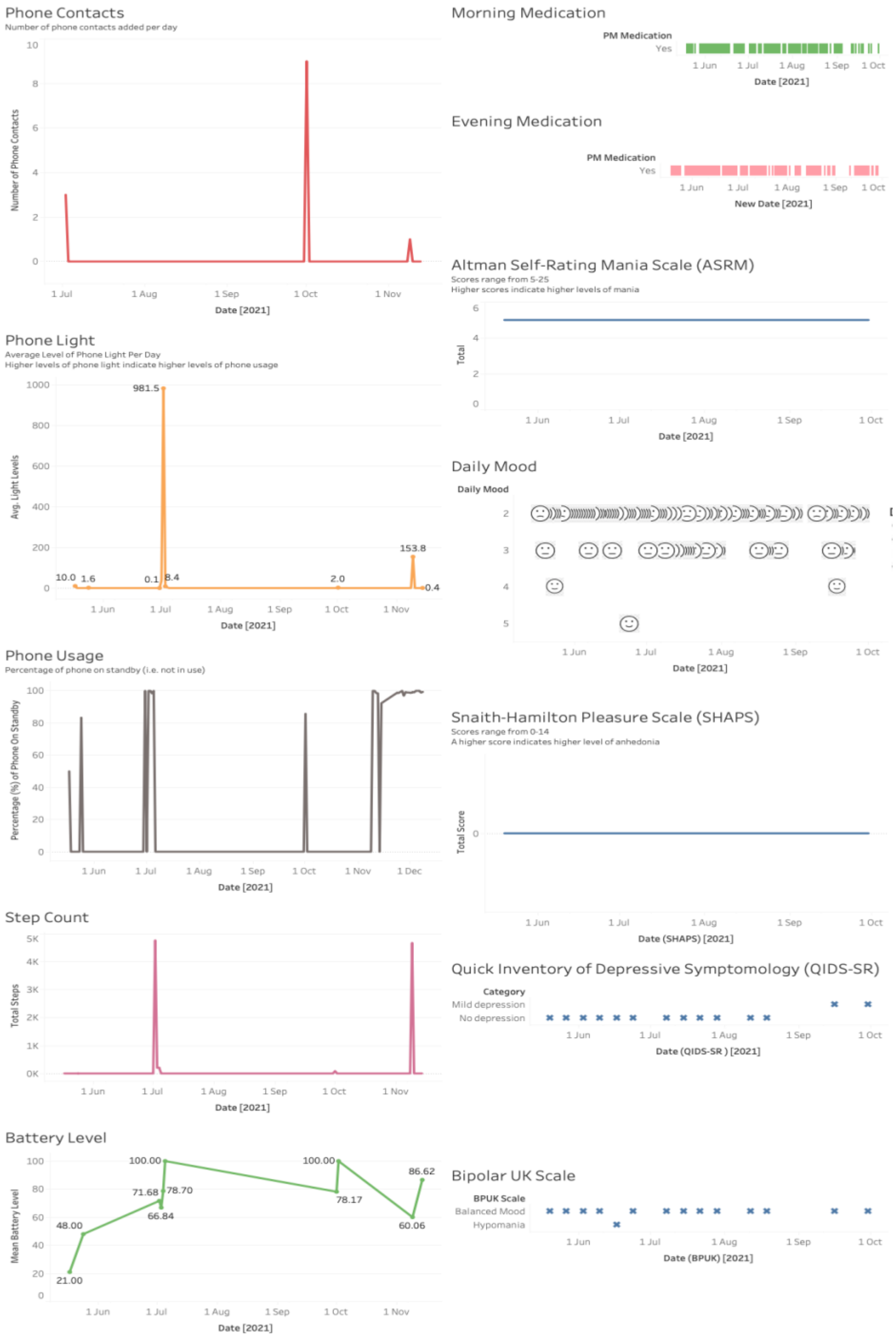
Phone Usage

Percentage of phone on standby (i.e. not in use)

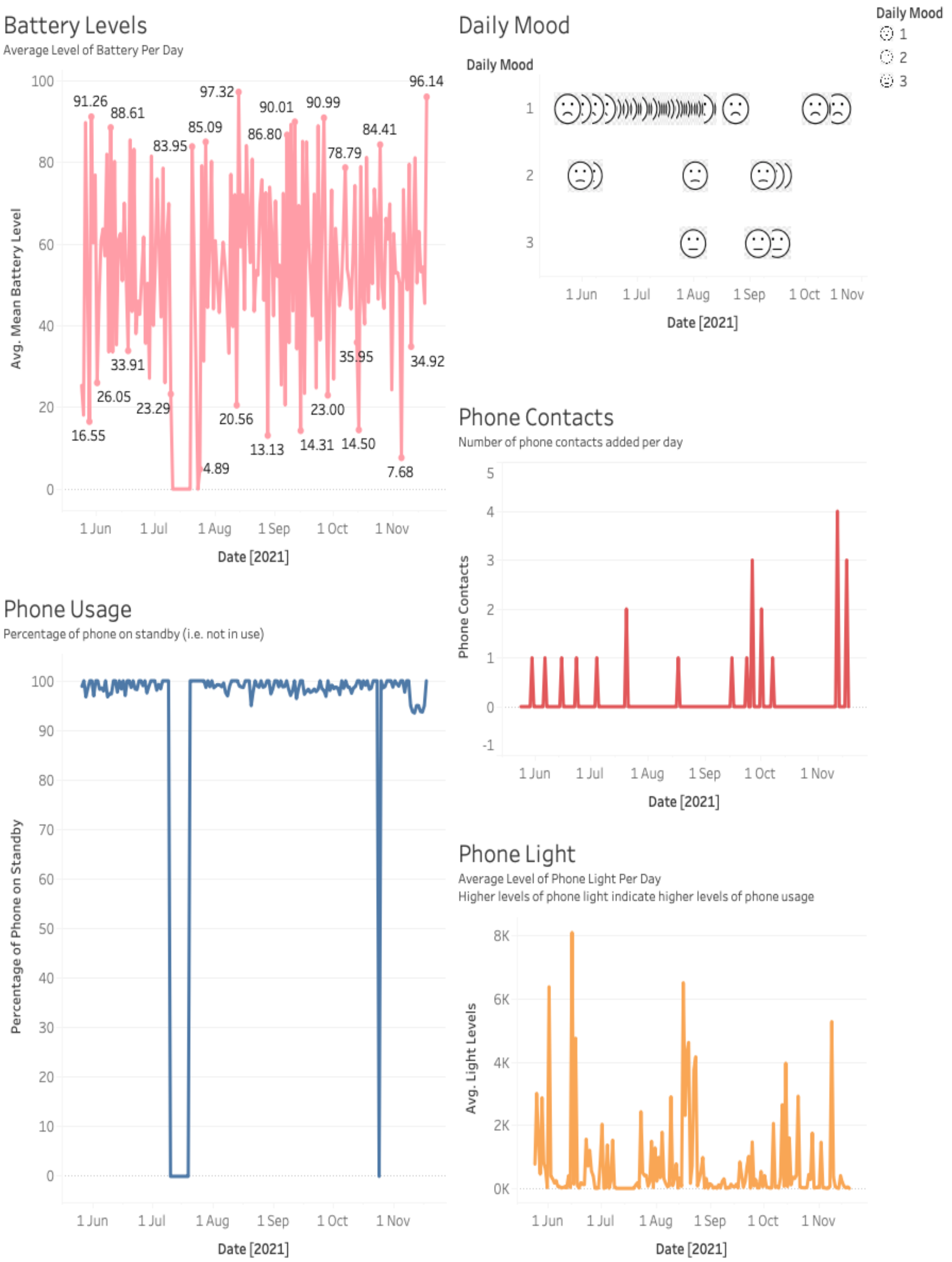
Step Count

Total steps taken per day

7.2 Dashboard for P002 (Presented in Tableau Reader 2021.4)

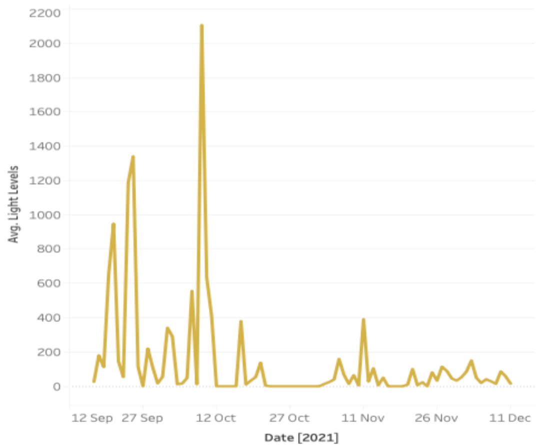


7.3 Dashboard for P003 (Presented in Tableau Reader 2021.4)



7.4 Dashboard for P004 (Presented in Tableau Reader 2021.4)

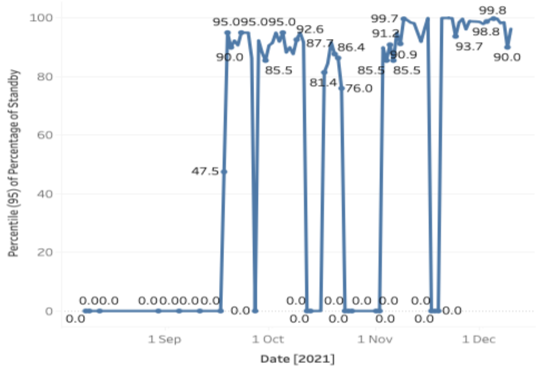
7.5 Dashboard for P005 (Presented in Tableau Reader 2021.4)



Phone Light

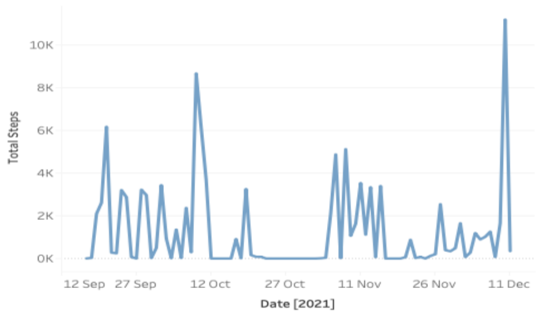
Average Level of Phone Light Per Day

Higher levels of phone light indicate higher levels of phone usage



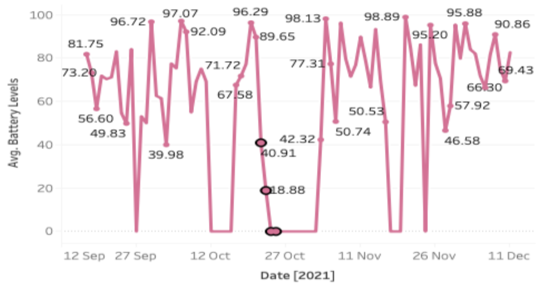
Phone Usage

Percentage of phone on standby (i.e. not in use)



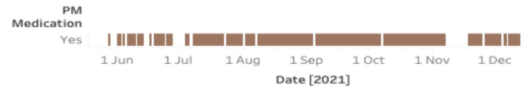
Step Count

Total steps taken per day



Battery Levels

Have you taken your AM Medication?



Have you taken your PM Medication?

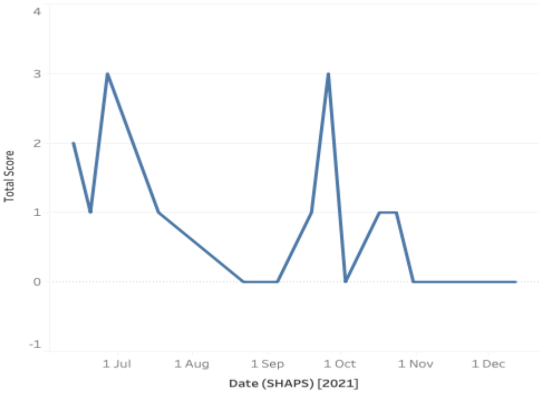
Bipolar UK Scale

Altman Self-Rating Mania Scale (ASRM)

Scores range from 5-25

Higher scores indicate higher levels of mania

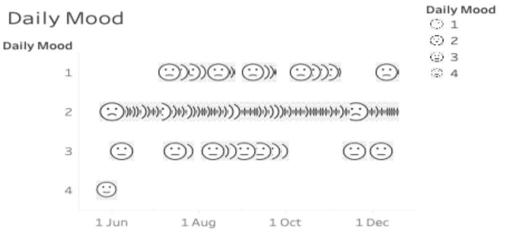
Quick Inventory of Depressive Symptomology (QIDS-SR)



Snaith-Hamilton Pleasure Scale (SHAPS)

Scores range from 0-14

A higher score indicates higher level of anhedonia



Daily Mood

7.6 Dashboard for P006 (Presented in Tableau Reader 2021.4)