University of Bergen
Department of Informatics

Western Norway University of Applied Sciences
Department of Computer Science, Electrical Engineering
and Mathematical Sciences

Ecological Momentary Assessment in Internet-Delivered Psychological Treatments using Wearable Technology

Authors: Jonas Haneborg and Tellev Sundt

Supervisors: Svein-Ivar Lillehaug and Yngve Lamo

May, 2022

Abstract

The growing prevalence of mental health problems is a global concern. Current psychological treatments are effective for a wide range of mental health problems. Yet, treatments today fall short with regards to scalability and struggle to meet the demand for help. To treat patients in a more cost-effective, accessible, and scalable manner, Internet-Delivered Psychological Treatment (IDPT) has posed as a promising solution. Although, IDPT has shown encouraging results, the technology falls short in some regards. One such shortcoming is low user adherence. Adaptive IDPT that allow for personalizing treatment to patient needs may help solve the issue of high drop-out rates in IDPT as they are thought to aid in increasing user adherence. Yet, to adapt and personalize treatment there is a need of meaningful data about patients.

In this thesis, we have created an artifact for the use of wearable data in IDPT. More specifically, our artifact can be split in two parts: (1) an extension of an IDPT framework that serves as a general component and allows for the utilization of wearable data to support Ecological Momentary Assessment (EMA) and (2) a demonstrative component that provides an example of how wearable data may be utilized in interventions to support adaptation. We have created an artifact, comprised of these two components, according to the design science research methodology. Through semi-structured interviews with domain experts of electrical engineering and psychology our artifact has been evaluated. As a result of this evaluation, we have learned that our artifact can serve as a basis for future research.

Contents

List of Figures

List of Listings

CHAPTER 1

1 Introduction

1.1 Motivation

The number of mental health ailments is increasing at a concerning rate and is a major cause for global worry (WHO, 2014). In fact, mental health ailments globally make up the largest single source of economic burden (Mental Health Foundation, 2016) through its effect on productivity losses and heavy use of resources on treatment (Knapp & Wong, 2020). However, there is limited accessibility to mental health care (Bucci et al., 2019). This is due to the current psychological treatments being costly with regards to resources and time, as well as a lack of trained clinicians (Jacobson et al., 2019). As a result, patients often have to wait for a prolonged amount of time before receiving treatment. This is unfortunate, due to early intervention being critical for a better prognosis (McGorry & Mei, 2018).

Utilizing digital technologies in mental health care shows great promise to help alleviate mental distress in a greater number of patients (Stroud et al., 2019). There is a wide variety of explored applications, including immersive virtual reality therapy (Freeman et al., 2017), Internet-Delivered Psychological Treatments (IDPT) (Andersson, 2016), and smartphones and wearables for monitoring, as well as providing alerts and information services (Balcombe & de Leo, 2021). A major benefit of using technologies such as smartphones, personal computers and wearables, is because of how prevalent and widely used they are. As such it is believed that they can be used to monitor, predict, identify, and treat mental health ailments in a more inexpensive, time-sensitive, and scalable manner compared to traditional treatment (Gutierrez et al., 2021).

IDPT has been shown to be effective in treating mental health problems (Mukhiya et al., 2020). Yet, current IDPT struggle with high drop-out rates (or low user adherence). It is believed that low user adherence is largely due to insufficient personalization of treatment (Fernández-Álvarez et al., 2017). Using wearable technology to monitor patients during interventions have the potential to provide additional data to assess patients (e.g., data on stress, activity levels and sleep), outside of clinical settings in a less invasive manner. Such assessment could instantiate a more personalized and hopefully more effective treatment.

Even though there already are existing solutions that utilize wearable monitoring in IDPT, none have a publicly available code base (are open-source). An open-source IDPT framework that allow for the integration of wearable data in treatments could enable reuse and improvement of the framework, and ultimately facilitate further research.

1.2 Problem Description

Suresh Kumar Mukhiya, has for his doctorate developed a framework for creating adaptive IDPT systems (2022). The framework has had many contributors, including INTROMAT (INtroducing personalized TREeatment Of Mental health problems using Adaptive Technology). The framework is an open-source project with the aim of supporting user interaction and increasing user adherence. Through modularizing the project as a collection of plugins, the framework is intended to be useful in any area of mental health care.

To enable adaptability there is a need of data regarding users (Mukhiya et al., 2020). There are several ways of gathering such data, e.g., through questionnaires. However, patients can become fatigued from answering too many questions and answers often vary depending on their state of mind. It is therefore believed that gathering data in a passive and less invasive manner, such as data gathered from wearables, can be advantageous (Griffin & Saunders, 2020). With the prevalence of wearables today, there is an opportunity to access a previously untapped stream of a patient’s activities, moods, and behaviours (Griffin & Saunders, 2020). Furthermore, several types of data collected from these devices have been shown to be correlated with symptoms of mental health problems (Hickey et al., 2021; Lu et al., 2020; Zamkah et al., 2020).

We have created an artifact that is comprised of a general component and a demonstrative component. The general component serves as an extension of the IDPT framework described previously and enables the utilization of wearable data in IDPT. Whereas the demonstrative component serves as an example of how such wearable data can be utilized in interventions. More specifically, how stress measurements can be used to support adaptation in interventions.

1.3 Research Questions

In this thesis we will answer research questions regarding the use of wearable data in IDPT. More specifically, the following two research questions will be addressed:

RQ1 How can an extension that allows for the use of wearable data be implemented to support EMA in interventions?

RQ2 How can such an extension be used to assist in adapting interventions?

1.4 Research Methods

For the research method of this thesis, we have decided to use design science. Design Science encompass iteratively designing, developing and evaluating an artifact to produce information for a knowledge base inside of a problem domain (Hevner et al., 2004). This artifact will serve as a means to answer the research questions. A further explanation of design science and our iterative process can be found in Chapter 3.

1.5 Terminology

This section includes a brief explanation of the commonly used terminology of the thesis.

Internet-Delivered Psychological Treatment (IDPT) is a term that covers all variations of psychological treatment that are provided through the Internet (Andersson, 2016). Whereas an IDPT system refer to a system of software that serves as the technological foundation for IDPT.

Ecological Momentary Assessment (EMA) denote the methodology of patients self-reporting in daily life (Doherty et al., 2020). Moreover, technology-based EMA refers to the use of digital technologies as tools to foster assessment, usually involving mobile or wearable devices (Colombo et al., 2019).

Intervention is used throughout this thesis, as to specifically refer to mental health interventions. Mental health intervention is a broader term, in comparison to mental health treatment. Mental health interventions contain therapy, yet also assessment of symptoms and psychoeducation (Benjenk & Chen, 2018). An IDPT is a form of a mental health intervention administered through the internet.

Wearable is a smart device worn on the body with the possibility of recording a range of different data through various sensors. Although the term representative is of any conveniently portable device that can gather data, the most common wearable devices are smartphones, smartwatches and activity trackers (Guk et al., 2019).

1.6 Thesis Overview

In this section we will biefly present the contents of each chapter of the thesis. As this thesis has been a collaboration between two authors, we want to proclaim the shared ownership of the content. Furthermore, the work put into this thesis in the form of research, implementation and evaluation has been split evenly between both authors.

Chapter 1 is an introductary chapter where main concepts of the thesis are explained. This includes the research questions, our motivation and research methods, to give the reader an overview of our problem domain and what the goals for this thesis are.

Chapter 2 is a collection of background materials for the thesis. Here we explain core concepts such as IDPT, EMA and wearables. In addition, we touch on subjects that are assosciated with our demonstrative component. More specifically, stress and measuring stress with wearables. Last, we will discuss some existing solutions.

Chapter 3 explain our research method and what its guidelines entails. Furthermore, our design process can also be found at the end of this chapter.

Chapter 4 describe our development process from the first iteration to the final version of our implementation. This description is split into two distinct parts. One regarding our general component and one regarding our demonstrative component.

Chapter 5 explain the evaluation process of our implementation.

Chapter 6 present our findings. This chapter answers the research questions in detail and touch on the limitations of our project.

Chapter 7 includes our conclusion as well as future work.

CHAPTER 2

2 Background

This chapter details relevant subjects of our artifact. First, we will present topics associated with our general wearable component, such as internet-delivered psychological treatments, ecological momentary assessment, and wearables. Second, we will introduce topics that are associated with our demonstrative component of how stress measurements can be used to support adaptive interventions. These topics include stress and how to measure stress using wearables. Last, we will discuss some existing solutions that are similar with regards to our intended artifact.

2.1 Internet-Delivered Psychological Treatment (IDPT) Systems

IDPT is psychological treatment administered through the internet. The term was initially used by Andersson (2016). There are numerous other terms that are used in similar contexts, such as web-based treatment, Internet-delivered cognitive-behavioural therapy or e-therapy (Andersson, 2016). We distinguish between two types of IDPT, guided or unguided (Morgan et al., 2017). Guided IDPT involves a clinician aiding the patient through an intervention (e.g., in the form of emails or phone calls). Whereas unguided IDPT is carried out by patients themselves without the assistance of clinicians (Morgan et al., 2017).

To provide interventions for patients, a software platform is required. Such a platform is used to present treatment material, assessment materials such as questionnaires, exercises, and facilitate patient-clinician communication (Andersson et al., 2019). These platforms are known as IDPT systems and may include applications on the web and on mobile, or make use of augmented or virtual reality.

IDPT propose a viable solution to the some of the problems that arise in traditional face-to-face treatments, such as cognitive behavioural therapy. While traditional treatments are effective for a wide range of mental illnesses, they fall short with regards to accessibility (Morgan et al., 2017). As a result, a great number of people have difficulty accessing psychological treatment to aid with their mental health issues (Bucci et al., 2019). One important reason being, that there are not enough trained clinicians compared to the amount of people needing help.

IDPT could play an important role in helping with the shortcomings of traditional treatments, as IDPT is highly scalable and can be accessed from anywhere with an internet connection. Furthermore, IDPT can be an option for people who do not attend treatment because of cost (Mukhiya et al., 2020), stigma (Bharadwaj et al., 2017) or anxiety (Langley et al., 2018).

2.2 Adaptive IDPT Systems

In traditional therapy, therapists adapt their practice with regards to their patients, with the goal of improving outcomes (Mukhiya et al., 2020). Dynamic treatment that changes in response to the development of the patient is important both to optimize treatment and ensure that patients do not drop out early from treatment (Gibbons et al., 2019). Adaptive IDPT systems expand on the concept of regular IDPT by further focusing on personalization (Mukhiya et al., 2020). A more personalized treatment is meant to increase usefulness and relevance, thus hopefully resulting in shorter treatment times and higher user adherence. Mukhiya et al. have proposed a model illustrating how entities in such a system should work together to achieve adaptability.

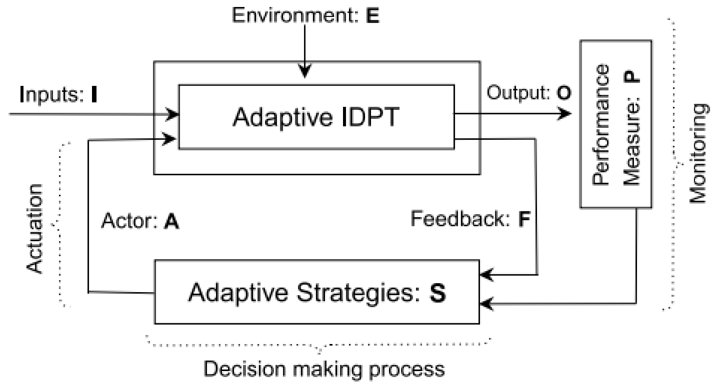

Figure 2.1: Components of an adaptive IDPT system. From Mukhiya et al. (2020)

The paper by Mukhiya et al. (2020) describes a model that presents IDPT systems as a function in an environment, which applies inputs and generate outputs, as seen in Figure 2.1. The system measure performance of these outputs and use some feedback function to determine adaptive strategies. Finally, actors trigger the adaptation. Depending on what kind of actors (e.g., therapists or algorithms as part of the IDPT system) trigger adaptation, we differentiate between the following adaptive systems:

(i) automatic systems that can self-adapt,

(ii) semi-automatic systems where some the adaption happens automatically and some is instantiated by therapists, and

(iii) manual systems, where all adaptation is facilitated by therapists.

There's a wide range of what can be personalized or tailored to patients during treatment. Such elements that can be tailored to patients is referred to as adaptive elements. An example of a typical adaptive element can be content presentation. More specifically, content presentation regards what type of content a patient receives and how it is represented. Notifications, alerts and reminders can also be adaptive elements. Another example of an adaptive element is feedback. Feedback are reflections of data collected from the IDPT displayed to the patients (e.g., a graph visualizing a patient's stress progression over time). Such feedback can help demystify progress and allow patients to get a better understanding of themselves (Resnick et al., 2020).

To allow for adaptation, data regarding patients' behaviour and symptom development are of great importance. Such data can be acquired from several sources. System interaction data can be used to gauge a patient's engagement during treatment. Such data is regarded as behavioural data. Psychometric tests or questionnaires may also be used to assess patients. Additionally, wearable devices may be utilized to monitor valuable measures of sleep, activity and stress, or other similar measures (Griffin & Saunders, 2020).

2.3 An Open-Source Adaptive IDPT Framework

An adaptive open-source IDPT framework has been developed, based on the principles mentioned in Section 2.2. The framework was primarily built by Mukhiya (2022) in association with his PhD thesis. However, the framework has also had a collaborative effort, with contributions from various interested parties. The framework allows for the creation of interventions with treatment components such as cases, modules and tasks. A case is related to a specific mental health issue. Examples of cases can be depression or ADHD. Further, each case contains one or more modules that target a particular dimension of the case. E.g., a case for depression can contain modules on sleep issues and concentration issues. A module may be connected to several cases. Lastly, each module contains one or more tasks. Tasks can either be informative or interactive.

Informative tasks, or learning materials, can either be text, video, or audio. Whereas interactive tasks encompass various exercises that require user participation. Such exercises can be, among other things, mindful exercises, or exercises regarding physical activity. Both informative and interactive tasks may record behavioural data of users.

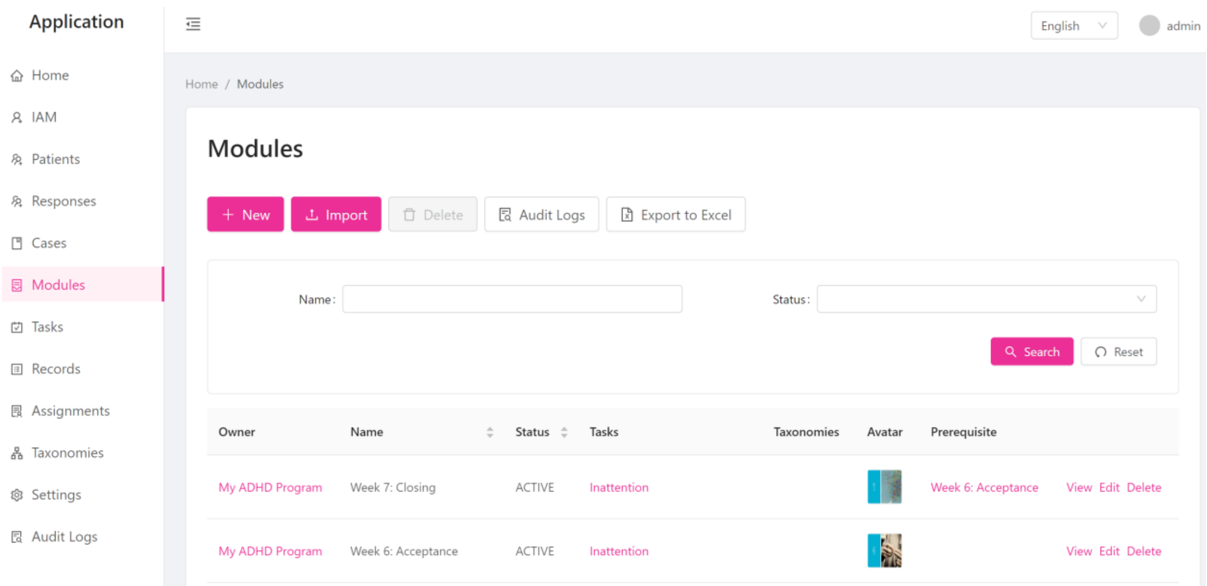

Figure 2.2: A screenshot of the open-source adaptive IDPT framework, showing some examples of modules connected to a case named "My ADHD Program".

Our project consists of extending this framework. More specifically, our contribution will be adding an extension that allows for the integration and utilization of wearable sensor data, whereas this is something that currently does not exist. This extension is what comprise our general component. In addition, we will develop a demonstration of how wearable data can be utilized in interventions. More specifically our demonstrative component. More on the implementation of these two components in Section 4.2 and Section 4.3.

> The codebase of the framework can be viewed at https://github.com/sureshHARDIYA/idpt

2.4 Ecological Momentary Assessment

Ecological momentary assessment (EMA) refers to the methodology of patients' self-reporting in daily life, capturing "life as it is lived, moment to moment, hour to hour, day to day" (Shiffman et al., 2008). "Ecological" refers to the environment from where the data is collected, and "Momentary" refers to the assessment being close in time to the experiences of the patient. EMA allows for measure-based care, defined as "assessment in which patient-reported outcome measures are used to track progress in care as part of a clinical process", which has been shown to be effective in treating psychological problems (Resnick et al., 2020). The method of EMA is closely related to developments in technology (Wilhelm & Perrez, 2013). Previously timed bleepers, personal digital assistants and now smartphones and wearables have been used to improve reach, usability, and reliability of EMA (Doherty et al., 2020).

Assessment in the context of psychological treatment has often been solely based on accounts of the patient. Such traditional clinical assessments are rooted in retrospective self-reports of the patient, where the patient summarizes symptoms and experiences of the past few weeks. However, the accuracy of such reports may vary. Self-reports are affected by biases such as recall bias (Colombo et al., 2019) and patients have been found to misjudge experiences of past events when retrospectively retrieving them (Ben-Zeev et al., 2009). As such there is a need of ways to assess patients in a more objective manner during treatment.

Self-assessment questionnaires such as Patient Health Questionnaires (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), Kessler Psychological Distress Scale (K-10) and Perceived Stress Scale (PSS) are thought to be reliable measures of depression, anxiety, general distress, and stress. Although, the validity of some of these questionnaires have been found to require further evaluation (Lee, 2012; Plummer et al., 2016). Furthermore, time and effort required to fill out these questionnaires limits their use (Staples et al., 2019).

The inability to monitor what happens in the daily-life of patients has been a shortcoming when assessing the effect of interventions (Griffin & Saunders, 2020). EMA emerged as an assessment strategy to better understand behaviour dynamics in patients' daily life (Shiffman et al., 2008). The first studies used this approach with paper-and-pencil daily diaries, but found them not very efficacious, due to low compliance, discomfort, and backfilling (Stone et al., 2007). A growing body of research has begun to explore using digital technologies as potential tools for assessment. More specifically, technology-based EMA has been proposed as a viable strategy to evaluate patients in naturalistic settings (Colombo et al., 2019). Technological tools to assist in EMA, could enhance the ease of use, reduce cost and expand the methods capabilities (Stone et al., 2007).

With the prevalence of wearables today, there is an opportunity to access a previously untapped stream of a patient's activities, moods, and behaviours (Griffin & Saunders, 2020). The data collected from these devices are not subjective, in contrast to self-reports. Another major benefit is that the data is collected passively, not requiring patients to report themselves. Passive collection of data may place less of a burden on users and in turn make EMA less invasive, which ultimately can increase the methodology's feasibility (Doherty et al., 2020).

2.5 Wearables

Digital technologies have the possibility to provide high quality, passive, reliable and continuous data collection (Stroud et al., 2019). Wearable technologies with the ability to assist and monitor are becoming increasingly more prevalent (*CCS Insight*, 2022; Muzny et al., 2020). Wearables, as they are commonly called, are smart devices worn on the body with the possibility of recording data on vital signs, among various other types of data. This data may then be processed, visualised and in many cases transmitted to other devices in an Internet of Things (IoT) environment. A wearable is often in the form of a smartwatch or an activity tracker. Although the names are somewhat interchangeable, smartwatches tend to feature bigger screens, network connectivity and other software and sensors not necessarily related to activity or fitness (Rawassizadeh et al., 2015).

In the domain of medical technology, wearables have a big potential to popularize personalized medicine. Personalized medicine is altering the treatment of a patient with regards to their needs in an effort to decrease the risk of disease, both during treatment of illness as well as a standardized health risk assessment to prevent common chronic diseases (Chan & Ginsburg, 2011). This differs from conventional medicine where the choice of best treatment is derived from a population average. There is a number of different wearables that may be utilized to assist in personalized medicine, ranging from the popular wrist-worn smartwatches to electronic patches placed on the skin (Yetisen et al., 2018).

Furthermore, new ways of measuring and monitoring mental health may help to prevent the usually long wait times between onset of symptoms of illness and diagnosis, which is important because earlier intervention may lead to better outcomes (Berk et al., 2011; Marshall & Rathbone, 2011). As such, there is a need of easy-to-use, unobtrusive, and inexpensive devices that can gather objective information on symptoms that patients fail to report accurately (Glenn & Monteith, 2014).

There is a wide range of different sensors found in wearables today. Of these sensors a diverse assortment of data that can be captured, and many of these types of data have correlations with symptoms of mental health illness (Sano et al., 2015; Zamkah et al., 2020). Examples of such sensor data are heart rate, heart rate variability, skin temperature and electrodermal activity. More about these types of data in Section 2.6.2.

2.5.1 Challenges Concerning Wearables

Wearable technology is an emerging field and although wearable technology is showing promise in multiple fields, the use of such devices poses some challenges that must be taken into consideration. Some important challenges include, but are not limited to, usability, cost-effectiveness, privacy, and interoperability. The following list will briefly touch on some of these important topics but does in no way serve as an exhaustive list.

Usability regards "the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments" (Keogh et al., 2021). How a patient experiences an intervention in which wearable technology is used, is in many ways tied to the design of the device, and thus it may be regarded as a deciding factor when it comes to the choice of wearable (Mathews et al., n.d.). The device must be comfortable to wear over extended periods of time, regardless of the physiology of the patient. It is important to note that the aesthetic appearance of a wearable device may also have an impact on a patient's opinion of the device. In addition, any software the patient is interacting with must be easy to learn for a wide range of users, taking into account mental or physical challenges they may have. Battery life, user experience, quality of data recordings, functionality, price, comfortability, and overall appearance are some of the most important concerns for consumers when deciding on a wearable device (Wen et al., 2017).

Interoperability describes how well separate systems or devices communicate with each other. Even though, numerous wearable devices are being developed, such devices have no guarantee of being interoperable with one another. Wearable manufacturers do often have proprietary solutions regarding both collection and transfer of data. Such proprietary solutions make it difficult for third parties to utilize the device and its data. Within the wearable industry, there is often greater focus on implementing new features rather than working towards establishing standards and promoting interoperability (Muzny et al., 2020). For wearable technology to be truly valuable for health care, improving interoperability is an important step. The most crucial factors regarding interoperability is that the data collected must be in a standardized data format, be easily accessible and generated from a reliable source (Casselman et al., 2017).

Data standards can be described as “documented agreements on representation, format, definition, structuring, tagging, transmission, manipulation, use, and management of data” (*EPA*, 2022). To promote interoperability, exchangeability and safety in health care, standardized formats data is vital (Schulz et al., 2018). This is just as important with regards to wearable sensor data. There exist a lot of different wearable devices from numerous manufacturers and there is no agreed upon data format. As a result, a wide range of different data formats are used. Due to wearable manufacturers having proprietary solutions and unique data formats, it makes it inherently difficult to create applications that can work with multiple wearable devices. With different data formats being used, data from different wearables needs to be converted from one format to a common format if it is to be compatible.

Several attempts at standardizing data formats in health informatics have been made. An example of such a standardization is Fast Healthcare Interoperability Resources (FHIR) developed by Health Level Seven International (HL7), a not-for-profit, standards developing organization (*HL7*, 2022). FHIR introduces RESTful APIs for information exchange as well as a standardization of components also known as resources. The FHIR resources represent usual health care concepts such as observations, patients and appointments. A literature survey performed by Ayaz et al. (2021) concludes that “the FHIR standard is capable of providing an optimized solution for medical data exchange between two systems and will establish data-sharing trust among health care providers”. Yet, Ayaz et al. goes on to explain that a challenge FHIR is faced with is a low adoption rate.

Data quality concerns how useful data is for its intended purpose. To be able to utilize collected data for analysis, a certain quality threshold must be met (Cai & Zhu, 2015). Such a threshold changes depending upon the type of data, the amount of data acquired and the intended use of the data. Wearables that can be used for non-clinical self-monitoring purposes without any issues could prove unfit for research purposes (Degroote et al., 2018). According to a literary review done by Cho et al. (2021) there are several factors that affect data quality collected from wearable devices. They list some of the most common technical problems seen with the devices, as user error and lack of standardization. An example of user errors is watch placement, which could lead to a big loss of data quality (Kamdar & Wu, 2016). There are also technical-related factors that affect data quality related to both hardware and software, as is not surprising given the proprietary nature of wearables (Ometov et al., 2021).

Security and privacy regard how wearable devices may compromise a user's privacy or security. Through numerous sensors, wearables have the possibility to record all kinds of data, such as health data, data about a user's geographic location and data on general living habits (Wen et al., 2017). As these kinds of data can be very sensitive, concerns regarding data ownership and security are important. With wearable data having various formats, being large scale and having many mobile links the data may have an increased risk of leakage or tampering (Guk et al., 2019). Because of this, strategies to ensure security and uphold privacy is important to make wearable technology safe, in addition to gaining public trust (Lu et al., 2020).

2.6 Stress

In general stress defines the threats to our body's balance, equilibrium or "homeostasis" (from Greek and translates to "steady state") (Chrousos et al., 1988). In psychology, it is more specifically a term related to the negative emotional states such as anxiety, agitation, anger, unhappiness, and frustration (Giannakakis et al., 2019). The concept of "fight-or-flight", describes how the autonomic nervous system is activated when reacting to a stressful event (Fink, 2009), in an attempt to restore the body's homeostasis (Chrousos et al., 1988). Such a response includes a change in sweat gland activity, skin temperature and cardiac activity (Seoane et al. 2014). Therefore, these psychological activities can give a good indication of autonomic nervous system activity and as such are considered to be good indicators of stress (Karthikeyan et al., 2013).

Stress comes in two variants: acute and chronic stress. The line between acute and chronic stress is, however, not easy to discern. One of the problems with separating the two variants is how to conceptualize the difference between reoccurring acute stressors and chronic stress (Rohleder, 2019). Acute stress is meant to keep us away from harm (Fink, 2009). Chronic stress on the other hand has no discernible benefits and is regarded as one of the most prevalent illnesses found in the world. The symptoms of chronic stress may manifest physically in the form of muscular tension and back pain, and psychologically in the form of overarousal and emotional distress (Can et al., 2019). Chronic stress is associated with an increase in perceived acute stress and significantly associated with the onset of major depressive episodes (Hammen et al., 2009).

It should be noted that the stress response is a natural response and not inherently bad. Neither too little stress (calm), nor too much stress (distress) but instead a moderate amount of stress (eustress) is considered optimal. This correlation is more famously known as the Hebb's curve (Hebb, 1955) as shown in Figure 2.3. Still, most health researchers agree that stress plays an important role in human health. Stress is linked to mental well-being and stressful incidents often being precursors to various major psychiatric conditions (Cohen et al., 2007).

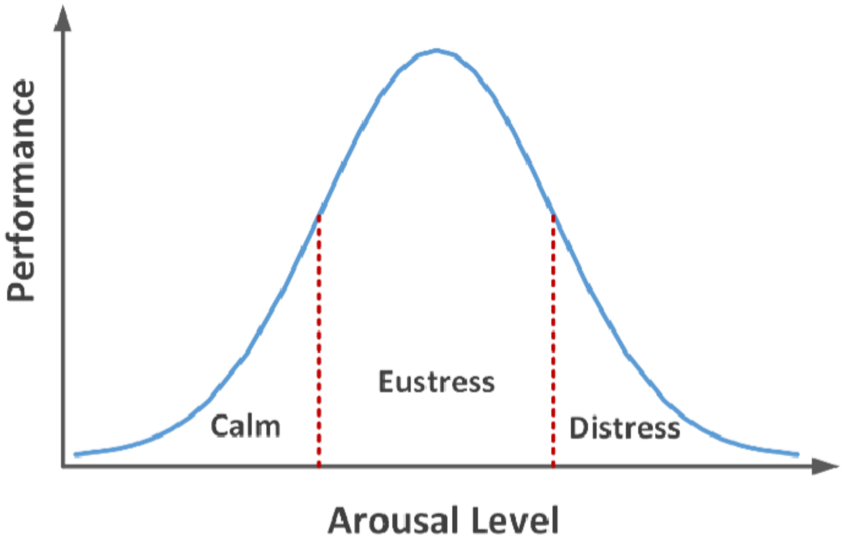

Figure 2.3: Association of arousal level and human performance. From Giannakakis et al. (2022).

However, stress can be difficult to objectively monitor. The psychological component of a stress response is inherently difficult to define and there may be a need to measure stress in multiple ways. There are different factors that contribute to the frequency in which individuals report stress when asked. Such factors can be how the individual defines stress, how the question is worded and even if it is a weekday or during a weekend (Zawadzki et al., 2019). Therefore, it seems advantageous to seek further information on how often and possibly how intense an individual experiences a moment of stress (MOS) through another source than reports from the individual itself.

2.6.1 The Stress Response

In a review on psychological stress detection using biological signals by Giannakakis et al. (2019) they explain that the stress response can be split into three main components: psychological, behavioural, and physiological. As mentioned in Section 2.6, capturing a psychological response of stress requires subjective recollections of a patient. However, such recollections are often prone to biases and inaccuracies. Capturing the behavioural response of stress is also possible. Body gestures and facial expressions are closely related to a response in stress, but even these behavioural body patterns are subject to partial control or manipulation. For these reasons, focusing on the physiological component of a stress response for detecting stress seems to be favourable.

The stress response ultimately occurs with the desired outcome to serve homeostasis, through regulating body functions such as heart activity, temperature, and respiration. All of which are functions, from an evolutionary standpoint, that are essential for survival. Other physical reactions to stressors are the production of sweat in sweat glands and a decrease in body temperature (Boucsein, 2012; Kataoka et al., 2002).

2.6.2 Measuring Stress with Wearables

To be able to easily monitor and detect stress and symptoms of mental ailments, researchers are trying to create compact and accurate devices (Hickey et al. 2021). Wearable devices are one example of such devices. These devices can hopefully reduce the economic burden on the healthcare system and reduce morbidity, e.g., by enabling intervening at an earlier stage (Steinhubl et al. 2015). There are currently numerous wearables that through an array of different sensors can measure stress. In the sections below we will outline some of the most common biological markers (biomarkers) derived from wearable sensors that may be used to measure stress.

Electrodermal activity (EDA), also known as galvanic skin response or skin conductance, is a marker of autonomic nervous system activity (Christopoulos et al., 2019), and commonly used for measuring emotional arousal. In events of high emotional arousal, the secretion of sweat is greatly increased, which can be measured on both hands and feet. In fact, unless the skin is fully saturated, there is a linear correlation between of skin conductance and emotional arousal (Boucsein, 2012). EDA can be further split up into an electrodermal response and an electrodermal level. The electrodermal response denotes a change in skin conductance. While the electrodermal level, also known as tonic skin conductance, is the base level of skin conductance. This level usually changes over time, but at a much slower rate than the electrodermal response. More on this in Section 4.3.1.

Skin temperature (ST) usually varies in a range between 32 and 35°C. It may however show an even greater variation because of factors such as fever, physical exertion, or extreme environmental temperatures (Vinkers et al., 2013). Furthermore, variations in ST are linked with conditions of stress and anxiety disorders (Kataoka et al. 2002). It is indicated that ST variations as a result of stress occur differently in different areas of the body (Giannakakis et al., 2019). A study by Vinkers et al. (2013) showed that during induced stress from a Trier Social Stress Test, subjects experienced a significant decrease in temperature of the hand palm, finger base and fingertips.

Heart rate variability (HRV) is a marker that reflects activity of the autonomic nervous system (Berntson et al., 1997). HRV is widely used for stress detection and is recognized as a sensitive and accurate indication of stress (Alberdi et al., 2016; Seoane et al., 2014). HRV can be derived from both photoplethysmography and electrocardiography sensors.

2.7 Existing Solutions

We have reviewed literature on IDPT using Google Scholar. Even though there were several IDPT systems described, there were very few that had an integration with wearable technology. The solutions we have described below are the ones we found that seemed most alike with our intended artifact. It should also be noted that none of the surveyed projects had their code base publicly available. Information regarding, e.g., the types of data formats used, would have been of great value. However, such details were not presented. As such, comparisons will consist of what is described and not how these projects are structured or designed. If possible, comparisons will be made regarding both components of our artifact. More specifically, a comparison regarding our general component and of our demonstrative component.

2.7.1 The CareWear Project

The CareWear Project is an online platform that is aimed at allowing mental healthcare professionals to use wearable data in their practice (Debard et al., 2020). The data collection is done by patients with the use of an Empatica E4. After patients have collected data, it can be uploaded onto the CareWear platform where the data will be transformed into interpretable indicators. This is typically done on a day-to-day basis. The indicators consist of moments of stress, number of steps, amount of physical exercise, as well as HR and HRV. A moment of stress is registered if patients verify the moment afterwards in the platform's interface. Additional details of the data (descriptive metadata) may also be added on the platform.

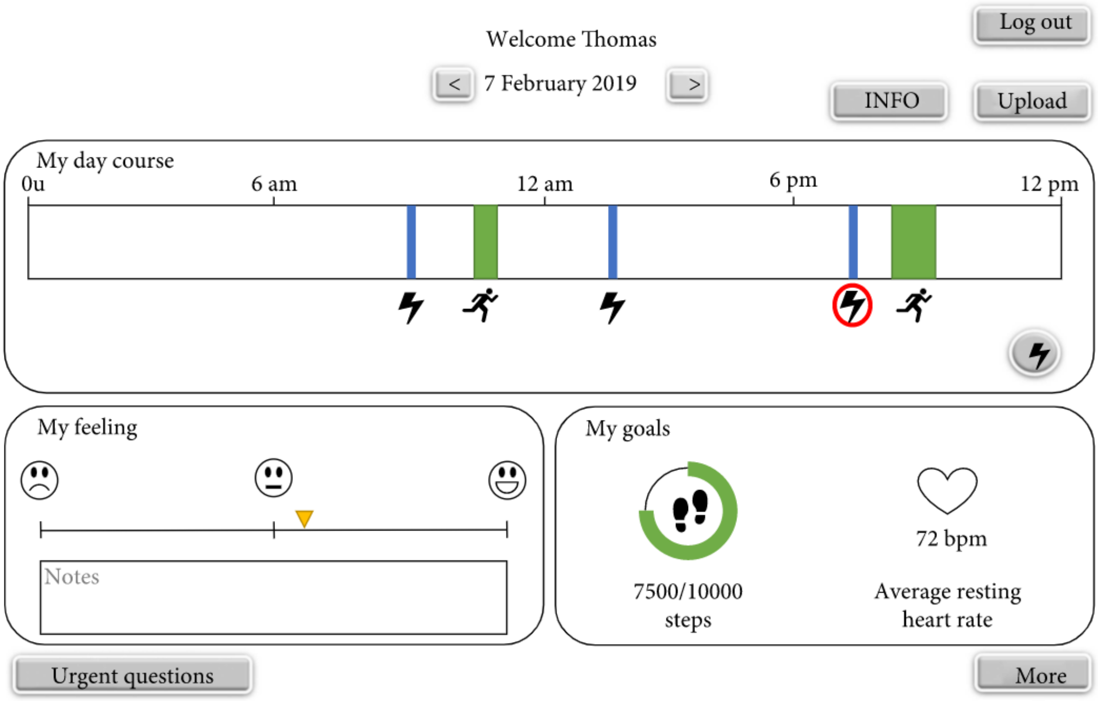

Figure 2.4: Screenshot of the The Carewear Project's platform viewing details of a patient's day. From Debard et al. (2020)

There is no available information regarding how the CareWear platform is structured or designed. As such, it is difficult to make any comparisons regarding our general component.

However, the Carewear Project has a lot of similarities as to how we intend to implement our demonstrative component. First, the project uses the Empatica E4, the same device we intend to use, for the collection of sensor data from patients. Furthermore, the project analyses EDA data to classify acute moments of stress.

The biggest difference between the project and our artifact is that the project functions as a tool to be used in addition to traditional treatment, rather than a part of a standalone online intervention.

2.7.2 The Innowell Platform

The Innowell Platform is a web-based platform that serves to assist in the assessment, management and monitoring of mental health problems, as well as the maintenance of well-being (Davenport et al., 2019). It is a platform that is focused on aiding young people, even though it is designed for and used by all populations (Iorfino et al., 2019). Mental health information is collected, stored and scored, reported back to the patient and clinician to provide a more genuine collective care (Chewning et al., 2012). The platform does not deliver treatment but instead aims to guide and support. As such, the platform excludes diagnosing, medical advice or direct treatment. With use of an activity tracker the platform records physiological data regarding activity levels and sleep quality. This is done so both the patient and clinician can make more informed clinical decisions.

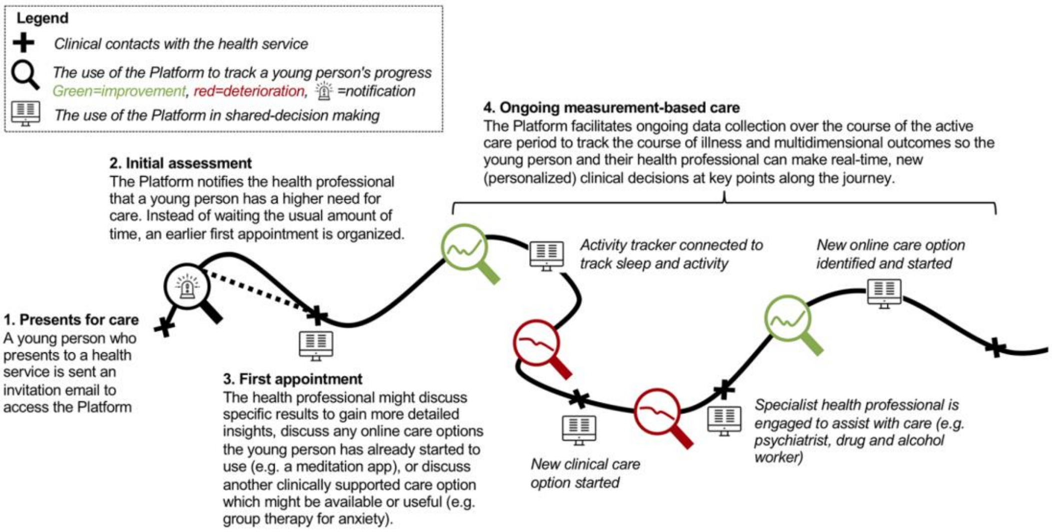

Figure 2.5: Illustrative journey of a patient through care. From Iorfino et al. (2019)

The Innowell Platform differs as it does not provide all parts of an intervention through the Internet. But rather, some parts of an intervention, such as treatment and diagnosing, is done through traditional face-to-face appointments. However, the platform utilizes measurement-based care, a term described in Section 2.4. Our intended artifact, too, will be developed with the goal of enabling measurement-based care.

However, there is no information on how the Innowell Platform is implemented to integrate wearable data, or documentation of how the platform is structured or designed. As such, no comparisons can be made regarding our general component.

As for our demonstrative component, we are not aiming to collect the same biological markers and choose to focus on stress. Yet, our component is in many ways similar. The goal of our component is also to provide continued assessment, as to provide more personalized treatment.

2.7.3 Youwell

Youwell is a platform designed to help clinicians build and deploy their own guided or unguided digital interventions (*Youwell*, 2022). The aim is to help clinicians adapt treatments to their patients without necessarily having a background in informatics. The interventions can include a wide range of material. Such material may include educational content, questionnaires, exercises and patient-clinician communication (*Youwell*, 2022).

The platform has been used as a foundation to create several online interventions. Some examples of interventions developed at INTROMAT with the use of Youwell's platform are:

- **eMeistring**, offers treatment for people with panic-disorders, social anxiety or depression above 18 years of age (eMeistring, 2022),
- **Stressmestring**, is a stress management tool for patients to learn about mental reactions to stress and anxiety, and learn how to deal with such reactions (*Youwell*, 2022), and
- **UngSpotlight**, is a digital self-help tool for young people with presentation anxiety (*UngSpotlight*, 2022).

According to Youwell, sensor data such as HR, sleep data and activity levels can be collected and used on the platform. However, there was no further explanation to as how this is carried out. After reaching out to Youwell they provided us a with a brief overview of how sensor data is integrated in the platform. They explained that the Youwell platform has an open API that can receive sensor data. As such, any wearable that exposes the collected sensor data through an API, can be integrated in interventions. What data is collected and how it is utilized is entirely up to those who are creating the intervention.

As for our general component, we also want to make it possible to use a range of different sensor data and various devices. However, we do not intend to fetch data from an API, as we believe this to be difficult with regards to interoperability, as mentioned in Section 2.5.1 and as will be further elaborated on in Section 4.2. Furthermore, Youwell does not have any information of how the wearable data are structured or stored, so no comparisons can be made here.

For our demonstrative component it is difficult to make any comparison, as there are no concrete interventions that we know of that has used the platform's ability to utilize wearable data.

CHAPTER 3

3 Research Method and Design

Throughout this chapter, we will elaborate on our chosen research method. Further, we will explain how we have designed our artifact through several iterations.

3.1 Research Method

Design science is a research method first introduced by Hevner et al. in the paper *Design Science in Information Systems* (2004). Design science as a research method is highly popular and according to Google Scholar, Hevner et al.'s paper (2004), has been cited in over 16,000 works (2022). In the book *An Introduction to Design Science*, Paul Johannesson and Erik Perjons describe design science as "the scientific study and creation of artifacts as they are developed and used by people with the goal of solving practical problems of general interest" (Johannesson & Perjons, 2014).

3.1.1 Guidelines of Design Science

Described by Hevner et al. (2004), are seven guidelines for anyone who wish to conduct effective design science research. These guidelines are based on the fundamental principles of design science and should therefore be sufficient in describing what design science seeks to achieve.

Guideline 1: Design as an Artifact

Guideline 1 emphasizes the development of an artifact, and that the artifact should be purposeful in the problem domain that intends to contribute to. Such an artifact may appear as:

- **constructs,** which provide us with the language needed to solve or to present a solution to a problem, and are also crucial for creating models and methods in the problem domain,

- **models,** which are ways of representing the problem domain itself,

- **methods,** which are collections of useful ways or processes researchers may employ to solve their problem, and

- **instantiations**, or implementation of artifacts, demonstrate a feasible solution in the problem domain.

In our thesis we aim to create an instantiation, namely an artifact implementation to extend the functionality of the IDPT framework. Our problem domain has pre-existing constructs and methods that provide a language and processes, respectively, for creating potentially useful artifacts. Therefore, we believe that an artifact instantiation would yield most value to the domain.

Guideline 2: Problem Relevance

The second guideline instructs that the artifact created will have to be relevant in reducing the difference between the goal state of a system and its current state of a system. The goal state is not necessarily a state where there are no more challenges to overcome in the system, but rather a satisfactory state. Design science aims to achieve this by creating an innovative artifact.

Guideline 3: Design Evaluation

The evaluation of the created artifact is an important part of the iterative process that is artifact design in design science. With frequent evaluation we may assist in finding a solution in the problem domain more aligned with the business environment's criteria, and thus following *Guideline 2* by being relevant.

We evaluated our artifact through a midway evaluation, where domain expert knowledge was incorporated into the feedback loop to enhance the value of the artifact. A further explanation regarding our evaluation process can be found in Chapter 5.

Guideline 4: Research Contributions

The fourth guideline explains that for design science research to be effective it must produce something of value in at least one of three areas. The first form of value is the artifact itself, which according to Hevner et al. (2004) is the most common contribution made by design science. We will summarize our own research contribution in Section 6.2.

The artifact may be innovative itself, extending the knowledge base of the domain, or it may be an original use of the pre-existing information in the knowledge base that solves a problem without a previous solution. The second form of contribution is through explicit extension and improvement of the knowledge base. Examples of this are formalizations of pre-existing models and design algorithms. The third and final type of contribution is the development of methodologies. Methodologies include metrics for quantification, which are especially important for design science research due to a focus on progress the problem domain to the mentioned goal state.

Guideline 5: Research Rigor

Guideline 5 emphasizes the use of rigor while conducting research within design science. Rigor in the context of behavioural and design science is the effective application of the methods and foundations that originate from the knowledge base.

Analysing the IDPT framework as an artifact within the domain of IDPT has given us an understanding of what technology is required as part of an IDPT system. Through discussion with one of the creators of the IDPT framework we have extended and evaluated our own knowledge to gain further insights. Through our interviews with domain experts, we have also supplied our knowledge of the relevant technologies with information about the domain and its actors.

Guideline 6: Design as a Search Process

When designing an artifact, according to *Guideline 6*, we should always try to find the optimal solution to our problem. However, the solution space when working on wicked problems, often encountered when designing information systems, is vast. Thus, formally listing the sets of methods, possible solutions and criteria and choosing the optimal solution is not feasible. As such, we are instead searching for a satisfactory solution to the problem.

When we are searching for these satisfactory solutions, attention must be paid as to how we set the threshold for a satisfactory solution.

In the context of our thesis, the solution we are searching for lives in a domain of emerging technologies and thus, comparing our artifact to existing solutions as to measure the value of our artifact is challenging. However, through qualitative evaluation we were able to get a better understanding of the usefulness of our artifact.

Guideline 7: Communication of Research

The last guideline of design science concerns the proper communication of the nature of the artifact. More specifically, with regards to developers tasked with evaluating, implementing or extending the artifact, as well as a thorough documentation of its intended utilization. Furthermore, the artifact needs to be presented in such a way that the novelty of it is apparent to a management-oriented audience, such that it may find use as a solution to an appropriate problem.

The IDPT framework, and the artifact we have created as an extension, are open-source. This should in-turn promote further evaluation and extension, by removing barriers related to the acquisition of source code which is a challenge in the field of IDPT systems. Furthermore, the presentation of our artifact within this thesis would hopefully detail its novelty in a clear manner, thus aiding in applying it to a relevant problem.

3.2 Design Process

This section is dedicated to illustrating the iterative process of our artifact development. Several shorter meetings were had with Mukhiya, the main contributor of the IDPT framework, throughout development. In addition to these meetings we also had a midway evaluation, presented in Section 5.1. These meetings were valuable for the design of our artifact, and how we ended up implementing the artifact. To present the design process in a clear manner, we will explain the process through three primary iterations. The iterations were focused on:

(i) designing a general wearable data model for the IDPT framework as our general component,

(ii) using our general component to develop a demonstrative component, and

(iii) improving upon our components by adding features based on the evaluation meeting.

First Iteration

The first iteration consisted of several attempts at conceptualizing the different components of our implementation. Initially we aimed to create an artifact that would allow the use of any type of sensor data from any device in an intervention. However, as later learned, this would prove to be challenging with regards interoperability. Thus, we decided to create a general data model for wearable data, rather than creating an interface for receiving wearable data from any device.

The foundation required for using wearable data in interventions, is the construction of data models to be used in both front-end and back-end of the system. The initial model was one-dimensional and based on the existing components of the IDPT framework, such as patients and tasks. We figured what attributes we thought be valuable for any type of wearable data and based our data model on these criteria.

Second Iteration

The second iteration started with the goal of creating an example of how our general component could be used in interventions, by implementing a demonstrative component. We decided on using stress as a measure to assess patients, due to the correlation of stress and mental health being well established. As such, we created a new data model, for types of scored data. The scored data concept was developed with the intention of being simple, and thus easy to visualize and represent.

A new module was implemented in the back-end to analyse and detect stress. This module would take the wearable data from the first iteration as input and create scored data based on the frequency of detected stress as output. The module was first prototyped in Python and later translated into JavaScript, as the back-end of the IDPT uses JavaScript executed using the Node.js runtime environment. The demonstrative component developed in this iteration would be further improved upon in the third and final iteration.

Third Iteration

Through the midway evaluation with Mukhiya, further explained in Section 5.1, several areas to improve with regards to both our components were illuminated. One such suggested improvement was the addition of a graph representation of scored data to better be able to interpret the data in a meaningful manner.

Furthermore, through insights of the midway evaluation, we decided to redesign the data models we had created to conform to a standardized data format. There is no consensus on a universal data format in health informatics. However, the standard of HL7 FHIR has shown potential as to solve interoperability issues in health informatics, as mentioned in Section 2.5.1.

CHAPTER 4

4 Implementation

When initially choosing a topic and a project to work on for our thesis, we decided that we wanted to contribute to the existing IDPT framework, as described in Section 2.3. Further, we had an interest in wearable technology and the possibility of these devices to assist within the domain of mental health care. Therefore, we started researching existing literature on wearable technology and IDPT. We found that wearables have the potential to monitor a wide range of physiological signals and that many of these have correlations to symptoms of mental health problems (Hickey et al., 2021; Lu et al., 2020; Zamkah et al., 2020). We also learned that IDPT has issues regarding low user adherence, largely due to a lack of personalized treatment (Fernández-Álvarez et al. 2017). By using wearables to conduct EMA of patients during an IDPT, one could hopefully capture measures that give an indication of how a patient is progressing in treatment. In turn this knowledge could be used to facilitate IDPT's adaptability to further personalize treatment.

Our primary contribution is an extension of the IDPT framework, as mentioned in Section 2.3, that allows for the utilization of wearable data. The intention behind making such an extension enable use of data that is captured outside of a clinical setting, through the use of EMA. Additionally, due to our contribution being open-source it would enable reuse and promote further research in this domain. Further, we also wanted to make a demonstrative to show an example of how wearable data can be utilized in IDPT. In our demonstrative component we look at biomarkers that can give an indication of a patient's frequency of stress. The reason for wanting to monitor stress, is because how the amount of stress a person experiences is greatly correlated to their mental well-being (Cohen et al., 2007; Daviu et al., 2019; Hammen et al., 2009).

In following sections, we will detail the implementation of our artifact. First, we touch on the architecture of the IDPT framework and technologies used. Further, we elaborate on our general component. Last, we describe our demonstrative component.

4.1 Technologies and Architecture of the IDPT Framework

4.1.1 Technologies

The following section will mention the most essential technologies of our artifact. As our artifact is an extension of an existing framework, the technologies were predetermined as we started development. For this reason, we will only briefly touch on the technologies used.

MongoDB is a NoSQL document database that is used for storing the IDPT framework's data.

Mongoose is an object data modelling (ODM) library for mapping objects in Node.js to MongoDB in the IDPT framework.

Node.js is an open-source runtime environment for JavaScript that is for the used for the IDPT framework's back-end logic.

React is an open-source front-end JavaScript library for creating user interfaces made up by components that is used for the IDPT framework's user interface.

GraphQL is an open-source query language that is used for communication between the back-end and front-end within the IDPT framework.

4.1.2 Architecture

This section contains a description regarding the architecture of the IDPT framework. As mentioned in the previous section, our artifact is an extension of an existing framework and thus the architecture was implemented prior to the development of our artifact. As such, the breakdown of the architecture will be short.

The architecture of the back-end is comprised of back-end endpoints, services, repositories and the database. A repository performs operations of the database, such as storing and fetching, for an entity. Examples of entities in the framework are users, modules, and wearable data. Services similarly perform logic operations related to the entities. These operations are invoked by the back-end's endpoint. To this endpoint, GraphQL operations can be sent from external sources, typically the IDPT's front-end.

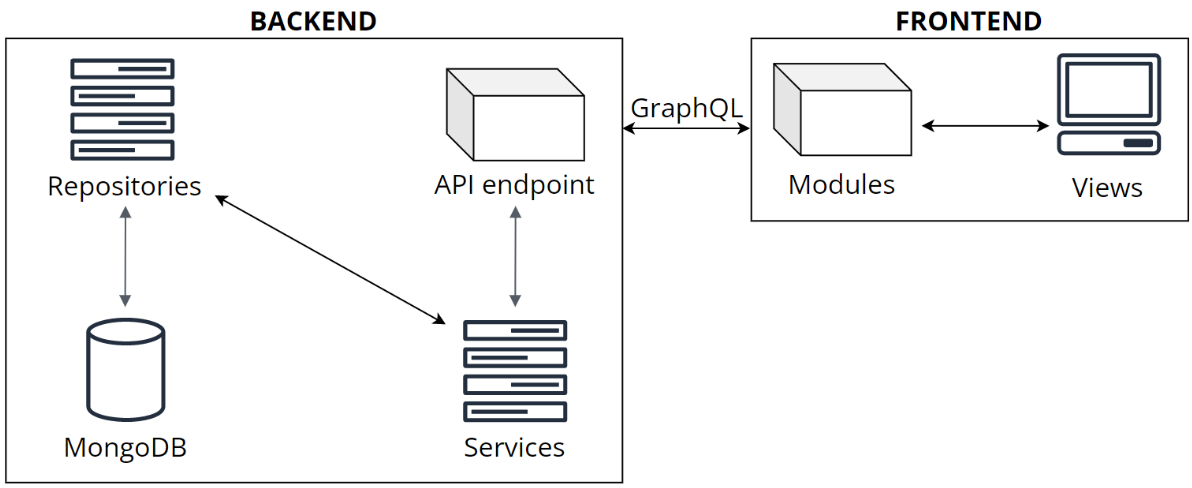

Figure 4.1: Illustration of the IDPT framework.

The structure of the front-end is alike that of the back-end. In the front-end too, each entity has a corresponding service in the modules component, as seen in Figure 4.1, that carries out logical operations on entities. Further, the front-end has a views component. Views is made up of React components which together makes up the user interface.

Further, the IDPT framework handles privacy of data by assigning privileges to different kinds of users. E.g., users assigned as patients are much more limited with regards to what can be accessed, compared to clinicians.

4.2 General Component

In this section, we will present the implementation of our general component. This component enables the use of wearable data in interventions, as an extension of the IDPT framework. The component is developed to be generic and reusable. There are obviously overlapping elements between our general component and demonstrative component as our demonstrative component is built using our general component. Still, elements regarding the user interface will be explained in Section 4.3 regarding our demonstrative component.

4.2.1 Additions to the Existing Architecture

As mentioned in Section 4.1, the technologies and architecture were largely predetermined as we started development. Still, in implementing our general component we added two new entities to the existing architecture, that are similar in structure to that of previous entities (e.g., patients and cases), and data-flow logic related to these entities. More specifically, we have developed entities for wearable data and scored data. These entities and their implementation will be further elaborated on in Section 4.2.2.

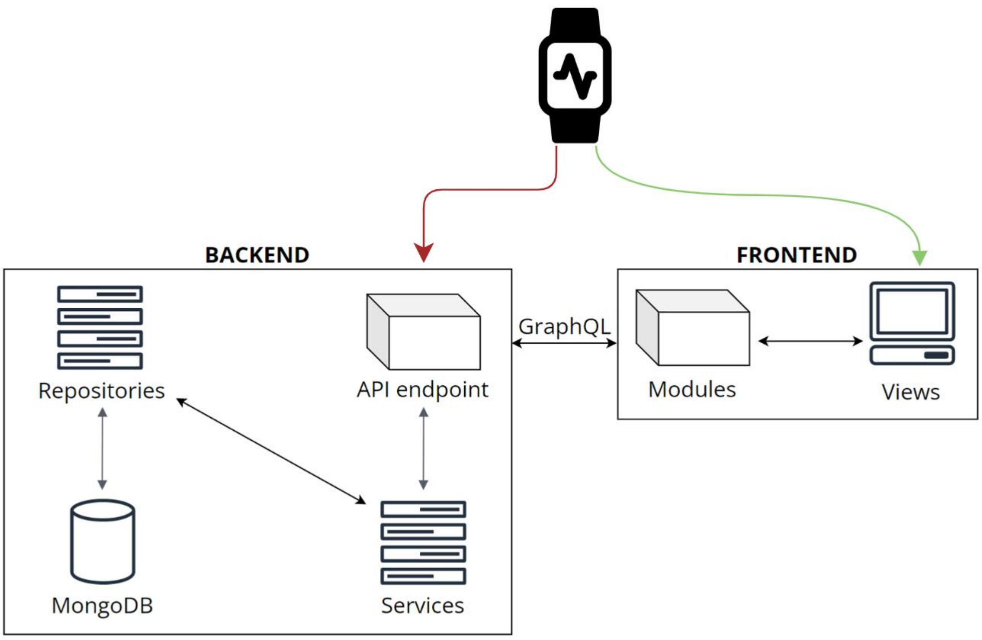

Figure 4.2: Wearable data integration in the architecture of the IDPT framework. The two arrows illustrate different scenarios of receiving wearable data.

Figure 4.2 illustrates how wearable data may be received in the IDPT framework. Or more specifically, which component may be responsible for receiving the data. The green arrow of the figure refers to the scenario of uploading wearable data to the framework through a user interface in the front-end, as is done in our demonstrative component. More on this in Section 4.3.5.

The red arrow refers to the scenario of automatically uploading wearable data, as is currently supported by the API endpoint of the framework. Although, for such a scenario to be possible, the uploaded wearable data would need to be structured according to our model for wearable data. This model is later described in Section 4.2.2. Additionally, the wearable device utilized would need an API suitable for transmitting the data from the device to an endpoint.

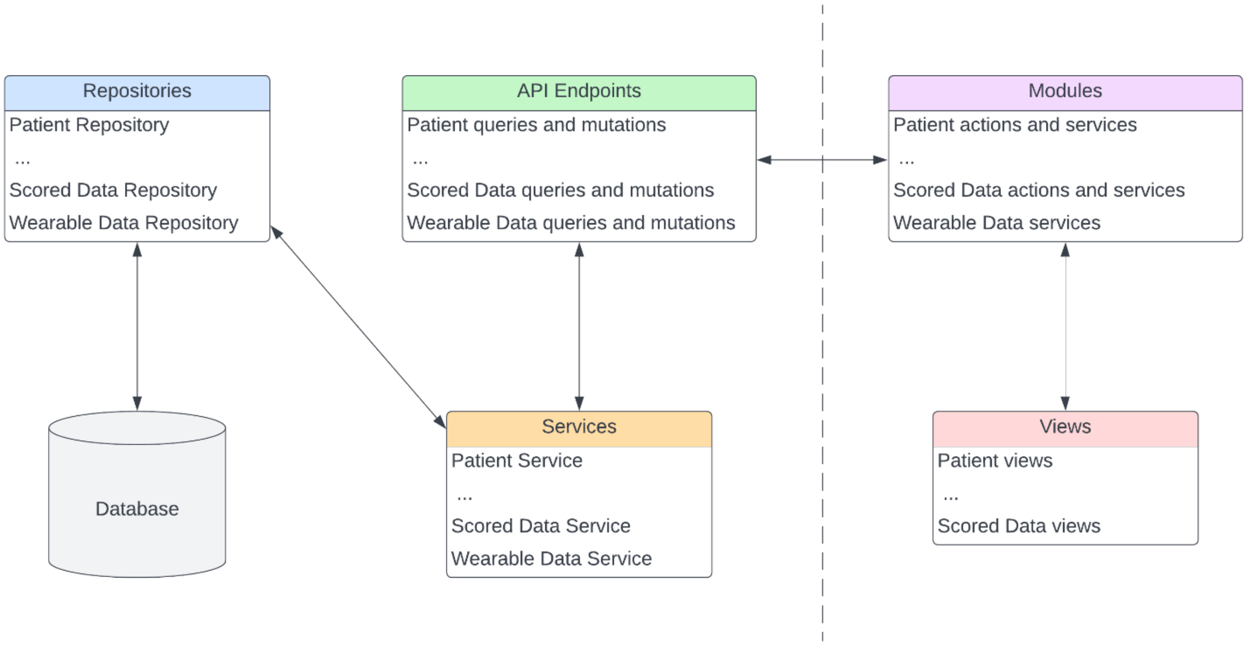

Figure 4.3: A more detailed version of Figure 4.2, showing components and their content related to entities, of the IDPT framework. More specifically, this figure is showing additions to each component by scored data and wearable data. Note that there are other entities that have been excluded as to keep the figure uncluttered.

4.2.2 Implementing a Wearable Data Integration

When referring to wearable data and scored data we are talking about unprocessed and processed or analysed sensor data from wearable devices, respectively. Scored data is a suggested way of formatting such processed wearable data. We designed this format with the intention of having a score that is easy to visualize and that emphasizes change or progression. However, the data format could easily be altered and structured otherwise if needed. More on scored data in relation to stress scores in Section 4.3.

To promote interoperability, we decided on using the HL7 FHIR specification for our models. This was as one of our takeaways from the midway evaluation. As such, we are opting for the use of an established and standardized data format rather than developing a proprietary format. More on this in Section 5.1. As mentioned in Section 2.4.2, a lack of standards is a challenge when it comes to utilizing wearable technology.

It should be noted that the scored data has, in addition to its FHIR specified model, a separate front-end data model similar to that of other entities, such as patients and tasks. As wearable data provides little value before it is processed and analysed there is no need for a front-end model for this entity. Therefore, we chose to separate scored data from wearable data as different entities, because their uses cases are different. To allow for the use of two separate models for scored data, we have implemented a module for transforming scored data between the FHIR specified model and the model used for representing scored data in the front-end, located within the modules component seen in Figure 4.3.

The FHIR specification defines its components as resources. In compliance with the HL7 FHIR's *Guide to Resources*, we chose to utilize the *observation* resource to represent wearable and scored data (*FHIR*, 2022). Following is a brief explanation of *resource* and *observation*.

Resource is a component of the FHIR data specification. Resources are hierarchical components that often have several optional sub-components. The modularity of the resources makes creating dynamic objects conforming to the standard easier.

Observation is a type of resource that embody the concept of patient observations. The observation comprises several required and optional fields for data and metadata. In our case, observations specifically encapsule sensor data, either raw data or analysed data.

We decided to omit some of the optional fields of the observation resource that we hypothesize will not be necessary for the IDPT, to reduce complexity. The same complexity reduction has been applied to child resources of the observation.

We have created two variations of the observation resource. First, a variation for wearable data including an attribute named *valueSampledData,* containing a list of raw sensor data. This data can e.g., be one-dimensional EDA data in microsiemens (μS) or three-dimensional gyroscope data. We denote the data with the name of the device used for measurements, the name of measurement type and a short description of the measurement type. More metadata regarding the measured data can easily be added to this observation variation if it is believed to be useful.

The other variation of the observation resource represents scored data. Such a score can easily be visualized and represented to a therapist or patient and may be used to facilitate adaptation (either manually or automatically). Scored data share most of its attributes with wearable data except for the *valueSampledData* attribute. Rather, the scored data has an attribute named *valueString*, which holds a single value representing the score. The scored data is used by the demonstrative component of our artifact and will be further elaborated on in Section 4.3.

All top-level attributes included in wearable as well as scored data are described below:

- ***resourceType*** describes the type of resource.
  In our case the type of resource is an observation.
- ***status*** indicates if the observation is complete or undergoing.
  We have set status to be final by default.
- ***code*** is a resource explaining what was observed. In the code resource we find metadata such as a systems code for a measurement (e.g., a number indicating a measurement of type ECG), a more readable representation of such a code and the system from where the code is defined.
- ***subject*** is a resource representing the patient that is measured.
- ***effectivePeriod*** indicates the timeframe where measurement has taken place.
- ***device*** is a resource containing information about the measurement device.
- ***valuedSampleData*** is a resource containing the measured data and related metadata (e.g., the frequency of samples and dimensions of the data array).
  This attribute is only found in wearable data.

- ***valueString*** is a value representing the result of analysed data.

  This attribute is only found in scored data and is represented as a numbered score.

- ***derivedFrom*** is a list of references to the wearable data that has been processed.

  This attribute is only found in scored data.

```
{
        resourceType
        status
        code {
                coding {
                        system
                        display
                }
                text
        }
        subject {
                reference
                type
                display
        }
        effectivePeriod {
                start
                end
        }
        device {
                display
        }
        valueString
        derivedFrom {
                references
        }
}
```

Listing 4.1: A representation of the structure of the model for scored data, complying with the FHIR specification. Levels of the hierarchy are uniquely coloured.

As we want to store both wearable data and scored data, we have corresponding Mongoose and GraphQL schemas for both wearable and scored data models. Schemas related to GraphQL are part of the API endpoints component, whereas Mongoose schemas are part of the repositories component. The Mongoose schemas and GraphQL schemas belonging to wearable and scored data contain an attribute named *fhir* which is an instance of the *observation* resource. Additional attributes found in these objects are *id*, *createdAt* and *updatedAt*, as seen in Listing 4.1. These were included to comply with the standard attributes that can be found in all main entities (e.g., patients and tasks) of the IDPT framework. In the future more of these entities and their attributes have the potential to be specified by the FHIR standard.

Figure 4.4: Scored data as represented in MongoDB Compass, a graphical user interface for MongoDB databases. The IDs which are listed in the array under the *derivedFrom* attribute, are references to the wearable data database entries in Listing 4.2.

```
_id: ObjectId('6275233d6a513d5a7087550c')
⌄ fhir: Object
    resourceType: "Observation"
    status: "final"
  ⌄ code: Object
    ⌄ coding: Object
        system: null
        display: "TEMP"
      text: "Skin temp. data in degrees celsius"
  › subject: Object
  › effectivePeriod: Object
  › device: Object
  ⌄ valueSampledData: Object
    › data: Array
    › origin: Object
      period: 250
      dimensions: 1
createdBy: ObjectId('619b7349fa0e462eccd3d4d0')
updatedBy: ObjectId('619b7349fa0e462eccd3d4d0')
__v: 0
```



```
_id: ObjectId('6275233d6a513d5a7087550a')
⌄ fhir: Object
    resourceType: "Observation"
    status: "final"
  ⌄ code: Object
    ⌄ coding: Object
        system: null
        display: "EDA"
      text: "EDA data in microsiemens"
  › subject: Object
  › effectivePeriod: Object
  › device: Object
  ⌄ valueSampledData: Object
    › data: Array
    › origin: Object
      period: 250
      dimensions: 1
createdBy: ObjectId('619b7349fa0e462eccd3d4d0')
updatedBy: ObjectId('619b7349fa0e462eccd3d4d0')
__v: 0
```


Figure 4.5: Wearable data for skin temperature and electrodermal activity.

```
const ScoredDataSchema = new Schema(
 {
   fhir: {
     type: FHIR.ObservationSchema,
     required: true,
   },
   createdBy: {
     type: Schema.Types.ObjectId,
     ref: 'user',
   },
   updatedBy: {
     type: Schema.Types.ObjectId,
     ref: 'user',
   },
 }
);
```

```
const ObservationSchema = new Schema(
  {
    resourceType: {
        type: String,
        required: true,
    },
    status: {
        type: String,
        required: true,
    },
    code: {
        type: CodeSchema,
        required: true,
    },
    subject: {
        type: SubjectSchema,
        required: true,
    },
    effectivePeriod: {
        type: EffectivePeriodSchema,
        required: true,
    },
    device: {
        type: DeviceSchema,
        required: true,
    },
    valueSampledData: {
        type: ValueSampledDataSchema,
        required: true,
    },
    valueString: {
        type: Number,
        min: 0,
        required: true,
    },
    derivedFrom: {
        type: DerivedFromSchema,
        required: true,
    },
  },
  {
    _id: false
  }
);
```

Figure 4.6: The Mongoose schema for scored data with associated *ObservationSchema* as subdocument.

```
type WearableData {
    id: String!
    fhir: FhirObservation!
    createdAt: DateTime
    updatedAt: DateTime
}
```

Listing 4.2: GraphQL type for wearable data.

Further, there are several GraphQL schemas associated with wearable data and scored data. An example of such a schema is the *WearableData* type as seen in Listing 4.4. This type sets restrictions for how wearable data is to be structured. Each *WearableData* must contain a *FhirObservation* type with metadata and data from a wearable measurement. Like for all other entities in the system, the *FhirObservation* maps directly to the Mongoose *ObservationSchema*. Due to the *FhirObservation* type being identical to the Mongoose *ObservationSchema* in structure, it is not included as a listing. The overlap in structure between GraphQL and Mongoose schemas leads to code redundancy in the IDPT framework. Moreover, this redundancy slows down implementation speed and allows for the risk of making inconsistencies between the corresponding structures, which may introduce bugs. Yet, as the structure of the framework was predetermined when we started development, our additions to the framework have followed the same structure. Such redundancy issues could be an area for future improvement.

4.3 Demonstrative Component

In this section we will detail the implementation of our demonstrative component. As mentioned in Section 4.2, there are obvious overlapping elements between this component and our general component. However, the topics included in this section we believe to be preferably explained in a practical setting, and therefore have been included here rather than in Section 4.2.

We will first present a brief overview of our additions to the IDPT framework's architecture regarding our demonstrative component. Next, we will elaborate on our algorithm for detecting moments of stress (MOS) and how we carried out an experiment to calibrate the algorithm. Further, we will touch on the device chosen for collecting the necessary biomarkers. Last, we will show how the scored data can be viewed and interacted with in the front-end.

4.3.1 Additions to the Existing Architecture

As mentioned in Section 4.2.2, we have added two entities and data-flow logic regarding these entities, to the existing architecture. Additionally, we have for our demonstrative component created a view for uploading wearable data, a view with a graph visualization for scored data and an analysis module related to stress detection.

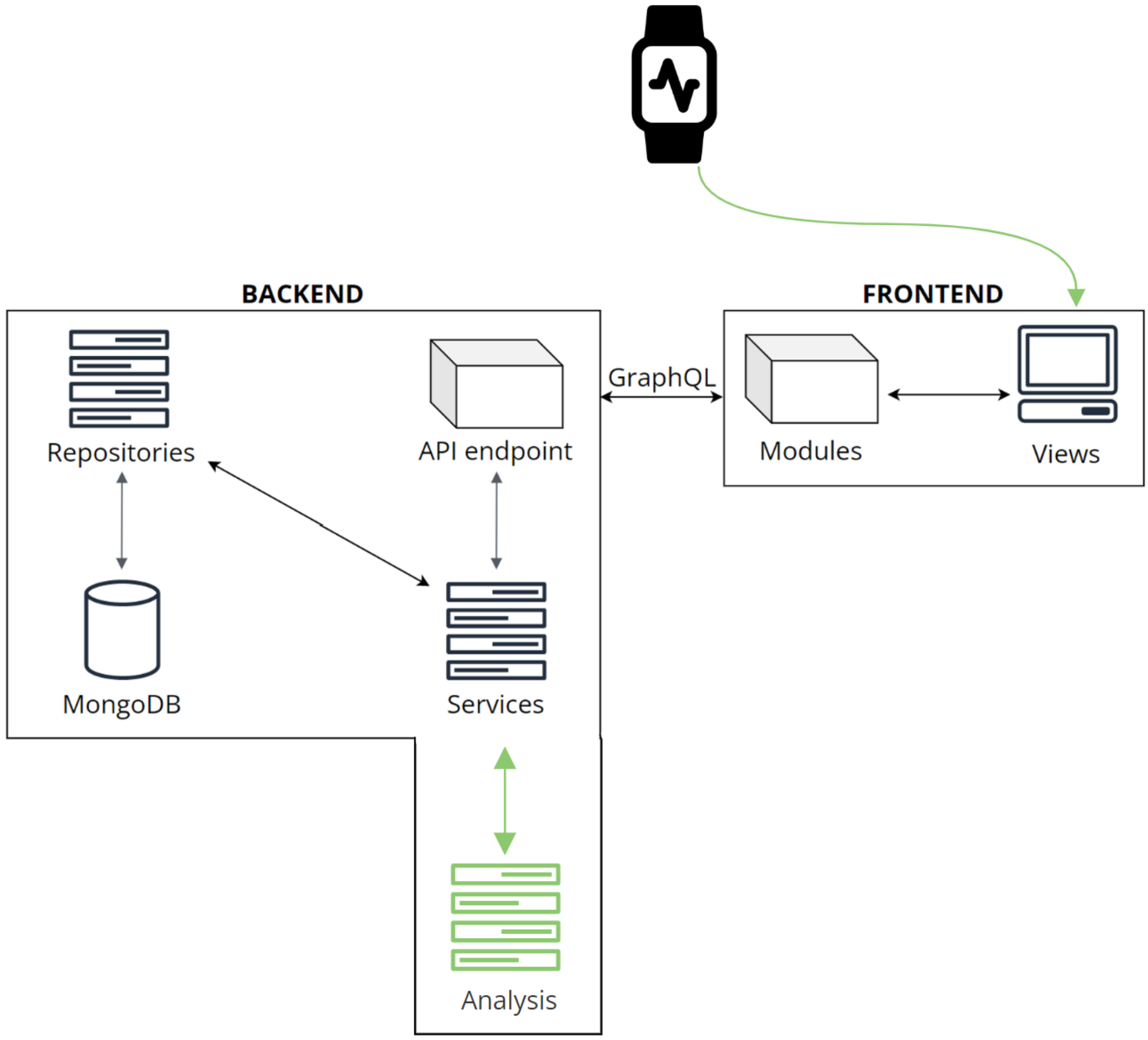

Figure 4.7: Additions of our demonstrative component to the existing architecture of the IDPT framework, highlighted in green.

The additions to the architecture by our demonstrative component revolve around the analysis of wearable data, and the creation and representation of scored data. Moreover, we have added a view for uploading wearable data, which can be seen in Figure 4.9. Further, we have implemented another view containing a graph for scored data, to visually emphasize the variations in scores over time. This view is later presented in Section 4.3.5 and illustrated in Figure 4.11.

Last, we have integrated a module for detecting stress using wearable data. More on this in Section 4.3.2. The module is utilized in the service component for wearable data, as seen in Figure 4.7. The analysis module receives wearable data as input and outputs scored data. The wearable and scored data entities have similar data-flows starting from their respective service components and ending up in the database, where they are stored.

Users assigned the role of patient only have access to their own scored data, whereas clinicians may access each of their patients' data. Besides these measures, our demonstrative component does not address any other concerns regarding privacy and security, as this is outside the scope of this thesis. Yet, such concerns are important and thus are mentioned as future work in Section 7.2.

4.3.2 Detecting Moments of Stress (MOS)

As we wanted to implement a demonstrative component to show how wearable data may be utilized in IDPT, we selected stress as a suitable measure to integrate. The reasoning for choosing this measure is due to its acknowledged correlation with mental health, as mentioned in Section 2.6.

Due to time constraints and the fact that our demonstrative component is a proof-of-concept, we decided to use an already established algorithm. In this way, we would be able to use an algorithm that has been evaluated, as we would not have the resources nor time to properly evaluate an algorithm ourselves.

We searched for scholarly papers regarding the use of biomarkers for detecting stress. This is a relatively well researched topic. Still, when finding relevant papers, the papers rarely describe what kind of algorithm was used or include description regarding their implementation. When searching for papers we primarily used Google Scholar and employed relevant search terms and operators for stress detection and wearables.

There are various different types of sensor data that has been used to detect stress. Among the most common types of sensor data for stress detection are heart rate, heart rate variability, electrodermal activity (EDA) and skin temperature (ST) (Giannakakis et al., 2019), all of which has been shown to be possible candidates for detecting stress (Chandra et al., 2021; Hui & Sherratt, 2018; Liew et al., 2016; Vinkers et al., 2013).

However, among the search that resulted in numerous scientific papers, we decided to base our stress detection algorithm on a paper written by Kyriakou et al. (2019). The algorithm uses EDA and ST and has been evaluated through showing high accuracy (84% on average) in a mixed method approach in a real-life setting. Furthermore, the paper includes a description of how the algorithm is implemented, which most other papers lack. It should be noted that even though it may have been beneficial to include other types of sensor data, we have chosen to narrow our use to EDA and ST, as our demonstrative component serves as a proof-of-concept. More on the addition of other sensor data in future work, in Section 7.2.

Description of the Algorithm

The algorithm from Kyriakou et al.'s paper (2019) is making use of sensors in an Empatica E4 smartwatch (more on this device in Section 4.2.3) for continuously measuring electrodermal activity (EDA) and skin temperature (ST) for the detection of moments of stress (MOS). The algorithm is rule-based and uses weights to decide how much of an impact each rule should have. An advantage of rule-based algorithms is the low time and computational cost. Below is a detailed description of what the algorithm entails.

Step 1 – Pre-processing In the first step we apply filtering and down-sampling to denoise the data before applying the rules. The steps taken for pre-processing EDA data is almost identical to that of the pre-processing for ST. The only difference being the order and cut-off frequency of the filters for EDA and ST.

First, a low pass filter is applied. The low pass filter removes high frequency changes in the measurements. These high frequency changes occur because of sensor noise. Such rapid changes in skin conductance and skin temperature are unrealistic (Schumm et al., 2008).

Next, the data is passed through a high pass filter. The high pass filter filters out slow changes in the data that are not indications of MOS. Finally, we down-sample the data from four hertz to one hertz. To prevent data loss the down-sampled data is calculated by averaging every four data points from the original data.

Step 2 – Applying the rules Following is an explanation of the five rules that iteratively adds a score for every second of data. Through summing these scores and comparing this sum to a threshold, the algorithm determines whether a point in time is an MOS or not.

- **Rule 1 – EDA Amplitude Increase**

  An increase in amplitude of EDA ranging from two to five seconds after a stimulus that could indicate a MOS. This is due to the response from a stressful stimulus first having a short delay, followed by a short rise time. When the increase is between two to five seconds, the score of 1.0 is given. Furthermore, a longer increase in amplitude is indicative of a weaker response. Therefore, an increase longer than five seconds is less likely to be a MOS and is given a score of 0.5 points.

- **Rule 2 - Skin Temperature Decrease**

  A drop in skin temperature is expected to take place three seconds after the onset of an EDA amplitude increase. If the decrease lasts for three seconds or longer, a score of 1.0 is given. A skin temperature decrease that lasts for three seconds or longer, within three to six seconds of an EDA amplitude increase results in a score of 0.5.

- **Rule 3 – EDA Response Rising Time**

  The rising time of the EDA response is relevant for a potential MOS. A mentioned in *Rule 1*, a longer rise time is correlated with a weaker response. Thus, if the rise time from a local minimum to a local maximum is less than or equal to five seconds, we give a score of 1.0. Rise times that are longer but no longer than 10 seconds are given a score of 0.5 points.

- **Rule 4 – EDA Response Slope**

  The slope of the EDA response is correlated with the intensity of a stressful event. An angle greater than 10° gives the score of 1.0. An angle between 8° and 10° results in a score of 0.5, as this was also measured where MOS were induced in test subjects.

- **Rule 5 – MOS Duration**

  Kyriakou et al. argues that a moment of stress may only occur once every 10 seconds. This is based on the typical durations of EDA latency (1-5 seconds), rise time (1-5 seconds) and recovery time (1-10 seconds). Recovery time is the duration from the peak value of an EDA response until it has recovered to 50% of the amplitude before the response.

  We implemented this rule by first finding a target score, which starts out as the highest score. Further, we remove all potential MOS with a lower score than the target score, in the following 10 seconds. Then the process repeats with the target score iteratively being decremented in steps of 0.5. The process ends when the target score is less than the threshold.

Step 3 – Scaling rules using weights The rules are weighted such that the maximum summed score is 100.0. As neither the threshold nor the weights Kyriakou et al. used to scale their rules are described in their paper, we had to calibrate these based on an experiment performed ourselves. More on this in Section 4.2.2.

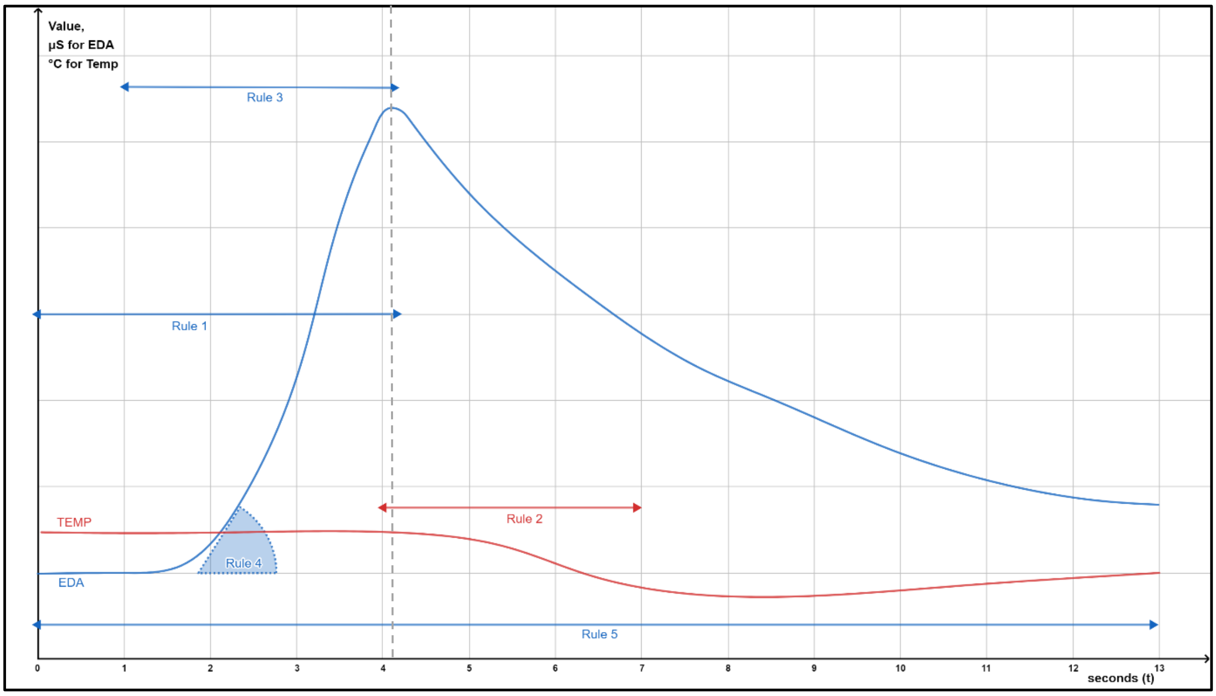

Figure 4.8: Graph showcasing the different time windows for each of the rules. A hypothetical stress event occurs at t = 0.

Stress Score

As mentioned in Section 4.2.1, we want to represent the frequency of perceived stress as a single number, namely a score. This score is calculated by dividing the number of moments of stress detected by the algorithm by the duration of the recorded data, in hours. The stress score indicates the frequency of MOS' a patient experiences in a day. It should be noted that the amount of stress a person experiences is highly individual and differ among people (Schabracq & Cooper, 2003). There is no conclusive research that defines a "healthy frequency of stress events" a person should experience in a day. Neither does a person who encounter a higher number of stress events per day necessarily face greater health problems. On the contrary, a person who experiences a far below average number of stress events, might show signs of disease, as mentioned in Section 2.6. Thus, we chose to emphasize the change of stress over time, rather than specific values.

4.3.3 The Empatica E4 Smartwatch

As we decided to use the algorithm by Kyriakou et al. (2019), we needed a device that was able to both record EDA and ST. We initially wanted to use a Fitbit Sense smartwatch, since it had both the sensors we needed and is a cheaper alternative compared to other devices. However, after doing some research we became aware of the fact that the Fitbit's API was very limited with regards to what kind of data could be exported. Specifically, the API did not allow for exporting EDA data. As such, the Empatica E4 proved to be only the viable alternative.

Still, the Empatica E4 is superior in the fact that it records EDA continuously while worn on the wrist, compared to the Fitbit Sense that requires its user to place their opposing hand on the bezel around the watch face to record EDA. Yet, it should be noted that the Empatica is considerably more expensive than other wearables on the market and that its design may be somewhat less appealing. The price of the Empatica E4 is 1,690 USD as of the time of writing (*Empatica*, 2022) and a study regarding The CareWear project concluded that participants found the device to be rather large and wished the device had a screen to provide direct feedback (Debard et al., 2020).

Figure 4.9: Picture of the Empatica E4 and some of its functionalities. From Empatica (2022)

The Empatica E4 is a medical-grade wearable device that offers real-time data acquisition. The device is equipped with a range of different sensors, including a photoplethysmography sensor, an infrared thermometer, and an EDA sensor. The data collected from the Empatica is typically accessed from Empatica Connect, which is a web portal for managing and downloading collected sensor data. The data is pushed to the servers from the Empatica E4 either via a smartphone connected using Bluetooth or by using a USB connected to a computer. The sensor data is downloaded from the web portal as a compressed folder containing several comma-separated values (CSV) files. Unfortunately, the Empatica platform is missing a public data endpoint, and thus data is not able to be automatically retrieved.

4.3.4 Laboratory Experiment to Calibrate the Algorithm

In this section we will cover an experiment we conducted as part of calibrating the MOS detection algorithm. Presented is the experiment in theory, our motivation, the process of conducting it and the results garnered from the experiment.

The algorithm described by Kyriakaou et al. (2019) did not include information regarding a threshold nor information regarding weights given to each rule. Therefore, we wanted to gather some data in a controlled experiment that hopefully induce distinct MOS, such that we could tweak the threshold and weights of our algorithm. In addition, we wanted to gather information regarding the subjects' stress levels through self-assessment questionnaires to see if this information would in any way be correlated to the subjects' recordings. However, it should be noted that this is in no way a conclusive experiment and does not serve as an evaluation of the algorithm. Rather the experiment served primarily to calibrate our algorithm.

We chose to induce stress in the subjects through auditory stressors, as has been carried out before (Grundlehner et al., 2009; Wijsman et al., 2013). In this way, we would know at which specific times of recorded EDA and ST data we should expect to see a response indicative of a MOS.

The Process

The subjects consisted of five students. Prior to the experiment the subjects filled out a Perceived Stress Scale (PSS) questionnaire. The questionnaire is "designed to measure the degree to which situations in one's life are appraised as stressful" (Cohen et al., 1983). The PSS questionnaire is among the most used self-reporting questionnaires for stress, is believed to provide acceptable accuracy of a subjects perceived stress, and is fairly easy to use (Lee, 2012).

The subjects were seated in a room by themselves wearing an Empatica E4. The device was placed on their non-dominant hand and the subjects were told to keep that hand still throughout the experiment, in order to collect data with the least noise as possible (as touched on in Section 2.5.2). The subjects were told to sit still and relax for ten minutes, with no distractions. For the first five minutes nothing happened, and we used this recorded data to measure a subject's baseline of ST and EDA. However, after five minutes we started playing abrupt and loud sounds at planned intervals from a speaker hidden in the subject's room. As the subjects were not aware of the sounds on beforehand, the sounds would startle the subjects and serve as auditory stressors. In the span of the last five minutes, a total of five auditory stressors were played at seemingly random intervals.

Immediately after the experiment, the participants were asked to answer the following three questions:

- **(Q1)** How stressed did you feel as the experiment began? (0/10),
- **(Q2)** How stressed did you feel upon hearing the first sound? (0/10),
- **(Q3)** Did you experience the sounds as less stressful after multiple sounds were played? (Yes/No).

Acknowledgements

There are several factors that may lessen the integrity of the results from an experiment like this. We were made aware from roughly half of the participants that they were familiar with our thesis being related stress and wearables. In turn, we were told, this led to the participants expecting to be subjected to a stressor during the experiment.

Another factor that may have affected our results is physical strain. The experiment was carried out on the second floor, which meant most participants had to walk up a flight of stairs before the experiment. Some participants were also arriving at the premise on foot, meaning their core temperatures where likely higher than usual, and likely their skin conductance too. We tried to mitigate this by letting the subjects rest for the first five minutes of the experiment, before the auditory stressors started playing.

Important to mention is that, for privacy concerns, all recorded data of subjects are not linked to the subjects' identity. Each set of recorded data associated from a subject was given a number and only identified by this number after recorded. Further, these sets of data are not to be published, but rather their results are presented in this thesis.

Results

In the PSS questionnaire subjects scored 6, 10, 13, 16 and 23. For the post experiment questionnaire, the average of all subjects' answers on Q1, Q2, was 1.6, 4.8, respectively. As for Q3 all subjects answered "Yes".

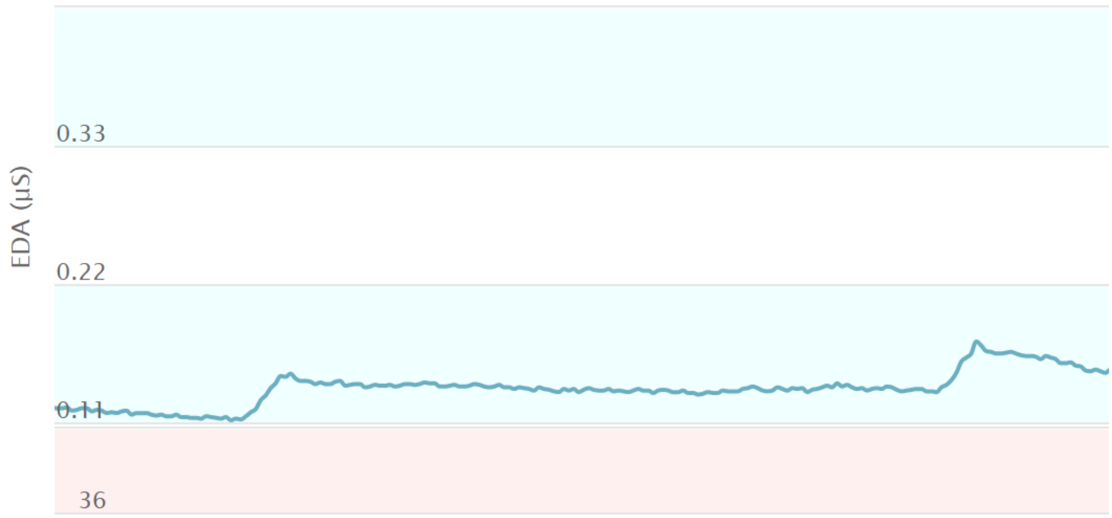

Figure 4.10: Example of a subjects EDA response to two auditory stressors (indicated by the two peaks with the highest amplitude)

As for the recorded data from the Empatica E4, we often saw what we thought to be clear correlations between the data and expected moments of stress, as seen in Figure 4.5. However, our algorithm did not perform as well as anticipated. Even though the algorithm usually indicated a moment of stress at an expected moment of time, the algorithm also found many more moments of stress that we believe to be false positives. Even after tweaking both weights and threshold, we did not get the algorithm to perform as we had hoped. Because of this, we could not draw any meaningful correlations between the questionnaires results and the wearable data either.

The exact reason for why the algorithm did not perform as expected we have not been able to conclude. It may be due do not finding the best set of weights or the correct threshold, as done by Kyriakaou et al. Alternatively, the false positives might be a product of erroneous data. EDA measurement especially, is sensitive and variables such as varying pressure on the EDA electrodes or motion artifacts have been found to distort data, which in turn may lead to false readings (Hickey et al., 2021).

4.3.5 Interacting with Scored Data

Through the front-end's user interface scored data can be viewed and manipulated, as seen in Figure 4.8. Wearable data is however only stored in the database in the back-end, analysed and not visualized for the users of the web application. User interactions with the scored data include operations for viewing and manipulating, searching for or filtering data e.g., by type, score, or patient. The data can also be exported or imported as Microsoft Excel files. Audit logs, that contain a log regarding the creation and deletion of data, can also be viewed.

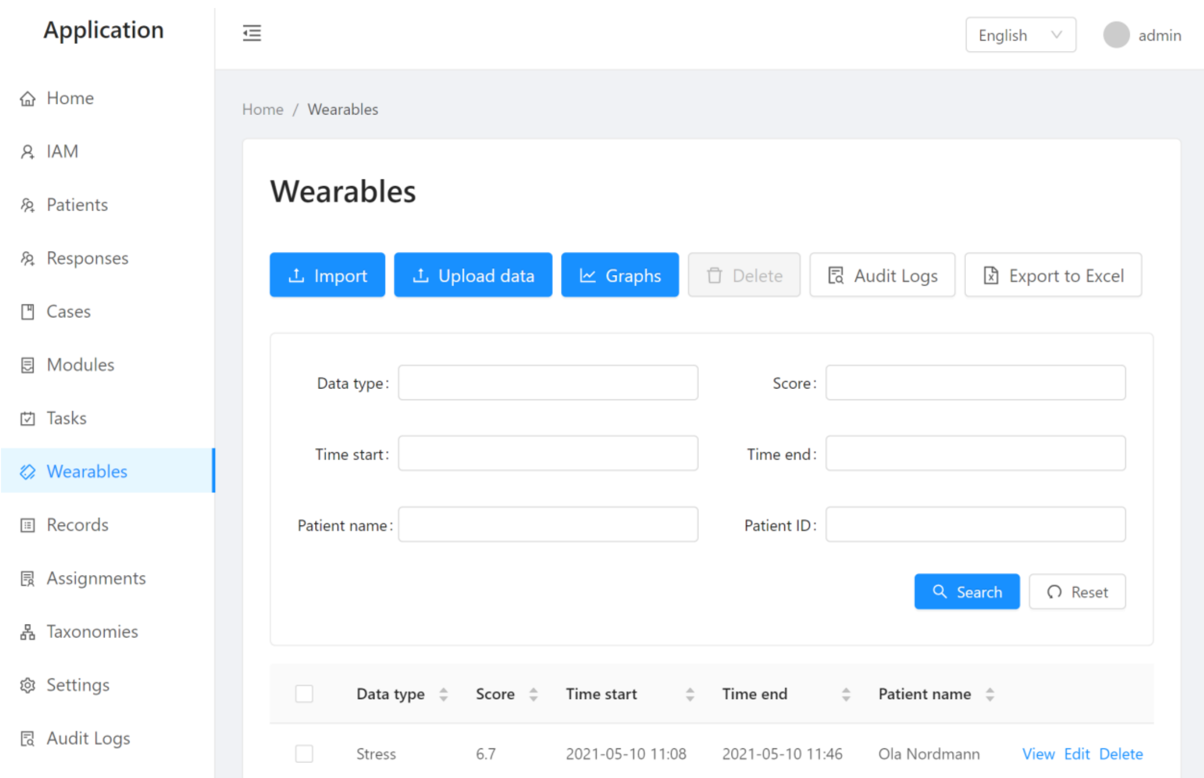

Figure 4.11: User interface for interacting with analysed wearable data in the IDPT system.

An upload button (labelled "Upload data" in Figure 4.8) is brings the user to page where wearable data can be uploaded (EDA and ST). This is seen in Figure 4.9. After the data is uploaded it is automatically stored in the database, analysed, and lastly displayed as scored data on the front-end. The uploading of data is only included as a part of our demonstrative component and not our general component. This is due to the fact that different wearable devices and types of sensor data have proprietary data formats which makes it inherently difficult to create a generic solution for uploading of data (as mentioned in Section 2.4.2).

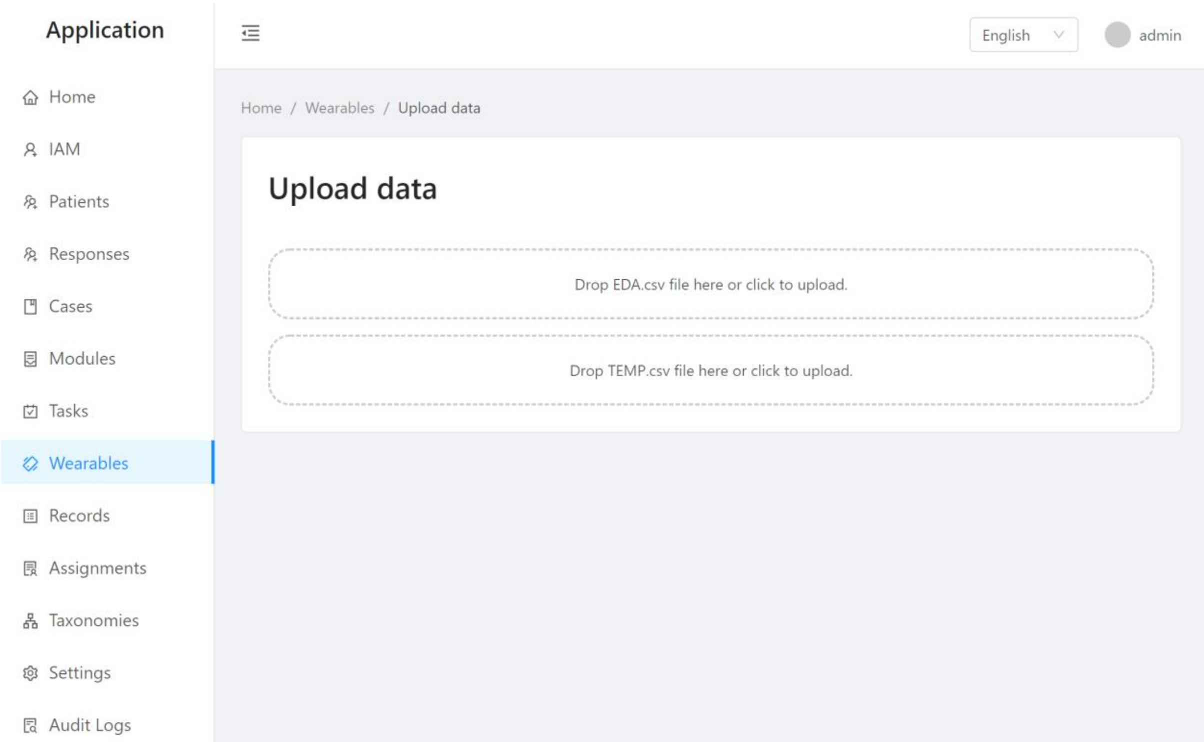

Figure 4.12: User interface for uploading wearable data to the IDPT system.

Beyond the table of analysed data as seen in Figure 4.8, we wanted a better visualization of how a patient's stress progressed over time. This was also suggested by Mukhiya in one of our unstructured midway evaluations. More on this in Section 5.1. By clicking a button (labelled "Graphs") the user is brought to a page where a patient can be selected, either by searching or from a drop-down menu, and a graph of the following patient's stress scores are displayed.

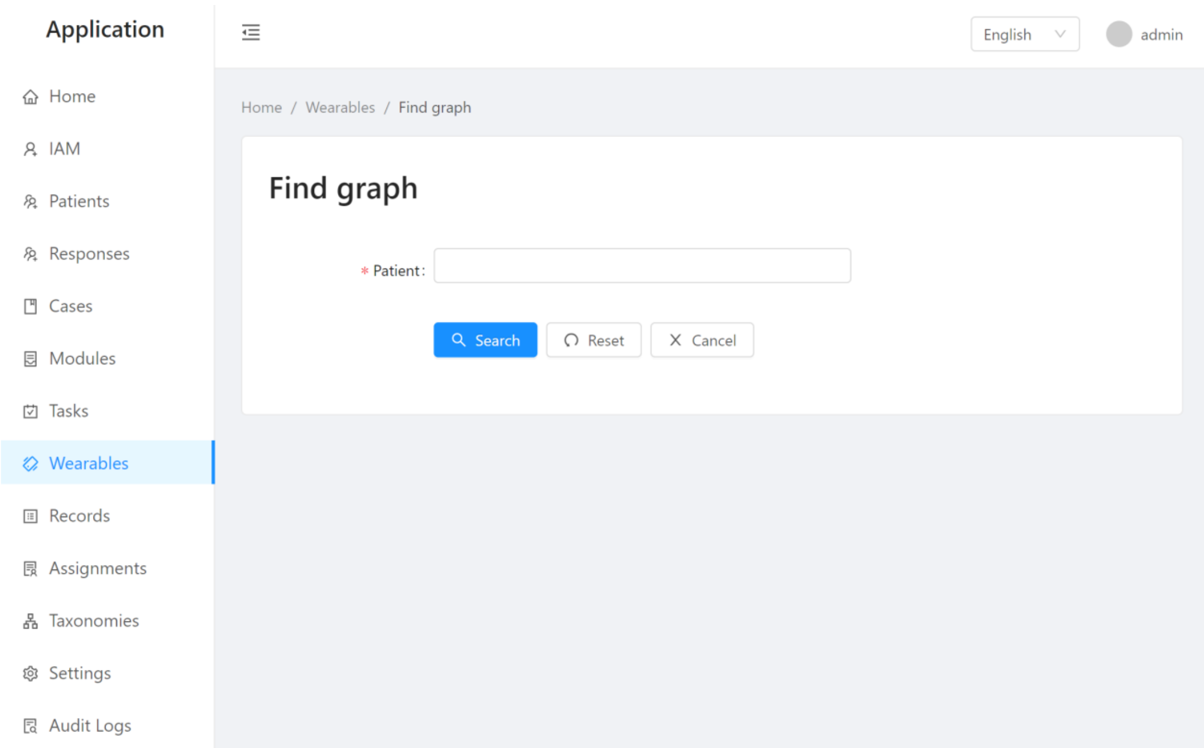

Figure 4.13: User interface for finding a patient's graph illustrating stress scores over time.

The graph representation of a patient's stress scores, presents the stress scores (on the y-axis) of the patient in chronological order (along the x-axis). The labels on the ticks of the x-axis are minimal, showing only the date when collection of data for the adjoining score started. This is done intentionally, as we wanted to keep the graph simple and easy to understand. We also wanted to emphasize overall progression rather than details regarding each score. As mentioned in Section 4.1, the values of the stress scores are of less importance, compared to the change between scores. When hovering over a data point in the graph a more detailed start and end time for the collected data (within minute precision) will be presented.

The stress scores of a patient are not directly connected to adaptive features of the IDPT framework, such that adaptation can be triggered automatically. More on this in Section 7.2. However, the scores may still be used to facilitate adaptation by serving as a measure for therapists to manually adapt the IDPT, as mentioned in Section 2.2. Additionally, the scores may serve as feedback for patients to help them get a better understanding of themselves and their progress.

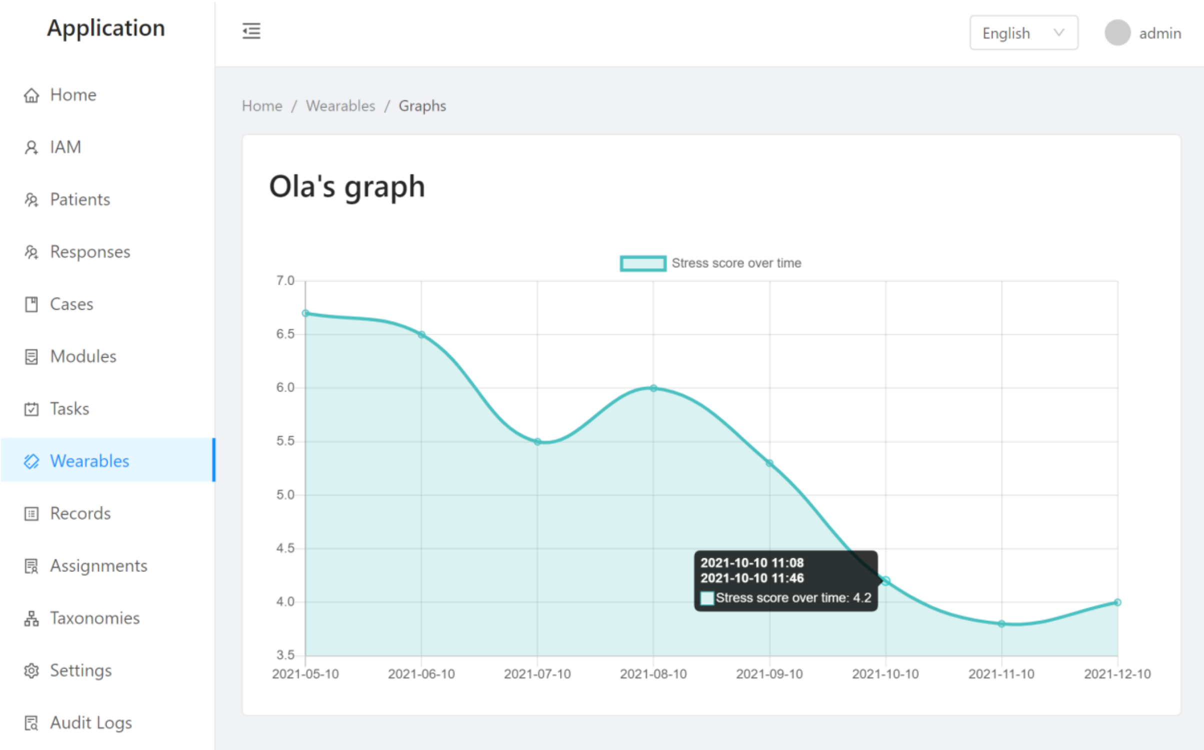

Figure 4.14: A graph showing a patient's hypothetical stress scores, based on synthetic data, over time, accompanied by an info box giving more details regarding the sixth data point.

It should be noted that even though the user interface that has been described in this section is explained in terms of being related to our demonstrative component and specifically scores of stress, it is just as much related to our general component and would work similarly with other types of scored data as well, as mentioned in Section 4.2.2.

CHAPTER 5

5 Research Evaluation

This chapter contains an explanation of the evaluation process of our artifact and the feedback received from domain experts. All throughout development we had several brief meetings with Mukhiya. However, as these meetings were short and educational in nature, they were not regarded as evaluation. Thus, we conducted an unstructured midway evaluation in the middle of our project's timeline. Towards the end of our project, we carried out two semi-structured interviews with domain experts.

5.1 An Unstructured Midway Evaluation

To discuss the design of our artifact we had a meeting with Suresh Mukhiya, the main contributor to the IDPT framework. This meeting consisted of us presenting our work so far and discussed how we should continue our development. As Mukhiya has a great deal of knowledge in the domain of IDPT and especially regarding the framework we were extending, we thought his feedback would be insightful.

Because of time constraints, we did not get to complete a structured or semi-structured interview. Instead, the meeting was in the format of an unstructured interview. Looking back, it might have been beneficial to perform several interviews with greater structure. We learned from this and carried out two semi-structured interviews towards the end of the project, as described in Section 5.2. However, we still received valuable feedback on changes we could make to our artifact.

The main takeaways from our unstructured midway evaluation were:

i) utilize a standardized data format,

ii) provide patients with the ability to upload data themselves, and

iii) visualize the progression of scored data for patients.

Previously, we had developed a proprietary format for wearable data and scored data that we thought fit and would work for a wide range of applications and solutions. However, after our meeting with Mukhiya we figured that using an established and standardized data format was favorable, as this would promote interoperability and reusability. Thus, we settled on implementing the HL7 FHIR standard as part of our general component, as mentioned in Section 4.2.1.

In our demonstrative component, it was previously only possible for clinicians to upload data on behalf of patients. This was an oversight from our part, and it made sense for patients to be able to upload their own data. In our final version of the artifact, users granted patient privileges will still have the option to upload wearable data in the user interface.

As for the representation of scored data this was previously only done in a table as seen in Figure 4.7 and made it difficult to get a picture of a patient's score progression. However, after the feedback from Mukhiya we implemented a way of selecting a patient and displaying the patient's scored data over time in a graph as seen with stress scores in Figure 4.8.

5.2 Semi-Structured Interviews with Domain Experts

As to evaluate our artifact we performed two semi-structured interviews with domain experts. One interviewee was a domain expert of psychology, whereas the other was a domain expert of electrical engineering. The domain expert of psychology is part of the research project INROMAT and had prior experience with IDTP systems. The domain expert of electrical engineering, however, did not. Rather, the expert had experience regarding monitoring patients with the use of wearables.

As software development and research tend to have qualitative features, qualitative evaluation is often well suited for such research (Hove & Anda, 2005). For this reason, we chose to perform semi-structured interviews as they can produce qualitative data. More specifically, such qualitative data can be produced through asking a set of already defined questions, while simultaneously allowing for unexpected types of information (Seaman, 1999).

Our interviewing process can be split in two distinct parts. First, a part where we presented our artifact to the domain experts. This part included a verbal description, as well as showing figures related to both components of our artifact. Last, a part where we asked the domain experts our already defined questions, including a few spontaneous follow-up questions.

As our interview was of a less structured format, feedback and questioning throughout the interview process was both allowed and encouraged.

A list of all questions and summarized answers from each of the domain experts can be found in Appendix A.

5.2.1 Interview Results

The domain experts' overall perception of our artifact was positive. Yet, they also provided some scepticism and raised a few topics of concern.

Both the domain experts expressed fondness for our artifact and sees the evaluation of patients outside of clinical settings to be important. Expert 1 elaborated that assessment in clinical settings is often too brief, primarily done with questionnaires, and such an assessment may give an incomplete picture of a patient. However, both experts noted that using wearable based EMA may provide a more holistic picture of patients. Expert 2 further explained that EMA could be useful with regards to predicting intervention outcomes at an earlier stage, and as such be able to adapt interventions or cut an intervention short for a patient showing little benefit. The expert briefly mentioned that the use of wearables for assessment has the potential to be less invasive. Moreover, they expressed that patients who feel responsible for their own progress are more likely to experience a better outcome. Thus, it could be beneficial for patients, in addition to clinicians, to be able to view their own EMA data and monitor their own progress.

Still, they noted that there are challenges regarding using wearable technology for EMA and that wearable-based EMA is a field that requires further research. Data collected by wearables are rather prone to errors and can be of varying quality. Further, Expert 2 has had prior experience with an excess amount of false positives in patient assessment, resulting in overburdening patients with assessment-based notifications. Both experts mention that it may be necessary to supplement quantitative data from wearables with qualitative data.

The experts further elaborated that supplementing with qualitative data might be especially important for complex measures, such as stress. Expert 1 noted that an example of how qualitative data could be helpful, is with the use of annotated data. Annotated data could provide more value to measures such as stress, by both increasing accuracy through weeding out false positives and providing a more detailed view of a patient's overall stress. Such annotation would require patients to label and detail data after it has been uploaded and analysed. The expert further noted, that even just a simple timeline of a patient's day, could help increase the value of a measure such as stress.

Expert 2 expressed that if the mentioned challenges are resolved, stress can be an especially useful measure to facilitate intervention adaptation for patients with anxiety disorders, as stress is a primary symptom of anxiety. Whereas for other diagnoses, there are likely other measures that are more beneficial. To represent the data, both experts agreed that both a table and a graph representation of the scored data are helpful. Expert 2 elaborated that a crucial factor of data representation is for clinicians to be able to interpret the data in an effortless manner.

Other concerns raised by both domain experts were the issues of data security and privacy regarding ownership of data. Among many other challenges, Expert 2 noted that correctly managing permissions to patient data is a challenge. In addition, Expert 1 mentioned that ownership of data might especially be a problem regarding commercial wearable devices. The overall concern of interoperability of wearables is raised, as there is no consensus on standards. For the technology to achieve its potential there is a need of interdisciplinary cooperation between manufacturers and health care.

The interviews resulted in valuable feedback and insights of how our artifact might be helpful in interventions. In general, the experts thought our artifact served as a suitable starting point for further research. However, further testing and improvement is needed. More on future work in Section 7.2.

CHAPTER 6

6 Discussion

This chapter contains discussions of our project. This includes the development of our artifact, findings attained of the design process and research evaluation. First, we will answer the research questions posed in Section 1.3. Second, a presentation of our contribution to the problem domain. Third, a reflection on the use of design science, qualitative techniques and principles, and our artifact. Last, a walkthrough of the limitations put on our project and the impact it has had on our results.

6.1 Answering the Research Questions

Following are the research questions posed in Section 1.3, and associated answers.

RQ1: How can an extension that allows for the use of wearable data be implemented to support EMA in interventions?

This research question has been largely answered in Section 4.2 and touched on in Section 5.1 and Section 5.2.1. The IDPT framework, as described in Section 2.3, did previously not have any way of integrating wearable data in interventions. As mentioned in Section 2.4, it is possible to conduct EMA in interventions without the utilization of wearable technology and this has been carried out before. However, with the use of wearables there is an opportunity to access a previously untapped stream of a patient's activities, moods, and behaviours. Moreover, wearables have the potential to place less of a burden on patients compared to traditional ways of assessment such as with questionnaires and enable EMA to be less invasive, as expressed the semi-structured interview in Section 5.2.1, which in turn could make EMA more feasible to use in interventions. Or if not less invasive, also mentioned in the interview, wearable data can help to provide a more holistic picture of patients.

As explained in Section 3.2, we started by creating an extension that enabled the use of generic wearable data in the framework. There are obviously different ways of implementing such an extension. Yet, there is little information that is publicly available regarding how current solutions are implemented. More on this in 6.4. However, after conducting the midway evaluation, as presented in Section 5.1, we gained the insight that using a standardized data format is favourable in many regards, such as to promote interoperability and reusability. Thus, this resulted in us utilizing the HL7 FHIR standard as a part of our general component. Yet, we opted to create a general structure as to fit a wide range of sensor data from many different wearable devices. This is described in Section 4.2.1. As such we have implemented an extension that allows for the use of wearable data in interventions, and that ultimately support EMA.

RQ2: How can such an extension be used to assist in adapting interventions?

This research question is addressed in Section 4.3, Section 5.1, and Section 5.2.1. As to demonstrate how such an extension explained in *RQ1* could be used to assist in adapting interventions, we created a demonstrative component, as described in Section 4.3. There are numerous types of data recorded from wearables that could be used to assess patients. Which types of data are most useful with regards to assisting in adapting interventions, needs further research. However, as a demonstration, in our demonstrative component we record EDA and ST with an Empatica E4, to detect MOS and derive a stress score. The stress scores serve as an illustration of how one may assess patients during an intervention in every-day life.

Data gathered through EMA, could be helpful in adapting interventions and personalizing treatment, as was concluded in the interviews from Section 5.2.1. As of now, there is no link between the stress scores and adaptive features of the IDPT system. Still, the scores could be used to instantiate manual adaptation, as described in Section 2.2. As mentioned in Section 5.2.1, stress could be a valuable measure to adapt interventions for patients with anxiety disorders. Whereas for other diagnoses, there are probably other measures that are more suitable. Yet, there is an issue of false positives. More on this in Section 6.4. As such, our demonstrative component needs to be extensively tested and evaluated. Yet, our component stands a demonstration of how our extension can be used to assist in adapting interventions.

6.2 Research Contributions

As part of the design science methodology and a part of Hevner et al.'s (2004) fourth guideline, our contributions should yield value to the problem domain in the form of an artifact, as described in Section 3.1.1. In our case the problem domain regards IDPT. The answers to our research questions from Section 6.1 should yield an overview of our artifact's contribution. Still, a brief summary of our contributions is described below.

We set out to implement an extension to the adaptive IDPT framework as to enable the use of wearable data in interventions, to support EMA. Our goal for creating such an extension, was to create an open-source solution, that would promote reuse and further research. Results from semi-structured interviews presented in Section 5.2.1 suggest that our problem domain requires further research, and that our artifact may prove valuable in this regard. We have laid a foundation that will allow the use of various sensor data from various wearable devices. Using an established data standard, more specifically HL7 FHIR, we are additionally encouraging interoperability, as suggested in our midway evaluation described in Section 5.1.

The extension of the IDPT framework, namely our general component described in Section 4.2, serves as a large part of our research contribution. Furthermore, our demonstrative component described in Section 4.3, serves as an example of how our extension may be used to assist towards adaptive interventions. The two components comprise our artifact and together contribute to the knowledge base. Still, our artifact remains a hypothesis at this stage and remains to be empirically tested.

6.3 Reflections

In this section we are going to present some important discoveries we have made during the reflective process of our project. In addition to reflections on the choice of research method and the resulting artifact, we also made discoveries regarding our design process. One such discovery being that incorporating a frequent evaluation in the iterative process would be beneficial for reaching the desired result by iteratively evaluating the design, as mentioned in Section 5.1. Another discovery we made was to conceptualize the goal for the project early as to reduce the time consumed by prototyping and increasing the value of meetings throughout development.

6.3.1 Research and Evaluation Methods

Design science has proven helpful throughout our project, both for development and evaluation. Artifact creation derived from design science fit appropriately with our contribution to the IDPT framework. The guidelines developed by Hevner et al. (2004) set comprehensible criteria for the project and acted as a guide for conducting research. With comprehensible criteria in place, it became clear how to design an artifact of value. Although, due to the value of an artifact being challenging to measure, we combined design science with a qualitative evaluation method.

In addition to the midway evaluation with Mukhiya, described in Section 5.1, we also had semi-structured interviews with two domain experts, as described in Section 5.2. Even though these were held later in the project's lifecycle, they provided us with several good insights. First and foremost, as a means to evaluate our artifact, but also as suggestions for future work.

We believe that a greater number qualitative evaluations earlier in the implementation process would have been favorable in terms of garnering feedback and aiding in our development. Further, a midway evaluation in the form of interviews with several domain experts would have been beneficial, with regards to further improving the value of our artifact.

6.3.2 Reflection on our Artifact

As mentioned in Section 2.6, we were not able to review how existing solutions had implemented their integration of wearable data in IDPT. We have contributed to an open-source framework, and thus our implementation is publicly available. As such we hope our work can promote further research in this domain.

Furthermore, due to the project being open-source it enables our implementation to be used in other solutions. We designed our artifact to be well structured and inherently easy to grasp, used a recognized data standard and developed to fit various sensor data. Thus, we envision that our work can be reused in other solutions. More on this in Section 7.2.

6.4 Project Limitations

Throughout development, we encountered several limitations.

First, after researching existing solutions regarding IDPT with a wearable integration or stress detection using wearables, we found that there were very few that had an overview of how the solutions were implemented or had any source code available to the public. As such, we could not evaluate our artifact through comparison of similar solutions. If implementations of existing solutions were available, it would have aided in the development of our artifact.

Furthermore, we were limited by time constraints. Unfortunately, the shortage of time and a lack of information regarding implementation, culminated in us not being able to resolve the issue of false positives for our stress algorithm. Further, we could not include all features to the extent we may have wanted. An example of this is how the implementation of data models for wearable and scored data required manually creating several schemas for sub-components that make up a FHIR resource. We believe that an optimal solution for this would be to have the components be automatically generated and evaluated based on the FHIR specification. Additionally, we wanted to connect the stress scores to adaptive features of the IDPT framework as to enable automatic adaptation, even though the scores may still be used to actuate manual adaptation without such a connection.

Last, in order to best evaluate our artifact throughout development, we would ideally have interviewed a greater number of domain professionals. This would have provided us with valuable feedback to aid in the design of our artifact.

CHAPTER 7

7 Conclusion

Throughout this thesis, we have explored the use of wearable technology to enable EMA in interventions. Our work serves as a basis for further research in this field.

7.1 Summary

We have developed an artifact consisting of a general component that allows for integrating wearable data in IDPT, along with a demonstrative component suggesting how the general component could be used to assist in adapting interventions. The general component of the artifact serves as an extension of the IDPT framework. The demonstrative component exemplifies how analyzed sensor data may be utilized to assess patients' stress in day-to-day life, with the goal of adapting interventions. Our artifact, comprised of these two components, can facilitate further development and research in the domain of IDPT.

7.2 Future Work

Further research is needed to investigate the usefulness of utilizing wearable technology for EMA in IDPT.

Regarding our general component, and as mentioned in Section 6.4, we believe that the schemas for sub-components of a FHIR resource should be automatically generated and evaluated based on the FHIR specification. Further, as mentioned in Section 5.2.1, our interviews with the domain experts illuminated the value of supplementing quantitative data with qualitative data. Future work on our general component could make a bridge to link these two types of data more easily, as to further increase the value of EMA.

The work we believe to be most important for increasing the value of our demonstrative component, is solving the issue of false positives, as mentioned in Section 6.4. If this issue is to be solved, connecting the stress scores to automatically instantiate adaptive features of an intervention could be valuable. Beyond this, including other types of sensor data is an obvious addition. Including a wider range of sensor data, may aid in providing a more holistic assessment of patients. However, with adding other sensor data and from different wearable devices, challenges regarding interoperability will need to be solved. As brought up in the semi-structured interviews with domain experts, data privacy and security are important topics that needs attention as well.

Finally, our artifact needs extensive empirical testing before we can conclude its usefulness. The artifact stands largely as a hypothesis at this stage.

Acronyms

ADHD	Attention Deficit Hyperactivity Disorder
API	Application Programming Interface
CSV	Comma Separated Values
ECG	Electrocardiography
EDA	Electrodermal Activity
EMA	Ecological Momentary Assessment
FHIR	Fast Healthcare Interoperability Resources
GAD-7	General Anxiety Disorder-7
HL7	Health Level Seven
HR	Heart Rate
HRV	Heart Rate Variability
INTROMAT	INTROducing Mental health through Adaptive Technology
K-10	Kessler Psychological Distress Scale
MOS	Moment Of Stress
ODM	Object Data Modelling
PHQ-9	Patient Health Questionnaire-9
PSS	Perceived Stress Scale
REST	Representational State Transfer
ST	Skin Temperature

Bibliography

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. In *Journal of Biomedical Informatics* (Vol. 59). https://doi.org/10.1016/j.jbi.2015.11.007

Andersson, G. (2016). Internet-Delivered Psychological Treatments. *Annual Review of Clinical Psychology*, *12*, 157–179. https://doi.org/10.1146/annurev-clinpsy-021815-093006

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry*, *18*(1). https://doi.org/10.1002/wps.20610

Ayaz, M., Pasha, M. F., Alzahrani, M. Y., Budiarto, R., & Stiawan, D. (2021). Erratum: The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities (JMIR Med Inform (2021) 9: 7 (e21929) DOI: 10.2196/21929). In *JMIR Medical Informatics* (Vol. 9, Issue 8). https://doi.org/10.2196/32869

Balcombe, L., & de Leo, D. (2021). Digital mental health challenges and the horizon ahead for solutions. In *JMIR Mental Health* (Vol. 8, Issue 3). https://doi.org/10.2196/26811

Benjenk, I., & Chen, J. (2018). Effective mental health interventions to reduce hospital readmission rates: a systematic review. *Journal of Hospital Management and Health Policy*, *2*. https://doi.org/10.21037/jhmhp.2018.08.05

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. *Cognition and Emotion*, *23*(5). https://doi.org/10.1080/02699930802607937

Berk, M., Brnabic, A., Dodd, S., Kelin, K., Tohen, M., Malhi, G. S., Berk, L., Conus, P., & McGorry, P. D. (2011). Does stage of illness impact treatment response in bipolar disorder? Empirical treatment data and their implication for the staging model and early intervention. *Bipolar Disorders*, *13*(1). https://doi.org/10.1111/j.1399-5618.2011.00889.x

Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & van der Molen, M. W. (1997). Heart rate variability: Origins methods, and interpretive caveats. In *Psychophysiology* (Vol. 34, Issue 6). https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

Bharadwaj, P., Pai, M. M., & Suziedelyte, A. (2017). Mental health stigma. *Economics Letters*, *159*, 57–60. https://doi.org/10.1016/J.ECONLET.2017.06.028

Boucsein, W. (2012). Applications of Electrodermal Recording. In *Electrodermal Activity*. https://doi.org/10.1007/978-1-4614-1126-0_3

Bucci, S., Schwannauer, M., & Berry, N. (2019). The digital revolution and its impact on mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*, *92*(2). https://doi.org/10.1111/papt.12222

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, *14*. https://doi.org/10.5334/dsj-2015-002

Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors (Switzerland)*, *19*(8). https://doi.org/10.3390/s19081849

Casselman, J., Onopa, N., & Khansa, L. (2017). Wearable healthcare: Lessons from the past and a peek into the future. *Telematics and Informatics*, *34*(7). https://doi.org/10.1016/j.tele.2017.04.011

CCS Insight (2022). URL: https://www.ccsinsight.com/company-news/a-brighter-future-for-wearables-after-another-strong-year/ (visited on 2022-01-24)

Chan, I. S., & Ginsburg, G. S. (2011). Personalized medicine: Progress and promise. *Annual Review of Genomics and Human Genetics*, *12*. https://doi.org/10.1146/annurev-genom-082410-101446

Chandra, V., Priyarup, A., & Sethia, D. (2021). Comparative Study of Physiological Signals from Empatica E4 Wristband for Stress Classification. *Communications in Computer and Information Science*, *1441*. https://doi.org/10.1007/978-3-030-88244-0_21

Chewning, B., Bylund, C. L., Shah, B., Arora, N. K., Gueguen, J. A., & Makoul, G. (2012). Patient preferences for shared decisions: A systematic review. In *Patient Education and Counseling* (Vol. 86, Issue 1). https://doi.org/10.1016/j.pec.2011.02.004

Cho, S., Ensari, I., Weng, C., Kahn, M. G., & Natarajan, K. (2021). Factors affecting the quality of person-generated wearable device data and associated challenges: Rapid systematic review. In *JMIR mHealth and uHealth* (Vol. 9, Issue 3). https://doi.org/10.2196/20738

Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The Body and the Brain: Measuring Skin Conductance Responses to Understand the Emotional Experience. *Organizational Research Methods*, *22*(1). https://doi.org/10.1177/1094428116681073

Chrousos, G. P., Loriaux, L., Gold, P. W., & National Institutes of Health (U.S.). (1988). Mechanisms of physical and emotional stress. In *Advances in experimental medicine and biology v. 245*.

Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological stress and disease. In *Journal of the American Medical Association* (Vol. 298, Issue 14). https://doi.org/10.1001/jama.298.14.1685

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385. https://doi.org/10.2307/2136404

Colombo, D., Fernández-Álvarez, J., Patané, A., Semonella, M., Kwiatkowska, M., García-Palacios, A., Cipresso, P., Riva, G., & Botella, C. (2019). Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: A systematic review. In *Journal of Clinical Medicine* (Vol. 8, Issue 4). https://doi.org/10.3390/jcm8040465

Davenport, T. A., LaMonica, H. M., Whittle, L., English, A., Iorfino, F., Cross, S., Clinical, Mp., & Hickie, I. B. (2019). Validation of the innowell platform: Protocol for a clinical trial. *JMIR Research Protocols*, *8*(5). https://doi.org/10.2196/13955

Daviu, N., Bruchas, M. R., Moghaddam, B., Sandi, C., & Beyeler, A. (2019). Neurobiological links between stress and anxiety. In *Neurobiology of Stress* (Vol. 11). https://doi.org/10.1016/j.ynstr.2019.100191

Debard, G., de Witte, N., Sels, R., Mertens, M., van Daele, T., & Bonroy, B. (2020). Making wearable technology available for mental healthcare through an online platform with stress detection algorithms: The CareWear project. *Journal of Sensors*, *2020*. https://doi.org/10.1155/2020/8846077

Degroote, L., de Bourdeaudhuij, I., Verloigne, M., Poppe, L., & Crombez, G. (2018). The accuracy of smart devices for measuring physical activity in daily life: Validation study. *JMIR MHealth and UHealth*, *6*(12). https://doi.org/10.2196/10972

Doherty, K., Balaskas, A., & Doherty, G. (2020). The Design of Ecological Momentary Assessment Technologies. In *Interacting with Computers* (Vol. 32, Issue 3). https://doi.org/10.1093/iwcomp/iwaa019

eMeistring (2022). URL: https://helse-bergen.no/emeistring (visited on 2022-04-05)

Empatica (2022). URL: https://www.empatica.com/en-int/research/e4/ (visited on 2022-03-20)

EPA (2022). URL: https://www.epa.gov/data-standards/learn-about-data-standards (visited on 2022-05-07)

Fernández-Álvarez, J., Díaz-García, A., González-Robles, A., Baños, R., García-Palacios, A., & Botella, C. (2017). Dropping out of a transdiagnostic online intervention: A qualitative analysis of client's experiences. *Internet Interventions*, *10*. https://doi.org/10.1016/j.invent.2017.09.001

FHIR (2022). URL: https://www.hl7.org/fhir/resourceguide.html?fbclid=IwAR36AJ-8BuUmwUeettSbBvRMPwQlhbBCovuD1-Jwho11R0KcvHoJRRf0A1Q (visited on 2022-01-15)

Fink, G. (2009). Stress: Definition and history. In *Encyclopedia of Neuroscience*. https://doi.org/10.1016/B978-008045046-9.00076-0

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. In *Psychological Medicine* (Vol. 47, Issue 14). https://doi.org/10.1017/S003329171700040X

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing*, *13*(1). https://doi.org/10.1109/TAFFC.2019.2927337

Gibbons, M. B. C., Gallop, R., Thompson, D., Gaines, A., Rieger, A., & Crits-Christoph, P. (2019). Predictors of treatment attendance in cognitive and dynamic therapies for major depressive disorder delivered in a community mental health setting. *Journal of Consulting and Clinical Psychology*, *87*(8). https://doi.org/10.1037/ccp0000414

Glenn, T., & Monteith, S. (2014). New Measures of Mental State and Behavior Based on Data Collected From Sensors, Smartphones, and the Internet. In *Current Psychiatry Reports* (Vol. 16, Issue 12). https://doi.org/10.1007/s11920-014-0523-3

Google Scholar (2022). URL: https://scholar.google.com/scholar?q=%22design+science%22&hl=en&as_sdt=0%2C5&as_ylo=&as_yhi=

Griffin, B., & Saunders, K. E. A. (2020). Smartphones and Wearables as a Method for Understanding Symptom Mechanisms. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00949

Grundlehner, B., Brown, L., Penders, J., & Gyselinckx, B. (2009). The design and analysis of a real-time, continuous arousal monitor. *Proceedings - 2009 6th International Workshop on Wearable and Implantable Body Sensor Networks, BSN 2009*. https://doi.org/10.1109/BSN.2009.21

Guk, K., Han, G., Lim, J., Jeong, K., Kang, T., Lim, E. K., & Jung, J. (2019). Evolution of wearable devices with real-time disease monitoring for personalized healthcare. In *Nanomaterials* (Vol. 9, Issue 6). https://doi.org/10.3390/nano9060813

Gutierrez, L. J., Rabbani, K., Ajayi, O. J., Gebresilassie, S. K., Rafferty, J., Castro, L. A., & Banos, O. (2021). Internet of things for mental health: Open issues in data acquisition, self-organization, service level agreement, and identity management. In *International Journal of Environmental Research and Public Health* (Vol. 18, Issue 3). https://doi.org/10.3390/ijerph18031327

Hammen, C., Kim, E. Y., Eberhart, N. K., & Brennan, P. A. (2009). Chronic and acute stress and the prediction of major depression in women. *Depression and Anxiety*, *26*(8). https://doi.org/10.1002/da.20571

Hebb, D. O. (1955). Drives and the CNS. *Psychological Review*, *62*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hickey, B. A., Chalmers, T., Newton, P., Lin, C. T., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J., & Lal, S. (2021). Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. In *Sensors* (Vol. 21, Issue 10). https://doi.org/10.3390/s21103461

HL7 (2022). URL: http://www.hl7.org/about/index.cfm?ref=nav%27 (visited on 2022-01-15)

Hove, S. E., & Anda, B. (2005). Experiences from conducting semi-structured interviews in empirical software engineering research. *Proceedings - International Software Metrics Symposium*, *2005*. https://doi.org/10.1109/METRICS.2005.24

Hui, T. K. L., & Sherratt, R. S. (2018). Coverage of emotion recognition for common wearable biosensors. *Biosensors*, *8*(2). https://doi.org/10.3390/bios8020030

IDPT Framework GitHub (2022). URL: https://github.com/sureshHARDIYA/idpt (visited on 2022-04-23)

Iorfino, F., Cross, S. P., Davenport, T., Carpenter, J. S., Scott, E., Shiran, S., & Hickie, I. B. (2019). A Digital Platform Designed for Youth Mental Health Services to Deliver Personalized and Measurement-Based Care. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00595

Jacobson, N. C., Weingarden, H., & Wilhelm, S. (2019). Digital biomarkers of mood disorders and symptom change. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0078-0

Johannesson, P., & Perjons, E. (2014). An introduction to design science. In *An Introduction to Design Science* (Vol. 9783319106328). https://doi.org/10.1007/978-3-319-10632-8

Kamdar, M. R., & Wu, M. J. (2016). PRISM: A data-driven platform for monitoring mental health. *Pacific Symposium on Biocomputing*. https://doi.org/10.1142/9789814749411_0031

Karthikeyan, P., Murugappan, M., & Yaacob, S. (2013). Multiple physiological signal-based human stress identification using non-linear classifiers. *Elektronika Ir Elektrotechnika*, *19*(7). https://doi.org/10.5755/j01.eee.19.7.2232

Keogh, A., Argent, R., Anderson, A., Caulfield, B., & Johnston, W. (2021). Assessing the usability of wearable devices to measure gait and physical activity in chronic conditions: a systematic review. In *Journal of NeuroEngineering and Rehabilitation* (Vol. 18, Issue 1). https://doi.org/10.1186/s12984-021-00931-2

Knapp, M., & Wong, G. (2020). Economics and mental health: the current scenario. *World Psychiatry*, *19*(1). https://doi.org/10.1002/wps.20692

Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., Liedlgruber, M., Wilhelm, F., Osborne, T., & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors (Switzerland)*, *19*(17). https://doi.org/10.3390/s19173805

Langley, E. L., Wootton, B. M., & Grieve, R. (2018). The Utility of the Health Belief Model Variables in Predicting Help-Seeking Intention for Anxiety Disorders. *Australian Psychologist*, *53*(4). https://doi.org/10.1111/ap.12334

Lee, E. H. (2012). Review of the psychometric evidence of the perceived stress scale. In *Asian Nursing Research* (Vol. 6, Issue 4). https://doi.org/10.1016/j.anr.2012.08.004

Liew, W. S., Seera, M., Loo, C. K., Lim, E., & Kubota, N. (2016). Classifying Stress from Heart Rate Variability Using Salivary Biomarkers as Reference. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(10). https://doi.org/10.1109/TNNLS.2015.2468721

Lu, L., Zhang, J., Xie, Y., Gao, F., Xu, S., Wu, X., & Ye, Z. (2020). Wearable health devices in health care: Narrative systematic review. In *JMIR mHealth and uHealth* (Vol. 8, Issue 11). https://doi.org/10.2196/18907

Marshall, M., & Rathbone, J. (2011). Early intervention for psychosis. [Review][Update of Cochrane Database Syst Rev. 2006;(4):CD004718; PMID: 17054213] . In *Cochrane database of systematic reviews (Online)* (Issue 6).

Mathews, S. C., Mcshea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., & Cohen, A. B. (n.d.). *Digital health: a path to validation*. https://doi.org/10.1038/s41746-019-0111-3

McGorry, P. D., & Mei, C. (2018). Early intervention in youth mental health: Progress and future directions. *Evidence-Based Mental Health*, *21*(4). https://doi.org/10.1136/ebmental-2018-300060

Mental Health Foundation (2016). *Fundamental Facts About Mental Health*. URL: http://www.oecd.org/els/emp/MentalHealthWork-UnitedKingdom-AssessmentRecommendations.pdf (visited on 2022-04-22)

Morgan, C., Mason, E., Newby, J. M., Mahoney, A. E. J., Hobbs, M. J., McAloon, J., & Andrews, G. (2017). The effectiveness of unguided internet cognitive behavioural therapy for mixed anxiety and depression. *Internet Interventions*, *10*. https://doi.org/10.1016/j.invent.2017.10.003

Mukhiya, S. K., Wake, J. D., Inal, Y., & Lamo, Y. (2020). Adaptive Systems for Internet-Delivered Psychological Treatments. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3002793

Mukhiya, S. K., Wake, J. D., Inal, Y., Pun, K. I., & Lamo, Y. (2020). Adaptive elements in internet-delivered psychological treatment systems: Systematic review. In *Journal of Medical Internet Research* (Vol. 22, Issue 11). https://doi.org/10.2196/21066

Muzny, M., Henriksen, A., Giordanengo, A., Muzik, J., Grøttland, A., Blixgård, H., Hartvigsen, G., & Årsand, E. (2020). Wearable sensors with possibilities for data exchange: Analyzing status and needs of different actors in mobile health monitoring systems. In *International Journal of Medical Informatics* (Vol. 133). https://doi.org/10.1016/j.ijmedinf.2019.104017

Ometov, A., Shubina, V., Klus, L., Skibińska, J., Saafi, S., Pascacio, P., Flueratoru, L., Gaibor, D. Q., Chukhno, N., Chukhno, O., Ali, A., Channa, A., Svertoka, E., Qaim, W. bin, Casanova-Marqués, R., Holcer, S., Torres-Sospedra, J., Casteleyn, S., Ruggeri, G., … Lohan, E. S. (2021). A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges. In *Computer Networks* (Vol. 193). https://doi.org/10.1016/j.comnet.2021.108074

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic metaanalysis. *General Hospital Psychiatry*, *39*. https://doi.org/10.1016/j.genhosppsych.2015.11.005

Rawassizadeh, R., Price, B. A., & Petre, M. (2015). Wearables: Has The age of smartwatches finally arrived? In *Communications of the ACM* (Vol. 58, Issue 1). https://doi.org/10.1145/2629633

Resnick, S. G., Oehlert, M. E., Hoff, R. A., & Kearney, L. K. (2020). Measurement-based care and psychological assessment: Using measurement to enhance psychological treatment. *Psychological Services*, *17*(3). https://doi.org/10.1037/ser0000491

Rohleder, N. (2019). Stress and inflammation – The need to address the gap in the transition between acute and chronic stress effects. In *Psychoneuroendocrinology* (Vol. 105). https://doi.org/10.1016/j.psyneuen.2019.02.021

Sano, A., Phillips, A. J., Yu, A. Z., McHill, A. W., Taylor, S., Jaques, N., Czeisler, C. A., Klerman, E. B., & Picard, R. W. (2015). Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2015*. https://doi.org/10.1109/BSN.2015.7299420

Schabracq, M. J., & Cooper, C. L. (2003). The Handbook of Work and Health Psychology. *Science*.

Schulz, S., Stegwee, R., & Chronaki, C. (2018). Standards in healthcare data. In *Fundamentals of Clinical Data Science*. https://doi.org/10.1007/978-3-319-99713-1_3

Schumm, J., Bächlin, M., Setz, C., Arnrich, B., Roggen, D., & Tröster, G. (2008). Effect of movements on the electrodermal response after a startle event. *Methods of Information in Medicine*, *47*(3). https://doi.org/10.3414/ME9108

Seaman, C. B. (1999). Qualitative methods in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, *25*(4). https://doi.org/10.1109/32.799955

Seoane, F., Mohino-Herranz, I., Ferreira, J., Alvarez, L., Buendia, R., Ayllón, D., Llerena, C., & Gil-Pita, R. (2014). Wearable biomedical measurement systems for assessment of mental stress of combatants in real time. *Sensors (Switzerland)*, *14*(4). https://doi.org/10.3390/s140407120

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. In *Annual Review of Clinical Psychology* (Vol. 4, pp. 1–32). https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Staples, L. G., Dear, B. F., Gandy, M., Fogliati, V., Fogliati, R., Karin, E., Nielssen, O., & Titov, N. (2019). Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: The PHQ-2, GAD-2, and K-6. *General Hospital Psychiatry*, *56*. https://doi.org/10.1016/j.genhosppsych.2018.11.003

Stone, A. A., Shiffman, S., Atienza, A. A., & Nebeling, A. (2007). Historical roots and rationale of ecological momentary assessment (EMA). *The Science of Real-Time Data Capture: Self-Reports in Health Research.*

Stroud, C., Onnela, J.-P., & Manji, H. (2019). Harnessing digital technology to predict, diagnose, monitor, and develop treatments for brain disorders. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0123-z

UngSpotlight (2022). URL: https://intromat.no/2020/03/16/ungspotlight-for-ungdom-med-frykt-for-a-snakke-foran-klassen/ (visited on 2022-03-17)

Vinkers, C. H., Penning, R., Hellhammer, J., Verster, J. C., Klaessens, J. H. G. M., Olivier, B., & Kalkman, C. J. (2013). The effect of stress on core and peripheral body temperature in humans. *Stress*, *16*(5). https://doi.org/10.3109/10253890.2013.807243

Wen, D., Zhang, X., & Lei, J. (2017). Consumers' perceived attitudes to wearable devices in health monitoring in China: A survey study. *Computer Methods and Programs in Biomedicine*, *140*. https://doi.org/10.1016/j.cmpb.2016.12.009

WHO (2014). World Health Organization - Health for the world's adolescents a second chance in the second decade. *World Health Organisation*.

Wijsman, J., Grundlehner, B., Liu, H., Penders, J., & Hermens, H. (2013). Wearable physiological sensors reflect mental stress state in office-like situations. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 600–605. https://doi.org/10.1109/ACII.2013.105

Wilhelm, P., & Perrez, M. (2013). *A history of research conducted in daily life*. https://doi.org/10.13140/2.1.2661.3126

Yetisen, A. K., Martinez-Hurtado, J. L., Ünal, B., Khademhosseini, A., & Butt, H. (2018). Wearables in Medicine. In *Advanced Materials* (Vol. 30, Issue 33). https://doi.org/10.1002/adma.201706910

Youwell (2022). URL: https://www.youwell.no/en/ (visited on 2022-02-02)

Zamkah, A., Hui, T., Andrews, S., Dey, N., Shi, F., & Sherratt, R. S. (2020). Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors. In *Biosensors* (Vol. 10, Issue 4). https://doi.org/10.3390/bios10040040

Zawadzki, M. J., Scott, S. B., Almeida, D. M., Lanza, S. T., Conroy, D. E., Sliwinski, M. J., Kim, J., Marcusson-Clavertz, D., Stawski, R. S., Green, P. M., Sciamanna, C. N., Johnson, J. A., & Smyth, J. M. (2019). Understanding stress reports in daily life: a coordinated analysis of factors associated with the frequency of reporting stress. *Journal of Behavioral Medicine*, *42*(3). https://doi.org/10.1007/s10865-018-00008-x

Appendix A

Answers to the Semi-Structured Interviews

Below are the questions from the semi-structured interviews and the answers received from the domain-experts.

Q1: What kind of experience do you have with Internet-Delivered Psychological Treatment (IDPT)?

Domain expert 1

- None

Domain expert 2

- As a psychologist
- Treated adults with anxiety and depression with IDPT's
    - Cognitive behaviour therapy with IDPT
- Helped develop 4-5 interventions in IDPT for various psychological ailments
    - E.g., MinADHD
    - Often combination of face-to-face and IDPT
- Helped create content, not implemented the technicalities of interventions

Q2 (follow-up question to Q1): Would a general IDPT framework be beneficial for creating such interventions?

Domain expert 2

- Anxiety and depression e.g., could be developed through a shared framework
    - A lot of similarities between these two diagnoses
    - E.g., modules for sleep would be shared for both diagnoses
- Not so much for IDPT treating adults with ADHD
    - The needs of an ADHD patient do not have as much overlap as e.g. anxiety has with depression
    - Needs different ways of presenting information to the patient
- Would be greatly beneficial with regards to timesaving

Q3: What types of measurement methods have you typically used to assess patients during an intervention?

Domain expert 1

- HRV calculated from ECG sensor data
- Uses Polar chest strap as monitoring device
- Assessing stress in patients with PTSD due to loss of sight in adulthood
- Chronic stress is often overlooked in patients
- Evaluating stress in patients by having patients perform a stress inducing task
- Comparing results of stress detection with an interview conducted by a psychologist with patients afterwards

Domain expert 2

- With IDPT: only questionnaires
    - Both pre and post intervention
- Traditional intervention: less questionnaires
    - More face-to-face interview (using questionnaires as a basis)
    - Questionnaires with pen and paper

Q4: Are there any other forms of measurements that could be useful to clinicians for improving the effectiveness of an intervention that are typically not used?

Domain expert 1

- EEG to measure electric activity in the brain
- MRI for pictures of the brain
- However, these measurements may be impractical due to their invasiveness

Domain expert 2

- EMA, as it is typically not used
    - Could predict treatment outcome earlier
        - If there is little to no progress – change treatment

        - Today, patients who don’t benefit from treatment, still finish them. This has little value for the patient
    - Sensor data like sleep etc. can be useful
    - Benefit of therapists not having to ask e.g. depressed patient questions, as this can be burdening to patients, but rather get information directly
    - Get full picture of patients
    - Especially useful regarding recall bias of depressed patients
        - Depressed patients tend to have a negative bias

Q5: Can you think of benefits to ecological momentary assessment (EMA)?

Domain expert 1

- Overall important to measure patients outside of clinical settings
- Assessment in clinical settings is too brief
- A too brief assessment may result in misdiagnosis
- EMA could be useful to get a more holistic picture of patients

Domain expert 2

- For most diagnoses there will be benefits
- E.g., for drug abusers
    - If patient is in radius of place where drugs can be bought, someone is alerted

Q6: Are you familiar with any existing methods of performing EMA during interventions?

Domain expert 1

- No

Domain expert 2

- Virtual reality: ungSpotlight (fear of public speaking)
    - Monitoring stress and eye movement
    - Exposure therapy with data to track progress
- Young people with schizophrenia
    - Measure stress

- Anxiety with augmented reality and stress measurements
  - E.g., exposure therapy for fear of spiders

Q7 (follow-up question to Q6): If so, are there any challenges associated with these methods?

Domain expert 1

Even though the domain expert were not familiar with any existing methods of EMA in interventions as asked in Q5, they still had some thought regarding general challenges of EMA.

- Privacy regarding ownership of data, as this is a big concern for public health systems. Especially a problem with commercial devices.
- Data security is another important issue.

Domain expert 2

- Must be accurate enough
  - "I'm not stressed now"
  - Patients tired of receiving notifications with false positives
- Privacy and security concerns regarding patient data
- Don't want to overburden patients
- How are the measurements really correlated with the diagnosis?
  - And not e.g., a random moment of stress?
- Easier for some diagnosis
  - E.g., anxiety is strongly correlated with stress

Q8: What are your thoughts on using wearable technology for EMA in IDPT?

Domain expert 1

- Valuable for assessing patients.
- However, quantitative measures will likely need to be supplied with qualitative measures.
- Needs more research.

Domain expert 2

- If done right, it would be greatly beneficial

- Therapy can be burdensome for patients
    - Lots of information, therapy and practice in between
- Can be useful to nudge patients in right direction
    - "Now you're stressed" – try this mindful activity
    - Support patients and make treatment more manageable
- Potential to be less invasive
- Needs further research
    - EMA can be difficult to perform accurately
    - E.g., they tried to record tone of voice to detect manism in a research project
        - But turned out to be very difficult
- Need to combine several types of data to get the "big picture" and increase accuracy
    - More sensor data types
    - And maybe qualitative data as well

Q9: Do you see any specific use cases where the use of wearables in intervention would be particularly useful?

Domain expert 1

- Support of handicapped patient in interventions.
- Could be used to quantify efficacy rehabilitation.

Domain expert 2

- Anxiety patients
    - As part of exposure therapy
    - Very useful
    - For panic, social anxiety, claustrophobia
- EMA + intervention
- For general well-being to
    - Don't have to be tied to a diagnosis
    - Reduce stress, improve sleep, work out more
- Schizophrenia

Q10: Do you have any experience with a lack of data standards in interventions?

Domain expert 1

- With sensor data in general there's a lack of data standards
- This is an issue that needs to be cooperated on between health care and manufacturers

Domain expert 2

- Only regarding the use of smartwatches for sensor data collection
    - The application only worked for Android phones, not iPhone

Q11: Do you have any experience with HL7 FHIR?

Domain expert 1

- No

Domain expert 2

- No

Q12: Do you think stress measurements can be useful to facilitate the adaptiveness of IDPT?

Domain expert 1

- Yes
- Could serve as a preventive measure

Domain expert 2

- For anxiety patients
- More useful for diagnoses where stress is part of symptoms
- Other variables (types of sensor data) for other diagnosis

Q13: How do you think clinicians would benefit the most from having stress scores represented? "

Domain expert 1

- Tables and graphs are good representations
- However, it could be more useful if data was annotated

- E.g., by clicking on a stress measurement in the graph, more data is presented
- Information regarding what a patient's day entails could make stress more valuable
- As mentioned in Q7, stress measurements would likely need to be supplied with qualitative measurements (e.g., questionnaires)
- When a patient feels that themselves are responsible for improvement, outcomes are generally better
    - As such patients should be able to view progress themselves as well
    - However, displaying negative progress to patients can have the opposite effect

Domain expert 2

- Finding the best representation of the data is a challenge
- The data needs to be easy to understand at first glance
- In graphs: the change over time is valuable
- If something is over a "threshold" it should be easy to observe by a clinician
    - E.g., receive an alert of some sort
- Valuable to compare patients to a baseline

Q14: Are there any other particular measurements from wearables you think would be valuable in an intervention?

Domain expert 1

- Derivatives of ECG sensor data

Domain expert 2

- Sleep and activity
- Difference regarding it is for the purpose of prevention or for treatment
- For depression: less social before entering depression
    - Could measure the number of other phones in close proximity
    - Look at GPS data
- Sleep is a universal useful measurement for most diagnoses