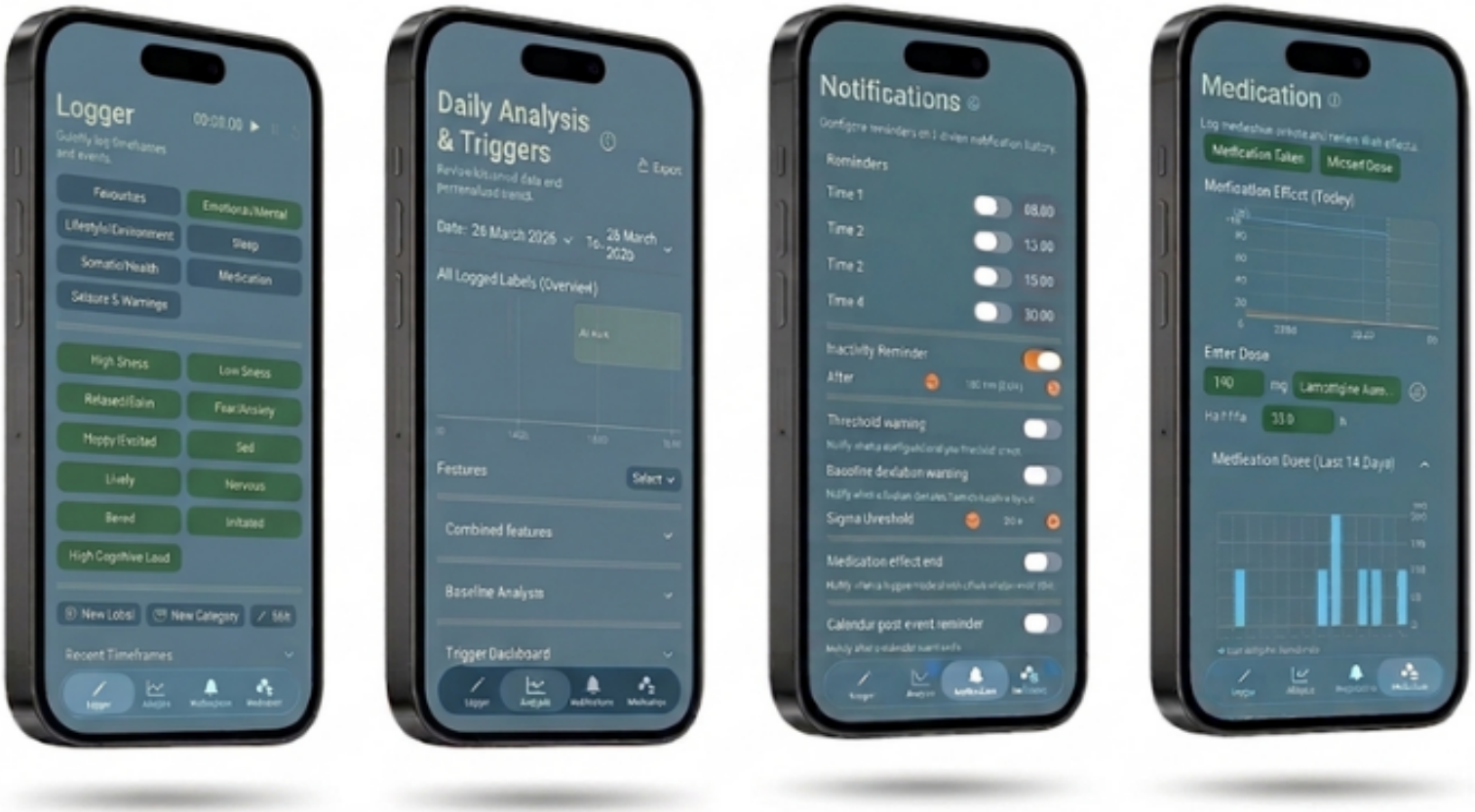
The Epilepsy Journal

Integrating subjective and physiological data to enable personalized understanding and prediction of epileptic events



Dirk Lagemaat
5099692

Graduation report

Integrated Product Design

TU Delft
Faculty of Industrial Design Engineering

Comittee

Chair: Dr. K. Song
Mentor: Dr. H. Verma

Executive Summary

Epilepsy is characterised by unpredictable seizures that significantly impact the quality of life. Although today's wearable technology is able to continuously monitor physiological data such as heart rate, respiratory rate, motion, and sleep, these data alone often lack the contextual information required to understand triggers related to an individual's epilepsy. There is an opportunity to record subjective experiences and accurately align these events with their corresponding physiological data to analyse epileptic triggers and generate preventive warnings related to an individual's historical physiological conditions.

This graduation project addresses this opportunity by designing and evaluating a journal application as a method to integrate physiological data from a wearable with subjective contextual information through a mobile application and Apple Watch. The goal is to create a time-synchronised data set that can be used to train personalised machine learning models capable of predicting seizures and generating preventive warnings.

The proposed design enables epilepsy patients to remotely log scenarios in real time while a wearable continuously measures physiological data types. These two data streams are synchronised through precise timestamps. Once the data is labeled, it is suitable for future machine learning purposes.

The research study of this project focuses on the nudging strategy of the developed mobile application. A usability study was performed with 7 participants to evaluate which type of nudging strategy yields the best user engagement and journaling compliance through the mobile application.

In a within-subjects study (N=7, three conditions for a total of 9 days), no significant differences were found between nudging strategies for the total amount of logged labels, total time covered with labels, or latency. However, a small positive trend has been observed, suggesting that a personalised approach yields the best journaling behaviour, indicating that in future research, with a larger sample size, the significant difference between nudge strategies could be proven.

In conclusion, this thesis describes the iterative development of a mobile application able to align subjective contextual information to continuously measured physiological wearable data. Classification of the physiological data with subjective contextual information enables future personalised machine learning applications that support the identification, detection, and forecasting of epileptic-related triggers and seizures. The design balances an epileptic safe interface, user engagement, data processing, and interpretable data visualisation.

Contents

1 Introduction

Epilepsy affected approximately 50 million people worldwide in 2024 [1], and despite advances in medication and monitoring, the unpredictability of seizures remains a major burden for patients. This study aims to develop a lightweight, transparent, continuous journaling system that enables patients to capture triggers and factors influencing their daily lives. Physiological data are combined with subjective journaling to provide valuable insights into recurring patterns for both clinicians and patients, covering activities, environmental contexts, and emotional or psychological states. This enables patients to actually show their neurologist what happened when they had a seizure, what kind of effect it had on their physiological types of data, or what events or factors before the seizure might have an impact on the individually differing diagnoses. Currently, there is no accessible way to trace back what might have had any influence on a seizure, it is mostly speculation based on generalised information. This approach requires an understanding of data extraction, visualisation, analysis, and feedback mechanisms through journaling.

1.1. Stakeholders

Motivation for this project originates from being a stakeholder myself, as an epilepsy patient. Key stakeholders include epilepsy patients, clinicians, and researchers. The project empowers patients to actively contribute to seizure forecasting through adaptive self-reporting, enhancing engagement and perceived control. Journaling contextual and subjective stimuli enables integration of diverse information types. For patients and clinicians, it offers a richer context to distinguish seizure precursors from external triggers, supporting personalised treatment and improved insight into a patient's personal diagnosis. For researchers, it links subjective experience with objective data, which can be used for epileptic event forecasting.

1.2. State-of-the-art

Although many studies demonstrate the feasibility of seizure prediction using physiological data (e.g., HRV, EDA, ECG) and machine learning [2, 3, 4, 5], few explore how human-generated inputs (e.g., journaling, mood logs, symptom reports) can be systematically integrated into personalised model training. The increasing availability of

wearable sensors, such as My Seizure Gauge and Empatica E4 [6], can capture physiological signals but often neglect how and when to collect contextual user feedback to maximise data value. Similarly, studies on self-prediction [7, 8] show that patients perceive premonitory symptoms, yet these insights are rarely incorporated into personalised models, and there has been no previous research dedicated to yielding the highest amount of subjectively classified physiological data.

1.3. Knowledge Gap

This highlights a gap: no empirically validated framework defines which journaling formats or nudging strategies yield the most accurate and interpretable context-valued and reliable user-generated data for personalised ML models in epilepsy. This is interesting for both epilepsy patients and clinicians to provide context for individual epileptic-related scenarios.

The current incorporation of subjective feedback from patients depends on manual digitisation afterwards, is subject to recall bias, missing physiological data, and unable to accurately link subjective events to objective data.

Future machine learning models require classified datasets to be able to identify which physiological behaviour represents which event. No accurate remote digital classification application is currently publicly available for subjective epilepsy-related events and their corresponding physiological changes.

This graduation project addresses this gap and extends current literature by developing and evaluating a mobile journal application which integrates subjective self-reports, physiological wearable data, an epilepsy-safe interface, and engagement strategies into one application. This design is guided by a research study focused on nudging strategy optimisation for long-term journaling compliance, and supports the interpretation and creation of personalised data sets through self-tracking, enabling future personalised seizure forecasting systems and a deeper insight into an individual's physiological behaviour. The research question corresponding to the study is elaborated in Chapter 5 and formulated as:

- *Which nudging strategy leads to the highest journaling compliance while preserving positive user experience and low perceived burden?*

2

Problem Definition

Epilepsy is characterised by unpredictable seizures [9], a major factor limiting quality of life and daily functioning. Prior knowledge of the problems epilepsy patients face on a daily basis originates from my own experience as an epilepsy patient. Currently, patients do not know for sure which types of events or triggers cause exactly which changes in physiological condition. No accessible method exists that enables patients to trace back which types of physiological data preceded a seizure, or which previous events may have contributed to it. Enabling linkage of self-reported relevant events and the corresponding physiological data using the developed app provides these insights.

Journaling and self-reporting are valuable tools in epilepsy treatment [8], providing contextual data beyond sensor capabilities. However, self-reports are often incomplete or biased due to forgetfulness, stigma, or emotional fatigue [10, 11].
Wearable biosensors can capture physiological signals (e.g., heart rate, electrodermal activity, temperature, motion) that change before seizures take place [12, 13, 14]. Yet, physiological data alone fail to account for contextual or subjective triggers such as stress or emotional state, which are personally related and vary over time[15].
Current seizure forecasting systems face three main challenges:

- **Incompleteness of data:** Physiological data lack personal context (e.g. activity, environment, emotions).
- **Low interpretability:** Models often act as black boxes [16], providing limited feedback to patients and clinicians [17].
- **Limited personalisation :** Algorithms rely on generalised datasets and fail to adapt to individual triggers or behavioural patterns [18].

These issues increase false positives and hinder clinical adoption. Integrating structured journaling and adaptive feedback mechanisms may address these gaps, improving both model accuracy and engagement. To date, such integration has not yet been established because there is no remote system publicly available yet that automatically and accurately links and classifies subjective events of a patient to the physiological data during these events. Future ML model accuracy relies on data from individual patients, to train them to recognize individual related physiological changes.

Previous research is mostly conducted with large sample sizes to provide generalized insights, which in this project's specific case is not applicable. The statistical results of ML models are difficult to interpret for the average patient, limiting the perceived value of the insights.

The goal is to provide a more concrete context for epilepsy patients' personal diagnosis by delivering a functional prototype of a wearable-connected journaling app that effectively nudges users to engage in the journaling activities, aligns with stakeholder needs, labels timeframes accurately, generates usable output data for future ML modelling, and enables identification of epileptic-related scenarios. A secondary goal is to implement a mock-up of a preliminary personalised ML model to demonstrate how physiological and journaling data could be integrated for future forecasting purposes.

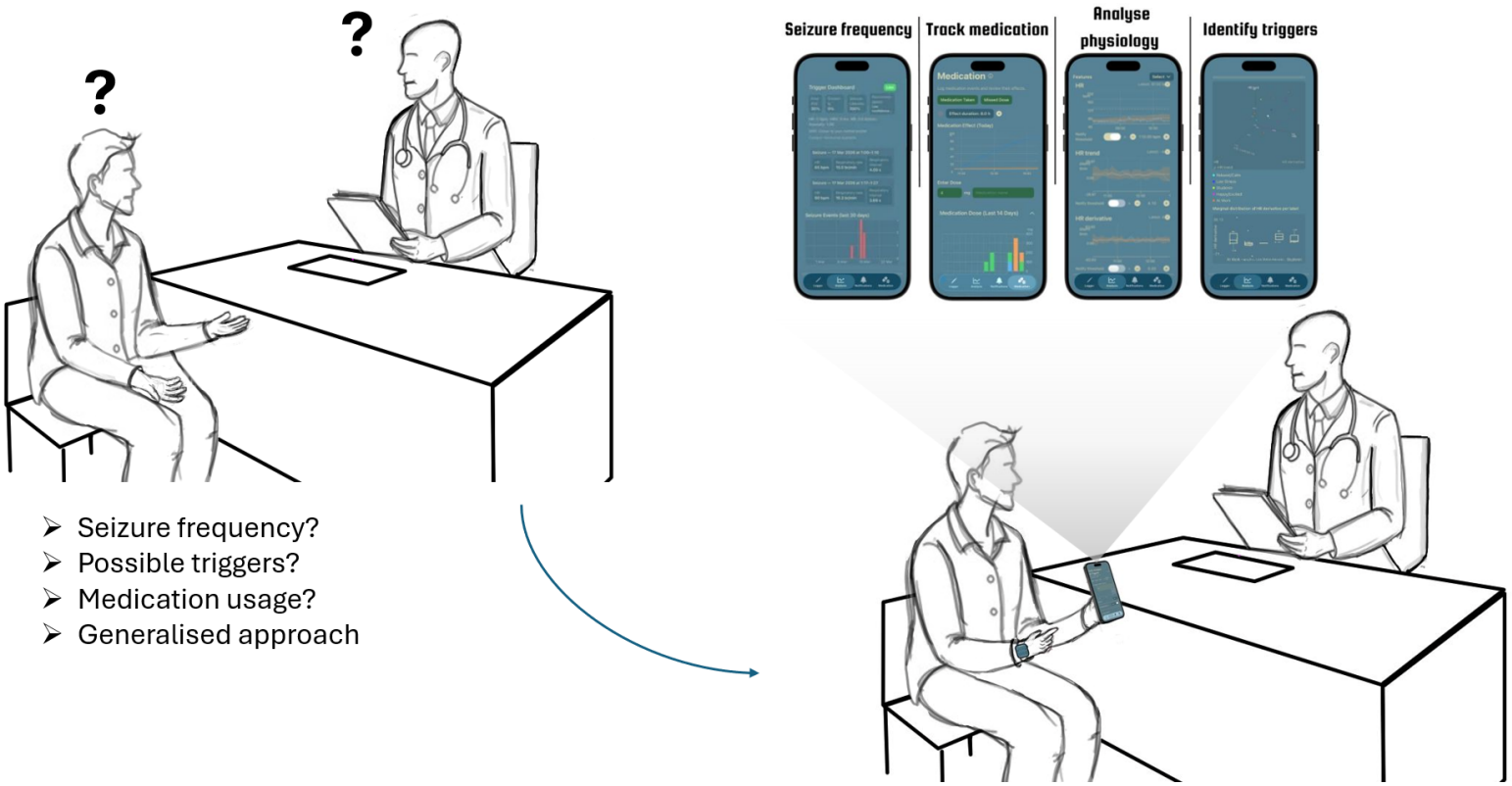


Figure 2.1: The developed journal application enables clinicians to access relevant data to improve and personalise epileptic treatment

3 Literature Research

In this chapter, all conducted literature research related to the project is summarised in subchapters.

3.1. Journaling

This subchapter explores previous research related to journaling epileptic related events and factors that can cause triggers for epilepsy patients. Journaling is a critical mechanism to capture subjective, contextual data that can be transformed into features for personalised ML models.

3.1.1. Digital journaling

Earlier studies have demonstrated the clinical value of capturing potential triggers using paper diaries, but also highlighted the recall bias and temporal inaccuracy [19, 20]. Recent research has explored smartphone app and web-based seizure diaries, which improve timestamp accuracy and data completeness compared to paper records [21]. A smartphone app has previously been used to capture self-reported seizures to train forecasting models [22]. Harms et al. [23] tested different interface designs and concluded that a scrolling layout did not perform as well as tabs, menus, and collapsible fieldsets. Zhang and Adipat [24] stated that to minimise cognitive load, long lists should be avoided, and content can be better structured into categories fitting the screen size.

Digital methods could be the foundation for combining self-reported data with continuous physiological streams from wearables (e.g., heart rate, EDA) to build probabilistic seizure forecasting models [12]. In their work, Brinkmann et al. demonstrated the relevance of combining subjective diaries with wearable-derived features, but they also highlighted the challenges related to data sparsity, inter-individual variability, and limited predictive performance. Existing models rely primarily on physiological signals, as they insufficiently capture and fail to account for subjective, individually related triggers. No model has yet been established which integrates subjective self-reports into ML models. Fisher et al. acknowledge the advantages of electronic diaries, specifically because of the real-time transmission of data, time-stamping of entries, and

reminders [25]. This indicates that while feasibility has been shown in prior research, the reliable and clinically meaningful integration of subjective and objective data of epilepsy patients has yet to be established.

3.1.2. Journaling contextual data

The research objective to use journaling to generate features for predictive models is strongly supported by the need for personalised seizure prediction [10]. Even when a clear diagnosis exists, the limited availability of accurate, high-quality information hinders effective therapy and forecasting [26]. Patients themselves possess unique insights into their triggers and prodromal symptoms, and leveraging these self-reports is seen as a key component in the new era of seizure prediction [27, 28]. Some patients experience precursors of seizures [29]. The journaling methodology, by capturing factors like stress, sleep, and aura, provides the ”ground truth” labels that can be transformed into effective features for ML analysis [30]. This feature transformation step is crucial, and advanced techniques, such as multi-scale feature learning, are used to effectively analyse the complex, non-linear nature of physiological signals in this context [5, 31, 32].

3.1.3. Journaling timeframes

Accurate timestamp alignment of user annotations with raw sensor data is critical for feature extraction, particularly to mitigate synchronisation issues during annotating events [14]. The ability to transform and classify journaling data automatically into effective ML features relies on this digital alignment, which affects the quality and interpretability of the labelled data [33]. Previous studies mostly rely on patients' subjective indication of seizure likelihood in the near future, which is captured through apps or own platforms, but do not state how this subjective information is synchronised with objective clinical data [21, 12, 18]. Stewart et al. used their own platform to collect subjective patient data remotely [34]. However, Quilter et al. specifically address this issue by using a Garmin wearable and a mobile application to align subjective and objective data for forecasting purposes [35].

Online diaries facilitate the complex analysis necessary to characterise repetitive patterns and clusters in patient data over long periods [19]. These insights directly support the project's aim of improving model interpretability and personalisation.

Continuous data can be divided into fixed-length, overlapping time windows (e.g., 30s to 10 min segments). This transformation creates discrete samples suitable for analysis and aligns the data temporally for multimodal fusion [36].

3.1.4. Label categories

To be able to classify physiological data, the following types of labels have been implemented as buttons to the application so scenarios can be journaled throughout the day:

Seizure Characteristics and Events

- **Seizure occurrence and severity:** The most basic label, recording the time, date, type, and perceived severity of the seizure [19]. This specific type of journaling must be done post-event, based on the wearable data to ensure accurate annotations [14].
- **Premonitory :** blurred vision, light sensitivity, dizziness, feeling emotional, concentration difficulty, hunger/food cravings, noise sensitivity, tired/weary, thirst, difficulty with thoughts, hyperactivity, headache, and difficulty reading/writing have been included due to their association with seizure self-predictions [8].
- **Aura/Prodrome symptoms:** Recording the presence and type of premonitory symptoms (prodrome) or immediate warnings (aura) experienced hours or minutes before a seizure [37]. Patient reports of these symptoms can be used to label the 'pre-ictal' state in ML models.
- **Medication:** The medication intake moments are an essential factor in seizure control. Additionally, a missed intake moment is a significant risk factor that may trigger seizures [26]. Therefore, two buttons have been made to track the medication intake. Besides these buttons, tracking adjustments in the doses over time is also a fundamental variable in seizure frequency [11].

Physiological and Somatic triggers

- **Sleep quality and Deprivation:** Subjective ratings of sleep quality, quantity, and whether the patient woke up feeling refreshed, as sleep disruption is a well-established seizure trigger [38, 39]. A 'Sleep deprived', 'Nap', and 'Woke up' button have been added because deprivation is a known high-risk state, nap data can be used to analyse the transition of physiological data types into and out of sleep states (which are periods of altered seizure risk), and the wake-up time is used to determine the circadian cycle and because it is a period of heightened seizure risk [40].
- **Illness/infection, overexertion, stress, missed pill, diet, and lack of sleep** have been self-reported symptoms by participants in previous research. Therefore, all these labels have been added to the interface of the app [41, 11].
- **Pain:** Documentation of pain level and type, especially migraine or pre-ictal headaches, which are highly prevalent comorbidities [10, 42, 8, 43]. These symptoms have been used before in journaling studies related to epilepsy [35, 19].
- **Hormonal cycles:** Documentation of the patient's menstrual cycle day is critical, as hormonal fluctuations are known to modulate seizure [44].
- **Physical activity and Exercise:** To ensure that the extracted features are correctly contextualised, an exercise label has been added to differentiate activity-induced spikes of physiological data and seizure-related autonomic changes. This includes recording the type, duration, and intensity of recent physical exertion [45].

Emotional and Psychological Context

- **Moods** Multiple emotional states (happy, relaxed, lively, nervous, sad, irritated, anxious, and bored) have been included due to their significant relation to seizure self-prediction [8, 39]. These moods, anxiety for instance, have been included as they are common self-reported symptoms of an ictal aura [42]. The calm/relaxed button is used to capture non-stressed physiological states for creating a baseline.
- **Stress levels:** Subjective ratings of perceived stress or recent acute stressors, which are strongly correlated with seizure onset for many patients [46]. Low/high stress buttons have been included to enable the modeling of the patient's stress-seizure link and analysis of the physiological stress response [47].
- **Cognitive load:** A high cognitive load button may also provide context for scenarios of mental exertion. This can also be reported by using a mental exhaustion or difficulty concentrating button [39].

Environmental and External factors

- **Diet:** Reporting recent meal consumption, caffeine, alcohol, or specific dietary changes, as metabolic factors (e.g., blood sugar level) can influence seizures [48].
- **Environmental context:** Recording the physical location or activity, providing context for the physiological data being streamed (e.g., at work, watching a bright screen, driving), including environmental light, temperature and audio. Patterns in these parameters can characterise the lifestyle of a user, enabling opportunities for efficient interventions by nudging [49].
- **Sensory stimuli:** Reporting exposure to known sensory triggers like flashing lights or certain sounds [11].

Conclusion

Previous literature emphasises the importance of capturing seizure events, premonitory symptoms, and contextual factors. However, the gap of a systematic method for converting subjective information into physiologically aligned labels ready for machine learning models still remains. This project extends this literature by introducing a methodology to temporally align these subjective reports with physiological data streams, usable for interpretation, analysis and personalised machine learning applications.

3.2. Physiological data

This section examines all physiological data related to epileptic seizures and highlights their importance in providing objective measurements that complement the subjective reported labels.

Enabling patients to access their personal health data can benefit both clinicians and patients, as it increases engagement and keeps both informed [50].

3.2.1. Single variables

Studies have demonstrated the efficacy of using electrocardiography (ECG) and respiratory rate (RR) [51] to derive measures like Heart Rate Variability (HRV) for seizure detection [52, 53], with expanded approaches incorporating Electrodermal Activity (EDA) and electromyography (EMG) [5]. The selection of these physiological streams aligns with the established finding that generalised tonic-clonic seizures are accompanied by distinct changes in the interrelations within the autonomic nervous system (ANS) [54]. The integration of this physiological data with self-reported contextual data is crucial, as the physiological signals alone often lack the necessary context to differentiate between a non-epileptic event (e.g., stress) and a pre-ictal state [5].

3.2.2. Wearable data

The project requires the collection of physiological data alongside contextual journaling to achieve seizure forecasting [18, 37, 55]. Some seizure detecting wearables that are currently available on the market are the Embrace 2, Epi-Care free, Nightwatch, Smart Monitor, or EpiMonitor (Figure 3.1) [56]. Wearables and smartphones enable remote patient monitoring in a non-invasive way [34].

Figure 3.1: Currently available epilepsy wearables

To transform physiological parameters from symptoms being journaled afterwards, to predictors being measured before journaling is possible, the following data types are being measured by the wearable:

Physiological data types

- **Sleep**: Alterations in sleep quality are known to be relevant to epilepsy [38]. The Apple Watch, through its accelerometer and heart rate sensor, collects data for sleep tracking and motion.
- **Exercise and Motion**: Heart rate (HR) and motion tracking (accelerometry/activity) are directly linked to seizure triggers [45]. The Apple Watch features an optical heart sensor and an accelerometer, collecting both data streams. Critical slowing can also be a biomarker for seizures [57].
- **Stress**: Stress is a known risk factor for seizures [58, 46], and its physiological manifestation is tracked via measures like Heart Rate Variability (HRV) [45].
- **Autonomic Changes**: Changes in autonomic function (e.g., heart rate and EDA) are strongly associated with seizure onset [45]. While the Empatica E4, mentioned in the brief, is a specialised research device capable of measuring Electrodermal Activity (EDA) [59, 60], newer Apple Watch models (Series 6 and

later) incorporate an electrical sensor capable of measuring EDA/skin conductance (used in various ways to infer stress and autonomic changes), making it potentially adequate for this project's initial phase. For this project an Apple Watch is used without an EDA sensor, inferring stress via indirect metrics such as HRV.

- **Increased sympathetic indicators:** An increase in Low-Frequency (LF) power or a rise in the LF/High-Frequency (HF) ratio immediately before a seizure indicates increased sympathetic activity relative to parasympathetic activity. This sympathetic surge is a strong physiological candidate for labelling the pre-ictal state in the ML model [5, 51].

Environmental data types

- **Environmental light:** Previous research by Baud et al. (2018) [61] explored the relationship between the average risk ratio and 24 h circadian and 8-day multidien cycle of epilepsy patients. To quantify the circadian cycle and extract features, continuous analysis of the environmental light has been included. This enables analysis of multiday rhythms [40, 62, 63].
- **Audio:** The perceived exposure to noise has been associated with an increased seizure risk [64]. This suggests that environmental noise can decrease the seizure-risk threshold.

Conclusion

This section demonstrated that physiological data types are relevant for seizure detection and forecasting. The features that have been extracted and implemented during development of the application can be found in Appendix B.1, including their formula. However, physiological data alone do not provide sufficient information to estimate what might have caused a seizure, leaving a gap in how the physiological data can be meaningfully interpreted in relation to the user's life. This project addresses this gap by supporting the interpretation of those physiological measurements by linking them to user-journaled labels. The app extends the literature by giving meaning to the physiological data through their relation to subjective events and contextual factors.

3.3. Forecasting

This section explores the latest developments how multimodal data fusion and machine learning techniques enable seizure detection and prediction, highlighting the transition from afterwards symptom analysis to proactive forecasting.

Pre-ictal periods

Chen et al. (2024) [65] describe how data from the pre-ictal period, the phase before an epileptic seizure, can be used in deep learning techniques to build seizure detection and prediction systems.

Multi-modal data fusion

The ability of wearable devices to facilitate reliable forecasts of seizure likelihood is acknowledged, providing the potential for patients to modify activities or take fast-acting medications [18, 37, 66]. Seizure detection and forecasting efforts increasingly rely on

combining signals that reflect changes in the autonomic nervous system (ANS) [54].

The project's methodology, which integrates self-reported data with physiological signals, is consistent with the current consensus in prior wearable epilepsy research, that seizure detection and forecasting benefit from multimodal data fusion [5].

The correlation of the physiological data types can be calculated using a Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) [54, 67]. The strength of the correlation between any two modalities can be estimated with an overall correlation (Appendix B.1.1).

Integrating multimodal data fusion into healthcare platforms enables predictive modelling, preventive interventions, and personalised healthcare strategies. [68].

3.3.1. Personalisation

The project's goal of developing a personalised model aligns with the latest advancements in this field, which prioritise adaptive models over generalised solutions. Recent deep learning approaches have demonstrated that personalised models are significantly more effective at pre-ictal seizure prediction than non-personalised alternatives, confirming that individual variability in seizure dynamics requires tailored modelling [69, 30]. Such models require long-term datasets [70] and incorporate individually related parameters to determine a baseline and be able to identify irregular physiological conditions corresponding to an individual patient [71]. By focusing on the acquisition of high-fidelity, user-specific contextual data through the journaling system, the project facilitates the creation of such adaptive models. This overall methodology directly supports the widely recognised goal of using wearables to facilitate reliable forecasts of seizure likelihood, empowering patients and caregivers to modify activities in anticipation of a seizure [18, 72].

Classification of data;

There are several models that can be used to classify data [73]. RNN and CNN models are used for data classification for physiological multi-signal predictions [74]. Poincaré plot and k-means are used in HRV analysis to identify oscillations of the heart rate in partial epilepsy [75].

Conclusion

The forecasting literature supports the use of pre-ictal analysis, multimodal data fusion, and seizure prediction. Although literature suggests that personalised and multimodal methods are promising, there has been no data collection system to generate labelled data sets needed to train such models in remote, real-life settings. The project contributes to this area by capturing timestamped subjective events and aligning them with the continuous physiological data for future forecasting models.

3.4. Nudging strategies

This section reviews literature on nudging strategies for mobile applications, with a focus on notification timing, frequency, and personalisation as determinants of usability

and user engagement to ensure consistent journaling behaviour.

There are challenges in monitoring seizure frequencies over a long-term period [76] and keeping patients engaged in eHealth services [77]. And to be able to provide any insights about which label causes which physiological changes, a dataset of historical journaled labels must be established. Therefore, to create this dataset as fast as possible, a nudging method which impacts the journal behaviour in the most positive way (in terms of amount of labels, total covered time and response time) is investigated for the final concept design.

Strategies

1. Personalised: This method explores the nudging of participants at personally chosen timeslots for receiving notifications. These times are determined based on their personal schedule and preferences [78, 79].
2. Fixed: In previous research, data were collected twice a day with a time interval of 12 hours, using an e-diary [8]. This method includes notifications at predetermined times (10:00 and 20:00) each day, regardless of the participant's physiological state or scenario.
3. After Inactivity: This method includes the period of time since the last journaled label. When leaving the app, a reminder is sent after. Predefined app-based reminders are suggested to maintain compliance over a long-term period [80].
4. JIT: Another nudging technique that has been identified to test during the research is by nudging the user just-in-time (JIT), when a user switches from task or scenario, which can reduce mental effort [81, 82]. Personalised, just-in-time nudges yield better outcomes than generic ones. De Vries et al. [79] show that personalising both content and timing of JIT nudges produces measurable improvements in behaviour/choices compared to non-personalised prompts. For journaling, this means tailoring when and how prompts are shown (e.g., immediately after a detectable physiological change, or at user-preferred windows). Chiam et al. describe an algorithmic nudging algorithm which uses health data from the user through a mobile app, which is synchronised with the user's wearable [83]. JIT increases the probability that the user will record accurate context near the event and converts physiological blips into labeled data.

Time

Freyne et al. [78] used the following notification moments during their study: In the morning at 10:00, in the afternoon at 14:30, and in the evening at 20:30.
Many users were immediately spurred to action, or took action within 1–2 hours, for many the receipt of a second or even third prompt for an alternative task can be a motivator to comply.

The response times for tasks decreases from 4.87 hours in the morning to 1.25 hours in the evening. During the free-living period, the completion rates for the morning and afternoon prompts increase to about 61 percent. This argues that to achieve the fastest response time, a period where the users knows he has time in the evening yields the most accurate label measurements.

During their study, people postponed their tasks after receiving a prompt to a moment later that day when they had more time. This depends on the individual's agenda and is mostly related to their job/ study. This indicates that a personalised notification schedule would suit the journaling method better due to a faster response time (and thus a better recall of events and timeframes).

The response times for tasks appeared to decrease when the window of opportunity for task completion decreased, and the number of reminders increased. This indicates that when receiving a reminder after one hour of receiving the push notification, with a time limit, would boost the amount of logged events.

Amount

A frequency of three push-notifications per day has been identified as the user tolerance limit [78]. The added value of regularly sending push-notifications to users has also been confirmed [84, 85, 86]. Frequent exposure to nudges increases intrinsic motivation, but only if the nudges are perceived as helpful and not as coercion [87]. However, there is also existing literature indicating that push notifications can cause user disruption [88]. Wohllebe et al. [89] demonstrated that notification frequency beyond user-specific tolerance thresholds leads to decreased app engagement. Therefore, a limit of 3 push notifications has been identified for the graduation project, with a reminder after a minimum of one hour of inactivity when no events have been logged.

Conclusion

The literature suggests that reminders such as fixed, inactivity, Just-In-Time, or personalised schedules can improve engagement in digital health tools. Nevertheless, the gap remains in understanding which type of nudging strategy is most effective for long-term self-tracking instead of only general engagement. This project addresses this issue by integrating nudging as a part of collecting data. The research study extends the literature by framing nudging as a method for improving the quality of journaled subjective labels.

3.5. Conclusions from literature research

This section synthesises the key findings from the literature research and design context of the project.

Key findings

The central unresolved gap that remains is that current tools and methods study subjective and objective data separately, while a practical method to collect and align these data types remains underdeveloped. Tools can detect seizures, but do not capture contextual triggers and connect them to physiological changes that support deeper analysis and personalised treatment.

1. Although research indicates that digital journals are more effective than manual/paper journaling, and the fact that wearables can measure physiological data that can be used to identify epileptic-related events, a method to automatically integrate subjective triggers into physiological data, or analyse triggers afterwards, has not yet been established.

2. Neither all physiological nor subjective events can be captured using only a wearable device or diary. Many triggers cannot be captured sufficiently by a smartwatch or an individual's indication alone. Multiple categories of types of labels can be used to analyse their influence on a patient's epilepsy. Because epileptic triggers differ per individual, all possible epileptic-related labels are included for a multimodal and personalised approach to represent patient-specific patterns accurately.
3. Seizure prediction relies on identifying patterns across combinations of physiological data, and such models require data sets of synchronised data and correlations linked to contextual labels. Collecting, synchronising, and creating these data sets remains a challenge in current research. A personalised ML model is significantly more effective than a generalised model in the identification/prediction of seizures related to an individual's specific epileptic-related changes in physiological conditions.
4. To achieve long-term data collection, the app must balance usability, safety, and engagement. Three types of nudges that have been identified to test their effect on journaling behaviour are fixed, inactivity, and personally chosen nudges.

Design context

The design context (Figure 3.2) is constructed by integrating three key domains: clinical relevance, data processing, and usability.

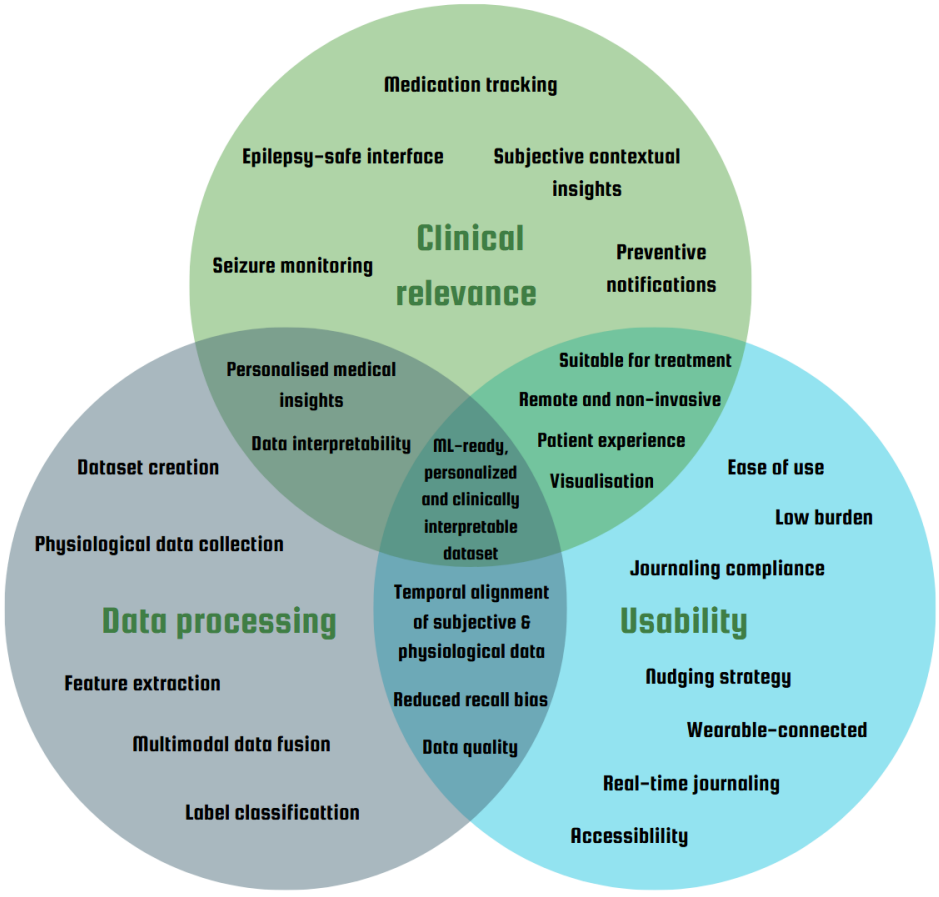


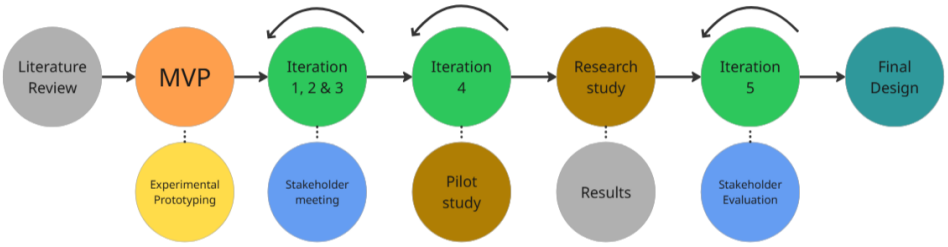
Figure 3.2: The project's design context

Clinical relevance is driven by the need to capture subjective seizure-related events and contextual triggers, which are essential for personalised insights and treatment. The data processing domain is established by research on transforming and temporally aligning subjective and physiological data into interpretable, ML-ready datasets. And usability is shaped by mobile health and journaling literature, emphasising low user burden, real-time data capture, and effective nudging strategies to ensure long-term engagement. The overlap of these domains addresses the identified gaps of linking subjective and objective data into a coherent, usable system for personalised seizure forecasting.

4

Prototype Development

This chapter discusses the iterative prototyping process of the app in anticipation of the research study of the project, to be able to test different nudging strategies.



Based on the literature research and conclusions drawn in Chapter 3, a list of requirements has been drawn up, and the choice of designing an app which connects to a user's Apple Watch has been made to mitigate the extra step of manually digitising subjective-related factors captured during the day of an epilepsy patient and directly linking them to the physiological Apple Watch data. A smartphone and Apple Watch application are very accessible and enable real-time remote patient monitoring.

To be able to train an ML-based personalised forecasting system, high-quality classified datasets of subjective journaled events are required.

During the prototyping process, multiple types of stakeholders have been approached to be able to identify and fulfil their needs (Figure 4.1).

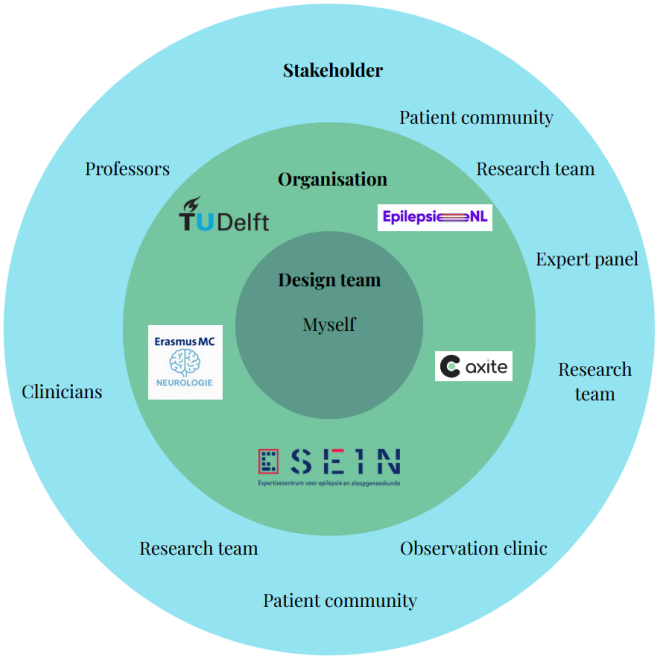


Figure 4.1: Stakeholders engaged during the project.

List of requirements drawn from literature research

Journal Logging System

1. The application must support structured multi-category labeling of seizure events, physiological triggers, emotional states, and environmental context.
2. The application must allow multiple and overlapping labels within a defined timeframe.
3. The application must support three logging modes:
    - Instant snapshot logging (max. two interactions, automatic timestamp),
    - Manual timeframe selection (including retrospective logging),
    - Stopwatch-based logging (start, pause, stop).
4. Completing a journal entry must require minimum interaction steps, with the logger page as the default home screen to increase accessibility.

Wearable Integration

1. The application must collect synchronised physiological data including heart rate, HRV, motion/activity metrics, and sleep indicators.
2. The application must collect environmental exposure data including light and audio levels.
3. All wearable and journal data must share a unified timestamp reference to enable synchronised analysis and ML preprocessing.

Timeline and Visualisation

1. The application must visualise labels and medication data within a horizontal timeline, supporting overlapping events and real-time indicators.
2. Users must be able to review, edit, and delete logged entries.
3. The application must provide a structured overview of historial entries including category, label, and timeframe.

Nudging Strategies

1. The application must support configurable daily notifications.
2. The application must support inactivity-based reminders following predefined non-logging intervals.
3. The application must support predefined threshold notifications triggered by detectable physiological changes.

Data Management

1. The application must export synchronised journal and wearable data in CSV/Excel file format.
2. The application must prevent data loss through local buffering and maintain timestamp precision suitable for 30-second window segmentation.

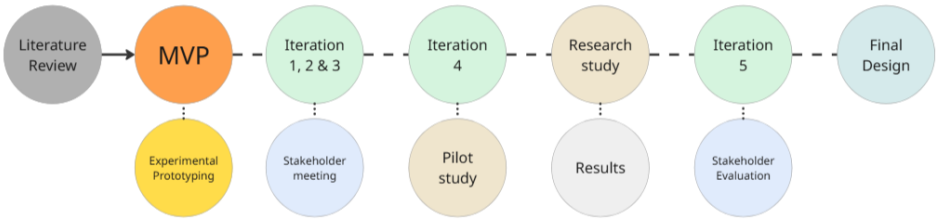
Usability and Safety

1. The interface must minimise cognitive load through grouped labels, limited scrolling, and a clear navigation structure.
2. The interface must avoid visual elements that may trigger photosensitive seizures (e.g., flickering, rapid high-contrast transitions).

Research Readiness

1. The application must enable measurement of notification response time, logging frequency, and labeled time coverage.
2. The prototype must operate continuously for at least nine consecutive days and run on iPhone and Apple Watch.

4.1. Minimal Viable Product



4.1.1. Connected application

The connection between iPhone and web app, and the visualisation of the interface was tested using a WebSocket web server made with a developed Python and an index.html file (Appendix B.8.2), to verify the collection of label input from an iPhone to the server. The web app is accessible through an internet browser on the user's iPhone and runs on either a private laptop or a Raspberry Pi, connected to the same network as the iPhone.

4.1.2. Interface

To minimise cognitive load and complexity, the app must be easily accessible. In addition, the registration of events must contain a minimal number of consecutive steps to complete. This enhances the total number of logged events by users. Therefore, the homepage of the app is the 'logger page' where users can immediately log their events during the day when entering the application.

A smooth, fading transition minimises the high-contrast colour changes of the interface. Visual patterns or repetitive geometries have been avoided.

Buttons

The interface includes all of the identified labels in Chapter 3.1.4, which are colour-coded based on their corresponding label category (Figure 4.2):

1. Mental/Emotional state
2. Environmental/Sensory factors
3. Activities/Behaviour
4. Sleep
5. Medication
6. Seizure-related
7. Physiology/Health

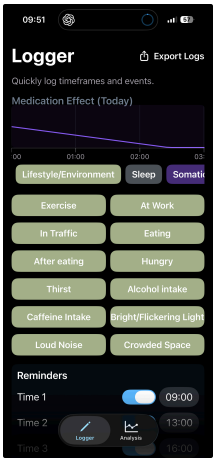


Figure 4.2: Colour-coded label buttons.

Potential User Scenarios

As discussed in Chapter 3.1.3, accurate determination of the timestamps of a journaled event is crucial for data classification and future ML purposes. Three potential user scenarios were presented as possible ways to journal an event on the app (Figure 4.3).

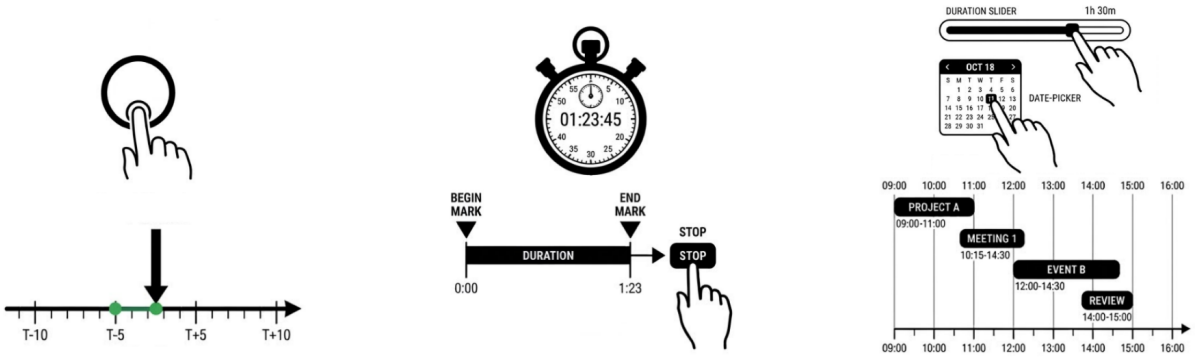


Figure 4.3: The fixed, stopwatch and flexible journaling methods

A 'snapshot', when a user immediately wants to journal an event. Pressing a button registers a predetermined, fixed, small timeframe (5 min, for instance) to capture a single-label scenario.

Using a stopwatch tool, when a user specifically wants to capture data from a preplanned event. The user hits the stopwatch button in real time when the event starts and finishes to determine the timestamps more precisely.

Journaling events afterwards requires flexible timeframe determination, where the user selects start/stop timestamps of the timeframes manually at any time.

These methods provided the idea to include an overview of the logged timeframes, where the labels are represented in a timeline of the day.

Timeframe Deteremination

When pressing a label button, a window pops up asking to determine the length of the timeframe of the corresponding label, which includes a start and end time (Figure 4.4).

The total number of steps to complete a journal measurement is limited to 4: clicking the label, determining the timeframe length, determining the starting moment, and confirming.

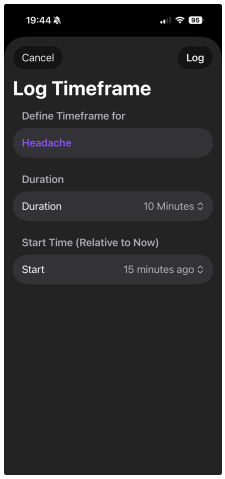


Figure 4.4: Timeframe determination.

Analysis page

To prevent scrolling, an extra page has been added to the interface of the app. Besides the 'logger' page, intended for directly accessible journaling, an 'analysis' page has been added so users can view their logged labels and timeframes, physiological data from the wearable, and determined thresholds for trigger-related notifications (Figure 4.5).

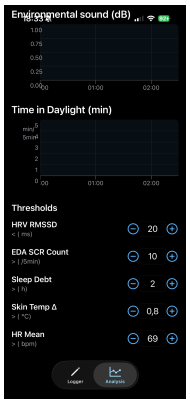


Figure 4.5: Analysis page and threshold determination.

Label Timeline

To create a clear overview of all logged labels, a timeline has been added on the second 'analysis' page of the app. The choice of separating the timeline from the labels has been made to minimise cognitive overload when meeting the primary objective, journaling labels.

The labels are visualised on the timeline with their corresponding label colour, and beneath each other when overlapping (Figure 4.6). The frame of the timeline is scrollable from left to right to go back and forth in time. A clear, dotted vertical line is presented on the x-axis, indicating the current time. When tapping one of the labels on the timeline, at any moment, the label can be adjusted or deleted (Figure 4.7).

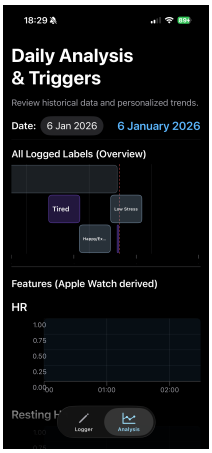


Figure 4.6: Label timeline.

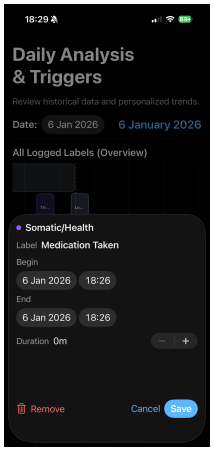


Figure 4.7: adjusting/checking labels by pressing them.

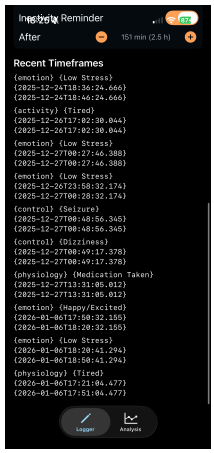


Figure 4.8: Recent logged labels overview.

Exporting Data

Another section has been added at the bottom, for a total overview of all recently logged labels in CSV-file form (Figure 4.8), defined as (category) (label) (begin-time) (end-time). An export button has been added on the logger page. This button enables users to export files with their logged data to other devices.

Medication Timeline and Suggested Triggers

From my own experience as an epilepsy patient, I know situations can occur where you forget to take your medication. A missed dose can highly affect seizure risk. Therefore, a medication-level graph linked to a reminder could help prevent a missed medication intake (Figure 4.9). In addition, an interview was conducted with a clinician (Appendix A.1) to identify important needs for stakeholders. Questions were asked about the type of information a practitioner would need to treat a patient more effectively. Key findings that have been applied as new design requirements include an overview of a patient's medication doses throughout the day, an overview of suggested epileptic-related situations, and the types of physiological data related to these situations were considered helpful during treatment. This should help a clinician to analyse the relationship between epileptic events, medication level/usage, and the corresponding physiological conditions.

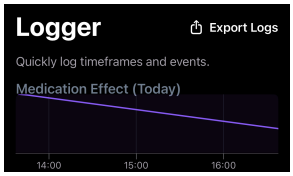


Figure 4.9: Medication timeline.

Nudging strategies

To include the options of all three identified nudging strategies in the research study from Chapter 3.4, multiple notification reminders have been incorporated that can be enabled and chosen manually at predetermined fixed times or after a period of inactivity (Figure 4.10).

The messages that are sent as push notifications contain the following text:

1. Please open the app and log any recent events.
2. Please log any recent scenarios.
3. The medication effect has worn off, please log your next dose.

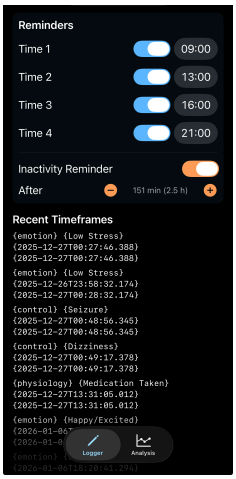


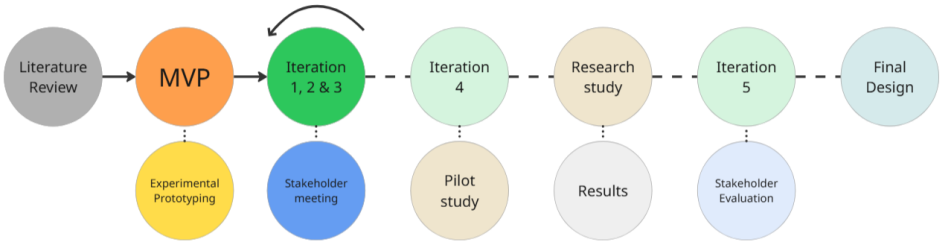
Figure 4.10: Fixed and inactivity notifications.

4.1.3. Research direction

After experimenting with the App interface, a functional prototype has been established, collecting and synchronising subjective journal and objective wearable data. After the MVP was developed, the research questions arose about how the prototype would align with the needs of all stakeholders related to the interpretability, aesthetics, and usability of the interface. And most of all, how the different types of identified nudging strategies impact a patient's journaling behaviour, to ensure user engagement and data collection over a long time period.

4.2. Iteration prior to pilot study

The design process of the prototype before the pilot study (Figure 4.11) consists of three iterations. From this phase, the application has been further developed using Swift in Xcode (Appendix B.8.1), to be able to test, distribute, and download the app on an iPhone remotely.



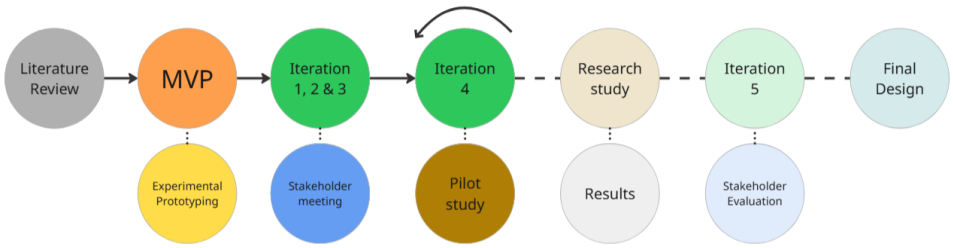
The iterations include:

1. Addressing requirements from the literature research and stakeholder consultations, to establish a usable and accessible interface, reducing complexity and minimising cognitive load using a S.C.A.M.P.E.R. method (appendix B.3).
2. Incorporating a baseline analysis functionality to support the interpretation of the physiological data and to support future ML model purposes (appendix B.4).
3. Lastly, the design has been refined based on stakeholder feedback (Appendix A.2, A.3, and A.4), aligning with their needs and adding personalised attributes such as custom labels and improved visualisation of data (Appendix B.5).

Figure 4.11: The concept design prior to pilot study.

4.3. Pilot study - iteration 4

Before starting the research study discussed in the next chapter, two participants have tested the functionalities of the iPhone application and the Apple Watch application. Improvements were made based on the key takeaways from the pilot and integrated into the design for the research study (Figure 4.12).



Goal

The pilot study is used to test the following functionalities of the app:

- Correct data collection and synchronisation
- Data visualisation and interpretation
- Perceived usability

Method

The pilot study was conducted to collect subjective feedback and test if all functionalities operate correctly, excluding any quantitative data collection and focusing on the qualitative insights from an open interview after the study.

Participant recruitment

Participants were recruited through community outreach and leveraging personal networks. The participants were given an informed consent form, and a real-life showcase was conducted on location to explain the functionalities of the app. The setup location differed per participant, but did not have any consequences. Participants required an iPhone with a minimum version of iOS 18 and an Apple Watch with a minimum version of WatchOS 8 to conduct the pilot study.

Participant procedure

The participants were asked to use the app to journal events during the day for a total of nine days, using the iPhone or Apple Watch application. The three identified nudging strategies from Chapter 3.4 that are being investigated have each been tested for three consecutive days. Not to track any outcome measures, but to check if the different nudging strategies operate correctly.

After the pilot, the participants were asked to participate in an interview where open questions were asked related to the usability and interpretability of the app's functionalities (Appendix C.1).

Key takeaways

1. The sleeping period and timestamps are difficult to estimate afterwards in the morning.
2. The dark-green theme colours cause difficulty reading and confirming label buttons. Especially for people with poor eyesight or advanced age.
3. Some of the buttons' colours depended on the 'light' or 'dark' mode of the user's phone. The 'light' mode made the text on the buttons hard to read due to the low contrast.
4. The journal analysis data was not correctly stored. Overnight response time was also added to the total latency.
5. Not all features appear in the app, as not all data types can be collected with older WatchOS versions.
6. Keep the analysis page available for users, as personal data is perceived as interesting. Initially hiding sections prevents confusion and minimises cognitive load.

Improvements

1. Suggested 'Sleep?' label fetched from the sleep-phase feature graph. This data is uploaded to the app in the morning after the user's Apple Watch connects with the iPhone. The sleep phases and data from the Health app of the iPhone are collected by the app.
2. A suggested 'Interruption' label is generated when an 'awake' state appears on the 'sleep phases' graph or when a feature's predetermined threshold is met.
3. The theme colours have been adjusted to increase colour contrast, with a lighter background and text colour, and added shades to the text on the logger page.
4. Forced the application to appear in 'Dark' mode.
5. Refresh the latency-timer every morning after the first notification of the day.

Figure 4.12: Research study design

5

Research Study

A research study was conducted to evaluate the effectiveness and acceptability of three nudging strategies and the overall usability of the interface of the app.

This is tested through a custom SUS questionnaire and journal compliance metrics. The journal compliance refers to the extent to which participants consistently log subjective factors in the journaling application when nudged. The compliance is quantified through three behavioural metrics: the number of journaled labels, the latency between a nudge and journal entry, and the total temporal coverage of the journaled labels. These metrics reflect not only whether users engage with the journal application, but also how timely and complete their self-reported data are. High journal compliance reduces recall bias and improves timestamp accuracy, resulting in more reliable data. This is essential for both clinical interpretation for effective treatment and for the development of accurate machine learning models, as inconsistent data directly limits predictive performance of future models.

Epileptic triggers are highly individual, and not all participants are diagnosed with epilepsy, making it inappropriate to judge label relevance at this stage. The effectiveness of interpretation and future machine learning models depends first on the availability, consistency and temporal accuracy of data before being able to determine which specific subjective labels are most predictive. Without proper compliance, any journaled subjective factor would not be suitable for accurate analysis and treatment. If a nudging strategy improves engagement, it will do so regardless of what type of label is journaled. Therefore, tracking the types of labels people have journaled is excluded from this research.

5.1. Research question

The goal is to evaluate which nudging strategy is most suitable for the final design of the journal application and for future ML model training purposes. The research question guiding this study states:

- *Which nudging strategy leads to the highest journaling compliance while preserving positive user experience and low perceived burden?*

Null hypothesis

H0: There is no significant effect of the nudging method on the journaling behaviour.

Alternative hypothesis

H1: The nudging strategy has a significant effect on journaling behaviour.

Sub-hypotheses

H1a: The nudging method significantly affects the number of journaled labels.
H1b: The nudging method significantly affects the total time coverage of journaled labels.
H1c: The nudging method significantly affects the response latency after a nudge.
H1d: The nudging method significantly affects the perceived user experience.

5.1.1. Variables

Independent variables

Three different nudging strategies have been tested as independent variables. The strategies consist of receiving fixed nudges at predetermined times, after a period of inactivity, or personally chosen reminders.

Dependent variables

The effectiveness of the nudging techniques is evaluated across two types of dependent variables, the journaling compliance and user experience. The variables are measured using a total of four outcome measures to answer each sub-hypothesis:

Journaling Compliance
a) The total amount of journaled labels.
b) Latency between a nudge and a journaled label.
c) The total temporal coverage of journaled labels.

User experience
d) The overall perceived usability of the app and nudging strategies has been evaluated using a custom SUS questionnaire.

5.2. Method

A within-subjects study design was conducted in which every participant tested all 3 nudging conditions, each for 3 consecutive days. To counterbalance potential learning effects, randomised orders of the conditions have been applied across participants using a Latin square method. All six possible orders of the conditions have been tested,

with one order being assigned to two participants due to a sample size of seven participants. In addition to reducing the inter-individual variability, the within subject design increases statistical sensitivity by allowing participants to act as their own controls. Each participant completed repeated tasks across multiple days and conditions, yielding a high number of observations per individual, partially compensating for the limited sample size. This approach enables the detection of within-person behavioural trends that may not be observed in a between-subject design.

Three nudging conditions are being evaluated as independent variables:

1. Fixed-time nudges, where participants receive 3 nudges a day (literature research from Chapter 3.4 concluded that this is the maximum amount of nudges before users perceive them as a burden). The predetermined times are set at 11:00, 17:00, and 22:30. These moments have been chosen without external validation, to give participants time in the morning to experience scenarios before being nudged to journal any events, and to have a buffer of 5,5 and 6 hours to prevent repetitive nudges and to allow participants to reflect on their encountered events throughout the day after working hours and before going to bed, as these are the moments people are most likely to have spare time to complete a journal task and minimises the latency and perceived burden.
2. Inactivity-based nudges, which trigger after 3 hours when a participant has not engaged in any journal activities on the application, meaning no journal entry of an occurred event. The 3-hour threshold was selected based on two sources from the literature review (Chapter 3.4). Which reported that response times varied between 1.25 and 4.87 hours and that notification frequency has a measurable negative effect on continued app engagement when the threshold of user tolerance is exceeded. A 3-hour inactivity reminder remains within the suggested tolerance range, balancing risk of recall bias and perceived burden.
3. Personalised nudges, in which participants are asked to reflect on their daily schedule, and make an indication at which times it would be most effective to receive nudges to remind them to journal any labels.

Participant recruitment

The participants were adults in the Netherlands who all have a university degree. The recruitment was conducted using personal networks and community outreach through stakeholder meetings.

All participants were provided with an informed consent form prior to their participation. The consent form includes a written elaboration of the purpose, procedures, data handling, and voluntary nature of the study. The study did not involve the collection of clinical or physiological data from participants. Participants were also informed that they could withdraw from the study at any moment without consequences.

5.2.1. Setup

The participants used the developed journaling application, which is installed on their personal iPhone through TestFlight. This requires participants to enroll in an internal team on Appstore Connect. Participants must add an account using an email address

to be able to download the application through Testflight. An Apple Watch was provided as part of the system to support wearable-based interaction with the application. However, no physiological data were captured during the study, as the study is focused on the journaling behaviour and not the collection of physiological data.

All nudges trigger both vibration and ringtone on the iPhone and Apple Watch. The nudges' notifications state: "Please open the app and log any recent events". However, participants retained full control over their device notification settings. Therefore, it could not be guaranteed that all notifications were perceived by participants. While this approach reflects realistic usage conditions, it introduces variability in notification exposure and response behaviour, which may increase noise in the data.

Requirements

To deploy the app on both iPhone and Apple Watch, participants must own an iPhone with a minimum version of iOS 18 installed and a watch with a minimum version of WatchOS 8, which are required to run the application properly.

5.3. Participant procedure

Participant procedure 1 - Deployment

After installing the app, participants receive a short in-person instruction explaining how to journal events accurately, how to determine a label's timeframe (Figures 5.1 and 5.2), and how to adjust the nudging strategy (condition). Participants were instructed to journal relevant events throughout their day, while also being nudged reminding them to do so. After finalisation of each condition, participants were contacted briefly to ensure that they switched to the correct condition.

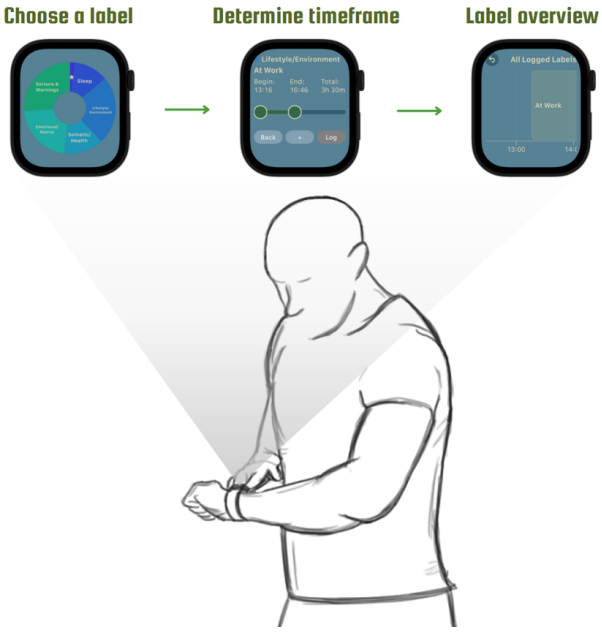


Figure 5.1: Journaling a scenario using the Apple Watch application.

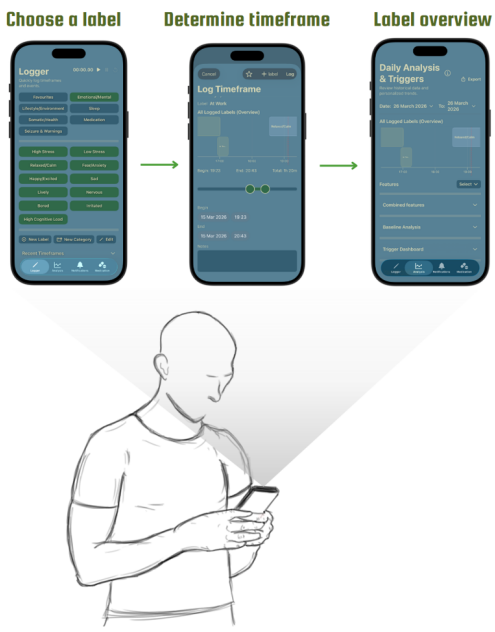


Figure 5.2: Journaling a scenario using the iPhone application

Participant procedure 2 - Quantitative data collection
During the research study, the application automatically registers:

1. The timestamps of all nudges
2. Timestamps of journaled labels.
3. Start and end times of the labels.

From this data, the following objective outcome measures were derived:

1. The total amount of journaled labels
2. The total amount of time covered by labels
3. The journal latency (minutes between nudge and first logged label)

Participant procedure 3 - Qualitative data collection
After completing all nudging conditions, participants evaluated the application and nudging strategies using a custom 5-point System Usability Scale (SUS) questionnaire (Appendix C.3). The standard 10-item SUS was extended to 29 items to capture dimensions specific to this study, including nudging preference, perceived burden, and data interpretability, which fall outside the scope of the original SUS questionnaire. Subsequently, a semi-structured interview was conducted to capture any deeper thoughts on the application and reflect on the participant's journal behaviour during the different conditions. The interview questions and the corresponding responses are presented in Appendices C.4 and C.5.

5.4. Results

A total of 7 participants were recruited between the ages of 24 and 70, including 3 participants diagnosed with epilepsy. None of the participants had prior health app experience.

The main hypothesis is further specified into 4 sub-hypotheses related to the outcome measurements described in Chapter 5.1.

Each sub-hypothesis corresponds to a distinct dependent variable and is tested with a significance level of $\alpha = 0.05$.

5.4.1. Journaling compliance

The following descriptive and inferential statistics have been used to summarise outcomes per nudging strategy.

- Regular descriptive statistics show the mean values and error bars representing ±1 standard deviation, and the total differences between the nudge strategies. This provides initial insights to interpret differences between the nudging strategies and trends, before testing the statistical significance.
- Pairwise comparisons are used to test if specific nudge strategies differ from each other and to support the interpretation of the trends before testing the statistical significance. Because the risk of errors increases when testing multiple conditions, the Bonferroni correction is used to lower the chance of false positives (error type 1).

- Mauchly's test is used to check if the variance in differences between the nudges is equal, indicating whether the assumption of sphericity was violated, which is necessary for a repeated-measures ANOVA test to determine which results must be used to interpret the findings. Because a repeated measures ANOVA assumes that the variance of differences between all conditions is equal.
- A repeated-measures ANOVA test suits this research best due to the three conditions that each participant has tested, and by measuring the outcomes across a continuous scale. This test has been executed separately for each dependent variable. The F-value indicates a ratio of the explained/unexplained variance in the results; a low F-value indicates that the variance between nudging strategies is small compared to the variance within the participants. The p-value indicates the significance of the differences, and the effect size (η^2) shows how much of the variance in the dependent variables can be explained by the independent variables (type of nudges).

Measurements across the 3 days per condition were averaged per participant per nudging strategy to obtain stable estimates and reduce within-condition variability.

All corresponding tables with detailed results of the statistical analyses of each sub-hypothesis are presented in Appendix C.2.

H1a: The nudging method significantly affects the number of journaled labels.

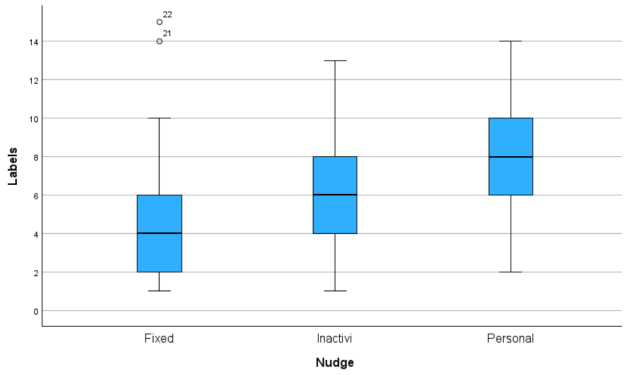


Figure 5.3: Mean daily logged labels per day for each condition (N=7, averaged across 3 days per condition). Error bars represent ±1 standard deviation.

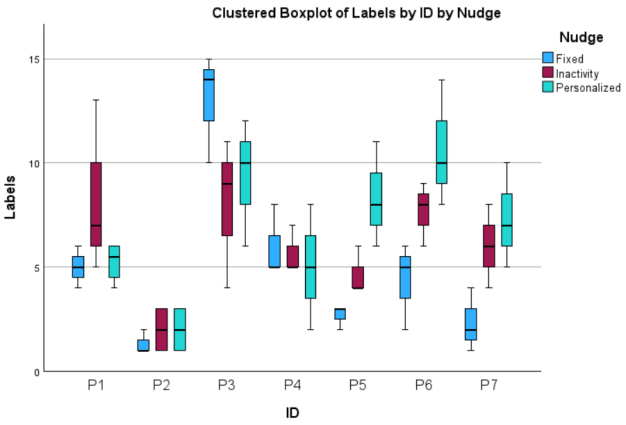


Figure 5.4: Total number of logged labels per condition, shown per participant across three days per condition. Boxplots represent the median, interquartile range, and full range of values.

Analysis

Observed mean differences suggest a directional trend, particularly for the personalised nudges (Figure 5.3). Individual participant data reveal variability in journaling behaviour across conditions (Figure 5.4), suggesting that the effectiveness of the nudging strategy may differ per individual, rather than a uniform effect across the sample size. A pairwise comparison does not suggest significant differences between the nudges, and due to a non-significant Mauchly test, the assumed sphericity results

of the repeated measures ANOVA were used. No significant difference between the type of nudge and the total amount of logged labels was found using a repeated measures ANOVA test, with F(2,12) = 1,386 and p = 0,287. The partial eta squared (η^2 = 0,188) shows an effect, suggesting that 18,8 percent of the variance in the number of journaled labels could be explained by the nudging strategy.

H1b: The nudging method significantly affects the total time coverage of journaled labels.

The total time coverage reflects the extent to which participants provide continuous subjective data throughout the day. A higher temporal coverage indicates more complete datasets, essential for aligning subjective and physiological data. However, this metric does not account for the relevancy or accuracy of the logged labels. Therefore, an increased coverage does not necessarily imply higher-quality data.

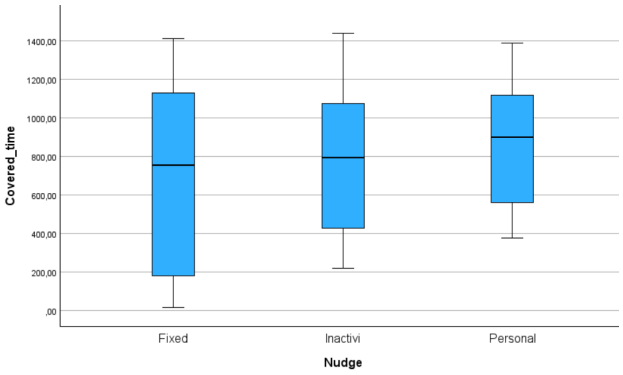


Figure 5.5: Mean of total amount of time covered with labels per day for each condition (N=7, averaged across 3 days per condition). Error bars represent ±1 standard deviation.

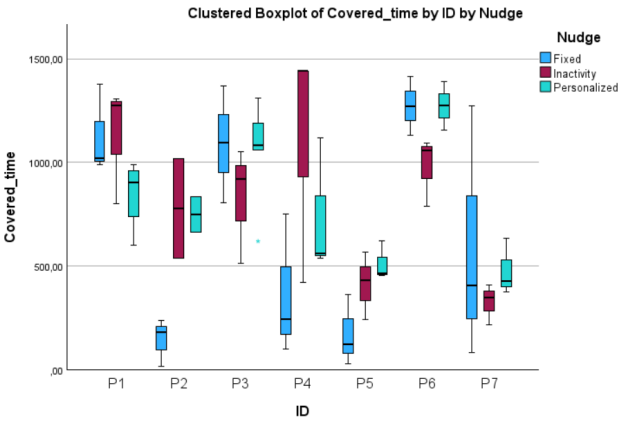


Figure 5.6: Total temporal coverage of logged labels (in minutes) per condition, shown per participant across three days per condition. Boxplots represent the median, interquartile range, and full range of values.

Analysis

The mean results show a small difference (Figure 5.5), but this effect is very weak. The distribution of time coverage across participants shows a considerable spread (Figure 5.6), suggesting differences in how individuals interpret and use the journaling system. The Pairwise comparisons showed no significant differences between the nudges, but indicate that, on average, the personalised nudges yield the highest amount of total time covered with labels. The Mauchly test was not significant, indicating that the assumption of sphericity has not been violated. Therefore, the sphericity assumed results have been used. A repeated measures ANOVA showed no significant difference between the type of nudge and the total time coverage of labels, with F(2,12) = 0,214, p = 0,811, and η^2 = 0,034.

H1c: The nudging method significantly affects the response latency after a nudge.

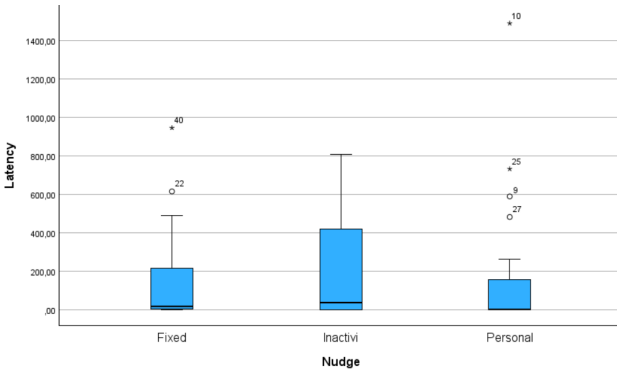


Figure 5.7: Mean of total latency per day for each condition (N=7, averaged across 3 days per condition). Error bars represent ±1 standard deviation.

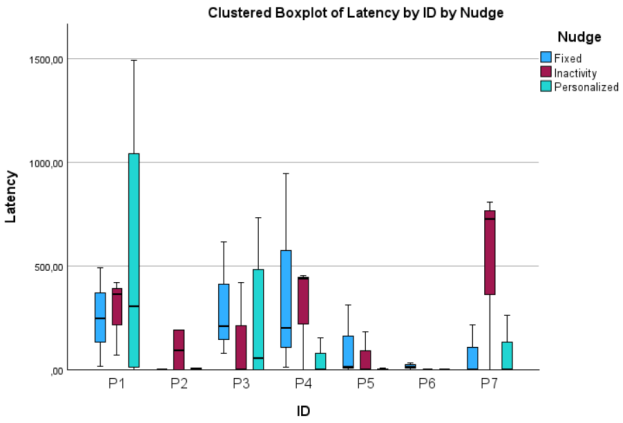


Figure 5.8: Latency (in minutes) between nudge and logged label per condition, shown per participant across three days per condition. Boxplots represent the median, interquartile range, and full range of values.

Analysis

The observed mean differences point out that the latency of responses of the personalised nudges is less than half (mean=79,19 and standard deviation = 82.63) of the fixed and inactivity nudges (Figure 5.7). Individual participant data show variability in latency both within and between participants (Figure 5.8), indicating inconsistent response times across days. While some participants demonstrate consistently low latency across all conditions, others exhibit large fluctuations, suggesting that response behaviour is context-dependent and influenced by individual routines rather than a consistent effect of the nudging strategy. The pairwise comparisons showed no significant differences in total latency between the different nudging conditions, but do indicate that the personalised nudges yield a shorter latency. Due to a non-significant Mauchly test, the sphericity assumed results were used for interpretation. The repeated measures ANOVA did not indicate a significant difference between the nudging strategy and the total latency, with $F(2,12) = 1,721$ and $p = 0,22$. The partial eta squared ($\eta^2 = 0,223$) shows that 22,3 percent of the variance in latency could be explained by the nudging strategy. This suggests a promising pattern, but it has not been statistically proven.

Summary

Although no statistically significant differences were observed across the conditions, visual inspection of the data and individual patterns suggests variability in responses and small directional trends. These findings indicate that the effect of nudging strategies may be subtle, context-dependent, and influenced by individual behaviour, rather than producing a strong uniform effect across users. The observed trends are further explored in the discussion.

5.4.2. User experience

There were 29 questions that were asked in the custom SUS questionnaire (Appendix C.3), which are ranked on a scale from 1 (strongly disagree) to 5 (strongly agree) by each participant. Mean scores per question are presented in Figure 5.9. Due to the descriptive nature of data collection, no inferential statistical test was applied across nudging conditions.

H1d: The nudging method significantly affects the perceived user experience.

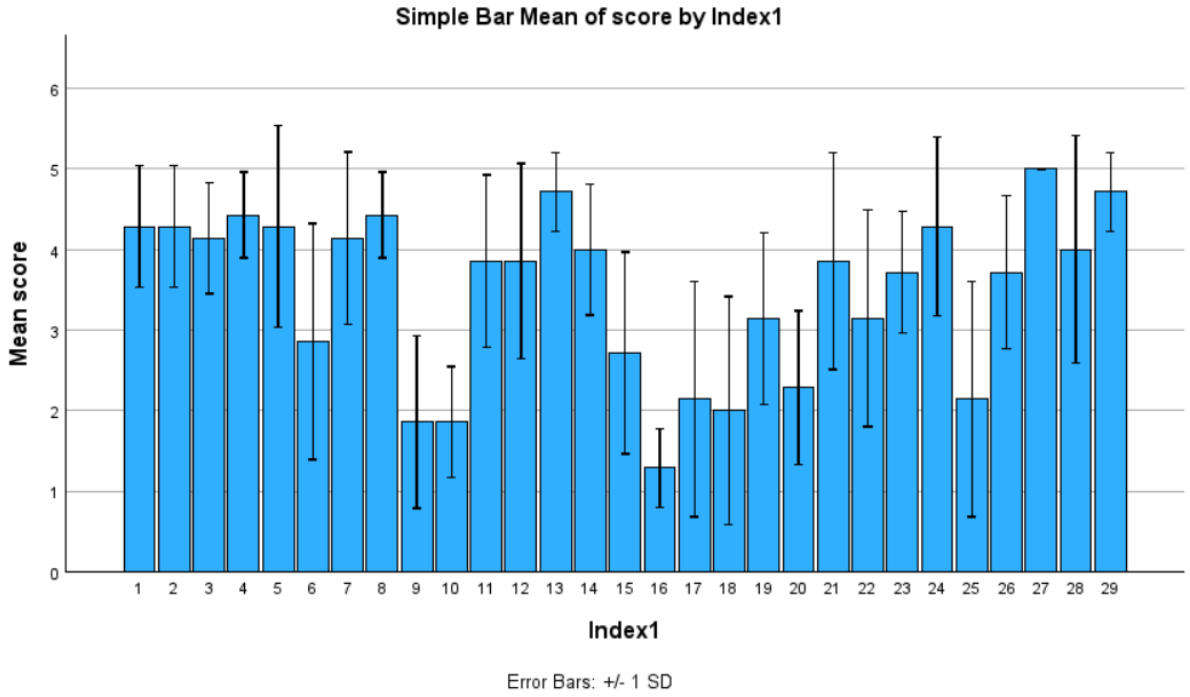


Figure 5.9: Mean values of SUS questionnaire results, with error bars indicating ±1 standard deviation across the participants

Analysis

Questions related to interface clarity, navigation hierarchy, and ease of journaling received consistently high mean ratings, indicating strong perceived usability across participants

No participant rated nudges as counter-effective (question 16) or annoying or frustrating (question 17), with both items receiving mean scores below 2.

The question regarding scrolling received a mean rating of 2.71, suggesting that scrolling remained a more significant interaction burden than anticipated beforehand.

The results related to nudging effectiveness showed a consistent trend, where personalised nudges (question 13) were perceived as the most effective nudging strategy, followed by inactivity nudges (question 12), and then fixed nudges (question 11).

The question assessing whether nudges were overlooked (question 15) received a

mean rating of 2.71, indicating that a moderate proportion of nudges were missed, which is confirmed by the qualitative interview data.

The perceived usability may influence journaling compliance indirectly, as a higher usability potentially increases user engagement. Differences in perceived user experience may also reflect how disruptive the nudging strategies are experienced. Even if usability scores are similar across conditions, differences in perceived burden may still affect long-term user engagement.

5.4.3. Qualitative insights

Method

In addition to the quantitative findings, qualitative insights were collected through semi-structured interviews after participants completed all nudging conditions (Appendix C.5). Interview data were analysed using thematic analysis. Transcripts were reviewed and categorised into five broader themes: interface usability, data interpretability, perceived accuracy, daily routine and nudging preferences, and clinical relevance. The feedback has been incorporated into the design requirements for the final design (Chapter 6).

The interviews consisted of a predefined set of open-ended questions covering key themes related to the usability of the journal application (Appendix C.4). The interviews were conducted in a controlled setting, were systematically documented, and took approximately 30-45 minutes to complete. This approach provided consistency across participants while remaining flexible to explore participants' experiences in greater depth.

Results

Interface

Overall, participants reported that the application was easy to use, with low perceived complexity. Journaling labels was perceived as straightforward, and the interface was considered accessible.

The hierarchy of the interface, navigation bar, and tab menu's were perceived as intuitive and easy to understand. However, minor concerns included low interpretability of several feature graphs, missing additional elaboration, and inconsistent behaviour of one of the time selection tool to determine a scenario's timestamps.

Interpretability

The semi-structured interview showed that participants indicated that the graphs on the analysis page did provide insights to a certain level. Participants were able to identify patterns in graphs. However, the absence of a deeper elaboration made it difficult for participants to understand the significance of observed patterns or peaks in their data, which decreased the interpretability. Also, without prior knowledge, some features were not intuitively understood (e.g. HR derivative, SpO2)

When a peak occurred in a graph, participants wondered if that was 'normal' or not.

The lack of a reference baseline or comparison to a users 'normal' physiological condition further limited the interpretability.

Perceived accuracy
The participants estimated that the accuracy of the timestamps of their journaled labels was approximately 95 percent, suggesting that the application supports reliable self-reporting, although this could not be objectively validated against ground truth data. Participants indicated that journaling afterwards could improve accuracy, as it allows more time to reconstruct scenarios.

The perceived accuracy decreased when journaling after a longer period (+4 hours) after occurrence, where planned activities were easier to recall than spontaneous events.

Daily routine and nudging preferences
Participants justified their preference for flexible personalised notifications that suit their daily routines rather than relying on a single predetermined strategy. Fixed nudges were sometimes perceived as disruptive when poorly timed. Inactivity nudges were considered less intrusive.

Daily routines strongly influenced journaling behaviour. The most suitable moments for journaling were during breaks, transportation or in the evening after work. Difficulties journaling were mostly encountered in the morning due to busy schedules or during planned activities.

During the 'personalised' condition, participants often chose a hybrid approach of fixed and inactivity nudges.

For some participants, a substantial proportion of nudges was overlooked, often due to disabled notifications or work-related scenarios.

Clinical relevance
The potential clinical value of the application was recognised by the participants, particularly in providing an integrated overview of physiological and behavioural data.

However, the perceived clinical value is highly dependent on interpretability. The usefulness of the data is limited by the ability to interpret it meaningfully.

Two of the participants diagnosed with epilepsy highlighted limitations in medication tracking. They noted that the dose of medication can fluctuate frequently and could be more accurately tracked in milligrams. Medication types and half-life of the medications may also vary, and the type of medication can vary, which could be visualised more realistically in the medication effect graph when integrated.

Summary
The qualitative findings provide context for the quantitative results, particularly in explaining why certain nudging strategies did not produce significant differences. Even

though the findings indicate that the usability is strong, further elaboration of the visualised data in the application could improve the interpretability by explaining the relevance related to epilepsy. The interviews made clear that the interpretability directly impacts user engagement. When the significance of data was not understood, the perceived value of that data decreased.

Overall, the findings suggest that self-reported data is generally reliable, while timeliness and context significantly influence accuracy, highlighting the importance of an effective nudging strategy.

The qualitative findings highlight key areas for improvement of the final design, particularly in terms of interpretability and clinical value, while confirming the application's strength in usability and ease of use. The results indicate that the alignment of the nudging strategy with individual routines and schedules is more important than the nudging strategy itself (fixed, inactivity), supporting the observation that nudging effectiveness is highly individualised and context-dependent.

5.5. Discussion and Limitations

5.5.1. Discussion

This research examined whether different nudging strategies influence journaling compliance and user experience. The results did not provide sufficient evidence to reject the null hypothesis (H0) for any of the sub-hypotheses ($p > 0{,}05$). The absence of statistically significant differences between the nudging strategies suggests that either the true effect size is small or that the variability in user behaviour outweighs the effect of the nudges. While the limited sample size reduces statistical power, it does not fully explain the results. The nine-day within-subject design increased sensitivity to within-behavioural changes. Therefore, alternative explanations related to behavioural variability must be considered.

Nudging effectiveness is likely to be highly individual, depending on daily routines, personal preferences, and responsiveness to nudges. Averaged results may have masked individual-level effects. Additionally, behavioural adaptation over time may have influenced the results. The exposure to different nudges for a period of nine days may have led to habituation or notification fatigue, possibly resulting in decreased responsiveness, which reduces observable effects. Also, notification delivery does not guarantee that it is also being perceived, meaning delays in latency may reflect overlooked nudges rather than intentional postponement. Another possible explanation is the presence of a ceiling effect, where if participants demonstrated a relatively high journaling compliance across all conditions, differences between conditions become difficult to detect.

Despite the lack of statistical significance, small observed trends may still be practically relevant. Even modest improvements in engagement can accumulate over time and contribute to higher-quality datasets. Therefore, the interpretation of the results should not rely only on significance testing, but also consider the consistency and di-

rection of the observed trends.

Furthermore, the compliance metrics primarily capture the quantity and timing of journaling behaviour, but do not account for the relevance of the journaled data. While increased frequency and temporal coverage enhance dataset completeness, they do not guarantee that the collected data is usable for clinical interpretation or machine learning purposes.

Although the statistical tests indicate no significant differences, the user experience results from the SUS questionnaire and semi-structured interview indicate that the perceived effectiveness of the personalised nudges surpassed that of both the fixed and inactivity nudges. The qualitative insights also showed that participants prefer personalised nudges adapted to their personal agenda. A likely explanation is that personalised nudges reduce disruption and cognitive burden by occurring at favourable moments, increasing the probability of any journal activities. The fixed and inactivity nudges may suffer from a situational mismatch, resulting in decreased journal compliance.

Relevancy for clinicians and future research

The findings align with the literature research in Chapter 3.4, which emphasises the importance of context-aware and personalised nudges based on patients' personal routine, which results in faster responses, decreased recall bias and decreased perceived burden. The study extends prior research by showing that personalised nudges may improve journal compliance, while highlighting the importance of individual variability.

Stronger evidence requires future research for a longer period with a larger sample size to validate these findings and to further explore types and effectiveness of personalised nudge strategies.

The results suggest that adaptive, user-driven nudge strategies and data interpretability are essential for the application's final design. By minimising the latency, recall bias is less likely to happen, which improves the accuracy of the logged labels' timestamps, and thus the accuracy of the data set to train a future ML model. Enhancing the total amount of labels and covered minutes with labels increases the size of the data set, which is crucial to train an accurate ML model. The perceived effectiveness of the personalised nudges indicates that users tend to engage more in active journaling, providing a broader, more complete contextual data set for clinicians, which could help during a patient's treatment to identify relevant triggers and high-risk periods for epilepsy patients.

5.5.2. Limitations

Methodological limitations

The absence of significant findings related to the different nudging strategies can be attributed to methodological constraints.

The participant requirements and a single available Apple Watch made it difficult to recruit participants for the study, resulting in a small sample size (N = 7). The sample size limited the statistical power, increased the likelihood of type 2 errors, and made it difficult to detect any significant effects. The short duration (three days per condition) constrains the analysis to short-term behaviour, excluding long-term adaptation or habit formation to nudging strategies. Variability in nudge perception and experiences introduces internal validity concerns, as it remains uncertain if the participants were equally exposed to the nudges.

The custom SUS questionnaire used to evaluate the user experience extended the standard validated 10-question SUS tool without formal validation of the additional questions.

A standard significance level at $\alpha = 0{,}05$ and a repeated-measures ANOVA test were applied to mitigate type 1 errors of false positives. However, type 2 errors (false negatives) have been acknowledged as a limitation due to the observed variance in the data and limited sample size. Results from the quantitative data showed outliers, suggesting that the journal compliance is highly individual-dependent and sensitive to real-life contextual variability, which may cause inconsistent user engagement throughout the day.

The journal compliance metrics do not capture the quality or accuracy of journaled data, limiting conclusions about data usefulness.

Data and measurement limitations

The study relied on self-reported data and did not include the physiological or subjective contextual data, limiting the depth of the analysis and the ability to assess the real-world effectiveness of generating personalised data for ML purposes.

Measurement bias is present, as self-reports tend to be subject to recall bias and social desirability bias. The accuracy of the self-reported timestamps could not be validated in the uncontrolled setting of the study, where the actual events are not being monitored and compared to the timestamps of the self-reported labels. Furthermore, subjective workload and usability ratings may have been influenced by recent events the participants encountered, introducing response bias.

Averaging data across days for the repeated-measures ANOVA test may also affect temporal variability and sensitivity to specific conditions.

Validity and reliability

Several sources of systematic and random errors may have influenced the results. The systematic errors include notification perception variability, as participants retained control over their phone's notification settings, meaning not all nudges were guaranteed to be perceived. Additionally, device usage bias may have influenced interaction behaviour and generalisability, as interactions differ between iPhone and Apple Watch usage, which may cause nudges to be perceived differently. The short testing duration per condition further limits the reliability of long-term behavioural ob-

servations.

The study was conducted in real-life conditions, which increases external validity but introduces uncontrolled contextual factors such as varying schedules and environments.

Random errors were introduced through daily routine variability between participants and unpredictable contextual interruptions, which may have influenced participants' journal behaviour independently of the nudging strategy.

Internal validity is strengthened through the within-subject design, the counterbalanced condition order, and automated timestamp registration. However, the influence of uncontrolled real-life situations cannot be eliminated.

5.6. Conclusion and Recommendations

5.6.1. Conclusion

The research study focused on answering the following research question:

- *Which nudging strategy leads to the highest journaling compliance while preserving positive user experience and low perceived burden?*

The results of this research study indicate that no single nudging strategy significantly outperforms others in improving journaling compliance under the tested conditions. However, observed trends and variability between participants suggest that nudging effectiveness is highly context-dependent and influenced by individual user behaviour. Personalised nudging shows potential as a direction for future development, because it allows alignment with individual routines and preferences. The findings highlight that improving the journaling compliance is not only a matter of selecting a single strategy, but rather of designing adaptive strategies suitable for user-specific contexts.

While nudging strategy alone is not a determining factor, personalisation appears to enhance journaling compliance and minimise perceived burden in practice.

5.6.2. Recommendations

Future research

For future research, it is recommended to use a larger sample size and use exclusively epilepsy patients as participants to improve generalisability, and to increase the duration period of the study to capture long-term behavioural effects.

To actually investigate the relevance and predictability of the journaled labels, the physiological data from participants' wearables and their subjective labels should be incorporated in future research to enable deeper analysis and assess data usefulness. This would enable the evaluation of the quality of the data (accuracy) and clinical relevance of the application. Unfortunately, this requires the collection of privacy-sensitive and medical data, which was left out of the scope of this study.

Design recommendations

The most important recommendation is to implement an adaptive, personalised nudging strategy in the final design, as this nudging strategy showed the most promising results. This should include multiple options to enable a hybrid strategy combining different types of (context-aware) nudges, tailored to an individual's preferences.

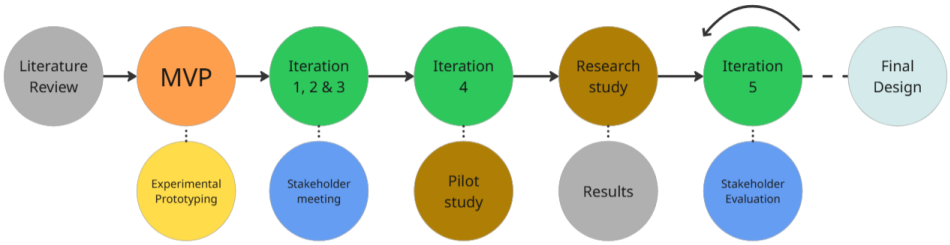
Secondly, improving the data interpretability by adding personal baselines and contextual explanations enables users to derive meaningful insights more easily from their data.

6

Final Concept

6.1. Iteration 5

Based on the results from the quantitative and qualitative insights from the research study in Chapter 5.4.3, a list of improvements has been drawn up and implemented. A final stakeholder meeting has been arranged to evaluate the clinical value (Appendix A.10).



Improvements

1. Personalised nudges outperformed fixed or inactivity nudging strategies in terms of perceived effectiveness. Accordingly, more personalised nudging strategies have been added via integration of the user's Apple Calendar app for JIT-nudges, enabling context-aware nudges, addressing the situational mismatch participants identified as a primary limitation of the fixed and inactivity nudging strategies. Also, a new type of warning has been added, which occurs when a feature deviates from the user's baseline (Figure 6.4).
2. Increased graph interpretability, by adding an orange graph representing the user's baseline and a shaded area visualising its standard deviation. Insufficient graph interpretability occurred multiple times during the semi-structured interviews. These improvements enable users to visually distinguish their own 'normal' physiological range from irregular states (Figure 6.6).
3. Information icons at the top of each page have been added to provide additional elaboration of all functionalities of each page and their relevance related to epilepsy, improving interpretability and minimising cognitive load.
4. Custom categories have been added to accommodate individual variability, improving both completeness and personalisation of the journaled data (Figure

6.2). Also, medication tracking has been extended to support medication type, dose, and half-life, aligning better with an individual user. The medication details are integrated into the medication level graph for a more realistic indication (Figure 6.3).

5. Including canonical correlation feature from literature research (Chapter 3.3). These correlations have been integrated into the combined feature graph, where different pairs of features can be selected (Figure 6.6).
6. A 'Trigger dashboard' section. This section includes a mock-up risk algorithm comparing the user's physiological data from the last 30 minutes and continuously compares it with previously logged seizures and pre-ictal periods to provide a risk score. This mock-up algorithm has not been validated and is constrained by the Medical Device Regulation and is not approved for clinical purposes. When a user logs a seizure, it is registered in a bar chart for an overview, including physiological data from each seizure and multiple types of correlations (Figure 6.6).

Journaling a subjective label

Figure 6.1: Steps required to journal an event.

Figure 6.1 demonstrates the required steps to journal an event. The number of steps is minimised by 4 in accordance with the list of requirements in Chapter 4: selecting a label, determining timestamps, and confirming.

Logger page

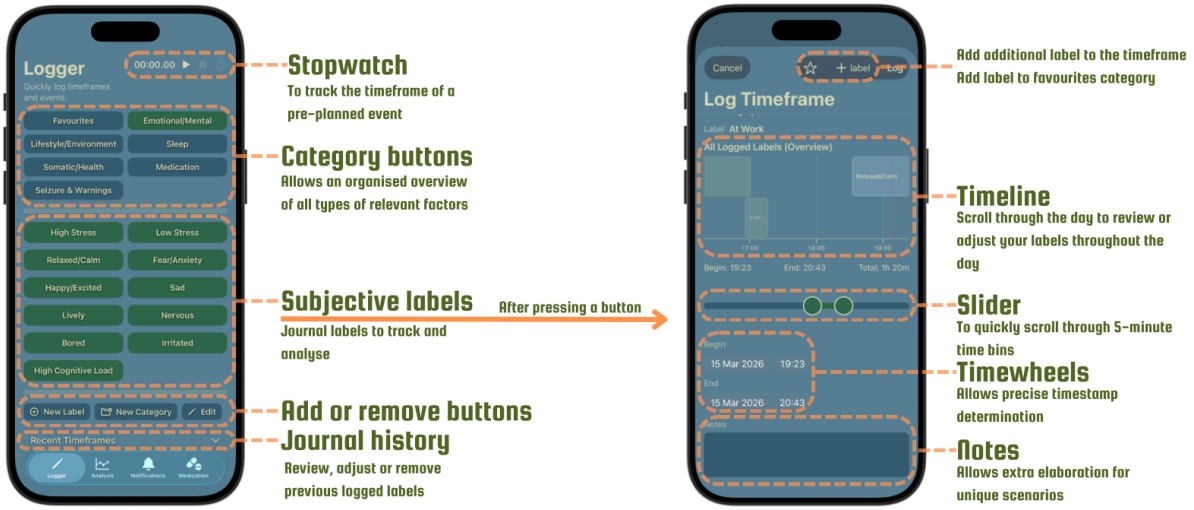


Figure 6.2: The logger page and timeframe determination window.

Figure 6.2 shows the logger page with categorised label buttons and the timeframe determination window, supporting multiple input modes: snapshots, planned events using a stopwatch, or afterwards logging. This multi-modal approach accommodates a range of journaling scenarios from urgent seizure-related logging to afterwards journaling in the evening.

Medication page

Figure 6.3: The medication page.

Figure 6.3 shows the medication page, which supports multiple medication types entered by name, dose tracked in milligrams, and a user-configurable half-life value that determines the medication effect graph. This enables the analysis of the correlation between medication timing, dosage fluctuations, and corresponding changes in physiological or subjective data.

Notification page

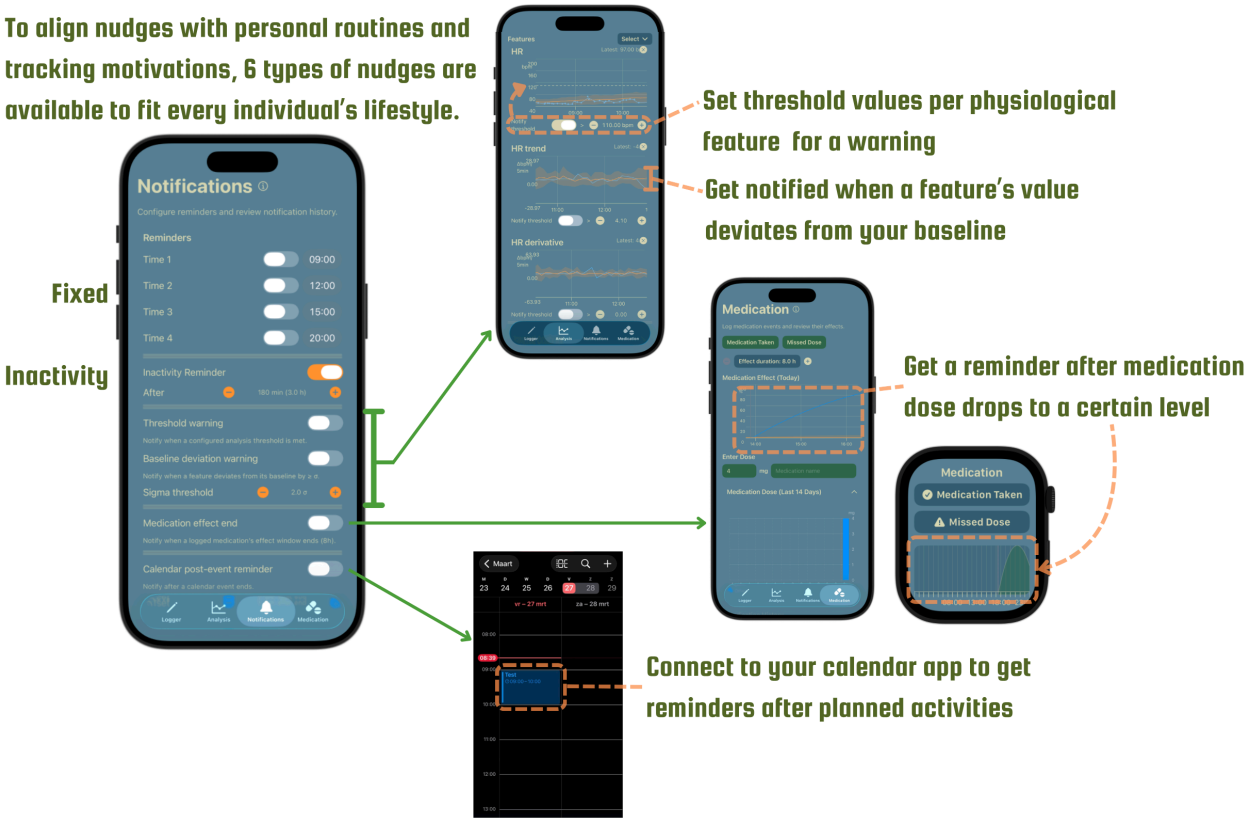


Figure 6.4: The notification page.

Figure 6.4 shows the notification page, allowing a hybrid nudging strategy including fixed, inactivity, calendar-connected, threshold, and baseline deviation notifications.

Suggested interruption during sleep

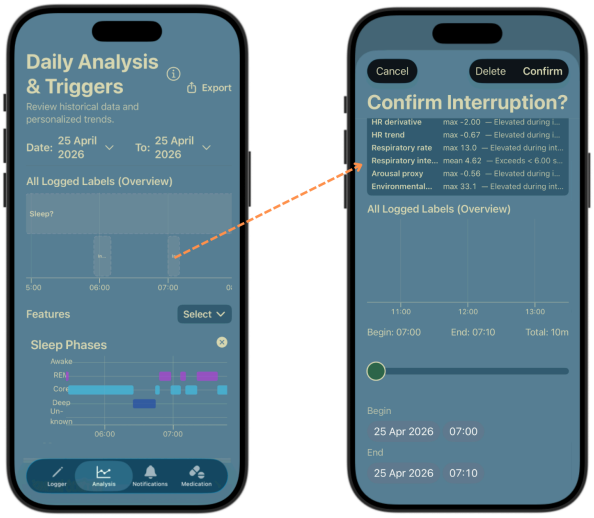


Figure 6.5: Interruption labels can be pressed to see the details.

Figure 6.5 shows a suggested 'sleep?' and 'interruption' label. When an irregular physiological condition occurs during sleep, users can trace back which physiological features were related to this behaviour and analyse their values. This enables users to trace back whether this physiological condition was related to a seizure.

Analysis page

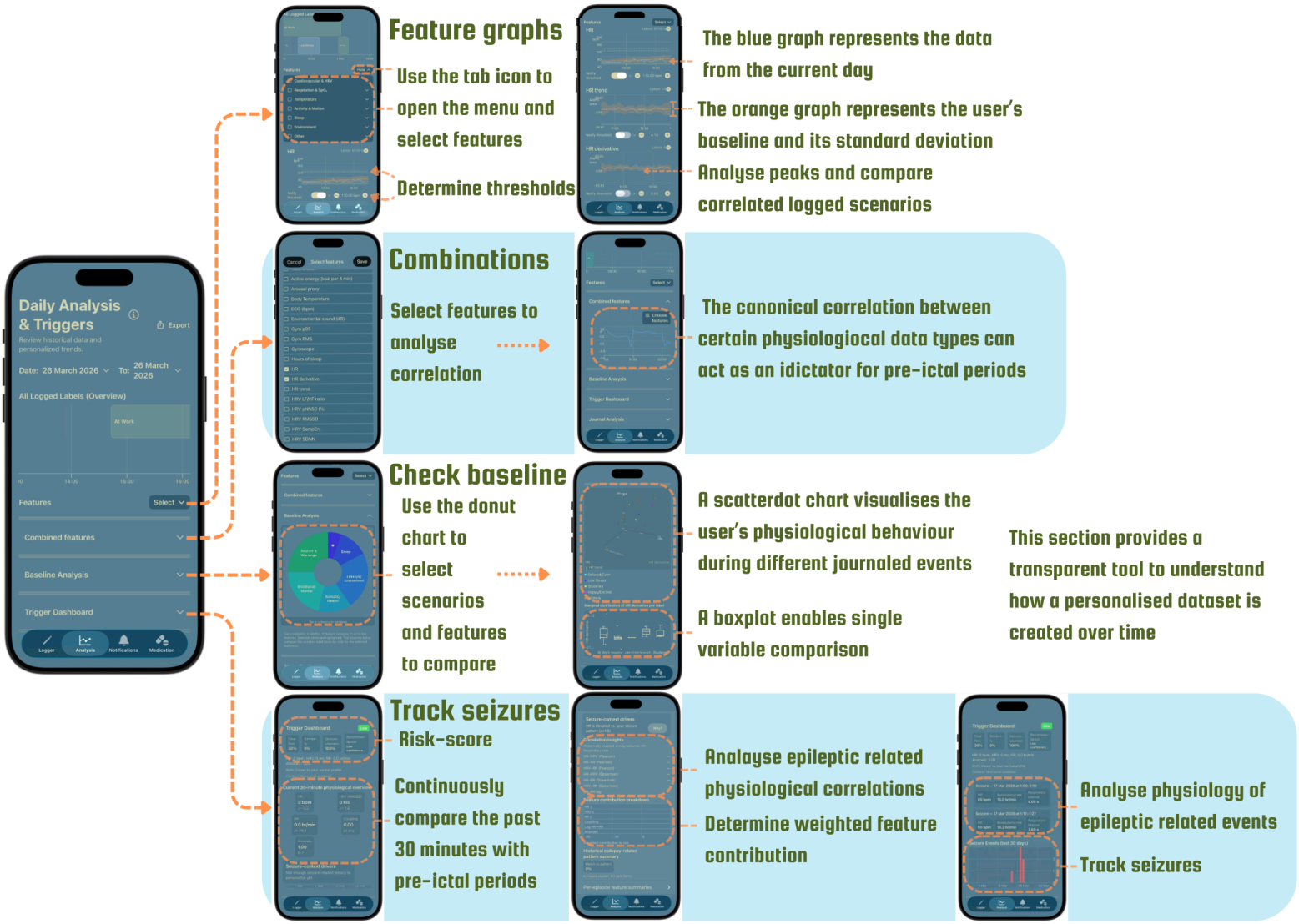


Figure 6.6: The analysis page, its subsections, and all functionalities.

Figure 6.6 presents the analysis page and its subsections. This page is structured to provide both patients and clinicians with clinically relevant data during treatment. Feature graphs of physiological data types, and correlations between types of data are visualised in graphs, including their baseline (deviation). Logged events can be compared across multiple physiological features. Users can keep track of their seizures and analyse physiology related to epileptic scenarios.

Watch interface

Figure 6.7: The interface of the Apple Watch application.

Figure 6.7 visualises the Apple Watch interface, containing three pages: a medication page to log and track medication usage and dose, a logger page to journal events using a doughnut chart, and a page providing an overview of all journaled events throughout the day on a timeline.

7

Final Design: The Epilepsy Journal

The final design (Figure 7.1) represents the result of five iterative design cycles, guided by literature research, stakeholder meetings, and findings of the research study. The design addresses the identified gaps from the study in Chapter 5 and challenges identified in Chapter 2, regarding low interpretability, data incompleteness, limited personalisation, nudging effectiveness, and clinical relevancy.

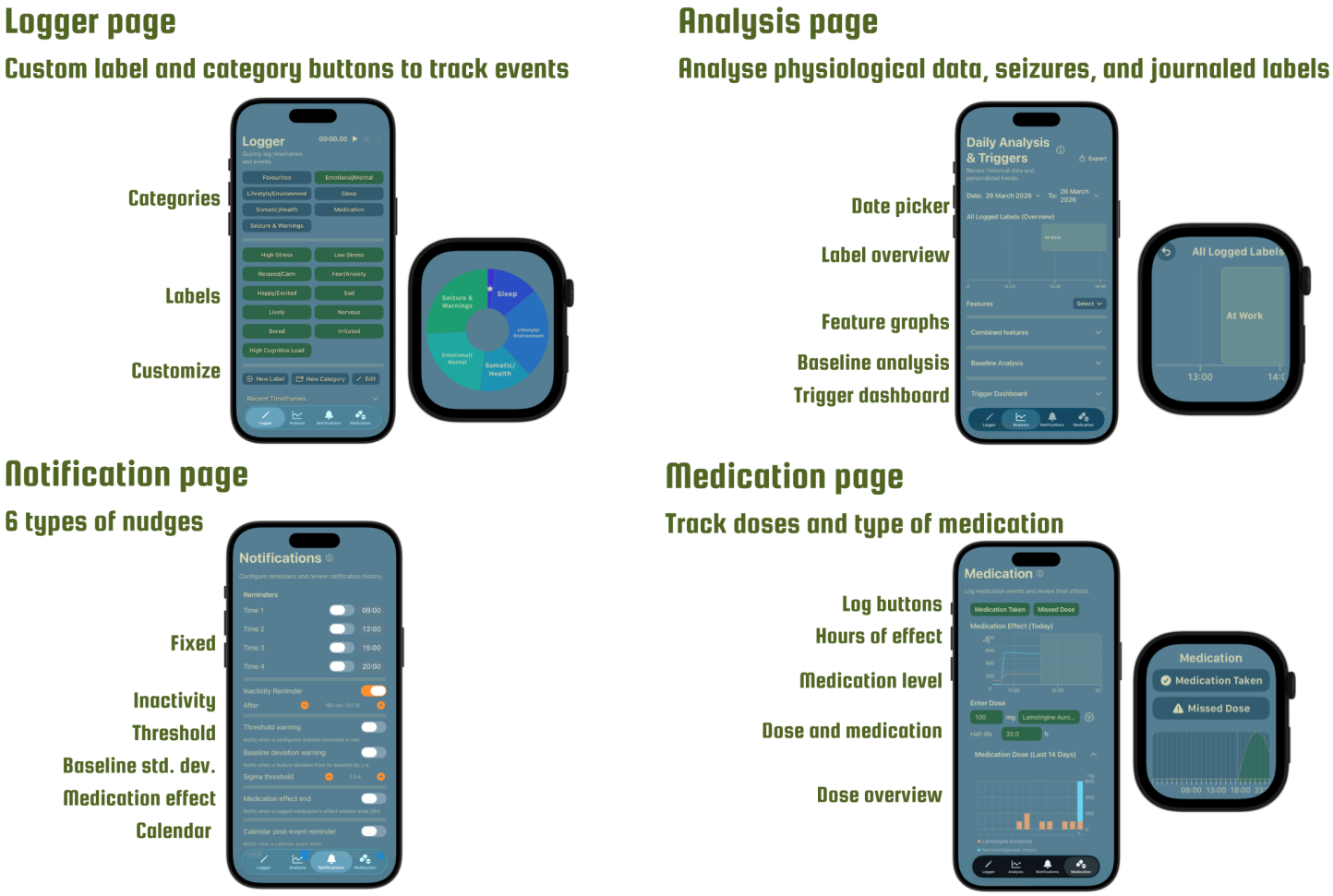


Figure 7.1: The total interface of the final design.

8

Reflection and Conclusion

8.1. Reflection

The goal of this project was to develop a functional prototype of a wearable-connected journaling application, capable of creating a personalised dataset to address three challenges: the incompleteness, low interpretability and limited personalisation of data used to improve the treatment of epilepsy patients.

Regarding the data incompleteness, the application provides an integrated system for real-time, timestamped alignment of subjective contextual labels with continuously measured physiological wearable data in a remote context. Tracking scenarios can be adjusted to suit an individual's relevant events or triggers related to their epileptic diagnosis throughout the day. Manual backward digitisation of subjective data is subject to recall bias and timestamp inaccuracy, which the application addresses by supporting multiple journaling methods (i.e., snapshot, stopwatch, post-event registration), automated timestamp registration, and a nudging strategy to minimise the response latency.

Low interpretability of raw physiological data was confirmed by the research study findings. Participants reported limited interpretability due to a lack of elaboration on the clinical relevance of certain features. This has been addressed by implementing information icons on each page and a baseline and shaded standard deviation graph for each physiological feature to distinguish normal physiological conditions from irregular states. However, these improvements and their impact on clinical interpretability have not been empirically evaluated and should be assessed to determine whether they enhance understanding in personal epilepsy treatment.

The application provides personalisation through custom labels and categories, adjustable medication tracking, and individual baselines to adapt to a patient's specific triggers and context, addressing the limitation of generalised treatment, which fails to capture individual variability. The hybrid nudging strategy allows a combination of user-determined fixed, inactivity, and calendar-integrated notifications and warnings for physiological thresholds or baseline deviations. This strategy allows users to align notifications with their individual routines, reducing situational mismatch, latency, and

perceived burden, enhancing the journaling compliance and data quality. The mock-up risk algorithm provides an indication of individual physiological-epileptic relationships, but does not provide a validated personalised model, which requires long-term data from individual patients across multiple seizure events.

8.1.1. Limitations

The application currently exclusively operates within an Apple ecosystem, requiring an iPhone with iOS 18 or later, and an Apple Watch with WatchOS 8 or later, which limits the accessibility for patients who do not own or have access to these compatible devices. The collected physiological data does not include electrodermal activity (EDA), which is a biomarker for the detection of pre-ictal states.

The absence of physiological data collection during the study means that the data pipeline has been validated structurally, but not yet with real epileptic events. Also, the length of the research study was insufficient to train or validate a machine learning model using epileptic-related data. Additionally, the clinical applicability of any output of the application remains constrained, as any generated output must be classified under the MDR (Medical Device Regulation) to avoid unauthorised medical recommendations. The design requires formal regulatory approval before clinical deployment may be initiated.

8.2. Conclusion

In this thesis, the development of a mobile application is described that captures continuously measured physiological wearable data and aligns it with subjective contextual scenarios, to generate a labelled dataset suitable for future personalised machine learning applications in epilepsy treatment. Three core challenges that have been addressed to generate a personalised dataset are data incompleteness, low interpretability, and limited personalisation of a patient's data. This has been achieved through an iterative design process of 5 iterations, guided by literature research, repeated stakeholder meetings with clinicians, epilepsy patients, and researchers, and a within-subjects research study of nine days.

The research study included a within-subjects comparison of three nudging strategies (fixed, inactivity, and personalised) across seven participants over nine days. While the study did not produce statistically significant differences between the three nudging strategies (all $p > 0{,}05$), the analysis indicates descriptive trends that suggest personalised nudges may improve journaling compliance compared to fixed and inactivity strategies. Qualitative insights provided support for this conclusion, with all participants preferring a personalised strategy aligning with daily routines. These findings extend previous literature on context-aware nudging as a data quality mechanism. Minimising recall bias reduces response latency and improves the temporal accuracy, which benefits the quality of a dataset suitable for machine learning purposes.

The final design addresses the identified gaps from the research study in Chapter 5 and challenges identified in Chapter 2, providing patients and neurologists with

an integrated overview of subjective events and corresponding physiological conditions, supporting identification of individual epileptic triggers and enabling more personalised treatment and monitoring. Integrating medication timing, seizure events, and physiological changes into an accessible and interpretable interface to support clinical relevance and increase understanding of a patient's diagnosis. The application implements a data pipeline from real-time wearable measurements through timestamped journal labelling to an Excel or CSV file with timestamp precision suitable for 30-second window segmentation, to create a personalised dataset suitable for machine learning purposes. Integration of a personalised baseline, canonical correlation features, and a risk indicator further demonstrates the feasibility of integrating predictive functionalities within a medical application.

8.2.1. Recommendations

For future research, the most important recommendation is to conduct a longitudinal study with a larger sample size of epilepsy patients with an active diagnosis, conducted over a minimum of six to twelve months to be able to evaluate long-term user engagement, journal compliance, training of a personalised machine learning model on a clinically meaningful dataset, and an assessment of the predictability, value, and significance of specific labels for physiological changes.

8.2.2. Next steps

SEIN and EpilepsieNL have announced that they will conduct a follow-up study in 2027, in which the app will be tested for one year with 50 epilepsy patients, addressing the recommendations precisely and building directly on the application developed in this thesis. The research is set to start in January 2027 and focuses on identifying relevant timeframes and biomarkers to define the detection of high-risk periods for seizures (Figure 8.1).

With the developed application of this thesis, synchronisation of physiological and subjective data is possible. The next step would be to identify relevant label timeframes related to high-risk periods of seizures, and what kind of effect the subjective logged labels have on the physiological changes. Also, for a more realistic user acceptability evaluation, for usage during a long-term period, user adherence and satisfaction scores will be used.

Further development of the application will continue in collaboration with EpilepsieNL and SEIN, guided by the findings from their research study.

EpilepsieNL

Aims.

- To develop a user-friendly seizure tracking app
- To identify relevant timeframes and biosignals to define detection of high-risk periods for seizures

Methods.

We developed an iOS seizure-tracking app that integrates diary loggings with Apple Watch health data. This app will also record behavioural changes to determine the contributing factors at the granularity of individual seizures. We will recruit people with active epilepsy (at least one seizure per year) who use an Apple Watch and iPhone. Recruitment will be promoted by EpilepsieNL through their social channels. All participants will be asked to provide consent for extraction of the key epilepsy features from their medical files. Participants will be queried daily regarding their seizure frequency over the preceding 24 hours and to document potential precipitants (e.g., stress or poor sleep) or other relevant changes (e.g., medication adjustments) in their seizure journal. These entries are categorized into several domains—emotional/mood, lifestyle/environment, sleep, medication, and somatic/health—each comprising multiple prespecified labels (see Figure 1). Domains and labels were defined based on multiple focus groups. The user can indicate the time-period for each label (e.g., duration of the seizure or the stress period). Apple Watch health data (eg activity, heart rate, HRV (SDNN method), respiratory rate, sleep estimates, and menstrual cycle tracking) are streamed every three minutes to the diary logging app. We aim to record for up to 1 year. We aim to include at least 50 users. We will examine acceptability using app adherence and user satisfaction scores (including responses for all scales and the individual items). We will examine the proportion of subjects reporting triggers, the number of triggers, and consistency of triggers across seizures. As a second step we will select the most common precipitant and apply linear mixed models to examine whether their presence alters the physiological output. To allow these comparisons, the time periods will be aligned with the same clock times on days without precipitants or seizures. We will apply similar models to examine the association between physiological data and self-reported seizures while accounting for

Figure 8.1: The follow-up research study set to begin in 2027

References

[1] World Health Organization. *Epilepsy*. 2024.

[2] C. E. Elger and C. Hoppe. “Diagnostic challenges in epilepsy: Seizure under-reporting and seizure detection”. In: *The Lancet Neurology* 17.3 (2018), pp. 279–288. DOI: 10.1016/S1474-4422(18)30038-3.

[3] S. Beniczky et al. “Machine learning and wearable devices of the future”. In: *Epilepsia* (2020). DOI: 10.1111/epi.16555.

[4] M. Eberlein et al. “Evaluation of machine learning methods for seizure prediction in epilepsy”. In: *Current Directions in Biomedical Engineering* 5.1 (2019), pp. 109–112. DOI: 10.1515/cdbme-2019-0028.

[5] D. Zambrana-Vinaroz et al. “Wearable epileptic seizure prediction system based on machine learning techniques using ECG, PPG, and EEG signals”. In: *Sensors* 22.23 (2022), p. 9372. DOI: 10.3390/s22239372.

[6] Empatica S.R.L. *Empatica E4 Revision 001*. 2014.

[7] G. Erba et al. “Patients’ and caregivers’ contributions for differentiating epileptic from psychogenic nonepileptic seizures: Value and limitations of self-reporting questionnaires: a pilot study”. In: *Seizure: European Journal of Epilepsy* 53 (2017), pp. 66–71. DOI: 10.1016/j.seizure.2017.11.001.

[8] S. R. Haut et al. “Modeling seizure self-prediction: An e-diary study”. In: *Epilepsia* 54.11 (2013), pp. 1960–1967. DOI: 10.1111/epi.12355.

[9] R.S. Fisher et al. “A practical clinical definition of epilepsy”. In: *Epilepsia* 55.4 (2014), pp. 475–482. DOI: 10.1111/epi.12550.

[10] S. R. Haut et al. “Can patients with epilepsy predict their seizures?” In: *Neurology* 68.4 (2007), pp. 262–266.

[11] H. Egenasi et al. “Current norms and practices in using a seizure diary for managing epilepsy: A scoping review”. In: *PMC* ().

[12] B. H. Brinkmann et al. “Seizure diaries and forecasting with wearables: Epilepsy monitoring outside the clinic”. In: *Frontiers in Neurology* 12 (2021), p. 697920.

[13] C. Chang et al. “Necessary for seizure forecasting outcome metrics: Seizure frequency and benchmark model”. In: *Epilepsy Research* 208 (2024), p. 107474. DOI: 10.1016/j.eplepsyres.2024.107474.

[14] E. Bruno et al. “Wearable devices for seizure detection: Practical experiences and recommendations from the Wearables for Epilepsy And Research (WEAR) International Study Group”. In: *Epilepsia* 62.10 (2021), pp. 2307–2321. DOI: 10.1111/epi.17044.

[15] M. O. Baud and V. R. Rao. “Gauging seizure risk”. In: *Neurology* 91.21 (2018), pp. 967–973.

[16] B. Letham et al. “Interpretability With Accurate Small Models”. In: *PMC - NIH* (2020).

[17] O. Wyrich et al. “Recommendations for reporting machine learning analyses in clinical research”. In: *PMC* (2021).

[18] S. B. Dumanis et al. “Seizure forecasting from idea to reality: Outcomes of the My Seizure Gauge Epilepsy Innovation Institute Workshop”. In: *eNeuro* 4.6 (2017), ENEURO.0349–17.2017. DOI: 10.1523/eneuro.0349-17.2017.

[19] R. S. Fisher et al. “Use of an online epilepsy diary to characterize repetitive patterns”. In: *ScienceDirect* (2015).

[20] A. Pandher et al. “Smartphone applications for seizure management”. In: *SAGE Journals* (2016).

[21] N. Zabler et al. “High precision in epileptic seizure self-reporting with an app diary”. In: *Scientific Reports* 14.1 (2024), p. 15832. DOI: 10.1038/s41598-024-66932-y.

[22] W. Xiong et al. “Forecasting seizure likelihood from cycles of self-reported events + wearables”. In: *ScienceDirect* (2023).

[23] J. Harms et al. “Navigation in long forms on smartphones: scrolling worse than tabs, menus, and collapsible fieldsets”. In: *Human-Computer Interaction – INTERACT (2015)*. Ed. by Julio Abascal et al. Cham: Springer, 2015, pp. 333–340.

[24] D. Zhang and B. Adipat. “Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications”. In: *International Journal of Human-Computer Interaction* 18.3 (2005), pp. 293–308.

[25] R. S. Fisher et al. “Seizure diaries for clinical research and practice: Limitations and future prospects”. In: *Epilepsy & Behavior* (2012).

[26] R. S. Fisher. “Bad information in epilepsy care”. In: *Epilepsy & Behavior* 67 (2017), pp. 133–134. DOI: 10.1016/j.yebeh.2016.10.022.

[27] L. Kuhlmann et al. “Seizure prediction - ready for a new era”. In: *Nature Reviews Neurology* 14.10 (2018), pp. 618–630. DOI: 10.1038/s41582-018-0055-2.

[28] E. Gleichgerrcht et al. “Machine learning on patient seizure diaries to forecast events”. In: *Epilepsia* (2022).

[29] A. Schulze-Bonhage and S. R. Haut. “Premonitory features and seizure self-prediction: Artifact or real?” In: *Epilepsy Research* 97.1-2 (2011), pp. 231–235. DOI: 10.1016/j.eplepsyres.2011.09.026.

[30] S. Jaddu et al. “Pre-Ictal seizure prediction using personalized deep learning”. In: *arXiv:2410.05491v1 [cs.LG]* (2024).

[31] T. Yamakawa et al. “Wearable Epileptic Seizure Prediction System with Machine-Learning-Based Anomaly Detection of Heart Rate Variability”. In: *Sensors* 20.14 (2020), p. 3987. DOI: 10.3390/s20143987.

[32] M. J. Cook et al. “Forecasting seizures with wrist-worn biosensors and deep learning”. In: *Epilepsia* (2021).

[33] O. A. Toheeb et al. “The impact of data quality on machine learning interpretability”. In: (2025).

[34] C. L. Stewart et al. “RADAR-base: Major depressive disorder and epilepsy case studies”. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 2018, pp. 1735–1743. DOI: 10.1145/3267305.3267540.

[35] S. Quilter et al. “Co-designing the ATMOSPHERE app for seizure precursor logging”. In: *JMIR Research Protocols* (2024).

[36] C. Rossi et al. “Training datasets for epilepsy analysis: Preprocessing and feature extraction...” In: *MDPI* (2024).

[37] S. A. Janse et al. “Patient and caregiver preferences for the potential benefits and risks of a seizure forecasting device: A best–worst scaling”. In: *Epilepsy & Behavior* 96 (2019), pp. 183–191. DOI: 10.1016/j.yebeh.2019.04.018.

[38] M. N. Shouse, A. M. da Silva, and M. Sammaritano. “Circadian rhythm, sleep, and epilepsy”. In: *Journal of Clinical Neurophysiology* 13.1 (1996), pp. 32–50. DOI: 10.1097/00004691-199601000-00004.

[39] L. Pick et al. “Investigating psychobiological causes …EMA protocol”. In: *PLOS* (2024).

[40] W. A. Hofstra. “The circadian rhythm and its interaction with human epilepsy: A review of literature”. In: *Sleep Medicine Reviews* 13.6 (2009), pp. 413–420. DOI: 10.1016/j.smrv.2009.01.002.

[41] A. Ge et al. “Seizure triggers identified postictally using a smart watch reporting system”. In: *Epilepsy & Behavior* 126 (2021), p. 108472. DOI: 10.1016/j.yebeh.2021.108472.

[42] S. R. Haut et al. “Clinical features of the pre-ictal state: Mood changes and premonitory symptoms”. In: *Epilepsy & Behavior* 23.4 (2012), pp. 415–421. DOI: 10.1016/j.yebeh.2012.02.007.

[43] S. R. Haut et al. “Seizure occurrence: Premonitory symptoms and prediction”. In: *Epilepsia* (2012).

[44] A.G. Herzog. “Catamenial epilepsy: Update on prevalence, pathophysiology and treatment from the findings of the NIH Progesterone Treatment Trial”. In: *Seizure* 28 (2015), pp. 18–25. DOI: 10.1016/j.seizure.2015.02.024. URL: https://doi.org/10.1016/j.seizure.2015.02.024.

[45] C. Sevcencu and J. J. Struijk. “Autonomic alterations and cardiac changes in epilepsy”. In: *Epilepsia* 51.5 (2010), pp. 725–737. DOI: 10.1111/j.1528-1167.2009.02479.x.

[46] S. R. Haut, M. Vouyiouklis, and S. Shinnar. “Stress and epilepsy: A patient perception survey”. In: *Epilepsy & Behavior* 4.5 (2003), pp. 511–514. DOI: 10.1016/S1525-5050(03)00182-3.

[47] J. S. Van Campen et al. “Cortisol fluctuations relate to interictal epileptiform discharges in stress sensitive epilepsy”. In: *Brain* 139.6 (2016), pp. 1673–1679. DOI: 10.1093/brain/aww071.

[48] P. E. Schauwecker. “The effects of glycemic control on seizures and seizure-induced excitotoxic cell death”. In: *BMC Neuroscience* 13.1 (2012), p. 94. DOI: 10.1186/1471-2202-13-94.

[49] G. M. Harari et al. “Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges”. In: *Perspectives on Psychological Science* 11.6 (2016), pp. 838–854. DOI: 10.1177/1745691616650285.

[50] K. A. Mikk, H. A. Sleeper, and E. J. Topol. “The pathway to patient data ownership and better health”. In: *JAMA* 318.15 (2017), pp. 1433–1434. DOI: 10.1001/jama.2017.12145.

[51] M. K. Moridani and H. Farhadi. “Heart rate variability as a biomarker for epilepsy seizure prediction”. In: *Bratislavské Lekárske Listy* 118.1 (2017), pp. 3–8. DOI: 10.4149/bll_2017_001. URL: https://doi.org/10.4149/bll_2017_001.

[52] J. Jeppesen et al. “Seizure detection based on heart rate variability using a wearable electrocardiography device”. In: *Epilepsia* 60.10 (2019), pp. 2105–2113. DOI: 10.1111/epi.16343.

[53] J. Pavei et al. “Early seizure detection based on cardiac autonomic regulation dynamics”. In: *Frontiers in Physiology* 8 (2017), p. 765. DOI: 10.3389/fphys.2017.00765.

[54] S. Vieluf et al. “Generalized tonic-clonic seizures are accompanied by changes of interrelations within the autonomic nervous system”. In: *Epilepsy & Behavior* 124 (2021), p. 108321. DOI: 10.1016/j.yebeh.2021.108321.

[55] A. Halimeh et al. “Explainable AI for wearable seizure logging”. In: *Seizure Journal* (2023).

[56] T. Rukasha et al. “Evaluation of wearable electronics for epilepsy”. In: *Electronics* 9.6 (2020), p. 968. DOI: 10.3390/electronics9060968.

[57] M. I. Maturana et al. “Critical slowing as a biomarker for seizure susceptibility”. In: *bioRxiv* (2019), p. 689893.

[58] N. R. Temkin and G. R. Davis. “Stress as a risk factor for seizures among adults with epilepsy”. In: *Epilepsia* 25.4 (1984), pp. 450–456. DOI: 10.1111/j.1528-1157.1984.tb03442.x.

[59] M. Garbarino et al. “Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition”. In: *2014 4th International Conference on Wireless Mobile Communication and Healthcare—Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. 2014.

[60] S. Van de Voorde. “Detectie van epilepsieaanvallen op basis van verschillende fysiologische signalen”. MA thesis. Universiteit Gent, 2018.

[61] M. O. Baud et al. “Multi-day rhythms modulate seizure risk in epilepsy”. In: *Nature Communications* 9.1 (2018), p. 88.

[62] P. J. Karoly et al. “Circadian and circaseptan rhythms in human epilepsy: A retrospective cohort study”. In: *The Lancet Neurology* 17.11 (2018), pp. 977–985. DOI: 10.1016/s1474-4422(18)30274-6.

[63] P. J. Karoly et al. “The circadian profile of epilepsy improves seizure forecasting”. In: *Brain* 140.8 (2017), pp. 2169–2182.

[64] M. Tigistu et al. “Frequency of seizure attack and associated factors among patients with epilepsy at University of Gondar Referral Hospital: a cross-sectional study, Gondar, North West Ethiopia, 2017”. In: *BMC Research Notes* 11.1 (2018), p. 652. DOI: 10.1186/s13104-018-3761-3.

[65] Q. Chen et al. “Epilepsy seizure detection and prediction using an approximate spiking convolutional transformer”. In: *ISCAS 2024 - IEEE International Symposium on Circuits and Systems*. IEEE, 2024. DOI: 10.1109/ISCAS58744.2024.10558341.

[66] A. Schulze-Bonhage and A. Kuhn. “Unpredictability of seizures and the burden of epilepsy”. In: *Seizure prediction in epilepsy: From basic mechanisms to clinical applications*. Ed. by B. Schelter, J. Timmer, and A. Schulze-Bonhage. Wiley-VCH Verlag, 2008, pp. 1–10.

[67] Y. Song et al. “Canonical Correlation Analysis of High-Dimensional Data with Very Small Sample Support”. In: *Signal Processing* 128 (2016), pp. 449–458. DOI: 10.1016/j.sigpro.2016.05.020.

[68] S. Chaabene et al. “An Overview of Methods and Techniques in Multimodal Data Fusion with Application to Healthcare”. In: *International Journal of Data Science and Analytics* 20.4 (2025), pp. 3093–3117. DOI: 10.1007/s41060-025-00715-0.

[69] T. Guo et al. “Empowering Wearable Seizure Forecasting with Scheduled Sampling”. In: (2023).

[70] L. Miron et al. “Seizure forecasting based on AI-supported analysis”. In: *SpringerLink* (2024).

[71] J. Tang et al. “Seizure detection using wearable sensors and machine learning (review)”. In: *PMC* (2021).

[72] C. L. Grzeskowiak and S. B. Dumanis. “Seizure forecasting: Patient and caregiver perspectives”. In: *Frontiers in Neurology* 12 (2021), p. 717428. DOI: 10.3389/fneur.2021.717428.

[73] Md Monir Khurram et al. “Epileptic seizures classification in EEG using PCA based genetic algorithm through machine learning”. In: *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*. IEEE, 2021, pp. 135–136. DOI: 10.1109/AIKE52691.2021.00029.

[74] Wenbin Zhou et al. “Multi-scale progressive gated transformer for physiological signal classification”. In: *Proceedings of the 2nd Conference on Health, Inference, and Learning*. Ed. by Yuan Luo, Eugene Agichtein, and Krishna Ratakonda. Vol. 189. Proceedings of Machine Learning Research. PMLR, 2023, pp. 406–423. URL: https://proceedings.mlr.press/v189/zhou23b.html.

[75] B. Wang et al. “Three-Dimensional Poincaré Plot Analysis for Heart Rate Variability”. In: *Complexity* 2022 (2022), pp. 1–9. DOI: 10.1155/2022/3880047.

[76] R. Miller et al. “Long-term seizure diary tracking habits in clinical studies (HEP1 analysis)”. In: *PMC* (2024).

[77] F. Idrees et al. “Persuasive technologies design for mental and behavioral health”. In: *PMC* (2024).

[78] J. Freyne et al. “Push Notifications in Diet Apps: Influencing Engagement”. In: *International Journal of Human-Computer Interaction* 33.10 (2017), pp. 833–845. DOI: 10.1080/10447318.2017.1289725.

[79] R. De Vries et al. “Personalizing digital nudges …” In: *ScienceDirect* (2025).

[80] S. Cohen et al. “Characterizing patient compliance over six months in remote digital trials of Parkinson’s and Huntington disease”. In: *BMC Medical Informatics and Decision Making* 18.1 (2018), p. 138. DOI: 10.1186/s12911-018-0714-7.

[81] P. D. Adamczyk and B. P. Bailey. “If not now, when? The effects of interruption at different moments within task execution”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2004, pp. 271–278. DOI: 10.1145/985692.985727.

[82] T. Okoshi et al. “Reducing users’ perceived mental effort due to interruptive notifications in multi-device mobile environments”. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015, pp. 475–486. DOI: 10.1145/2750858.2807517.

[83] J. Chiam, A. Lim, and A. Teredesai. “NudgeRank: Digital algorithmic nudging for personalized health”. In: *arXiv* (2024).

[84] T.-H. Hsu and J.-W. Tang. “Development of hierarchical structure and analytical model of key factors for mobile app stickiness”. In: *Journal of Innovation & Knowledge* 5.1 (2020), pp. 68–79. DOI: 10.1016/j.jik.2019.01.006.

[85] H. Malik, E. M. Shakshuki, and S. Katuku. “Exploring the Relationship Between Version Updates and Downloads of Asthma Mobile Apps”. In: *Procedia Computer Science* 109 (2017), pp. 624–631. DOI: 10.1016/j.procs.2017.05.368.

[86] A. Smith et al. “Developing smartphone apps for behavioural studies: The AlcoRisk app case study”. In: *Journal of Biomedical Informatics* 72 (2017), pp. 108–119. DOI: 10.1016/j.jbi.2017.07.007.

[87] A. Balaskas and et al. “The Psychology of EdTech Nudging”. In: *MDPI* (2025).

[88] A. Sahami Shirazi et al. “Large-scale assessment of mobile notifications”. In: *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI ’14)*. 2014, pp. 3055–3064. DOI: 10.1145/2556288.2557189.

[89] A. Wohllebe et al. “Mobile apps in retail: Effect of push notification frequency on app user behavior”. In: *Innovative Marketing* 17.2 (2021), pp. 102–111. DOI: 10.21511/im.17(2).2021.10.

A

Appendix A - Stakeholder Research

This appendix is related to all stakeholder meetings during the project. When an iteration was finished, or during the ideation phase of a new iteration, stakeholders have been engaged who may be able to help with challenges/problems faced during that specific iteration phase. The stakeholders were given a consent form (Appendix D.2.2) before the meetings started.

Stakeholder map

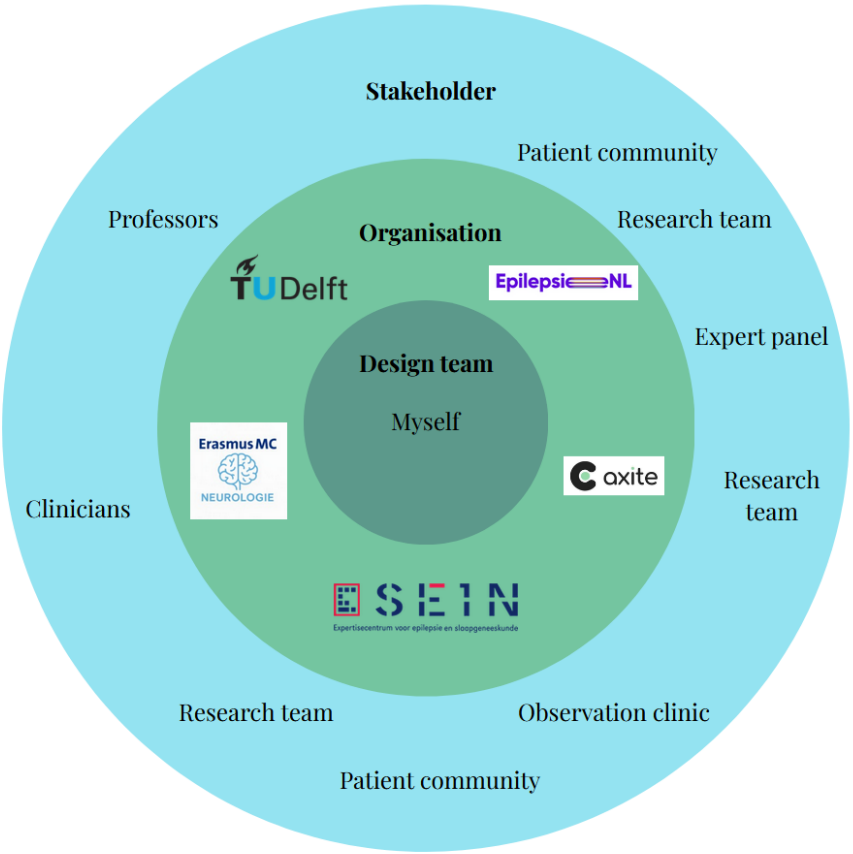


Figure A.1: Stakeholder map.

A.1. Interview 1 - Clinician

Date: 2nd December 2025

The participant mentioned that he mostly operates with EEG models to create a forecasting model himself, which, unfortunately, is not included in this project.
Current interface for neurologists: timeframe of last period (3 months) with all measured seizures on the x-axis, and another one stating the amount of medication taken.

Suggested additions: Labels with an expanding window below every measured seizure on the x-axis. The window contains a list of suggested relevant factors by the ML model. Update medication bar: Amount and moments of medication.

To transform 'sleep' (or other scenarios) from a symptom (as label input) measured manually before, to a predictor useful for ML training: let the smartwatch automatically measure sleep variables.

Follow-up questions will not be necessary to include in the journaling app, as too many steps influence the amount of logs. Also, if necessary, these questions will be asked in person by the neurologist.

Still no clear indication of how the output of the journal methods can be assessed. This is indeed not verifiable with data from such a small time period. However, when delivering a proof-of-concept using synthetic data, this could also be a big step for the future development of such a system.

Time lag of 'high-risk' parameters differs per individual, but the journal methods could identify this variable.

Suggested contacts for future research studies:

- Sein
- EpilepsieNL
- Brian@Home
- CS-Professor

EpilepsieNL could maybe distribute a future model of the app to their registered members. Their de-identified data can be used for test-modelling.

High-risk factors: sleep, stress, menstrual cycle, seasonal (light-dark contrast), lifestyle (consumption).

Gyroscopic feature: even slight movement can indicate a verbal seizure or other preictal periods.
Environmental audio and light are also useful. Pupil analysis is unrealistic for the project as well as implants.

HRV alone is already a good feature, but a combination of features would be better.

Some patients do not want to collect data about their disorder. Also, the use of a digital application won't make it accessible for the elderly or people without a phone.

A.2. Interview 2 - Computer Science Faculty, TU Delft

Date: 14th January 2026

This meeting was arranged with a professor in computer science with research expertise in machine learning and interpretable AI models, to explore the field of predictive ML models, and how to be able to implement a model into the developed application. During the interview, the fact that a lot of patients would not be interested in actively journaling their daily lives to create an ML model came up. This made me think about how other people perceive the idea of collecting their own data for an ML purpose. This might seem odd, or unknown, as a lot of people do not know what data or features can even tell something about their health. Applying all these derived, unfamiliar data types to another unknown model can be scary. To counteract this very understandable idea of ML models, I thought it would be more accepted by patients if there were a transparent way of showing which data is interesting for what purpose, and how it is incorporated into the personalised ML model.

Conclusions from participant feedback

Deep Neural Networks create black boxes, from which it is hard to interpret the results/calculations. A maximum of 2/3 layers is recommended to remain interpretability of the data/choices of the algorithm.

There are some publicly available databases: Physionet, Rosalind Picard (Empatica), TUH seizure dataset (EEG).

Proper user research is needed to validate the app's usability and the interpretability of the created insights. Patient feedback is crucial.

The applied resolution of the saved data (the intervals between measured values corresponding to a label) determines what kind of prediction can be made. If it is preferred to predict scenarios for multiple periods (10 min, 30 min, 1 hour) every prediction period must have its own ML model. A 10/15 minute period seems most realistic.
It depends on the period for which the ML model can predict. For instance, when measuring every 10 minutes, the differences of values between a period of 10 minutes are compared, and from those changes a prediction can be made for the value of the data values for the next 10-minute timestamp.

To visualise the differences in feature data between different scenarios, a box plot of two variables could be useful, visualising the mean, median, and also determining a P-value.
Histograms would also be a sufficient way to display differences between scenarios.

Make multiple designs for different target groups.

Sub-goal determined; Check and analyse medication effects, how does the patient respond, what changes when switching from one type of medication or adjust the doses. This would require more types of labels to match all types of subjective experiences related to medication effects or changes in doses.
To be able to compare different periods, it must also be able to select more than one period (from:... to:...) to compare averages of different weeks (and thus the effect of different periods).

A way to assess the quality of the app would be to hand out questionnaires to participants and, after a period of using the app, evaluate how their stress/quality of life has been affected.

False-Positive suggestions would be a way to determine the data quality. Also, the fact that a high-risk situation has been avoided by the app's warnings could be a determination.

The differences in change points would be good input for an ML model, and to compare the effects of different medications, as well as the length of a seizure, and the number of seizures.

A.3. Axite Workshop

Date: 4th February 2026

Axite is a spin-off team from SEIN, EpilepsieNL, Nightwatch, and the University of Leiden. They are developing a similar sort of app, but connected to an EEG headband.

During this workshop, they discussed the frequency of asking participants to complete a questionnaire and how to approach the participants. This correlated to one of the research studies for this project.

They had not thought of any nudging method yet. However, the workshop focused mainly on how to keep participants engaged. A peer-to-peer system has been suggested, enabling comparison of scenarios/labels/events with other patients' subjective journal data. Gamification is also mentioned with rewards after patients have reached their own predetermined goal. The type of goal could be chosen by the patients themselves, such as less stress, better sleep, or any other goal.

Their questionnaire has the following order:

1. A yes/no question whether something significant has happened in the meantime.
2. An open question of 50/100 characters.
3. An impact rating 1-10.

4. An importance rating 1-10
5. The date

A.4. Meeting Sein

Date: 9th february 2026

During this meeting, the objective was to discuss collaboration opportunities and to ask if Sein is possible to distribute questionnaires for patients and clinicians throughout their network.

Needs for treatment and research aligned with previously identified needs from clinicians, as the same data is relevant for treating a patient as well as for deeper analysis of patient data.

An introduction was given to their currently pending researches and their research teams for future meetings and collaboration.

A.5. Meeting EpilepsieNL

Date: 12th February 2026

During this meeting, I went to the office of EpilepsieNL in Houten to discuss the patient perspective and potential network distribution. With their knowledge from the patient's perspective and the current developments related to epilepsy detection and prevention, they were able to provide some valuable insights.

During this meeting, they were open to sharing a preview of the app and a small set of questions (related to the usability of the app from a patient's perspective) within their network.

The most valuable insights gained from this meeting were:

- Currently, another app is being developed by the AXITE team, which is focused on monitoring related subjective data with an EEG-headband. They are planning on asking the participants to fill in an open questionnaire once a week, asking the user whether a relevant situation has occurred since the last time they filled in a questionnaire. However, when their participants are not compensated with money, they tend to lose interest and do not engage with the app. This correlates with previous research suggesting that an open and non-personal questionnaire is not a very efficient way to keep users engaged.
- Do not immediately let patients access the analysis page. Enable users to track and analyse the features they are particularly interested in; the rest might be an overkill in terms of too much information. A suggestion was to let the patient add features step-by-step, which they find most interesting/important. They thought that 3 features would already be a lot of information for the average patient.
- To keep patients engaged in using the app, personal goals could be set which

the particular patient would like to gain more insight into their epilepsy. A goal could be to analyse how epilepsy affects their sleep, and what they can do to improve their sleep quality, or identify what causes are related to their current sleep behaviour.

- At a research centre where one of the employees of EpilepsieNL's son is currently staying, they are using an app to track the number of seizures of a patient. However, it only accounts for the total amount of seizures per week, and its UI is unclear. Colours are used inappropriately, which causes confusion. Therefore, a way to efficiently visualise the frequency of seizures could be a possible addition to the app.
- The data of the user is currently stored on the user's phone. The data can only be shared when exporting all journaled labels of a certain period of days (including the corresponding physiological data). But what happens when a patient loses their phone? This privacy-sensitive matter has not been addressed yet, but it is very important to implement it safely and correctly.
- They also warned me that this app is meant for a clinical purpose, which brings a lot of trouble in terms of approval/privacy/policy, etc. For instance, an algorithm is not static, but changes over time; every time a designer does a small iteration or adjustment, all requests for approval must be filed over again. It is not worth checking the regulations related to clinically implemented products/services, as it is too big of a burden to find an approach that aligns with all regulations which can be applied within the remaining time of the graduation project.
- Open questions can be useful, as they provide context beyond the objectively measured physiological data. But too open, not personally related questions only confuse users and do not generate that much valuable data. Therefore, a small follow-up question could be added to the current labels. When pressing a label, there could be an extra text field where the user can elaborate why they think this label has occurred (Appendix B.5.1). For instance, when a user journals 'high stress', the user can add a small note in the text field to address what the user thinks might be the cause of the 'high stress' label (Figure B.16).

A.6. Meeting Sein

Date: 16th February 2026

This meeting was with a clinical technologist who is a PhD candidate of SEIN in Heemstede. She was willing to ask her colleagues whether it is possible to distribute a short overview of the app and a usability questionnaire within SEIN's network.

Furthermore, we discussed the following topics:

- Multiple labels per timeframe: It is currently only possible to journal single labels. She suggested that when multiple scenarios occur at the same time, an adjustment could be made to prevent users from completing the same task multiple times. This required a small adjustment to the timeframe-determination window, which pops up after pressing a label on the logger page (Figure B.16), to enable

users to add extra labels when determining a timeframe of a single pressed label.

- Also, the addition of a visual with suggested epileptic-related situations was discussed, which correlated with the thoughts of an earlier interviewed clinician, who would like to see an overview of the frequency of occurring seizures.
- Again, the analysis page was perceived as too much information to show to a patient, as a lot of graphs are not directly related to what a patient might want to track.
- Another suggestion of setting a goal per patient was discussed, which also came up during an earlier meeting with EpilepsieNL. This might keep users more engaged in using the app. This could also minimise the amount of functionalities on the analysis page, as only the features can be shown which correlate to the user's predetermined goal. For instance, a user might be interested in improving their sleep quality and tracking which factors might have an impact on their sleep.
- To prevent users from searching for the right label, or when they do not need all labels but use a set of labels more often than others, a set of favourite labels could be added so the users can instantly access the labels he/she uses the most, without having to search within the current categories.
- Currently, there are labels included in the logger page that are seizure-related. However, there is no possibility of distinguishing types of seizures. Therefore, a text box could be added when determining the timeframe of a label to provide extra context with a small note.
- The resolution of the saved values of the physiological datatypes when saving a timeframe should be as low as possible. Meaning, that a future ML model would be more accurate when the time between saved measurements is shorter, enabling the identification of rapid changes of physiological data, and not skipping a peak value of a graph.
- To make the visualisations of all data clearer, a small note with an elaboration of what the visual represents could be added, so a clinician can instantly check the status quo, which is useful during a follow-up appointment with a patient.
- A suggested model was encouraged, but only if it aligns with a predetermined goal of a user, because otherwise it might cause confusion and maybe scare patients because they get a suggested warning without knowing where it is related to.

A.7. Meeting patient 1

Date: 17th February 2026

This meeting was very valuable, as the participant is diagnosed with epilepsy himself, has developed clinical services in the past, has participated in previous research studies related to epilepsy and trigger identification, and is an active member of Sein, Kempenhaeghe, EpilepsieNL, and the patient community.

Environmental pressure (specifically in a plane) influences the O2 level within the brain, which can cause a seizure.

To counteract high-contrast or loud noises, the participant has AI-earphones to stabilise surrounding noise and decrease the risk of a seizure.

Every so often, the participant wears a diabetes pad on his arm (Figure A.2). When tapping the pad with a phone, the blood-sugar values get uploaded to the LibreLink platform. This enables him to keep track of this parameter and to ensure that it remains stable across the day.

Figure A.2: Libre pad

After a focal seizure, the blood sugar level rises, which makes it an important indicator of epilepsy related situations.

It would also be interesting to have different food labels, as not only the fact that you're eating but also what you are eating is relevant for epilepsy patients.

A period without seizures can also be interesting to keep track of because it can also be a type of trigger for other seizures when a patient is experiencing frequent or periodic seizures.

Positive stress is a mental state that has not yet been implemented in the app. This distinction between types of stress is useful for a deeper analysis of triggers. When the adrenaline level rises due to positive situations, the effects on the physiological data can differ from stress-related scenarios, like pressure or fear, and have different impacts related to epilepsy. Particularly for this participant, positive stress-related situations did not cause any epileptic event. The option to add an extra note to a label could be a good solution.

The participant was interested in all types of features of the app, instead of feeling overwhelmed and disoriented.

A goal that the participant would set for himself when using the app would be to minimise peaks in data. The participant experiences more seizures when physiological values vary a lot.

Apple has done previous research with an Apple Watch to monitor epileptic events; the results of these studies could be very interesting for this project.

An overview of the amount and severity of seizure/epileptic-related events is suggested as a good addition.

The participant also has an NVS prosthesis, which sends pulses to the participant's brain. Whenever an epileptic related situation occurs, the participant can grab a magnet and rub it over the place in the chest where the NVS is located to increase the amount/severity of the pulses.

Narcosis during operations also affected seizure behaviour. In a very severe way of non-stop seizures.

The questionnaires should be in B2-level language, in the language of origin of participants, to make it understandable for everyone.

The biggest hurdle the participant encountered during medical, privacy-sensitive data handling is the fact that, between clinicians, the policy around sharing data is very complicated and more of a legal problem. However, when a patient is sharing data with his clinician, there are no laws or hurdles, as it is the patient's own choice.
Currently, this data is mostly shared in PDF files, therefore, the currently used CSV files could be better adjusted to a PDF format.

A.8. Meeting Sein

Date: 5th March 2026

During this meeting, a researcher from Sein contacted me about the progress of my project. They were wondering if the project could support their research proposal.

They are currently looking to develop a predictive model, in which they need to identify epileptic-related situations, which my project could be used for.

They suggested implementing the app I have designed during their research, which starts in 2027.

A.9. Meeting patient 2

Date: 8th March 2026

This meeting was set up to let another epilepsy patient conduct the research study described in Chapter 5. This patient lives at an epilepsy research centre, more specifically, in an observation clinic.

It was interesting to see how the observation clinic of Sein was organised. Knowl-

edge was gathered about how Sein operates and the struggles of patients who have to move to the clinic.

A.10. Clinical evaluation

Date: 25th March 2026

This meeting was conducted with participant 1, as a review compared to the prototype of the previous meeting (Appendix A.1). The clinical value of the app has been confirmed.

The app could be useful for clinicians to evaluate a patient's medical usage, seizure frequency, and analyse changes in physiological behaviour with subjective labels to identify triggers.

However, the output of the analysis of the wearable data must not be a recommendation or medical statement, as algorithms for medical purposes are high-risk rated in the MDR.

To meet the MDR standards, any output derived from the wearable data can only be used to visualise that output or to state a number, and not suggest any current risk, because that would be a medical suggestion which an algorithm may not make.

B

Appendix B - Design Process

B.1. Calculation of extracted features

From the variables measured continuously by the wearable, features are calculated to provide more insight into the fluctuating changes of the physiological parameters. From these features, we can indicate which events are linked to relevant changes. These features can be used to identify what events relate to personal triggers of an epilepsy patient.

HRV - mean, median, SD: The SD is derived from a baseline mean.
SDSD = std(diff(NN)).
$SD1 = \sqrt{(0.5) * SDSD.}$
$SD2 = \sqrt{2 \cdot SDNN^2 - 0.5 \cdot SDSD^2}$.

Poor sleep quality Sleep debt quantifies the difference between a patient's target sleep requirement (Hreq) over a period of multiple days (D).

$$\text{Sleep Debt} = \sum_{d=1}^{D} (H_{\text{req}} - H_{\text{actual},d})$$

Seizures can also occur in relation to the circadian cycle, with specific times being high-risk periods. This feature is crucial for identifying temporal patterns and optimising the monitoring and medication intake moments of patients [61, 63].

Heart rate (HR) mean and derivative: Heart rate changes are common physiological indicators of epileptic seizures, with these key indicators of ictal and pre-ictal events [54].

$$\overline{\text{HR}} = \frac{1}{N} \sum_{i=1}^{N} \text{HR}_i$$

HRV Time-domain metrics (SDNN, RMSSD, pNN50) quantify autonomic cardiac regulation. These features have been used in algorithms for predictive models for wearable

ECG devices [52].

$$\text{SDNN} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\text{RR}_i - \overline{\text{RR}})^2}$$

$$\text{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(\text{RR}_{i+1} - \text{RR}_i)^2}$$

$$\text{pNN50} = \frac{\text{Count}\{|\text{RR}_{i+1} - \text{RR}_i| > 50 \text{ ms}\}}{N-1} \times 100\%$$

HRV non-linear (SD1, SD2, entropy metrics) analysis can identify subtle pre-ictal shifts in autonomic state. This is known as the ”slowing” phase of an absence before a seizure event [57].

$$\text{SD1} = \sqrt{\frac{1}{2}\text{RMSSD}^2}$$

$$\text{SD2} = \sqrt{2 \cdot \text{SDNN}^2 - \text{SD1}^2}$$

The balance between low-frequency (LF) and high-frequency (HF) components of HRV is a key metric in determining autonomic regulation. It is used to understand the relationship between cardiac autonomic dysfunction and epileptic seizures [45].

$$\text{LF}_{\text{norm}} = \frac{\text{LF}}{\text{LF} + \text{HF}}$$

$$\text{HF}_{\text{norm}} = \frac{\text{HF}}{\text{LF} + \text{HF}}$$

$$\frac{\text{LF}}{\text{HF}} \text{ ratio} = \frac{\text{LF}}{\text{HF}}$$

The respiratory rate(RR) and its interval are also monitored as a pre-processing step to ensure the validity of the HRV calculations. Raw RR data can be influenced by motion or noise [**Bottcher2022Data**, 36].

$$\text{RR}_i = T_{R,i+1} - T_{R,i}$$

The SD1 parameter of the RR rate is used to show beat-to-beat rapid changes, and SD2 for long-term beat-to-beat changes. These provide useful information on short- and long-term fluctuations [51]. These parameters can be used to generate a Poincaré plot for nonlinear and geometrical analysis for HRV [75].

$$SD1 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}\frac{(RR_i - RR_{i+1})^2}{2}} \tag{B.1}$$

$$SD2 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}\frac{(RR_i + RR_{i+1} - 2\overline{RR})^2}{2}} \tag{B.2}$$

Electrodermal activity (EDA) reflects sympathetic nervous system arousal. The EDA mean, or skin conductance level (SCL) provides insights into the autonomic response.

$$\overline{\text{SCL}} = \frac{1}{N}\sum_{i=1}^{N}\text{SCL}_i$$

The EDA skin conductance response (SCR) count measures rapid peaks in response to stimuli, often related to stress and arousal. Analysing amplitude and frequency can provide information on nervous system activity and correlated journal labels. This feature is often used in previous research for the development of an ML model for seizure detection [71].

Skin temperature changes (ΔT) influence peripheral vasoconstriction and is an indirect physiological marker for changes in autonomic tone. Therefore, skin temperature is included in multiple seizure detection models [56, 6].

$$\Delta T = T_i - T_{\text{baseline}}$$

B.1.1. Combination of features

The correlation of the physiological data types can be calculated using a Principal Component Analysis (PCA) and Canonical Correlation Analysis (CCA) [54, 67]. The strength of the correlation between any two modalities can be estimated with an overall correlation: $\rho_c = 1 - \prod_{i=1}^{r}(1 - k_i^2)$
Where k_i denotes the i-th estimated canonical correlation, and r represents the number of components between the two physiological data types.

B.2. MVP

Label buttons derived from literature review:

1. Mental/Emotional state
2. Environmental/Sensory factors
3. Activities/Behaviour
4. (Sleep)
5. Medication
6. Seizure-related
7. Physiology/Health

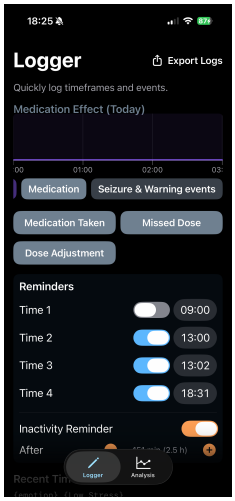


Figure B.1: Options for fixed and inactivity reminders.

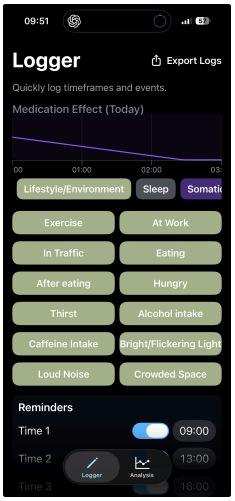
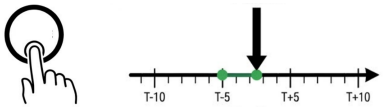


Figure B.2: Colour-coded categorised buttons.

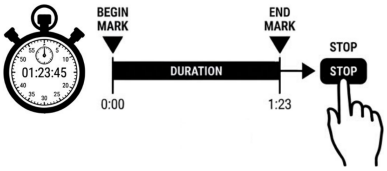
Figure B.3: Label buttons.

Journaling an event/label

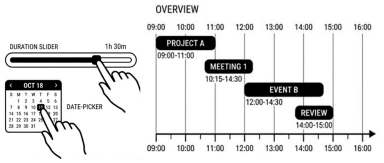
- A 'snapshot', when a user immediately wants to journal an event. Pressing a button registers a predetermined, fixed, small timeframe (+-3 min, for instance) to capture a single-label scenario. When journaling multiple labels in a short period (<10 min), short overlaps can occur. Advantages of this method are the minimal amount of required steps to complete a measurement, and the minimal cognitive load, which is essential for people with transient cognitive impairment. This is useful for journaling crucial seizure-related events or high-risk periods when there is not much time to label an event. This has been applied to all label buttons, where double-tapping a button enables users to instantly measure a label.



- Using a stopwatch tool, when a user specifically wants to capture data from a pre-planned event. The user hits the stopwatch button in real time when the event starts and finishes to determine the timestamps more precisely. This method allows single or multiple labels to be assigned to the measured period. However, there is no possibility for overlapping timeframes or determining other time periods afterwards.



- Journaling events afterwards requires flexible timeframe determination, where the user selects start/stop timestamps of the timeframes manually at any time. These periods can also be preset for the near future. This adds an option to restore important missed events and optimises the completeness of the total amount of labelled hours. This method enables the user to overlap multiple timeframes and to log single or multiple labels per timeframe.



The stopwatch function has been added at the top of the interface, above the buttons for instant access, so timeframes can be determined precisely. Start, pause and stop buttons have been placed besides of the stopwatch. After stopping a stopwatch measurement, it is possible to determine which labels will be logged and linked to the wearable data during this period.

Medication timeline

After interview 1 with a clinician, Participant 1, (Appendix A.1), new insights were shared to adjust the interface so it aligns better with the needs of clinical experts. This includes a timeline indicating all seizures or suggested sensitive periods by the ML model, with a small tab beneath each indicated seizure. When tapping this tab, a

small list of relevant features and parameters is shown. This allows the practitioner to quickly scan the seizure events and, if necessary, get a quick overview of the most important triggers. When the same relevant factors are suggested, these can be taken into account when personalising the thresholds for a patient.

Another addition that has been made in the interface of the app is a timeline with on the y-axis the medication dose and on the x-axis the moments of intake. The length of time before the graph decreases and reaches 0 can be determined manually based on the individual's required time between intakes (Figure B.4). For instance, if a patient needs at least 8 hours between their morning and evening intakes. After 8 hours, a notification is sent as a reminder for a new intake.

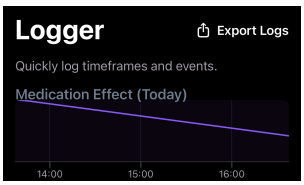


Figure B.4: Interface mock-up showing medication timeline and suggested seizure triggers.

B.3. Iteration 1

In this section, the first iteration of the functional prototype is elaborated.

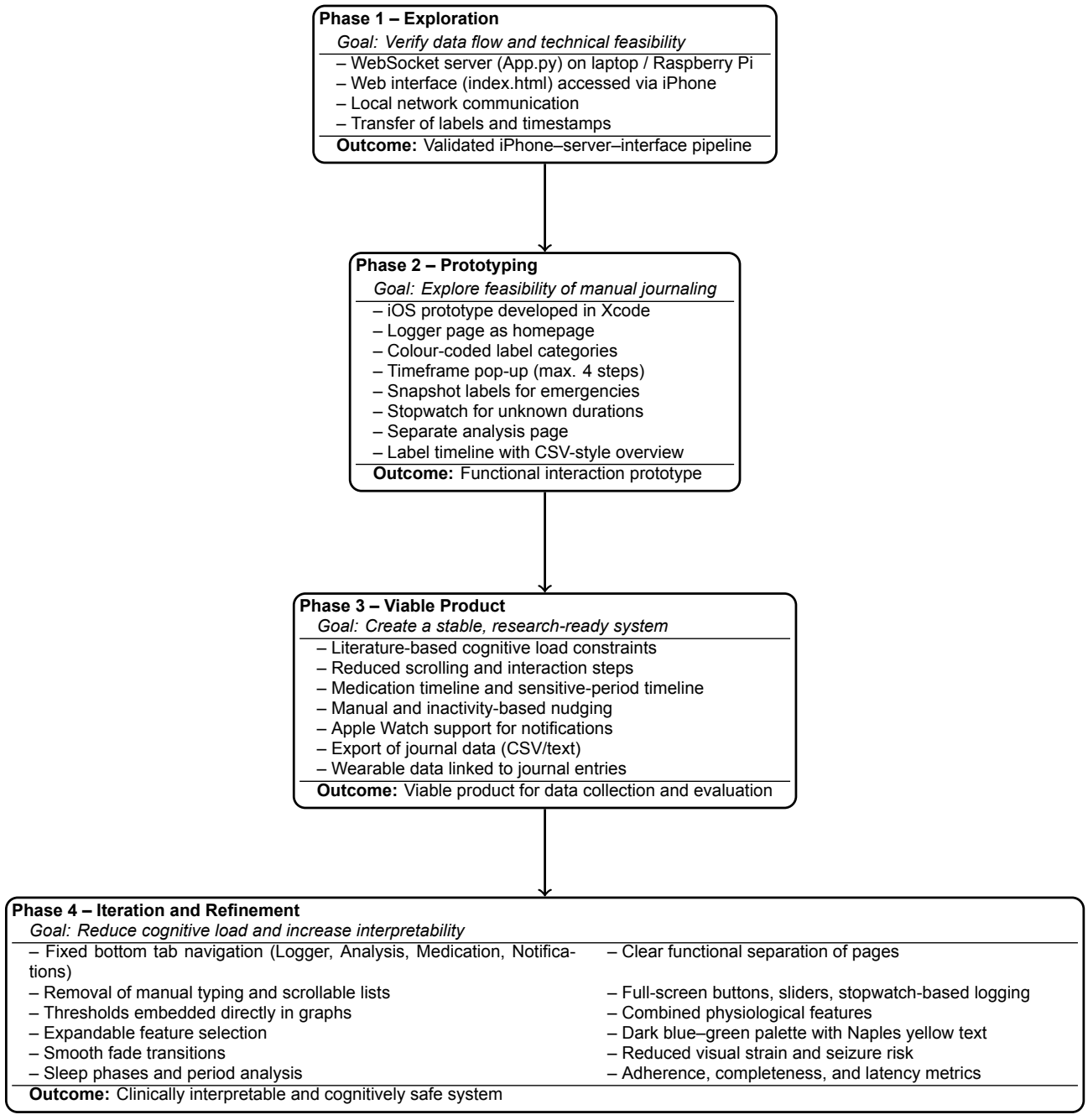


Figure B.5: Overview of the iterative app development process.

Figure B.6: System overview.

S.C.A.M.P.E.R. overview iteration 1

Substitute	**Navigation Method:** Replaced the fold-out list and swipe navigation with a fixed bottom tab menu to improve accessibility and enable faster page switching (Chapter 3.4).
Combine	**Thresholds & Graphs:** Integrated threshold values directly into physiological data graphs using striped horizontal lines, allowing real-time comparison between data and limits.
Adapt	**Clinical Coupling:** Adapted findings from Vieluf et al. (2021) by adding combined-feature analyses (HR–RR, RR–Temp, HR–Temp) at the bottom of the analysis page for deeper autonomic insight.
Modify	**Visual Comfort:** Modified the interface to a dark blue-green palette with Naples yellow text and smooth fade transitions to reduce visual strain and minimise seizure risk (Chapter 3.4).
Put to Another Use	**Apple Watch Nudges:** Repurposed the Apple Watch from a passive data collector to a notification device using vibrations and sound for discreet, accessible nudges.
Eliminate	**Manual Entry & Scrolling:** Removed manual timeframe entry and scrollable category lists, replacing them with fixed full-screen buttons and sliders to reduce unnecessary steps (Chapter 3.4).
Rearrange	**UI Hierarchy:** All separate functions have been given their own page. The Export button has moved from the Logger page to the Analysis page, and a Stopwatch has been added to the Logger page to prioritise real-time event tracking.

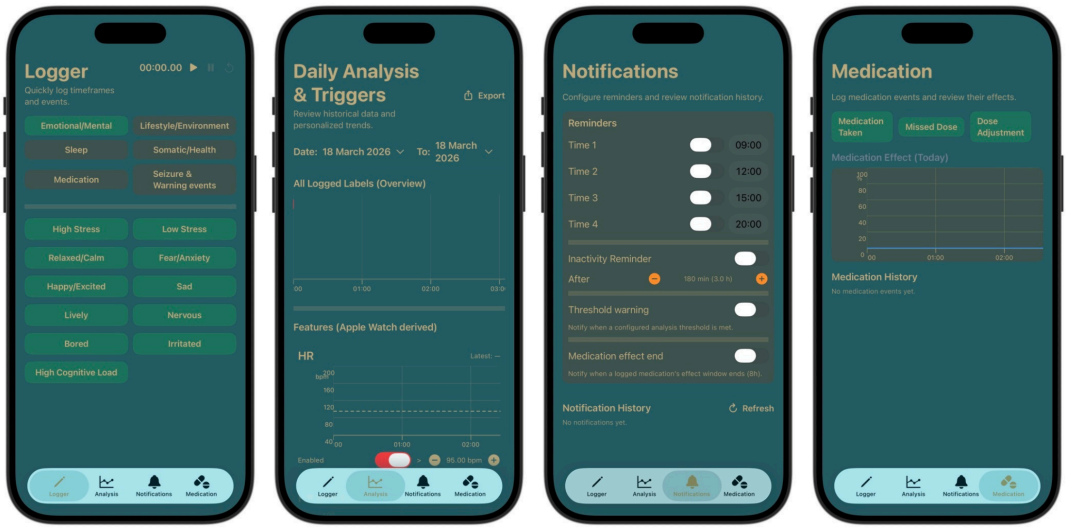


Figure B.7: Design after iteration 1

Colour changes

To align with the criteria found in previous research, the colours of the interface have been adapted to a dark blue-green theme to prevent high colour contrasts. The text colour (naples yellow) is also adjusted for this matter. Different sections have been divided with a divider line to indicate different parts of a page.

The colours of the theme have been inspired by two books; A Dictionary of Color Combinations Vol. 1 and Vol. 2.

Small stopwatch added and replacement export button

The export button, which was previously located on the top right corner of the logger page has been moved to the top right corner of the analysis page. A small stopwatch has been placed at the top right corner of the logger page instead.

Buttons fitting the whole page + removal of scrollable category buttons

Scrollable category list has been replaced by fixed buttons. This is done to fit all buttons in the whole screen and prevent unnecessary steps to select the correct label.

Labelling and timeframe determination

To log any journal data, there are several options to complete this task (Figure B.8):

1. Press label, which then shows a sliding tab for the label menu
2. Press and determine duration of the label
3. Press and determine when label started
4. Confirm log

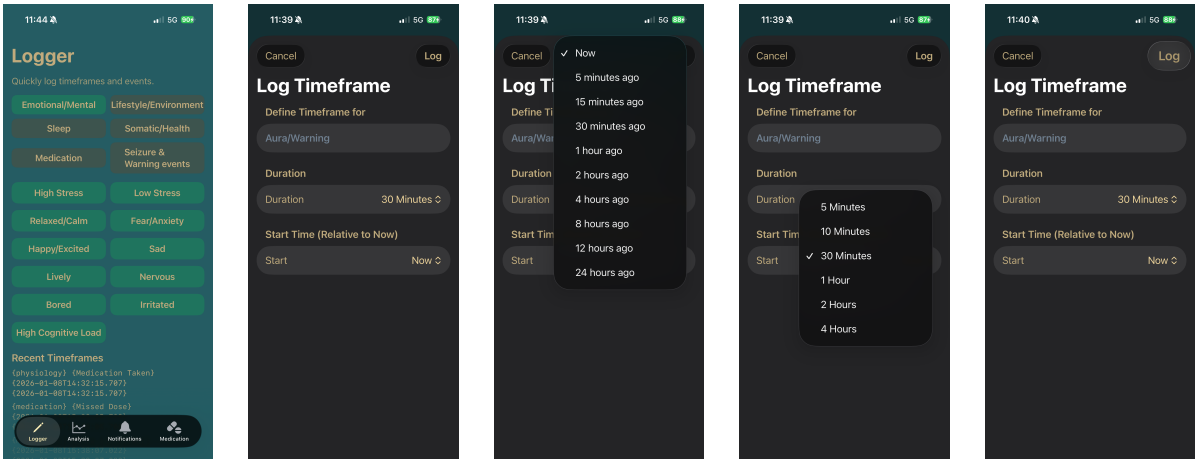


Figure B.8: Logging using fixed times

Or, using a sliding bar (Figure B.9):

1. Press label, pop-up slide page
2. Determine duration and time using the sliding bar
3. Confirm log

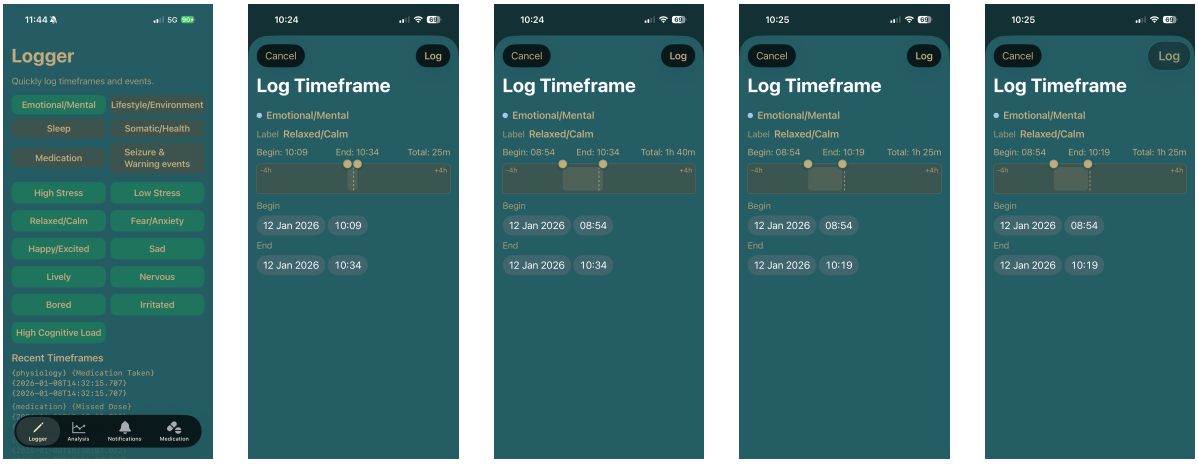


Figure B.9: Logging using a slider

To determine the timestamps more precisely, the start and end time have been stated for a second time, with an adjustable, scrollable digital clock to determine them per minute (Figure B.10).

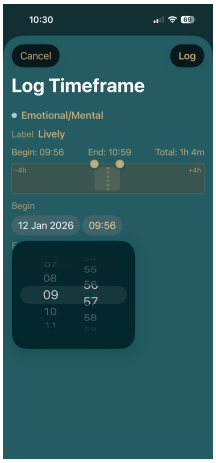


Figure B.10: Manual time determination.

The last option, manually typing the length and start of the label, will require more separate actions and will therefore require more time to log a label compared to the other two options. Therefore, this option has been ruled out.

Total amount of missing logged labels

Included for the analysis of the total amount of time covered by labels.

The purpose of this variable is to analyse the completeness of the participant's daily journaling. The amount of total hours that have not been logged (not counting any time sleeping) is stated on the analysis page, below the all logged labels timeline.

Registration of moments of nudging

For correlation analysis of nudges and journaling. A list of all timestamps related to reminders and notifications is located on the reminders page.

Nudging after activity/suggested triggers

A suggested trigger timeline has been added to the analysis page, this timeline will suggest when a scenario has happened which meets the determined threshold values. When the suggested scenario on the timeline is pressed, a window pops up with the suggested related physiological data types and their corresponding values at the time of the suggested scenario.

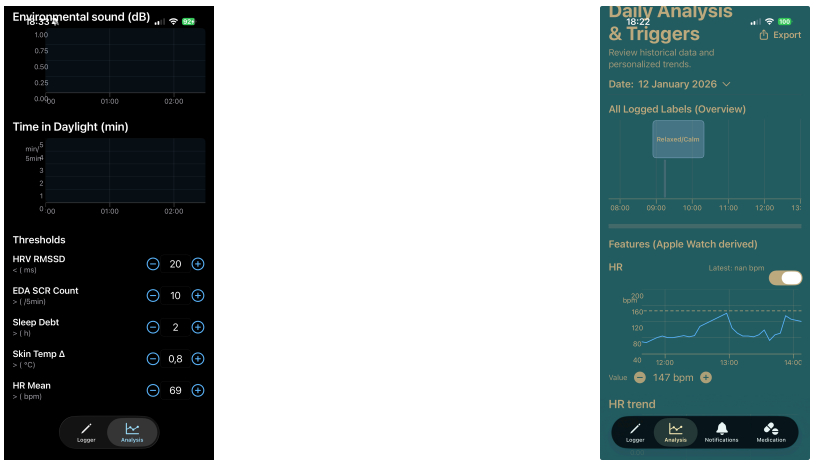
This adjustment generates an extra type of nudge, which can also be turned on or off on the reminder page, called 'Threshold warning'. For each feature, it can be determined separately whether it contributes to the threshold warning. This can be done at each feature graph on the analysis page.

Threshold value visualisation

To determine the thresholds, previously a '-' and '+' button had been applied besides the threshold names to determine the values (Figure B.11a).
To provide optimal clarity of the value of the corresponding, a striped horizontal line has been added to indicate the currently determined value of the threshold. This allows users to compare the threshold to the value of the specific physiological data

type during the day.

The striped horizontal line indicating the currently determined threshold is included into the graphs of all physiological datatypes (Figure B.11b). This minimises the total required space on the analysis-page, and thus the amount of scrolling.
Every threshold of a displayed physiological datatype can be (de)activated using an on/off button. This button determines whether the threshold activates a 'threshold warning' message.



(a) (b)

Figure B.11: Manual threshold determination (a) and threshold determination with striped line (b).

Include a menu for different functions of the app

To fit all components of a section into one whole page of the phone's screen, a tab-menu has been added, and can be used to switch from page to page. All different components of the app have been assigned to an individual tab. The pages that have been added to the tab menu are: Logger, Analysis, Medication and Notifications.

Multiple options of tab menu's have been explored:

1. Horizontally scrollable pages so the user can swipe between pages. Unfortunately, this method does not align with the determined criteria.
2. A tab-icon button at the top left corner of the screen. This icon can be pressed and folds out a list of the pages.
3. A fixed tab menu at the bottom of the screen, decreasing the amount of tasks to switch pages from 2 to 1 compared to the tab-icon and improves accessibility by always being visible on the screen.

The app still opens at the homepage, which is the logger page, for instant accessibility of journaling.

When switching from page, the transition is more smoothly, letting the pages fade into each other instead of an instant switch when pressed to prevent high-colour contrasts.

The medication timeline has been moved from homepage to the Medication page.

Sleep phases
For a more detailed analysis of the sleep quality besides the total hours of sleep, a feature graph with the different phases (awake, core sleep, deep sleep, and REM sleep) has been added. This data is directly being fetched from the sleep app on the Apple Watch.

Resolution graphs
The resolution has been set to 5 minutes. This has been done to obtain smoother graphs and reduce the noise of the physiological data types.

Combined features
At the bottom of the analysis page, below all feature graphs, an extra section has been added with the combined features from the Vieluf et al. (2021) study. This includes HR-RR, RR-temp, and HR-temp combination feature graphs.

Research output
To align with the outcome measurements of the research study, the following 'Journal analysis' features have been added at the bottom of the analysis page:

1. Time between nudge and journal activity (latency).
2. Total amount of labels per day/measurement (completeness).
3. Total time covered with labels (adherence rate).

Period analysis
To be able to analyse the average values of a particular number of days, a second date determination tab has been added at the top of the analysis page. The feature graphs will now present the average values throughout the day.
or
When a period of days has been determined, the feature graphs of the current day are still visualised, but the average graph (coloured c95 m54 y0 k0) of that period is also visualised on the feature graphs, enabling comparison of the current day.

Feature graph layout
The large number of graphs of all derived features caused the analysis page to extend beyond the preferred length to reduce scrolling. Therefore multiple options have been explored:

1. Fixed feature buttons with one graph representing the determined feature. Or when pressing more features, the other feature graphs are displayed below each other, extending the page length.
2. A foldable tab with a list of all features and one graph representing the determined feature/multiple graphs of determined features.
3. All features are listed, with a foldable tab besides their title. When pressing the tab, the corresponding graph and threshold determination buttons are visualised. '(Figure B.12)

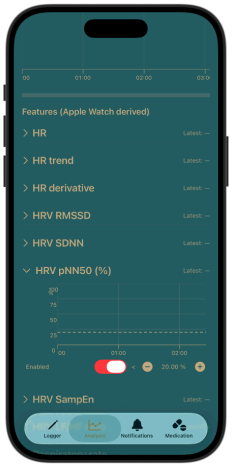


Figure B.12: List of features and their tab, unfolding feature graphs

4. Fixed category buttons, which list the coherent feature graphs that are shown beneath each other.
5. An expandable 'Features' tab is located beneath the label timeline (Figure B.13a). When pressed, a window unfolds with boxes to mark all features (Figure B.13b). When the features have been determined, the tab can be closed again and the graphs of the features which boxes have been checked, are shown beneath each other (Figure B.13c). Each graph contains a button to close/remove the graph from the page, and can be activated again in the feature tap on the top of the screen.

(a) Feature tab **(b)** Feature tab options **(c)** Feature tab graphs

Figure B.13: Single feature tab

Reduce amount of unnecessary shown graphs which the user is not interested in. Reduce amount of required space of the screen by using the expandable tab. Minimizes vertical scrolling by implementing 'close' button at each graph.

Suggested sleep timeframes

To instantly cover the time the user slept with a label, a suggested sleep label is added to the label timeline. This timeframe is marked with a question mark sign, waiting to be

confirmed by the user. These timeframes are based on the 'Hours slept' feature graph, when the determined threshold has been met. When the suggested sleep labels have been confirmed, they are registered and further processed on the app.

B.4. Iteration 2

To be able to identify irregular physiological states, a baseline model of the user has to be established (Chapter 3.3.1). To get inspiration on how to visualise the process of data classification for ML model purposes, Meetings with stakeholders have been arranged. One of them was a professor from the Computer Science faculty of the TU Delft, who is conducting research on epilepsy-related predictive models. Discussing types of suitable ML models, the struggle of interpretability, and understandable types of visualisations, supported this second iteration of the design.

New iteration requirements

1. Visualisation of data to enhance interpretation. Including classification and comparison between journaled events.
2. Whether to hide the analysis page for patients, or make separate interfaces for patient/clinician.
3. New purpose: medication effect monitoring, when patients are adjusting their medication (dose). This does not require a predictive model.

Analysis page update

The analysis page has been extended with a 'baseline analysis' section (Figure B.14). This originated from the idea to transparently visualise how an ML model classifies the journaled labels, and how it uses its data and features to create a personalised model. When unfolding this tab, a doughnut chart shows a follow-up menu of label categories, labels, feature categories, and features. When the label and feature have been selected, it shows the related statistics in box plots to compare metrics between different scenarios.

Figure B.14: Baseline analysis update

B.5. Iteration 3

After meeting with more stakeholders (Appendix A.5, A.6, and A.7), another iteration phase started, improving the prototype by meeting stakeholders' needs. These adjustments mainly focused on the completeness of all included labels.

B.5.1. Key adjustments

Custom labels

A '+ label' button has been added to the logger page. Triggers and experiences vary between patients. Allowing custom labels enables patients to track personally relevant events that are not included in the predefined categories and labels. This improves the completeness and personalisation of the data set.

Notes for labels

Users can now add notes to an entry of a label (Figure B.15). This allows documentation of details that may help interpret patterns or explain irregular physiological behaviour when labels alone do not capture relevant contextual information.

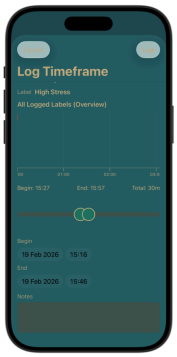


Figure B.15: Notes for labels

Multi-labels

Multiple labels can be assigned when determining a timeframe. This mitigates the step of needing to add more single labels for the same time period. When complex situations occur, they cannot be captured with a single label.

Favourite labels

Users who frequently journal the same labels, and are not interested in other predefined labels, can now prioritise these labels by marking them as 'favourite'. A new category button was added to the logger page. The labels can be accessed more quickly, reducing effort, interaction time and cognitive load when trying to find the correct label.

Extended baseline analysis

A scatter-dot plot is added to visualise relationships between selected labels and up to two physiological features. This enhances interpretation and supports identifying correlations between triggers and physiological behaviour.

Watch interface
The app has been extended with a coherent Apple watch app, which includes the logger, analysis, and medication page.

'+ label' button
This enables users to add their own labels, as patients can best determine which other, currently unavailable events are related to their epilepsy (Figure B.16).

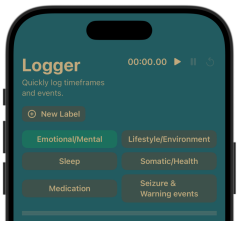


Figure B.16: New label button

Adjustable medication effect time
The timer of the medication graph has been made adjustable to suit a patient's medication routine (Figure B.17).

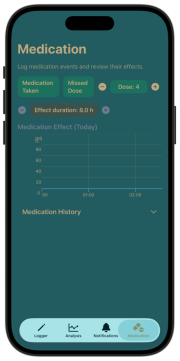


Figure B.17: Adjusting the effect time of the medication influences the graph

'Favourite' labels
A new label category has been added called 'Favourites', where users can add different label buttons to this category to access more commonly logged labels faster.

Repeating inactivity reminder
After an inactivity reminder has been sent, and the user does not perform any journal activities, no other reminders are sent. Therefore, from now on the determined interval of inactivity determined by the user will be repeated instead of sent once.

Logging double labels for a single determined timeframe
After pressing a label button, when determining the timeframe, it is now possible to add other labels to the same timeframe.

Baseline analysis section update

The baseline analysis has been extended with a scatter-dot plot visualising selected labels and features through the doughnut-chart in different coloured dots to be able to compare feature values between labels (Figure B.18).

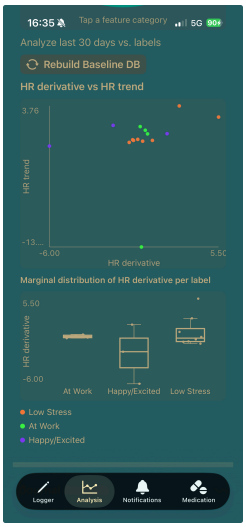


Figure B.18: A scatter-dot chart enables comparison of physiological features between different labels.

B.6. Iteration 4 - pilot study

During the pilot, it was noticed that a lot of time uncovered with labels is during sleep, as it is also hard as a user to guess at what time you actually fell asleep. Luckily, with the sleep features derived from the Apple Watch, it is possible to trace back when you fell asleep and woke up. This is when I thought of a suggestive sleep label. When the data related to sleep has been uploaded from the watch to the iPhone app, it is visualised in the sleep phases feature graph on the analysis page. When the graph does not hit the 'awake' value on the y-axis, it adds a label with striped edges to the all logged labels timeline, with the suggestion 'sleep?' stated on its visualised box. When tapping this label on the timeline, the pop-up window asks the user to either confirm the period of sleep or adjust the timeframe if they think the suggested label is inaccurate. This step enables immediate coverage of a large amount of time during a day. The number of hours of sleep has been visualised in a bar chart and added to the feature graphs.

Whenever the Apple Watch measures an state where the user is not sleeping but awake, the suggested sleep label in the timeline is interrupted by a 'awake' label, also visualised in a striped box (Figure B.19). The reason for the suggested awake-label has originated from the fact that seizures can also occur during the night, and when the watch measures an active/awake state, it is possible to trace back how all the other features have behaved during this period.

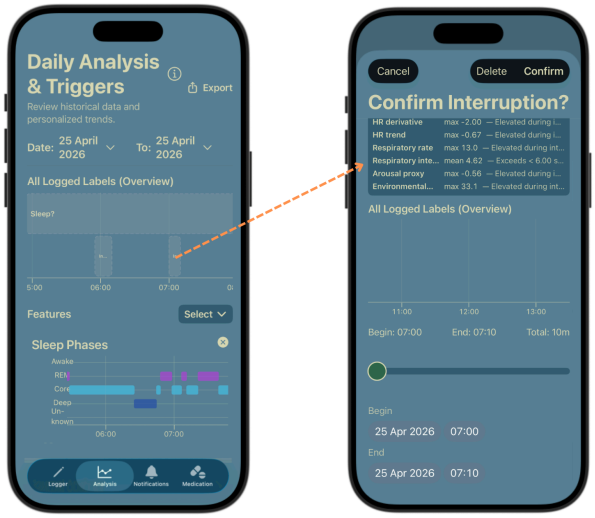


Figure B.19: During sleep, and during irregular physiological behaviour when sleeping, a 'sleep?' or 'interruption' label will appear, displaying the relevant features' values. The user is able to be confirm or delete the label.

Response time

Whenever a user receives a notification at the end of the day, and no label is logged after, the response time includes the time overnight in its calculation. This resulted in a response time of 1490 minutes, which is incorrect. Therefore, the calculation of the response time has been adjusted to exclude overnight time, by checking if a user has journaled any label after the last notification of the day. After 5:00 of the next day, if there has not been any new entries, this will be indicated with 'Response in the evening' at the journal analysis section on the analysis page. This value is presented as 'no response' or 'responded'. The time between the last nudge of a day and the first journaled label of the next day is excluded from the response time.

Next, when a notification is sent, it is saved in the notification history section. But, when a user accesses the app by clicking the push-notification, it is not saved on the app, only when manually accessing the app through an iPhone.

Theme colours

Colours of the background, buttons, and text have been adjusted to enhance readability (Figure B.20)

Scatter dot chart

The scatterdot has been extended by enabling 3 features to be selected for analysis at the same time, transforming the chart to a 3D plot (Figure B.21).

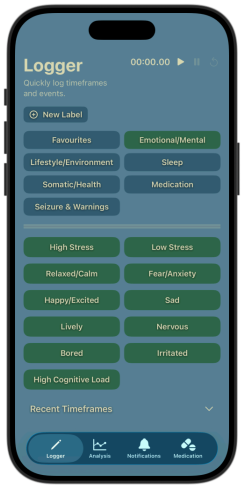


Figure B.20: New colour theme

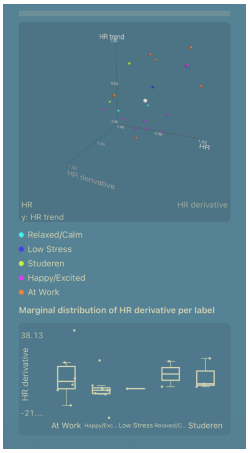


Figure B.21: 3D scatterdot plot

B.7. Iteration 5

Nudging strategy

From the results found in Chapter 5.4, the nudging strategy that has been most promising was the personalised strategy.

Export files update

Multiple options have been provided to be able to export all data, only data from logged labels, or select a specific set of labels for which data is exported.

Custom categories

To spread the total amount of labels per category and to preserve a clear overview of all labels, custom categories have been enabled.

Link to personal agenda

Just-in-time (JIT) notifications were found in the literature research (Chapter 3.4) as a fourth option to nudge users to log any events. An approach to achieve this option is to persist the app to access the user's calendar app. This allows the developed journal app to send notifications after a predetermined amount of minutes after a planned event in the calendar app has ended. Users must first allow the app to access their calendar.

+ buttons

currently, after adding a label or category, the timeframe determination window pops up, as if a user wants to journal the new label. This is an unnecessary step, the user must be able to add labels/categories without the obligation to immediately determine and log a timeframe.

Information icons

At the top of each page, besides the title, there is an 'i'-information button added with a small elaboration of each attribute of the corresponding page.

B.7.1. Improving clinical value

Baseline and shaded standard deviation graphs

The feature graphs are now also included with an orange baseline graph as a 'normal' indication of the user's physiological condition during the day. To be able to analyse the current state of the user compared to the baseline, the standard deviation of the baseline graph is rendered in a shaded orange colour.

Standard deviation warning

Looking at the baseline graph, we can determine when a user has entered a state crossing the standard deviation, and thus also an irregular state.

Custom medication

When patients test different types of medication, previously it was only possible to adjust the dose in an amount of pills. Now, users can manually enter multiple specific types of medication and the dose in milligrams, which is visualised in a bar chart for the last 14 days to track medication usage more precisely. The half-life of medication can also be registered for a more realistic graph of medication effects.

Risk algorithm

Pearson/Spearman correlations, Pc correlation-variance, and a logistic regression model have been added. When journaling an epileptic related label it will be analysed and all correlations are calculated. Next, the data of the past 30 minutes is measured and compared to the historical logged epileptic-events.

Canonical correlation feature graph

The default options to combine features are replaced with a single main graph where pairs of features can be displayed at the same time, representing their output using a canonical correlation feature from literature research (Chapter 3.3).

Additional Libre-pad

One of the patients has a Libre-pad to measure blood sugar levels (which also triggers seizures), which could also be integrated in the app using its Bluetooth connection.

B.7.2. Aligning with clinicians' needs

After evaluating the app with multiple clinicians, it was confirmed that during a treatment, the first things a clinician is interested in are the patient's seizure frequency and medication usage. Therefore, the 'Seizure events' and 'Medication dose' charts have been moved to a more prominent location on the analysis page, for instant accessibility for clinicians to the most relevant data of their patients during a treatment.

B.8. Repository

B.8.1. Mobile application files from Xcode

This Github repository contains the full Xcode project of the developed application:

- *https://github.com/dirkcrlagemaat/Graduation*

A Readme file is available for further explanation of the usage and structure of the application.

B.8.2. Web application files

Preliminary to the first prototype, the development of a working connection between iPhone and webserver was explored using a websocket server. The python and index files used for this experimental phase are located in the following Github repository:

- *https://github.com/dirkcrlagemaat/Experimental*

C

Appendix C - Research study

C.1. Interview questions pilot study

Visual design

How do you experience the visual design (colours, layout, simplicity)? Is the hierarchy of the app's components and the navigation menu easy to understand? What stands out positively? What appeals to you less?

Navigation

When you look at the different pages (Logger, Analysis, Medication, Notifications), does the structure seem clear and logical to you? Why or why not?

Complexity

Does the app seem simple or complex? Which elements contribute to that feeling?

Learnability

How easy do you think it would be to learn how to use this app? What might be confusing at first?

Accessibility of features

Do you think you would be able to easily find features such as logging events, viewing timelines, or checking medication schedules? Why or why not?

Cognitive load

Does the design seem to limit mental effort (for example, through clear buttons, visual separations, suggested labels)? Why do you think that?

Timeframe determination

Does it seem clear how to determine or adjust the timestamps of an event? What might make this difficult?

Completeness of labels

When you look at the available labels (e.g., sleep, medication, events), do you feel that important situations are missing? Which specific labels, scenarios, or types of data (e.g., heart rate, respiratory rate, etc.) would you find most interesting to track?

Interpretation of data

The app displays physiological data (sleep, heart rate, HRV, stress, movement). How easy are the graphs and visualisations to understand? Which functionalities would you remove or add to this page? What makes the analysis page (un)clear?

Interpretation of charts

Do the timelines and combined charts seem useful for recognising patterns? Why or why not? Which form of data visualisation would be clearer or easier to understand/interpret?

C.2. Statistical results research study

H1a: The nudging method significantly affects the number of journaled labels.

Descriptive Statistics

	Mean	Std. Deviation	N
Labels_fixed	4,9514	3,90425	7
Labels_inactivity	6,0486	2,23901	7
Labels_personalized	6,8924	3,01223	7

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Nudge	,731	1,568	2	,456	,788	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Nudge

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Pairwise Comparisons

Measure: MEASURE_1

(I) Nudge	(J) Nudge	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-1,097	1,163	1,000	-4,922	2,727
	3	-1,941	1,410	,653	-6,576	2,694
2	1	1,097	1,163	1,000	-2,727	4,922
	3	-,844	,871	1,000	-3,709	2,021
3	1	1,941	1,410	,653	-2,694	6,576
	2	,844	,871	1,000	-2,021	3,709

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Nudge	Sphericity Assumed	13,261	2	6,631	1,386	,287	,188
	Greenhouse-Geisser	13,261	1,576	8,416	1,386	,289	,188
	Huynh-Feldt	13,261	2,000	6,631	1,386	,287	,188
	Lower-bound	13,261	1,000	13,261	1,386	,284	,188
Error(Nudge)	Sphericity Assumed	57,406	12	4,784			
	Greenhouse-Geisser	57,406	9,454	6,072			
	Huynh-Feldt	57,406	12,000	4,784			
	Lower-bound	57,406	6,000	9,568			

The descriptive statistics indicated that the personalised nudging condition yielded the highest mean number of daily journaled labels, followed by the fixed and inactivity conditions. The full descriptive values per condition and per participant are presented in the table above. Mauchly's test of sphericity was not significant, indicating that the assumption of sphericity was not violated; therefore, sphericity-assumed results were used for interpretation. A repeated-measures ANOVA revealed no statistically significant effect of nudging condition on the number of journaled labels, $F(2,12) = 1.386$, $p = 0,287$. The partial eta squared value of $\eta^2 = 0,188$ indicates a medium-to-large effect size, suggesting that approximately 18,8 percent of the variance in the number of journaled labels may be attributable to the nudging strategy. However, given the small sample size (N = 7) and the associated low statistical power, this effect size estimate is likely upwardly biased. Pairwise comparisons with Bonferroni correction revealed no significant differences between any pair of conditions. The null hypothesis H1a cannot be rejected on the basis of this analysis."

H1b: The nudging method significantly affects the total time coverage of journaled labels.

Descriptive Statistics

	Mean	Std. Deviation	N
Coveredtime_fixed	671,0957	490,70440	7
Coveredtime_inactivity	730,7629	362,05673	7
Coveredtime_personalized	740,2729	427,85483	7

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Lower-bound
Nudge	,634	2,276	2	,321	,732	,910	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Nudge

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Pairwise Comparisons

Measure: MEASURE_1

(I) Nudge	(J) Nudge	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-59,667	143,349	1,000	-530,920	411,586
	3	-69,177	83,164	1,000	-342,575	204,221
2	1	59,667	143,349	1,000	-411,586	530,920
	3	-9,510	109,553	1,000	-369,660	350,640
3	1	69,177	83,164	1,000	-204,221	342,575
	2	9,510	109,553	1,000	-350,640	369,660

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Nudge	Sphericity Assumed	19684,199	2	9842,099	,214	,811	,034
	Greenhouse-Geisser	19684,199	1,465	13440,702	,214	,745	,034
	Huynh-Feldt	19684,199	1,819	10819,320	,214	,791	,034
	Lower-bound	19684,199	1,000	19684,199	,214	,660	,034
Error(Nudge)	Sphericity Assumed	552537,798	12	46044,817			
	Greenhouse-Geisser	552537,798	8,787	62880,349			
	Huynh-Feldt	552537,798	10,916	50616,598			
	Lower-bound	552537,798	6,000	92089,633			

Descriptive statistics indicate that the personalised nudging strategy yielded the highest mean total time covered with labels per day, though differences between the nudging strategies were small. Mauchly's test of sphericity was not significant, confirming that the sphericity assumption was not violated; sphericity-assumed results were therefore used. A repeated-measures ANOVA indicated no statistically significant effect of nudging condition on total time coverage, $F(2, 12) = 0{,}214$, $p = 0{,}811$. The partial eta squared value of $\eta^2 = 0{,}034$ indicates a small effect, suggesting that approximately 3,4 percent of the variance in temporal coverage may be attributable to the nudging condition. This is the smallest effect observed across all three compliance measures, indicating that temporal coverage is the least sensitive of the three outcome measures to variation in nudging strategy. Pairwise comparisons with Bonferroni correction confirmed no significant differences between any pair of conditions. The null hypothesis H1b cannot be rejected based on this analysis.

H1c: The nudging method significantly affects the response latency after a nudge.

Descriptive Statistics

	Mean	Std. Deviation	N
Latency_fixed	171,6200	140,97067	7
Latency_inactivity	185,3814	188,81334	7
Latency_personalized	79,1900	82,62699	7

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Lower-bound
Nudge	,711	1,706	2	,426	,776	,996	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Nudge

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Pairwise Comparisons

Measure: MEASURE_1

(I) Nudge	(J) Nudge	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-13,761	74,513	1,000	-258,721	231,198
	3	92,430	44,328	,246	-53,297	238,157
2	1	13,761	74,513	1,000	-231,198	258,721
	3	106,191	64,129	,446	-104,630	317,013
3	1	-92,430	44,328	,246	-238,157	53,297
	2	-106,191	64,129	,446	-317,013	104,630

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Nudge	Sphericity Assumed	46688,370	2	23344,185	1,721	,220	,223
	Greenhouse-Geisser	46688,370	1,551	30093,851	1,721	,230	,223
	Huynh-Feldt	46688,370	1,992	23442,119	1,721	,220	,223
	Lower-bound	46688,370	1,000	46688,370	1,721	,238	,223
Error(Nudge)	Sphericity Assumed	162816,693	12	13568,058			
	Greenhouse-Geisser	162816,693	9,309	17491,084			
	Huynh-Feldt	162816,693	11,950	13624,979			
	Lower-bound	162816,693	6,000	27136,116			

Descriptive statistics indicated that the personalised nudging condition was associated with the lowest mean response latency, with values less than half of those observed in the fixed and inactivity conditions. This pattern was the most numerically pronounced difference observed across all three compliance measures. Mauchly's test of sphericity was not significant; sphericity-assumed results were therefore used for interpretation. A repeated-measures ANOVA revealed no statistically significant effect of nudging condition on response latency, F(2,12) = 1,721, p = 0,220. The partial eta squared value of η^2 = 0,223 indicates a large effect size, suggesting that approximately 22,3 percent of the variance in response latency may be attributable to the nudging condition. As with H1a, this estimate is likely inflated given the small sample size and must be treated as an exploratory indicator rather than a reliable population-level effect. Pairwise comparisons with Bonferroni correction revealed no significant differences between conditions. The null hypothesis H1c cannot be rejected; however, the large descriptive difference in latency between personalised and non-personalised conditions represents the most practically meaningful trend observed in the results.

H1d: The nudging method significantly affects the perceived user experience. The corresponding custom SUS questionnaire can be found in Appendix C.3.

Descriptive Statistics

	N	Mean	Std. Deviation
Q1	7	4,29	,756
Q2	7	4,29	,756
Q3	7	4,14	,690
Q4	7	4,43	,535
Q5	7	4,29	1,254
Q6	7	2,71	1,380
Q7	7	4,14	1,069
Q8	7	4,43	,535
Q9	7	1,86	1,069
Q10	7	1,86	,690
Q11	7	3,86	1,069
Q12	7	3,86	1,215
Q13	7	4,71	,488
Q14	7	4,00	,816
Q15	7	2,71	1,254
Q16	7	1,29	,488
Q17	7	2,14	1,464
Q18	7	2,00	1,414
Q19	7	3,14	1,069
Q20	7	2,14	,900
Q21	7	3,86	1,345
Q22	7	3,14	1,345
Q23	7	3,71	,756
Q24	7	4,29	1,113
Q25	7	2,14	1,464
Q26	7	3,71	,951
Q27	7	5,00	,000
Q28	7	4,00	1,414
Q29	7	4,71	,488
Valid N (listwise)	7		

Figure C.1: Mean value of and standard deviation of the SUS questionnaire.

C.3. Custom SUS questionnaire research study

Instructions: Please indicate how much you agree or disagree with the following statements based on your experience using the app.
Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree

#	Statement	1	2	3	4	5
1	I thought the menu was easy and clear to use.	☐	☐	☐	☐	☐
2	It was easy to find something on the app.	☐	☐	☐	☐	☐
3	The UI hierarchy was clear.	☐	☐	☐	☐	☐
4	The app's use-cues were very clear.	☐	☐	☐	☐	☐
5	The expandable tabs correctly prevented too much information from being shown at once.	☐	☐	☐	☐	☐
6	I needed to scroll a lot during usage of the app.	☐	☐	☐	☐	☐
7	I found the app's theme colours pleasant.	☐	☐	☐	☐	☐
8	The transitions between pages or moving parts of the app were pleasant and not disturbing.	☐	☐	☐	☐	☐
9	It was hard to complete a journal task.	☐	☐	☐	☐	☐
10	It requires too many steps or tasks to properly use this app.	☐	☐	☐	☐	☐
11	The fixed nudges were effective.	☐	☐	☐	☐	☐
12	The inactivity nudges were effective.	☐	☐	☐	☐	☐
13	The personalised nudges were effective.	☐	☐	☐	☐	☐
14	The nudges stimulated me to journal data quickly.	☐	☐	☐	☐	☐
15	I have missed or overlooked the nudge notifications frequently.	☐	☐	☐	☐	☐
16	The nudges worked counter-effectively.	☐	☐	☐	☐	☐
17	I found the nudges annoying or frustrating.	☐	☐	☐	☐	☐
18	I have adjusted the fixed reminders multiple times.	☐	☐	☐	☐	☐
19	I found the stopwatch function useful and effective.	☐	☐	☐	☐	☐
20	I have used the stopwatch to determine a timeframe more precisely.	☐	☐	☐	☐	☐
21	I was able to determine the timeframes very precisely.	☐	☐	☐	☐	☐
22	It was easy to determine a label timeframe.	☐	☐	☐	☐	☐
23	The slider was a good tool to determine a timeframe.	☐	☐	☐	☐	☐
24	The scrollable digital clock was a good tool to determine a timeframe.	☐	☐	☐	☐	☐
25	I have frequently used the double-tap function to label events.	☐	☐	☐	☐	☐
26	The app included all labels of scenarios I encountered during the day.	☐	☐	☐	☐	☐
27	I would not have a problem sharing my physiological smartwatch data to enable further analysis.	☐	☐	☐	☐	☐
28	I would be more motivated to use the app if I had a specific goal I want to track.	☐	☐	☐	☐	☐
29	I would be interested in further physiological-related analysis, such as graphs, thresholds, and a personalised baseline model.	☐	☐	☐	☐	☐

C.4. Semi-structured interview

1. How do you feel about the visual design (colours, layout, simplicity)? Is the hierarchy of the app's components and the navigation menu easy to understand? What stands out positively? What feels less appealing?
2. Navigation Clarity: Looking at the different pages (Logger, Analysis, Medication, Notifications), does the structure seem clear and logical to you? Why?
3. Complexity: Does the app appear simple or complex? What elements contribute to that feeling?
4. Learnability: How easy do you think it would be to learn how to use this app? What might be confusing at first?
5. Accessibility of Functions: Do you think you would be able to easily find functions like logging events, reviewing timelines, or checking medication schedules? Why?
6. Cognitive Effort: Does the design seem to minimise mental effort (e.g., clear buttons, visual dividers, suggested labels)? What makes you think so?
7. Logging Events: Looking at the logging interface (buttons, stopwatch, snapshot labels), how clear is it how to log an event?
8. Time Registration: Does it seem clear how you would determine or adjust the time of an event? What might make this difficult?
9. Snapshot / Quick Logging: The app includes quick "snapshot" labels for urgent moments. Does this look helpful in stressful or seizure-related situations? Why or why not?
10. Completeness of Labels: Looking at the available labels (e.g., sleep, medication, events), do you feel important situations might be missing? What specific labels/scenarios/type of data (Heart rate, respiratory rate, etc.) would you be most interested in to track?
11. Understanding Physiological Data: The app displays physiological data (sleep, heart rate, HRV, stress, motion). How easy do the graphs and visualisations appear to understand? What functionalities would you delete/add to this page? What makes the analysis page (un)clear?
12. Graph Interpretation: Do the timelines and combined graphs appear helpful for identifying patterns? Why or why not? What type of visualisation of the data would be more clear/easier to understand/interpret.
13. Data Confidence and Trust: Based on the design and integration with Apple Watch, would you trust the accuracy of the collected and displayed data? What influences your level of trust?
14. Fit Within Daily Routine: Looking at the design, do you think this app could fit into your daily routine? Which moments during the day would you be able to journal events or scenarios that have happened that day?
15. Reviewing Past Events: Does the app appear to make it easy to review previous events and understand patterns over time? What would make this easier?
16. Personalisation: The app allows adjustments such as selecting labels, adjusting notifications, and hiding graphs (based on description). How important is personalisation to you in an epilepsy app? What would you want to customise? How do

you think you would feel about letting an ML model access all your physiological data to be able to provide personal insights?

17. Overall Value and Improvements: Do you think this app could support better understanding of your epilepsy? What improvements would you suggest before using it in real life?

Questions focused on nudge strategy and journal compliance

1. Can you recall a moment where you were frustrated due to a nudge? Which nudge?
2. Did you have a hard time trying to remember to journal any label?
3. Did you have a hard time remembering what (time) events had occurred earlier that day?
4. Indicate how accurate you think your determined timestamps are compared to reality (percentage).
5. Did you adjust the moments of the fixed reminders? Which times did you choose? Was it related to a certain scenario/event?
6. When journaling labels independently, without being reminded by a nudge, at what moments did you journal events on the app?
7. How often (percentage) did you actually act when seeing the notifications?
8. What times during the day are hard to journal?
9. How often did you lack any memory to journal labels accurately?
10. What type of notifications do you prefer? why?
11. Which type of notifications do you think is most effective? Why?
12. what labels did you miss during the usage? What type of scenarios/situations have happened that were not available in the label options, but are relevant to track?

C.5. Results semi-structured interview

C.5.1. Participant 1

Interface

Visually clear, consistent colour contrast.

But on first glance, due to the many attributes, it can look complex while in reality it is quite simple.

Some built-in label buttons were irrelevant for this participant to track.

Added a lot of own label buttons.

Seemed off that the label timeline was not located on the logger page.

Suggestion of 'remembering' what features the user has selected, so only the features that the user is interested in are permanently visualised on the analysis page.

Interpretability

Without any background knowledge, the feature names are unclear, such as RMSSD or standard deviation, which increases the uncertainty about what type of physiological behaviour is 'normal'.

Providing contextual elaboration increases motivation. The interpretability strongly impacts user engagement.

Perceived accuracy

High perceived accuracy of 95 percent. However, when planning activities beforehand, it can be hard to determine the timestamp at the end of the period. Therefore, backwards journaling was experienced as more accurate.

Some peaks in physiological data seemed off, but were mostly related to the low fetching resolution from HealthKit towards the application.

When being uncertain about the exact timestamp of an event, the physiological sensor data is trusted more than the participant's own memory when journaling afterwards.

Daily routine and nudging preferences

Best moments to use the journal application were during public transportation, breaks or just before or after an activity.

In the evening, the journaling behaviour was experienced as most challenging due to the battery life of both iPhone and Apple Watch, or during transitions between different activities, where the participant was uncertain when exactly one event stopped and the next started.

The fixed nudges were perceived as most disturbing, especially when they are sent immediately after the participant has journaled a label. The inability to adjust the moments causes light frustration.

A more context-aware nudging strategy, such as the inactivity nudge, was perceived as more useful and pleasant.

Suggestion to link nudges to personal calendar, so planned events are automatically synchronised to the journal application.

When having something to track, it motivates and supports consistent journaling. This difference was noted after day 2 of the study.

Clinical relevance

The advantages are clear; however, without proper elaboration and low interpretability, more data does not increase the value, as it stays unclear whether the physiological changes are normal.

A personal baseline would improve the interpretability to indicate normal/irregular behaviour.

C.5.2. Participant 2

Interface

A lot of information on the analysis page does not seem immediately necessary for each patient, but the tab menu preserves a clear overview of the analysis page.

Scrolling back in time when determining timestamps of a label can take a long time when journaling afterwards.

Interpretability

Some text on the analysis page is displayed in a small font, which was hard to read. An additional elaboration, further explaining the functionalities of each page, was missing.

Perceived accuracy

Without any indication, it is hard to trace back the exact starting moment of a seizure.

Daily routine and nudging preferences

The participant noted that sleeping with a wearable was experienced as counterproductive, as falling asleep with a wearable on the wrist is harder and distracted the participant from sleeping.

The participant stated that having a specific goal (tracking stability in this case) ensured engagement.

Clinical relevance

The seizure overview was perceived as a big advantage.

A simple addition could be made to indicate the severity of seizures.

No important labels were missing.

For future development, linking the app to a Libre Pad (which measures blood sugar levels) can extend the types of data integrated into the application.

Medication doses can differ in a short time; an overview of medication usage could provide a clear overview over a longer period of time.

C.5.3. Participant 3

Interface

The hierarchy was clear, no comments about the usability; it was pretty intuitive.

The navigation bar made it easy to understand. The medication page has not been used by the participant.

The first time using the app, the number of buttons can be overwhelming; however, after one time usage, all functionalities are clear after knowing which labels you want to track. It is easy to learn to use.

The theme colours minimise the cognitive load by their soft contrast. The minimal amount of text also helps to quickly scan all attributes of the interface.

The timeline provides an accessible and clear overview.

The time wheels to determine the timeframe of a label were lagging, as they returned to their previous value instead of the time the user has determined.

Interpretability

No important labels were missing. However, it would be useful to track a user's weight.

The graphs were easy to interpret.

An overview of the past week/14 days per feature instead of only the current day.

Perceived accuracy

After 4-5 hours, the perceived accuracy of the determined timestamps decreases. The accuracy was perceived with an accuracy between 90 and 95 percent. planned activities were perceived as easily determinable; however, emotional/mental events are more difficult to remember the timestamps precisely.

It was not hard to remember the events previously that day. It was recognisable that when journaling more frequently during the day it was easier to remember and accurately determine the timeframe.

Daily routine and nudging preferences

The mental effort to journal events decreases when using the app for a longer time, because using it needs to be integrated into the user's routine.

During breaks of after work in the evening were the most suitable moments to journal events.

Nudges did not cause any negative effect or frustration.

When a nudge was overlooked, it was hard to remember to journal a label. It is hard to remember to complete a task during work. The need for an effective nudging strategy to collect data was confirmed.

The fixed times were adjusted to 12:00, 16:00 and 20:00, times that the participant knew beforehand that there was time to journal events.

Around 40 percent of the nudges were perceived as overlooked. An explanation could be the fact that notifications were disabled during working hours.

The personalised nudges were perceived as most effective and preferred over inactivity or pre-fixed nudges. The participant still used fixed nudges, but aligned the times with their personal routine.

Clinical relevance

Tracking weight would be a good addition. The built-in labels are good, but maybe it would be good to display them as an example, before initial usage, then ask the user what their goal is and consult their clinician to decide which labels and physiological features are interesting to track to eliminate irrelevant buttons/features/information.

Training an ML model was perceived as the obvious follow-up step. However, an elaboration of how the ML model is being trained and how it is a must to prevent a 'black-box' form happening, where the users do not know what is happening with their data.

Automating tasks improves the total usability of the app. Also, the Apple Watch interface was perceived as a very useful addition, as it increases accessibility and lowers the threshold to journal events.

C.5.4. Participant 4

Interface

The application was perceived as intuitive and user-friendly. All required information was visible at once and was perceived as well-organised.

Added a lot of own labels. Also highlighted the fact that when setting a personal goal, it can result in more engagement.

However, there were some interface hurdles. The timewheels did not always work correctly when determining a timeframe, automatically returning to the previously determined time, and the purpose of the stopwatch was not clear.

The suggested 'interruption' labels that appear when irregular physiological behaviour is detected during sleep are also activated immediately after waking up, which is incorrect and needs to be fixed.

The ability to track medication is very useful, but a clear visual of a temporal overview is lacking.

Interpretability

There was a lack of elaboration on each feature graph's functionality. Participants know what the type of feature means, but not always why it is being measured, indicating a need for further elaboration to increase interpretability and usability.

Another noticeable and useful comment related to the medication-level graph, which does not account for half-life, and different doses of multiple types of medication.

Perceived accuracy

The accuracy of the determined timestamps of the logged labels was around 95 percent, with an uncertainty of 5 percent.

However, at the current state, the medication level graph seems too unrealistic, missing the ability to add further details of medication type.

Journaling afterwards was perceived as effective, due to more time to reconstruct the situations and their transitions.

Daily routine and nudging preferences

The best journaling moments were in the afternoon, when things had calmed down.

Directly after an event was not always perceived as useful, as there is not always time during the varying transitions between scenarios.

Nudges did not lead to any frustration and were not perceived as a burden. There was no significant difference observed between the types of nudges.

Regarding the timing of the nudges, the change of overlooking a nudge was the highest in the morning, due to a busy schedule. In the evening, the response rate decreased significantly. However, the fixed nudge at 22:30 was perceived as too late, causing less collected subjective data.

The personalised combination of fixed and inactivity nudges was most effective, with fixed reminders at 8:00, 17:00, 20:00 and 21:30, and after an inactivity of 3 hours.

The participant is interested in adding more features, such as blood-sugar levels, through a LibrePad.

Clinical relevance

Enabling an overview of all relevant data was perceived as very valuable.

The biggest limitation was the interpretability of multiple feature graphs. Also, a more detailed medication level graph and further data on medication usage were missing.

Providing types of goals for users to track could increase user engagement by only displaying data or features related to the predetermined goal.

C.5.5. Participant 5

Interface

The application was perceived as clear and easy to use, with no major usability issues or frustrations. The tab menu was experienced as efficient and fast.

However, some interface improvements were suggested. Adding more distinct visual sections or outlines per page could improve structure and readability. The purpose of the stopwatch was not immediately clear, and adding a small information icon was suggested to clarify its functionality.

It was also suggested to automatically display the ”favourite” label category buttons

upon opening, allowing quicker access to more commonly logged labels. Snapshot labels were not used at all.

Interpretability

There was a clear need for further elaboration on the feature graphs. Even though the advantages were acknowledged, the lack of explanation regarding what the features represent and why they are relevant limited interpretability.

It would also be useful if, after irregular physiological behaviour, a suggestion would be generated explaining the correlation to previous events.

Perceived accuracy

There were no concerns regarding the application's system accuracy or physiological data.

However, the accuracy of manually logged data depends on recall. While it was generally easy to reconstruct timestamps of past events, spontaneous events and transitions between activities were more difficult to remember precisely.

Journaling afterwards, particularly later in the day, was perceived as more reliable due to having more time to reflect and reconstruct events.

Daily routine and nudging preferences

There was no clear structure in journaling behaviour when nudges were absent; logging occurred at random moments throughout the day. The most suitable moments for journaling were during breaks or after work, when there was sufficient time and mental space.

Nudges themselves did not cause frustration. However, their effectiveness was limited by visibility. A large proportion (approximately 75 percent) of nudges were overlooked, mainly because notifications on the phone were often disabled. Nudges delivered via the Apple Watch were more effective and more frequently noticed.

Personalised nudging was preferred, particularly when aligned with the user's schedule. The combination of fixed and inactivity-based nudges was perceived as most effective. Additionally, integrating nudges with the Apple Calendar was suggested to improve timing and relevance, ensuring nudges occur at moments with a higher likelihood of immediate journaling

Clinical relevance

The clinical value of the application was mainly linked to improving interpretability. While the data itself was not questioned, its usefulness depends on providing sufficient context and insight.

A key opportunity lies in linking physiological changes to previously journaled events, enabling users to better understand cause-and-effect relationships. This would enhance the relevance of the data for both personal insight and potential clinical use.

C.5.6. Participant 6

Interface

The navigation menu at the bottom is easy to understand and clear. The overall structure helped to separate sections effectively, and tasks required minimal effort due to the low number of steps.

Some minor issues were identified. The colours were perceived as slightly dark, and when adding notes in the 'timeframe determination window' after pressing a button on the logger page, it was not possible to scroll downwards to properly review the text.

Also, the purpose of some feature graphs were not immediately obvious at first glance. Snapshot logging was understood but not found necessary during the 9-day study.

Interpretability

Basic features such as heart rate, sleep, and steps were clear and understandable. However, the exact function and meaning of more complex feature graphs were not always immediately clear, especially in the beginning.

Overall, the app provides a clear overview of tracked data, but an additional explanation of feature graphs could improve interpretability and user understanding.

Perceived accuracy

The data was perceived as highly reliable, mainly because it is sourced from the Apple Watch, which the user trusts.

The timestamps were generally perceived as accurate, mostly due to the daily routine. For longer events, an estimated maximum deviation of 10 minutes was reported. But overall, recalling events was not considered difficult.

Daily routine and nudging preferences

The app fit well within the daily routine, especially because the participant could decide what to track. Journaling was often done when already using the Apple Watch, making quick logging convenient.

Nudges were not perceived as frustrating and were helpful, particularly on the Apple Watch, where they enabled immediate action. Around 70 percent of nudges were followed, while others were postponed due to meetings or scheduled activities.

Personalised nudges were preferred over the fixed or inactivity nudging strategies. Mostly because the personalised nudges align better with the participant's schedule, which increases the likelihood of immediate logging. Personalised nudges were planned around breaks or after work.

Clinical relevance

The application was perceived as valuable due to the ability to track multiple aspects of daily life in one clear overview, providing useful insights.

Trust in the data was high, even though concerns were raised about privacy when using personal data for machine learning. The addition of features such as food intake tracking could further increase relevance.

C.5.7. Participant 7

Interface

The interface was perceived as very clear, with an intuitive navigation menu to quickly access the different pages.

The label categories helped to quickly access the labels that the participant wanted to journal. Sections were clear and the tab menu supported keeping an overview of the attributes the participant was interested in, decreasing cognitive load by excluding unnecessary information.

The colour theme supported the aesthetics of the interface and enhanced readability.

Interpretability

The graphs make it easy to identify peaks. But not all features' relationship to epileptic events was directly clear.

The medication level graph made it easier to track when a new dose intake was necessary and provided more insight into personal medication usage.

One suggestion was made to integrate the logged events into features' graphs' to be able to directly analyse the consequences of the logged events per physiological feature.

Perceived accuracy

Sometimes the behaviour of the physiological graphs on the analysis page seemed off, due to certain peaks. However, there were no concerns related to the accuracy of the wearable data.

The accuracy of the journaled events was high, around 95 percent. Journaling afterwards lowered the perceived accuracy when an event had taken place more than 3-4 hours before journaling the event.

Daily routine and nudging preferences

When being nudged during busy events, journaling activities were postponed until a more suitable moment occurred later that day.

The personalised nudges were preferred over the fixed or inactivity nudges. The participant noticed that the response latency decreased when being nudged at preplanned breaks or moments of spare time.

Clinical relevance

The medication effect graph could be improved by adding more detail related to the specific type of medication. This would enable a more realistic graph, which degrades

over time related to the medication, instead of only the advised number of hours between medication intakes.

A deeper elaboration of each feature graph's relation to epileptic events and how they can help analyse triggers would be useful.

The ability to adjust the type of label buttons improved the total completeness of the available labels to track, as participants can personally choose what they would like to track.

The ability to compare physiological behaviour with journaled scenarios was perceived as very useful.

D

Appendix D - Required documents

D.1. Project Brief

Project Proposal form
IDE Master Graduation Project

In this proposal the agreements made between student and supervisory team about the student's IDE Master Graduation Project are set out. This document needs to be prepared for the Kick-off meeting and should be submitted in MyCase.

Name student Dirk Christiaan Radjan Lagemaat **Student number** 5099692

Project title Designing a personalized wearable-connected journaling system for seizure in

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

MSc programme ☐ Design for Interaction ■ Integrated Product Design ☐ Strategic Product Design

☐ Other *(in case of a double degree outside IDE):* _____

Introduction

Describe the context of your project in the box below; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

Epilepsy affects approximately 65 million people worldwide, and despite advances in medication and monitoring, seizure unpredictability remains a major burden for patients (11).This study aims to develop a lightweight, transparent, continuous journaling system enabling patients to capture triggers and factors influencing their daily lives. Physiological data are combined with journaling to provide valuable insights, covering activities, environmental contexts, and emotional or psychological states. This approach requires understanding of data extraction, visualization, analysis, and feedback mechanisms through journaling.
Key stakeholders include epilepsy patients, clinicians, and researchers. The project empowers patients to actively contribute to seizure forecasting through adaptive self-reporting, enhancing engagement and perceived control. Journaling contextual and subjective stimuli enables integration of diverse information types. For clinicians, it offers richer context to distinguish seizure precursors from external triggers, supporting personalized treatment. For researchers, it links subjective experience with objective data in chronic disease forecasting.
Although many studies demonstrate the feasibility of seizure prediction using physiological data (e.g., HRV, EDA, ECG) and machine learning (9, 22, 17), few explore how human-generated inputs (e.g., journaling, mood logs, symptom reports) can be

→ *space available for images / figures on next page*

Introduction (continued): space for images

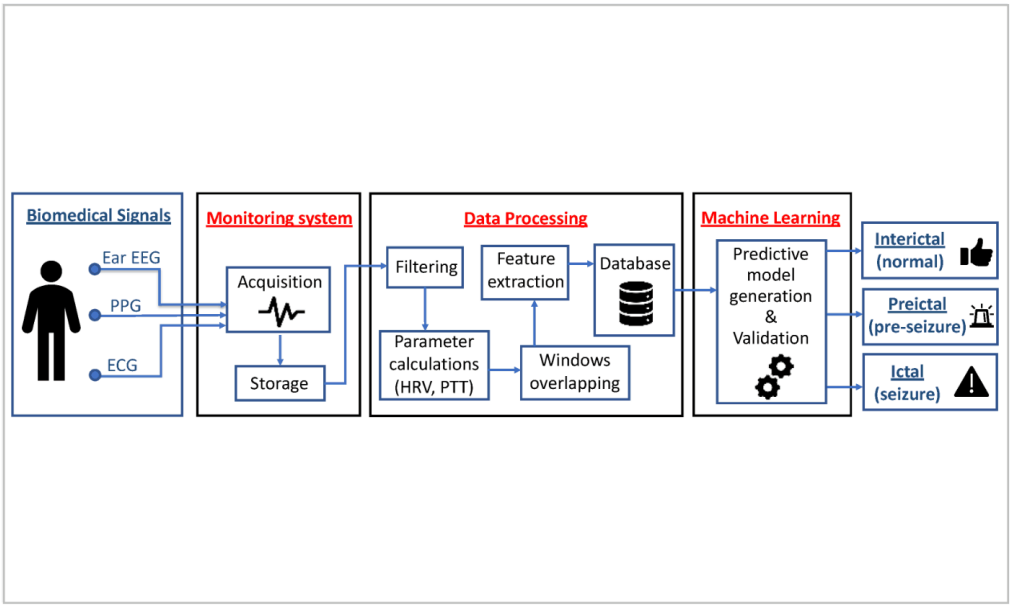


Figure 1: Data processing.

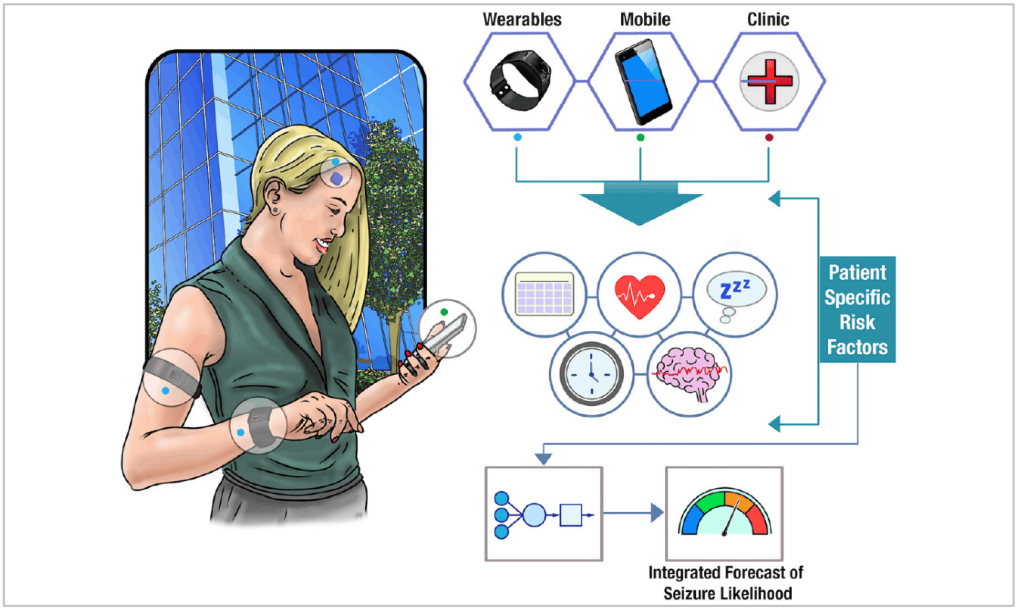


Figure 2: Wearable-connected seizure forecasting system.

Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice. (max 200 words)

Epilepsy is characterized by unpredictable seizures (16), a major factor limiting quality of life and daily functioning (17, 18). Journaling and self-reporting are valuable tools in epilepsy management (6, 27), providing contextual data beyond sensor capabilities. However, self-reports are often incomplete or biased due to forgetfulness, stigma, or emotional fatigue (28, 29). Wearable biosensors can capture physiological signals (e.g., heart rate, electrodermal activity, temperature, motion) that change before seizures (18, 24). Yet, physiological data alone fail to account for contextual or subjective triggers such as stress or emotional state, which are personally-related and vary over time (25, 26). Current seizure forecasting systems face three main challenges:
Data incompleteness: Physiological data lack personal context (e.g., activity, environment, emotions).

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project (1 sentence. As you graduate as an industrial design engineer, your assignment will start with a verb (Investigate/Design/Validate/Create), and you may use the format:
(Investigate/Design/Validate/Create) ***a*** *(what will be the deliverable → prototype/roadmap/process/intervention /approach/guideline/strategy/…)* ***to*** *(what should it do →(create/understand/evaluate/validate/improve/execute/ analyse/…) (the objective → experience/value/process/product/…)* ***for*** *(whom → target group/client/…)* ***in*** *(what context).*

Designing a journaling method for the integration of physiological sensor data from wearable devices with journaling strategies to enhance the personalization, interpretability, and clinical relevance of data, useful for machine learning models to forecast seizures for people with epilepsy (4, 5, 19).

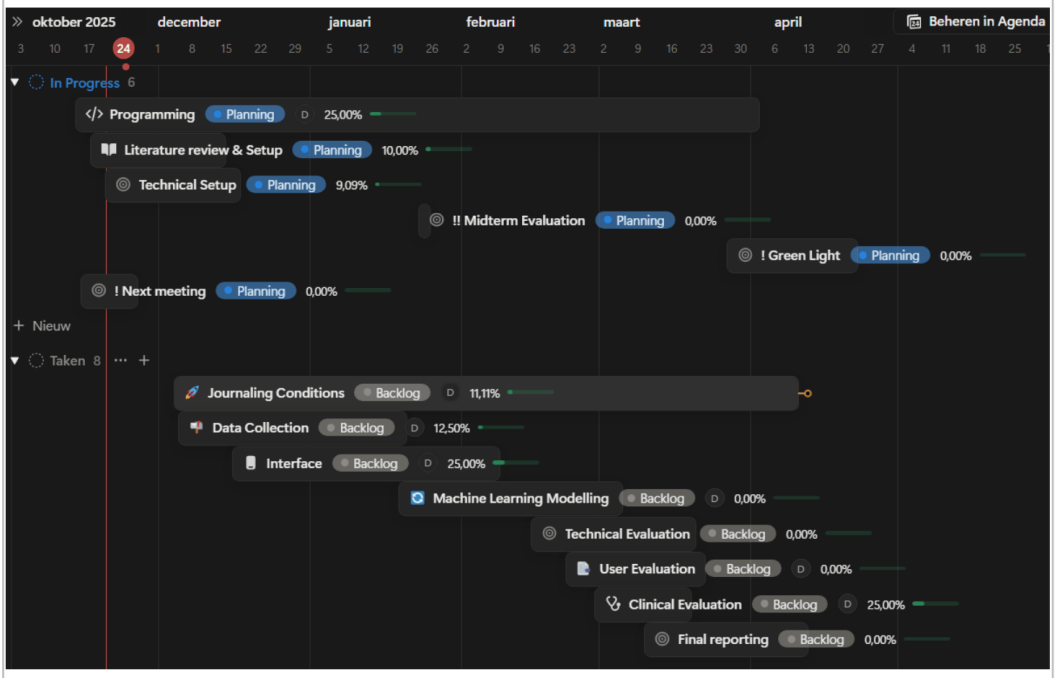
Explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words).

Designing a journaling method for the integration of physiological sensor data from wearable devices with journaling strategies to enhance the personalization, interpretability, and clinical relevance of data, useful for machine learning models to forecast seizures for people with epilepsy (4, 5, 19).

Data Collection: A wearable device (e.g., Empatica E4) will be deployed for 6+ weeks, collecting physiological data (HR, HRV, EDA, temperature, motion) while users log contextual factors (physiological, emotional, environmental, activity) through a combination of self-designed and existing journaling methods (an app, for instance) (18, 24, 25). These multimodal datasets address data incompleteness in forecasting models (1, 11, 19).
Journaling Conditions: Multiple methods and variations will be tested to assess how

Project planning

*To make visible how you plan to spend your time, you must make a planning for the full project. You are advised to use a Gantt chart format to show the different phases of your project, deliverables you have in mind, meetings and in-between deadlines. Keep in mind that all activities should fit within the given run time of 100 working days. Your planning should include the **Kick-off**, **Midterm Evaluation**, **Green Light** and **Finalisation** (ceremony). Please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any (for instance because of holidays or parallel course activities). Add (an image of) the planning in the box below. If it is not readable, you can add the planning as an attachment to My Case along with this Proposal.*



Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some (max 5) personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology (200 words max).

Within 100 days, the goal is to deliver a functional prototype of a wearable-connected journaling web app that generates notifications, labels timeframes, and identifies seizure-related scenarios. If feasible, a mock-up of a preliminary personalized ML model will demonstrate how physiological and journaling data can be integrated for future applications. This will be achieved through a mixed-methods evaluation of journaling methods and the combination of quantitative ML metrics with qualitative UX insights. Due to time constraints, developing a fully functional model remains an ancillary goal.

Sub-questions:
- How do different journaling strategies (event-triggered, periodic, modalities, interfaces) affect user adherence and the completeness of contextual data (response rate, temporal coverage, consistency, signal-context alignment)? This influences the quantity and quality of usable data, correlating to the interpretability and perceived value among patients, clinicians, and future research.
- Which minimal combinations of physiological and self-reported parameters yield interpretable input for simple forecasting outcomes? This sub-question is not a priority; it

D.2. Consent forms

D.2.1. Consent form for the research study

Delft University of Technology
HUMAN RESEARCH ETHICS - INFORMED CONSENT FORM

Research Study – Journaling method for epilepsy

You are being invited to participate in a research study titled 'Designing a personalized wearable-connected journaling system for seizure insights'. This study is being done by Dirk Lagemaat from the TU Delft.

The purpose of the research study is to collect data to create and improve a journaling method for epilepsy patients to track factors that might trigger a seizure. This study will take approximately 6-9 days to complete.

We will be asking you to evaluate the journal app, including different types of nudging techniques. This concerns receiving multiple notifications per day and logging events from your daily life using the designed app. Afterwards, we're interested in your personal/professional experience and opinion about the usability of the different nudging techniques. You'll be asked to fill out a SUS / NASA-TLX / UEQ form and conduct an interview.

The data will be used for evaluating the usability of the journaling method and assessing the generated output. These two aspects are an important part of the thesis to which this research study belongs. The final report of the thesis will be published on the TU Delft Repository.

As with any digitally documented research, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by only collecting personally identifiable information regarding your age and (location of) expertise. Opinions and professional evaluations will be anonymized, only stating your expertise in the report as validation for your input. The data will be stored safely in a locked location and will be used exclusively for evaluation and improvement purposes of the project to which this research study belongs.

Your participation in this study is entirely voluntary, and you can withdraw at any time. You are free to omit any questions. Any given data can be excluded and removed from the research study within a timescale of 14 days' notice.

Contact details:
Corresponding researcher: [redacted]
Responsible researcher: [redacted]

TEMPLATE 2: Explicit Consent points

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICPANT TASKS AND VOLUNTARY PARTICIPATION		
1. I have read and understood the study information dated 02/12/2025, or it has been read to me. I have been able to ask questions about the study, and my questions have been answered to my satisfaction.	☐	☐
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and that I can withdraw from the study at any time, without having to give a reason.	☐	☐
3. I understand that taking part in the study involves one of the agreed-upon evaluation methods as stated in the opening statement. This either means completing a paper questionnaire format or an interview. The interviews will either be audio-recorded (and transcribed as text) or captured using notes; one of the two can be agreed upon before the research study starts.	☐	☐
4. I understand that the study will end after the interview or questionnaire has been finished, with a maximum of 45 minutes.	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
5. I understand that taking part in the study involves a reputational risk and a data breach risk. I understand that these will be mitigated by the right to stop at any point, anonymization of collected data, and a secure, backed-up storage location of the data.	☐	☐
6. I understand that the following steps will be taken to minimise the threat of a data breach, and protect my identity in the event of such a breach: Data will be stored in a locked location and processed on a private laptop, and backed up in the “SURFdrive” of the TU Delft. .	☐	☐
7. I understand that information collected about me that can identify me, such as expertise or location of practice, will not be shared beyond the study team.	☐	☐
8. I understand that the identifiable personal data I provide will be removed from the responses and destroyed after the study.	☐	☐
C: RESEARCH PUBLICATION, DISSEMINATION, AND APPLICATION		
9. I understand that after the research study, the de-identified information I provide will be used for evaluation and improvement purposes of the accompanying thesis report, concerning the development of a journaling method, which will be published on the TU Delft Repository. These will either be referred to as quotes or statistical values from this research study.	☐	☐
10. I agree that my responses, views, or other input can be quoted anonymously in research outputs.	☐	☐
11. If you do not have any problem with stating your real name during the research study: I agree that my real name can be used for quotes in research outputs	☐	☐
D: (LONGTERM) DATA STORAGE, ACCESS, AND REUSE		
12. I permit the de-identified evaluation and opinions that I provide to be archived in the TU Delft repository as anonymized transcripts, so they can be used for future research and learning.	☐	☐

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
13. I understand that any of the collected raw data will be stored for 6 months after the research study and deleted thereafter.	☐	☐
14. I understand that access to this repository is openly accessible.	☐	☐

Signatures

Name of participant	Signature	Date

I, as researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands what they are freely consenting.

Researcher name	Signature	Date

Study contact details for further information:

Dirk Lagemaat
D.C.R.Lagemaat@student.tudelft.nl

D.2.2. Consent form stakeholder meetings

Delft University of Technology
HUMAN RESEARCH ETHICS - INFORMED CONSENT FORM

Research Study – Journaling method for epilepsy

You are being invited to participate in a research study titled 'Designing a personalized wearable-connected journaling system for seizure insights'. This study is being done by Dirk Lagemaat from the TU Delft.

The purpose of the research study is to collect data to create and improve a journaling method for epilepsy patients to track factors that might trigger a seizure. It will take you approximately *15-60* minutes to complete, depending on the type of evaluation.

We will ask you to complete several steps using the designed app, execute a clinical evaluation / NASA-TLX / SUS / UEQ / interview / Co-Design session. This regards your personal/professional experience and opinion about the usability of the journal application and its output.
The data will be used to evaluate the usability and interpretability of the application and to assess the generated output. These two aspects are an important part of the thesis to which this research study belongs. The final report of the thesis will be published on the TU Delft Repository.

As with any digitally documented research, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by only collecting personally identifiable information regarding your age and (location of) expertise. Opinions and personal/professional evaluations will be anonymized, only stating your stakeholder position regarding this study in the report as validation for your input. The data will be stored safely in a locked location and will be used exclusively for evaluation and improvement purposes of the project to which this research study belongs.

Your participation in this study is entirely voluntary, and you can withdraw at any time. You are free to omit any questions. Any given data can be excluded and removed from the research study within a timescale of 14 days' notice.

Contact details:
Corresponding researcher: [redacted]
Responsible researcher: [redacted]

TEMPLATE 2: Explicit Consent points

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICPANT TASKS AND VOLUNTARY PARTICIPATION		
1. I have read and understood the study information dated 02/12/2025, or it has been read to me. I have been able to ask questions about the study, and my questions have been answered to my satisfaction.	☐	☐
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and that I can withdraw from the study at any time, without having to give a reason.	☐	☐
3. I understand that taking part in the study involves one of the agreed-upon evaluation methods as stated in the opening statement. This either means completing a set of tasks and a paper questionnaire format or an interview. The interviews will either be audio-recorded (and transcribed as text) or captured using notes; one of the two can be agreed upon before the research study starts.	☐	☐
4. I understand that the study will end after the interview or questionnaire has been finished, with a maximum of 45 minutes.	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
5. I understand that taking part in the study involves a reputational risk and a data breach risk. I understand that these will be mitigated by the right to stop at any point, anonymization of collected data, and a secure, backed-up storage location of the data.	☐	☐
6. I understand that the following steps will be taken to minimise the threat of a data breach, and protect my identity in the event of such a breach: Data will be stored in a locked location and processed on a private laptop, and backed up in the "SURFdrive" of the TU Delft.	☐	☐
7. I understand that information collected about me that can identify me, such as expertise or location of practice, will not be shared beyond the study team.	☐	☐
8. I understand that the identifiable personal data I provide will be removed from the responses and destroyed after the study.	☐	☐
C: RESEARCH PUBLICATION, DISSEMINATION, AND APPLICATION		
9. I understand that after the research study, the de-identified information I provide will be used for evaluation and improvement purposes of the accompanying thesis report, concerning the development of a journaling method, which will be published on the TU Delft Repository. These will either be referred to as quotes or statistical values from this research study.	☐	☐
10. I agree that my responses, views, or other input can be quoted anonymously in research outputs.	☐	☐
11. If you do not have any problem with stating your real name during the research study: I agree that my real name can be used for quotes in research outputs	☐	☐
D: (LONGTERM) DATA STORAGE, ACCESS, AND REUSE		
12. I permit the de-identified evaluation and opinions that I provide to be archived in the TU Delft repository as anonymized transcripts, so they can be used for future research and learning.	☐	☐

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
13. I understand that any of the collected raw data will be stored for 6 months after the research study and deleted thereafter.	☐	☐
14. I understand that access to this repository is openly accessible.	☐	☐

Signatures

Name of participant	Signature	Date

I, as researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands what they are freely consenting.

Researcher name	Signature	Date

Study contact details for further information:

Dirk Lagemaat
D.C.R.Lagemaat@student.tudelft.nl