IDEA-FAST

Identifying Digital Endpoints to Assess FAtigue, Sleep and acTivities in daily living in Neurodegenerative disorders and Immune-mediated inflammatory diseases.

Grant Agreement No. 853981

WP4 –Device-specific Data Analytics and Performance Assessment

D4.1: Definition of assessment protocol for device-specific digital endpoints

Lead contributor	Teemu Ahmaniemi (VTT), Meenakshi Chatterjee (Janssen)
Other contributors	Emmi Antikainen (VTT), Mark van Gils (VTT), Anita Honka (VTT), Stefan Avey (Janssen), Francesca Cormack (Cambridge Cognition), Haneen Njoum (Sanofi)

Due date	30 SEP 2020
Delivery date	6 OCT 2020
Deliverable type	R[1]
Dissemination level	PU[2]

[1] **Use one of the following codes:**
R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
OTHER: Software, technical diagram, etc.

[2] **Please choose the appropriate reference and delete the rest:**
PU = Public, fully open, e.g. web;
CO = Confidential, restricted under conditions set out in Model Grant Agreement;
CI = Classified, information as referred to in Commission Decision 2001/844/EC.

Document History

Version	Date	Description
V0.1	7 Aug 2020	Initial Draft, continuous collaborative updates made in
V0.9	18 Sep 2020	Draft submitted to Steering Committee
V1.0	6 Oct 2020	Submitted version

Table of Contents

1 Abstract

The IDEA-FAST project assesses the value of modern wearable measurement devices and the data and information they generate in the context of sleep disturbance and fatigue assessments. The assessment is based on two studies: Feasibility Study (FS) and Clinical Validation Study (CVS). The subjects in the studies will be patients of chronic diseases such as neurodegenerative disorders (NDD) and immune-mediated inflammatory diseases (IMID) as well as healthy participants. Work package 4 (WP4) in the project is responsible for device specific data analytics. This document describes the assessment protocol for the FS which is the first deliverable of WP4.

The assessment protocol described in this document will be based on data that will be collected in the FS: clinical data, baseline and online questionnaire data and device data. As the absolute reference of the sleep- and fatigue related data will not be available, the assessment will be based on three factors: 1) Data quality and reliability of the digital measure, 2) Performance as compared to patient reported outcomes, 3) Performance across patient cohorts.

The protocol suggests that the outcome of device assessment will be provided as a recommendation based on the above 3 factors. The device performance will be represented in visual formats like Bland-Altman plots and radar charts. This allows reader to compare the devices across the above factors in an intuitive objective manner.

2 Introduction

2.1 *Context*

IDEA-FAST is a research project uniting clinical experts, technical research partners, pharma industry, device providers, patient experts and other stakeholders in an Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 853981. It receives support from the European Union's Horizon 2020 research and innovation program and EFPIA and PARKINSON'S DISEASE SOCIETY OF THE UNITED KINGDOM LBG.

IDEA-FAST aims to help patients with chronic diseases such as neurodegenerative disorders (NDD) and immune-mediated inflammatory diseases (IMID) by improving health-related quality of life (HRQoL) and activities of daily living (ADL). More specifically, the project assesses the value of modern wearable measurement devices and the data and information they generate in the context of sleep disturbances and fatigue. It is hoped that this advances the current state of practice, where such assessments rely mainly on subjective reports, typically using standardized questionnaires provided by patients every few months. Such evaluations are prone to recall bias, reliability issues and poor sensitivity to change. Changes in patient state measured by digital devices ('digital endpoints') are objective and can be continuous. However, it is important to assess how much they actually correlate with changes in the patient state as assessed by traditional clinical tools and practices, and, crucially, how reliable the data are. It is of similar importance to address user acceptance and robustness of the measurement set-up in real-life under uncontrolled daily living settings.

In IDEA-FAST, we address the challenge via two studies: the Feasibility Study (FS) and Clinical Validation Study (CVS). The detailed protocol of the FS and plans for the CVS can be found elsewhere, but shortly: from a set of candidate devices a subset of 10 devices was chosen that capture activity, physiological, neurophysiological, cognitive and social activity-related measures. These are being evaluated in the different clinical partners' sites for up to 5 weeks for each patient. Based on the results of the FS, a further selection of devices will be made towards a limited set that will be evaluated in the considerably larger CVS. This selection process takes into account viewpoints related to

usability, user acceptance as well as assessments of performance and quality of the signals. It is the latter parts, assessment of signal quality and performance, which are in the scope of this document.

Related documents:

- Protocol of the Feasibility Study in Kiel
- WP3 device selection deliverable
- WP4 initial data analysis pipeline
- Statistical Analysis Plan for the FS (available by end 2020)

2.2 Assessing the device-related outcomes of the Feasibility Study

As the FS will provide key information towards the CVS, it is important that an objective and transparent approach is defined on how to assess performances of devices and which measures to use to quantify concepts like data quality and reliability. The overall aim of the study is to identify candidate *digital endpoints* which may not necessarily be dependent on a specific brand or type of hardware or software. However, the task of estimating a universal digital device-agnostic feature (for example separating the value of "heart rate variability" as a device-independent measure from "heart rate variability measured with device X") is non-trivial. Moreover, obtaining a device-agnostic measure will depend on whether the same digital measure is obtained in parallel from multiple devices (e.g. acceleration, EEG) or from a single device (e.g. cognition, social activity). Therefore, we recognise that digital measures in the study will inevitably be affected by aspects of study design, device and other data acquisition-related characteristics, and thus we outline the performance assessment of measures in a 'device-specific' manner. To minimise bias, IDEA-FAST has put a great effort into the device selection process by requiring CE marking status of all devices, factoring in earlier experiences of IDEA-FAST partners in using the devices (or though published studies), and conducting an elaborate assessment of technical properties of devices in WP3 with a wide range of different experts. This minimises the risk of unjustly dismissing a potential digital measure because of poor device hardware or software.

This document primarily provides an initial plan to perform device-specific assessment of potential digital measures for HRQoL and ADL (and specifically sleep quality and fatigue) based on the FS. The outcome of assessment will then be used to provide a recommendation of the performance of each device.

2.3 Outline of this document

This document outlines the initial framework for quantitative assessment of the performance of digital measures for each device used in IDEA-FAST. Specific aspects of the data analysis can be modified accordingly upon detailed analysis of FS data when it becomes available. The document is outlined as follows. Section 3 explains the technical characteristics of the data available for analysis. Section 4 focuses on the assessment methodologies which will be primarily based on the following three factors:

1. Data quality and reliability of the digital measure,
2. Performance as compared to Patient Reported Outcomes (PROs)
3. Performance across patient cohorts

The outcome of above factors will finally be translated to an overall performance measure for each device. This device-specific scoring approach is meant to serve as a recommendation or guide to facilitate comparison of performance across different devices. Section 5 discusses the possible open questions, caveats, and future work. Section 6 concludes the document.

3 Data sources

The main objective of the FS study is to identify the most promising digital correlates of device-specific data with sleep and fatigue that can be further evaluated in the CVS. To achieve this goal, it is essential to perform a) detailed analysis on quality of the endpoints/measures obtained from the devices, and b) association between digital measures with traditional clinical outcomes (or surveys) and PROs. The devices used in the FS cover a broad range of predefined concepts of interest (COI) that are related to sleep quality and fatigue, including physical activity, general physiology and neurophysiology, cognition, social function and interaction (Figure 1). Details of the study protocol including the schedule of measurements is provided in the FS protocol. Table 1 presents a summary of the PROs and digital features.

Table 1. A simplified view of features derived from the devices and PROs collected during the study or at site visit. Each cell in the table corresponds to a possible correlation between a feature and a PRO. The most prominent correlations are marked with letters reflecting the category of the endpoints: A (Activity), S (Social/cognitive), P (Physiology) or N (Neurophysiology). The colours of the letters correspond to device categorisation in Figure 1 (reproduced here from feasibility study plan: link). HR: heart rate, HRV: heart rate variability, GSR: galvanic skin response, ReactionT: reaction time, Circ. R: circadian rhythm.

PROs / Endpoints		Stress Monitor App									Study Visit	
		1/night		4/day						1/day	1/week	
Categ.	Feature	Sleep Q	Sleep dur.	Sleepiness	P. Fatigue	M. Fatigue	Pain	Anxiety	Depr.	Activity	FACIT	MOS SS
Activity	Intensity	AS	AS	AS	AS	AS	A	AS	AS	A	AS	A
Phys + social	Steps	A			A					A	AS	A
	Type	AS	AS				A				AS	A
Sleep	Duration		AP	P	P	P			P		P	PN
	Quality	APN		P	P	P					P	PN
	Wakeups	APN									P	PN
Psycho-Phys.	HR/HRV			P	P	P	P				P	P
	GSR			P	P	P	P	P	P			
	SkinTemp			P	P	P	P					
Cogn. + motorics	Memory	S	S	S	S	S					S	S
	ReactionT	S	S	S	S	S					S	S
	Speech	S	S	S	S	S					S	S
Other	Circ. R										AP	AP
	WearTime											

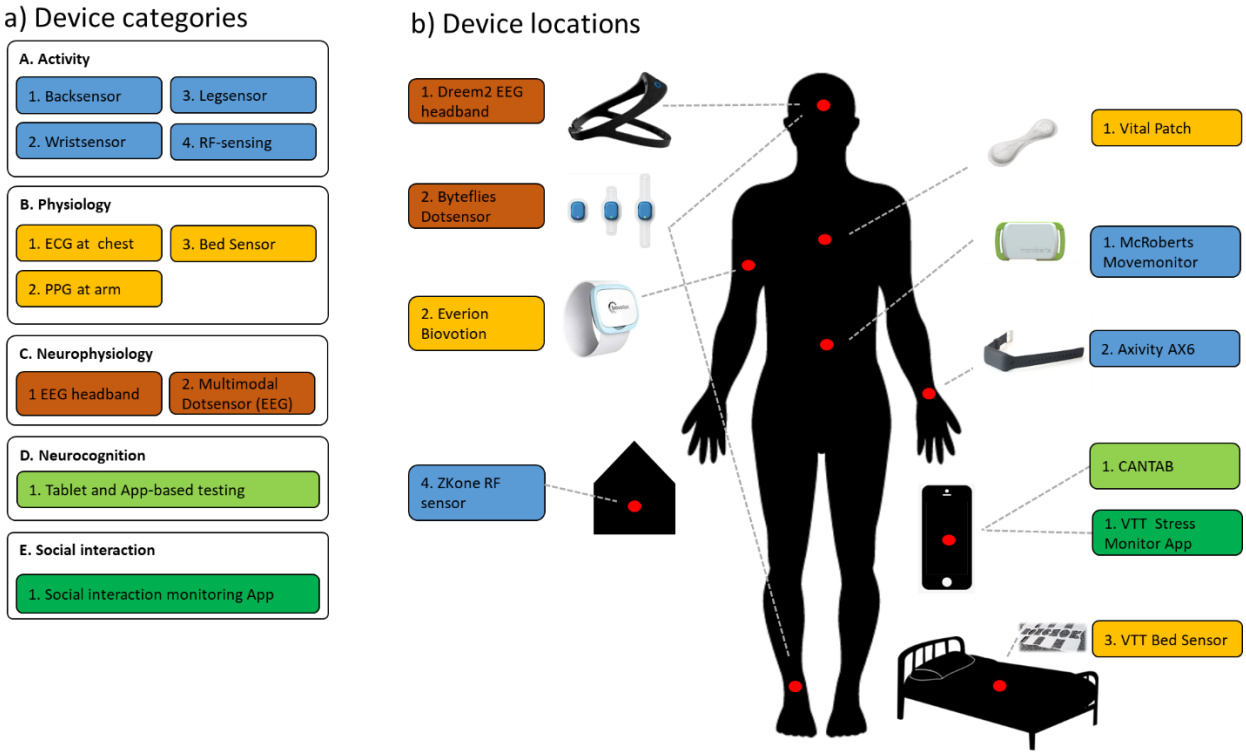

Figure 1. Devices, locations (b) and their categories (a) in feasibility study. The device categories and their colours correspond to the colours in Table 1.

3.1 Clinical data

Pseudonymised participant demographic, medical, clinical and PRO data will be captured in the eCRF system maintained by the Cambridge Clinical Trial Unit (UCAM) and entered by professionals at the different sites. It will be securely stored as tables in a relational database management which will eventually be linked to the IDEA-DAST data management system as developed in WP5.

3.1.1 Baseline data

Participants will attend a scheduled study visit at the recruitment centre in the beginning of the study, during which their basic demographics and medical data will be collected. Disease-specific clinical assessments will also be carried out for patient participants. Patients will be expected to complete several self-reported outcome measures within 1-3 days from the scheduled site visit. These self-reported outcome measures comprise a) fatigue and sleep assessments using existing measurement tools, and b) HR-QoL and/or potential confounders of fatigue and sleep disturbances (see appendix 1 for more details). The following baseline data will be especially relevant in the context of assessment of each device:

- Demographics (including age, sex, education, occupation, ethnicity)
- Height and body weight
- Resting Heart rate
- Blood pressure (lying/standing)
- Disease specific details (including questionnaires addressing severity of disease and medications)

This data will be especially important when assessing the disease specific effects, i.e. to evaluate for sensitivity of a digital measure from a device to detect disease specific change in fatigue or sleep. Furthermore, it may be interesting to explore the influence of medication on changes in measures like activity and sleep.

3.1.2 Sleep and fatigue assessment post device use

The sleep and fatigue assessment questionnaires include the Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F) and the Medical Outcomes Study Sleep Scale (MOS-SS) acute. The FACIT-F is a 13-item fatigue assessment questionnaire covering physical fatigue, functional fatigue, emotional fatigue and social consequences of fatigue using a 5-level rating scale. The MOS-SS acute is a 12-item questionnaire assessing sleep disturbance, sleep adequacy, somnolence, quantity of sleep, snoring, and awakening short of breath or with a headache using a 5-level rating scale. Both FACIT-F and MOS-SS questionnaires are filled out by participants after every device use period (~5 days).

3.2 Online questionnaires

Online questionnaires in the feasibility study are facilitated by the Stress Monitor application running on the participants' smart phone. The application launches a questionnaire pattern 4 times a day. The schedule and type of questions are listed in Table 2.

Table 2. Questions and scheduling of the questionnaires by the Stress Monitor application.

Question		**Format/Resolution**	**Launching time**
General feeling	Physical fatigue	7-level scale	9:00, 13:00, 17:00, 21:00
	Mental fatigue		
	Anxiousness		
	Depression		
	Pain		
Daytime sleepiness		10-level options scale (Karolinska Sleepiness Scale)	13:00, 17:00, 21:00
Sleep details	Bedtime	Android clock face time entry	9:00
	Wake up time		
	Sleep rating	7-level scale	
	Time to fall asleep	6-level options scale	
	Time awake during night	6-level options scale	
Daily activities	Physical activity	7-level scale	21:00
	Mental activity		
	Activity details	Free text entry	

3.3 Voice recordings

The Stress Monitor application also facilitates voice recording feature that is launched after each evening questionnaire. Voice recording analysis is not part of the FS protocol, but is an explorative addition. The main objectives of recording voice are 1) To assess the verbal fluency of the participants, 2) To assess articulatory/cognitive challenges caused by fatigue or sleep 3) Request qualitative data related to daily feelings/activities. The voice recording scheduling and tasks are stated in Table 3. At the moment of writing, the voice recording is still in a planning stage and the list of acoustic features is subject to update.

Table 3: Voice recording tasks and related details.

Voice tasks	Description	Duration	Launched	Potential features
Time/date entry	Subject is instructed to say the date and time	10 sec	Every day	Words/min, Pause length, Accuracy, pronunciation errors, acoustic characteristics (intensity, pitch, jitter, shimmer, harmonicity)
Daily diary	Subject is asked to describe her/his daily activities	Up to 2 min	Every day	Words/min, number of words, frequency and length of pauses and hesitations, acoustic characteristics (intensity, pitch, jitter, shimmer, harmonicity)
Articulation	Subject is asked to say "papapapa…" as rapidly as possible	15 sec	Every other day	Syllable rate (syllables / second), variability in rate, change in rate over attempt, acoustic characteristics (intensity, pitch, jitter, shimmer, harmonicity)
Sustained Phonation	Subject is asked to sustain "aaaaaah" vowel sound until the end of the recording	15 sec	Every other day	Duration of sustainment, acoustic characteristics (intensity, pitch, jitter, shimmer, harmonicity)

3.4 Device data

During the study period of up to 36 days, participants are asked to use various digital devices covering all COIs, with a simultaneous use of maximum 3 active devices (those requiring active effort by participant) in addition to a smartphone. Activity based sensors worn at the back, wrist and leg measure movement (via. accelerometer, gyroscope) while a radiofrequency-based sensing record movement in the surrounding. Physiological measures are assessed using a mobile electrocardiogram (ECG) device at the chest, photoplethysmography (PPG) at the arm and a pressure-sensitive bed sensor. Electroencephalography (EEG) data for measurement of sleep stages are collected using multimodal dots or a headband during sleep. Mobile applications are used to assess cognitive skills and social activity. All active devices will be worn for 5 consecutive days followed by at least 2 rest days. To evaluate association with clinical assessment and PROs, the objective digital endpoints will be obtained either directly from the device manufacturer or calculated from raw data using supporting software. Table 4 lists each device per COI along with the corresponding features available from it (see appendix 2 for detailed description of devices). All devices are CE-certified and GDPR compliant with acceptable technical performance specifications related to their ability to reliably capture, process, store, and transfer relevant digital data.

Table 4: List of digital devices and corresponding features

	Device Name	**Device type**	**Features**
Activity	AX6 Axivity	Wrist worn accelerometer	Activity type and intensity, energy expenditure
	McRoberts Movemonitor	Accelerometer worn on lower back	Activity type and intensity, energy expenditure, sleep duration and quality
	ZKONE (passive)	RF based contactless sensor	Heart rate, breathing rate in stationary position
Physiology	Vital Patch	ECG patch worn on chest	Heart rate, heart rate variability, breathing rate, body posture, skin temperature, activity
	Everion Biovotion	PPG device worn on arm	Heart rate, breathing rate, blood oxygenation, skin temperature
	VTT Bedsensor (passive)	Pressure sensitive bed sensor	Heart rate, breathing rate, breathing disturbances, movement on mattress
	Dreem 2	EEG headband	Sleep stages, heart rate, breathing rate, breathing rate variability

Neurophysiology	Byteflies Sensor Dots	Multimodal (ECG, EEG, EMG, EOG) sensor worn on arm, chest, behind ear	Sleep stages, heart rate, heart rate variability, muscle activity, eye movement, activity
Social & Cognitive Activity	VTT Stress Monitor	Android based mobile application for social behavior	Mobile usage, online sleep and fatigue PROs
	CANTAB	Battery of cognitive tests via touchscreen device	Cognitive skills -memory, attention, psychomotor speed, executive function

4 Data analysis methods

The data features, or digital measures, listed in section 3.4 originate from several different devices that operate on different sampling frequencies over different times of the day. Moreover, they are located at different places differently during measurements, with or without direct physical contact to the user. The objective of the digital endpoint assessment protocol is to facilitate the identification of digital endpoints that are able to capture clinically relevant changes in fatigue and sleep. Therefore, the data analysis methods presented here produce numerical and visual support to compare different digital measures.

As illustrated in Figure 2, the analysis will evaluate the performance of a candidate digital measure from several aspects: 1. the data quality and reliability of the digital measure (left box), 2. its performance as compared to reference, here PROs (right box) and (3) assessment of factors 1 and 3 within each patient cohort (middle box) to evaluate the sensitivity of a digital measure to detect a *disease-specific change* in fatigue or sleep. The quantitative and visual analysis results covering all three aspects enable comparison between digital measures, devices, and measurement modalities. The end result is a combination of these 3 aspects – it serves as a guide to the selection of digital measures to evaluate in the IDEA-FAST CVS.

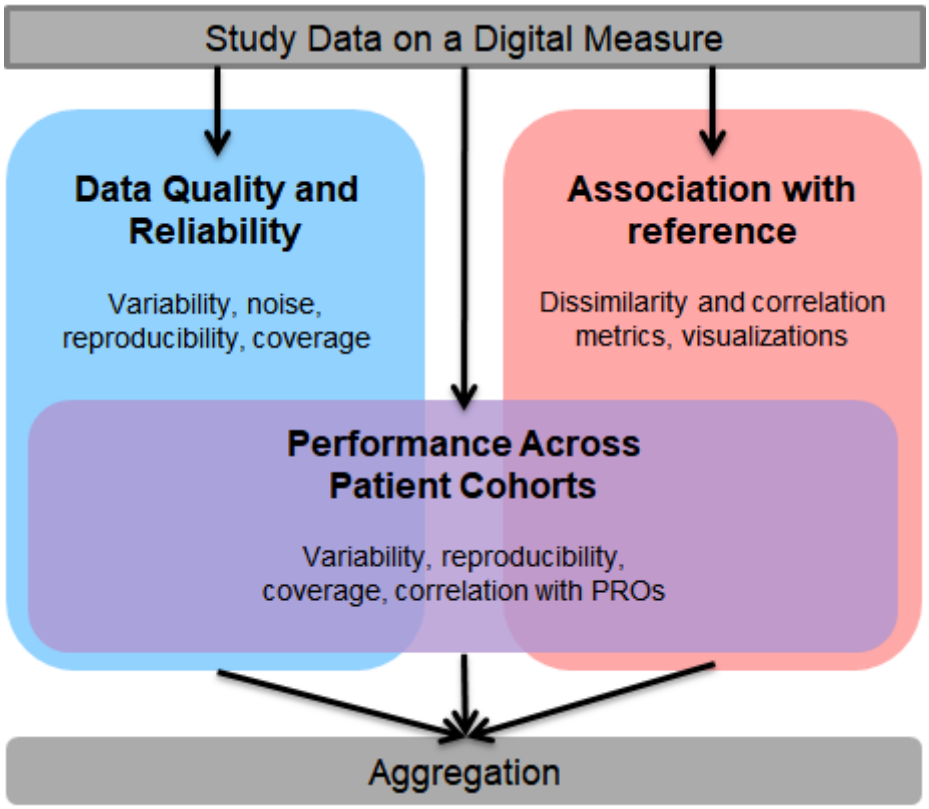

Figure 2. Each digital measure from a device is analysed in three sections: data quality and reliability (left), correlation to reference or patient reported outcomes (right) and performance assessment specific to patient cohorts (middle). Assessment across patient cohorts implements the first two analysis task but within disease-specific patient cohorts.

The effect of covariates such as age, gender, region, ethnicity, and time since diagnosis will be considered in the context of specific analysis methods, as identified in the following subsections.

4.1 Data quality and reliability

Data quality and reliability analysis includes investigating variability of the digital measure, sources of noise and disturbances, reproducibility when applicable, and typical data coverage over the measurement period. The analysis methods are summarized in Table 5. This section of analysis will study the digital measures within individual patients and over all patients, without reference metrics or distinction between cohorts. However, different covariates will be considered when analysing variability and coverage. For instance, age may correlate with activity and phone usage, thus affecting the variability and coverage of these measures.

Table 5. Summary of analysis methods for data quality and reliability assessment.

Variability	***Noise***	***Reproducibility***	***Coverage***
Statistical descriptors	Outlier analysis Longitudinal changes/trends/shifts	Correlation Statistical hypothesis tests	Device-specific timestamp-based algorithm

These analyses may also be conducted on different frequency bands of the signals when relevant, as the noise level and variability can differ from the results obtained over the full spectrum. This is

especially important for those signals whose features are located around specific frequencies. Sleep features in EEGs are a typical example, where the robustness of different features detection is very dependent on frequency. Besides, in situations where two different devices can be used simultaneously on the same patient to measure the same types of data, the resulting signals can also be analysed and compared simultaneously in time and frequency using time-frequency or time-scale transforms such a wavelet analysis.

4.1.1 Variability

Variability is evaluated using statistical metrics, such as standard deviation and interquartile range [1]. Mean and median can complement these results with more information on the data distribution, giving additional information on data quality and setting a baseline for further analyses. Especially for digital measures that are expected to be normally distributed (such as skin temperature) can be described through mean and standard deviation, whereas other measures (such as phone application usage) are better described through the non-parametric median and interquartile range. Unnaturally high or low variability of a digital measure (when the reference patient state is stable) may indicate difficulties, e.g., in the measurement setting. The amount of acceptable variability is defined individually for each digital measure but is the same across devices measuring the same quantity.

4.1.2 Noise

A low signal-to-noise ratio implies poor data quality with a small portion of reliable data. Thanks to the fact that this study uses CE marked well-tested devices, noise is mainly expected to be present in the time-domain (amplitudes, due to artefacts) rather than in the frequency domain (continuous noise). Outliers are a type of noise, possibly caused by disturbances in the measurement, and can disturb data analysis. If the digital measure data can be assumed to follow some distribution, the prevalence of statistical outliers, i.e. data points significantly different from the mean of neighbouring measurement points, can provide insights on data quality [2]. Prior to the outlier analysis, the signal mean is inspected to evaluate whether the signal is within a reasonable range to start with. The allowed variability and the required size of the inspected neighbourhood depend on the sampling frequency and the measure in question. Furthermore, the data is inspected for other disturbances, such as measurement technology-specific disruptions. For instance, a sensor response may experience some shift over time or due to changes in ambient conditions and cause bias to the digital metric. Such technology-dependent linear trends and other undesirable temporal periodical variations are evaluated over the full measurement period to distinguish such effects. It is good to keep in mind that several devices in IDEA-FAST do have in-built noise cancelling features and quality assurance of the signals they provide, and as such are expected to provide good quality output with respect to common noise characteristics. However, unusual artefacts (especially in uncontrolled real-life settings) may have characteristics that are not caught in in-built processing algorithms, and evaluation of their impact is especially important in the context of this project.

4.1.3 Reproducibility

Reproducibility is heavily connected to reliability and may be studied within patient for digital measures that are expected to remain reasonably similar between repeated measurements, such as average resting heart rate on a weekly basis [3]. The level of correlation between repeated measures and its variability across patients describes reproducibility. Removal of longitudinal trends in the data prior to the reproducibility analysis may be reasonable for some measures.

4.1.4 Coverage

Coverage describes the amount of valid data collected over the whole measurement period and reflects the device's practical operability and, implicitly, links also to user-friendliness and acceptability. For example, devices mounted in a room can experience data gaps when the subject leaves the room, whereas wearables similarly lack data over non-wear periods that occur, e.g., when the device is charging. However, some data gaps may even appear when the device is thought to be operating normally. Coverage is evaluated using device-specific algorithms that are based on the timestamps available in the data. These algorithms are developed as a part of the WP4 data analysis pipeline and may provide insights on the cause of data gaps, possibly revealing the most common reasons for data collection failures on each device.

4.2 Digital measure performance against subjective reference

Comparing subjective PROs to digital measures can indicate the digital measures that correlate to clinically relevant changes of sleep and fatigue. Yet, full one-to-one equivalence is not expected or necessarily even desirable, as subjective measures are affected by many factors. Moreover, the PROs reflect complex concepts that are represented in ordinal scale values with integer numbers – this highlights that the exact numerical comparison of association and error measures needs to be done with a realistic mindset. The associations between the two are studied through (linear and non-linear) correlations, regression error metrics, and visualisation methods, as summarized in Table 6. The set of covariates (listed in Section 4) will be considered at each step, concentrating on especially significant or non-significant associations.

One of the challenges in assessing the associations between digital measures and subjective patient-reported outcomes, is the dissimilar sampling frequencies. In order to compare differently sampled measures, the digital measures should either be resampled to match the frequency of the reference subjective outcome, or alternatively aggregated over comparable periods of time, into values such as mean, range, or quantiles. The aggregation window may depend on the nature of the digital measure (e.g. using time windows with stable value, 'stationary' values) as well as that of the reference measure. Furthermore, the digital measures likely require normalization to a suitable range to ease comparison with the fixed-range subjective measures.

Table 6 Summary of methods to analyse associations between digital measures and the patient reported outcomes.

Association with PROs	***Regression error***	***Visualization***
Spearman correlation Pearson correlation Kendall rank correlation Somers' D Prediction probability P_K	Mean absolute error Root mean square error Whiteness of the residuals	Bland-Altman plots Correlation scatterplots Box-and-whisker plots Violin plots

4.2.1 Correlation with patient reported outcomes

A correlation coefficient computed between a digital measure and a PRO describes the linearity or monotonicity of their relation, and can be evaluated using Pearson or Spearman correlation, respectively, or through other non-parametric measures in addition to Spearman correlation, such as Kendall's τ, Somers' D, or prediction probability P_K [4]–[6]. Highly positive or negative correlation signifies co-occurring change (concordance) in the measures and, therefore, is a good indicator for good digital measures. The correlation p-value further establishes whether the correlation is statistically significant from zero; a significant correlation yields a p-value below the selected

significance level. The significance level α is selected carefully for the data set in order to control the probability of falsely rejecting a significant correlation [7].

Correlation is studied both within subject and across subjects. The within-subject correlation can be explored with mixed effects models on repeated measures for further insights on the subject-level [8].

4.2.2 Regression error

If the digital measure and the PRO are reasonably correlated (p-value indicates statistical significance, or there is otherwise clinical relevance), a regression model can be formed and evaluated through error metrics generally used to evaluate regression, such as mean absolute error (MAE) and root mean square error (RMSE) [9]. Of the two, mean absolute error is more robust to outliers, whereas root mean square error weights the individual large error over small ones. Mean absolute error describes the average differences between the metrics, whereas root mean square error complements the previous by describing the magnitude of the most blatant errors. Examining the largest outstanding errors can reveal special conditions when either one of the two measures is unsuitable to describe changes in health. As with all regression modelling exercises, examination of validity of the model is of great importance. Hence analysis of randomness (whiteness) of the errors is an essential item to report.

4.2.3 Visualization

Visualisation often gives more information than mere numbers do. Methods to use include Bland-Altman plots [10], correlation scatterplots, box-and-whisker plots and violin plots [11]. The Bland-Altman plot describes both systemic and random differences, while giving a good visual presentation of the relationship between the measures. Correlation scatterplots visualize the linear relation, which is intuitive to interpret. Box-and-whisker plots, on the other hand, can be especially useful in analysing the data distributions over categorical covariates, e.g., gender or region. Violin plots expand this concept further by visualising actual distributions.

4.3 Digital measure performance across patient groups

One of the major aspects considered in both the feasibility study and the clinical validation study is the digital measure performance across different patient cohorts, including several disease cohorts as well as healthy controls. Characteristics of each cohort may have a significant effect on the performance of some devices and digital measures, and one measurement modality may suit one cohort but not another.

Performance assessment across patient categories includes both correlation analysis and data quality assessment while considering variations within and between groups. The analysis methods are outlined in Table 7. The effect of covariates will be considered in variability, coverage, and correlation studies. Especially the severity of the condition may be an interesting variable in the performance assessment.

Table 7. Summary of analysis methods used to evaluate digital measure performance across cohorts.

Variability	***Reproducibility***	***Coverage***	***Correlation with PROs***
ANOVA ANCOVA	Mixed effects model	Group-wise mean coverage with standard deviation	Group-wise correlation

4.3.1 Variability

Analysis of variance (ANOVA) and analysis of covariance (ANCOVA) will be employed to study the statistical variance between subject groups along with the variance within each group [12]. If variance within disease cohorts is larger than between cohorts, the digital measure can be considered insensitive to the type of disease, or whether the subject is healthy or not. Then again, by using subject categorization based on significance of change in PROs, good digital endpoints would be indicated by statistically significant differences between the categories. The probability of falsely claiming significant difference between groups is controlled by the significance level α, selected while considering the effect of sample size.

4.3.2 Reproducibility

Reproducibility within subject groups over selected periods can be analysed using linear mixed effects models [13]. Mixed effects models can describe both random and fixed effects in data and are considered especially useful when studying repeated measurements. Here, the disease group and time are the primary fixed effects, whereas subjects inflict variance to the data as a random effect. Comparing repeated measures using mixed models helps distinguish which variables cause variations in reproducibility and evaluate it within groups.

4.3.3 Coverage

Coverage can be studied within the different groups to identify device usage issues that may be typical within each group. Coverage will be computed using the device-specific coverage algorithms for each measurement, similarly as in quality assessment (see section 4.1). Mean coverage (with standard deviation) in each subject group may reveal data collection difficulties for specific groups. Additionally, ranking the reasons for data gaps (when available; varies between devices) may provide further insights to the digital measure performance on a given group of subjects.

4.3.4 Correlation with patient reported outcomes

Correlation of each subject group with the PROs can help distinguish the groups for which a digital measure performs well, even if it is unsuitable for some groups. The correlation metrics described in subsection 4.2.1 are used similarly in this sector of analysis. Also, similarly, the correlation p-value will be assessed to judge whether the correlation is statistically significant.

4.4 Aggregation of performance measures

The results regarding quality and reliability, association with reference, and performance across subject groups are aggregated to assist the selection of digital measures for the CVS. A radar chart representation provides a quick visual overview on the performance of any digital measure within a specified patient group. In the radar chart, performance metrics are presented as sectors extending from the centre of the circle towards the edges. The radial axis indicates the performance level. The included performance metrics are scaled to obtain values between zero and one. To simplify the chart, only a selected number of performance metrics are included from each analysis category. The overview chart is accompanied by more detailed analysis results to provide more insights into the performance of the digital measure. Figure 3a illustrates the visualization principle in an imaginary case, where three representative performance metrics have been selected from each analysis category.

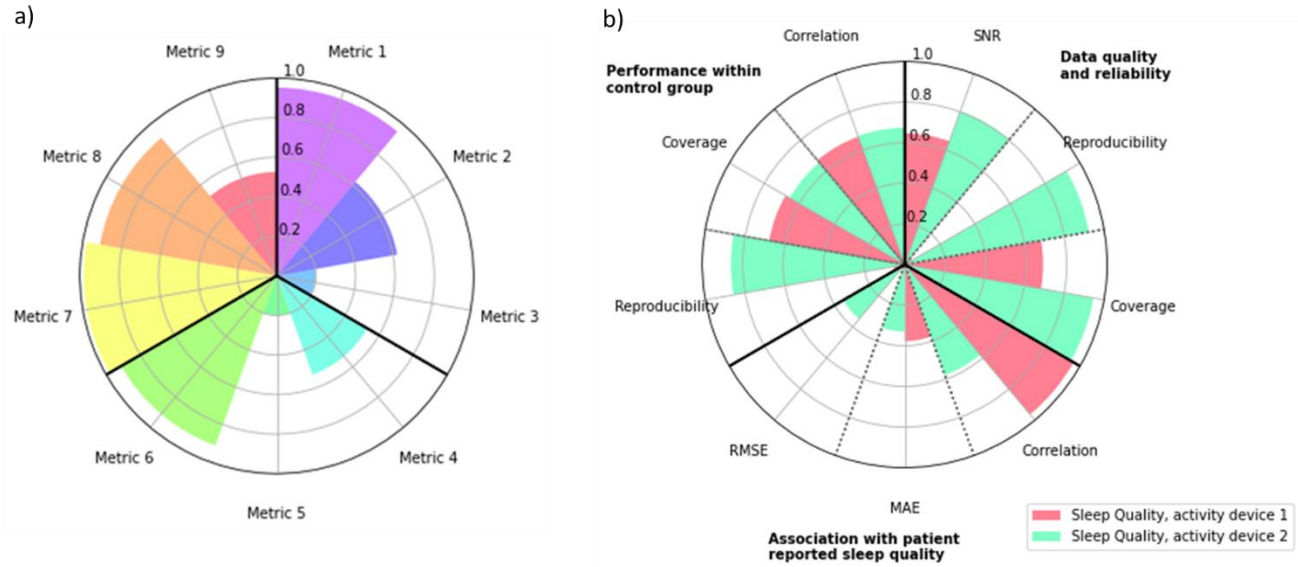

Figure 3: a) The overview on digital measure performance is provided through radar chart visualizations. In this imaginary example, three performance metrics have been selected to represent each analysis category (sector separated by black lines). Each metric is depicted in a different colour. The performance metrics are scaled from zero to one, corresponding to the worst and best possible score, respectively. b) A practical example of aggregated radar charts using imaginary data. Aggregated overviews may be useful when comparing performance metrics between digital measures (within a group of patients), for instance, between the digital measures originating from the same device, as exemplified here. The black solid borders distinguish the analysis categories and the different colours distinguish the digital measures under comparison.

The individual overview charts can be further aggregated or grouped based on, e.g., measurement modality or device to create visual representations on a larger scale. One can, for example, compare the performance of a digital measure when it is derived from different devices (example in Figure 3b), or multiple distinct digital measures originating from the same device. This supports higher level analyses and evaluation of suitable measurement approaches within each group of patients.

5 Conclusions

This document presented the preliminary protocol for assessing the performance of the devices measuring indicators of sleep and fatigue. The assessment is based on three factors: 1) Data quality and reliability of the digital measure, 2) Performance as compared to patient reported outcomes, 3) Performance across patient ('disease-specific') cohorts. The first factor concentrates on the statistical characteristics of the data obtained from each devices and attempts to reveal the reliability and robustness of the data. The second and third factors aim to find associations between the features of the objective data and subjective data including the baseline information of the participants. The associations are based on e.g. error estimates and correlation analysis. The overall performance is provided as a visual aggregate of the results accompanied with a recommendation of the device use in the further studies in the project.

The protocol presented in the current version of this document is a preliminary plan for data analysis and will naturally evolve when actual data will be available. Therefore, the details of the protocol are updated when more understanding of the data characteristics in the feasibility study is gained.

Furthermore, the final version of the protocol will serve as basis for analysis methods for the clinical validation study. This is expected to be in hand at M18 of the project.

6 References

[1] J. Han and M. Kamber, "Descriptive Data Summarization," in *Data Mining: Concepts and Techniques*, 2nd ed., San Francisco, CA: Elsevier Inc., 2006, pp. 51–61.

[2] J. Han and M. Kamber, "Outlier Analysis," in *Data Mining: Concepts and Techniques*, 2nd ed., San Francisco, CA: Elsevier Inc., 2008, pp. 451–453.

[3] J. W. Bartlett and C. Frost, "Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables," *Ultrasound in Obstetrics and Gynecology*, vol. 31, no. 4. John Wiley & Sons, Ltd, pp. 466–475, 01-Apr-2008.

[4] H. Akoglu, "User's guide to correlation coefficients," *Turkish Journal of Emergency Medicine*, vol. 18, no. 3. Emergency Medicine Association of Turkey, pp. 91–93, 01-Sep-2018.

[5] R. Newson, "Parameters behind 'nonparametric' statistics: Kendall's tau, Somers' D and median differences," 2002.

[6] W. D. Smith, R. C. Dutton, and N. T. Smith, "A measure of association for assessing prediction accuracy that is a generalization of non-parametric ROC area," *Stat. Med.*, vol. 15, no. 11, pp. 1199–1215, 1996.

[7] R. L. Lieber, "Statistical significance and statistical power in hypothesis testing," *J. Orthop. Res.*, vol. 8, no. 2, pp. 304–309, Mar. 1990.

[8] J. Z. Bakdash and L. R. Marusich, "Repeated Measures Correlation," *Front. Psychol.*, vol. 8, no. MAR, p. 456, Apr. 2017.

[9] T. Chai, T. Chai, and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature," *Geosci. Model Dev*, vol. 7, pp. 1247–1250, 2014.

[10] J. M. Bland and D. G. Altman, "STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT," 1986.

[11] J. L. Hintze and R. D. Nelson, "Violin plots: A box plot-density trace synergism," *Am. Stat.*, vol. 52, no. 2, pp. 181–184, 1998.

[12] R. L. Ott and M. Longnecker, *An Introduction to Statistical Methods and Data Analysis*, 5th ed. Duxbury, 2001.

[13] S. Nakagawa and H. Schielzeth, "Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists," *Biological Reviews*, vol. 85, no. 4. John Wiley & Sons, Ltd, pp. 935–956, 01-Nov-2010.

Appendix 1 – Case Record Form (for ALL participants) CRF_A

Instructions:

Part A/B: to be completed by research staff

Part C: to be completed by research participants with the exception of C3.1

Part A: Demographics

1. Subject Code: ☐☐☐-☐☐☐

2. Centre Code: ☐☐☐

3. Date of Assessment: ☐☐/☐☐/☐☐☐☐

4. Age: ☐☐☐ years/ ☐☐ months

5. Education (highest level) ☐ primary or no formal education

☐ secondary

☐ professional certificates or diploma (below degree levels)

☐ university undergraduate degrees

☐ postgraduate degrees or higher

6. Height: ☐.☐☐ m

7. Weight: ☐☐☐.☐ kg

8. Heart Rate: ☐☐☐ bpm

9. Blood Pressure (Lying): ☐☐☐ (systolic) / ☐☐☐ (diastolic) mmHg

10. Blood Pressure (Standing): ☐☐☐ (systolic) / ☐☐☐ (diastolic) mmHg

11. Occupation (to include paid or unpaid work):

☐ Retired

☐ Not working (unable to wrk)

☐ Part time

☐ 1 part-time job, ☐ 2-3 part-time jobs, ☐ 4 or more part-time jobs

Please specify _____

☐ Full time

☐ Full time carer

☐ Housewife/Househusband

12. Ethnicity: ☐ Black ☐ Caucasian ☐ Chinese ☐ Indian ☐ Hispanic

☐ Mixed ☐ Other, please specify: _____

Part B: Disease Characterisation

B1: Which subject group does the participant belong to?

☐ **Healthy volunteer**

☐ **Huntington's disease**

☐ **Parkinson's disease**

☐ **Inflammatory Bowel disease**

☐ **Crohn's disease**

☐ **Ulcerative colitis**

☐ **Rheumatoid Arthritis**

☐ **Sjogren's syndrome**

☐ **Systemic Lupus Erythematosus**

B2: Year of Diagnosis (if not known, enter "NK"): ☐☐☐☐

B3: Current treatments for the disease: (freetext)

B4: Other concurrent treatments for other conditions:

B5: Comorbidity

Age	☐ < 50 ☐ 50-59 ☐ 60-69 ☐ 70-79 ☐ ≥ 80	
	Yes	**No**
Myocardial Infarction (history of definite or probable MI (ECG or enzyme changes)		
Congestive Heart Failure (Exertional or paroxysmal nocturnal dyspnea and has responded to digitalis, diuretics, or afterload reducing agents)		
Peripheral vascular disease (Intermittent claudication or past bypass for chronic arterial insufficiency, history of gangrene or acute arterial insufficiency, or untreated thoracic or abdominal aneurysm (≥6 cm))		
Cerebral Vascular Accident or Transient Ischaemia Attack		
Dementia		
Chronic Pulmonary Obstructive Disease		
Connective Tissue Disease		
Peptic ulcer disease (Any history of treatment for ulcer disease or history of ulcer bleeding)		
Hemiplegia		
Moderate to severe Chronic Kidney Disease Severe = on dialysis, status post kidney transplant, uremia, moderate = creatinine >3 mg/dL (0.27 mmol/L)		
Leukaemia		
Lymphoma		
AIDS		
Solid Tumour	If yes, ☐ Localised ☐ Metastatic	
Liver disease ***Severe*** = cirrhosis and portal hypertension with variceal bleeding history, ***Moderate*** = cirrhosis and portal hypertension but no variceal bleeding history, ***Mild*** = chronic hepatitis (or cirrhosis without portal hypertension)	If yes, ☐ Mild ☐ Moderate to severe	
Diabetes mellitus	If yes, ☐ Uncomplicated ☐ End organ damage	

Part C: Generic Assessment

C1: Fatigue Assessment

FACIT-F ***Can be filled out at home after study visit on day 1-3***

Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue Scale (Version 4)

Below is a list of statements that other people with your illness have said are important. **Please circle or mark one number per line to indicate your response as it applies to the past 7 days.**

		Not At All	A Little Bit	Somewhat	Quite a Bit	Very Much
1	I feel fatigued	0	1	2	3	4
2	I feel weak all over	0	1	2	3	4
3	I feel listless ("washed out")	0	1	2	3	4
4	I feel tired	0	1	2	3	4
5	I have trouble starting things because I am tired	0	1	2	3	4
6	I have trouble finishing things because I am tired	0	1	2	3	4
7	I have energy	0	1	2	3	4
8	I am able to do my usual activities	0	1	2	3	4
9	I need to sleep during the day	0	1	2	3	4
10	I am too tired to eat	0	1	2	3	4
11	I need help doing my usual activities	0	1	2	3	4
12	I am frustrated by being too tired to do the things I want to do	0	1	2	3	4
13	I have to limit my social activity because I am tired	0	1	2	3	4

Multi-dimensional Fatigue Inventory (MFI) ***Can be filled out at home after study visit on day 1-3***

Instructions:

By means of the following statements we would like to get an idea of how you have been feeling **lately**.
There is, for example, the statement:

"I FEEL RELAXED"

If you think that this is **entirely true**, that indeed you have been feeling relaxed lately, please, place an **X** in the extreme left box; like this:

yes, that is true ☒1 ❑2 ❑3 ❑4 ❑5 **no, that is not true**

The more you **disagree** with the statement, the more you can place an **X** in the direction of "no, that is not true". Please do not miss out a statement and place only one **X** in a box for each statement.

1	I feel fit.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
2	Physically, I feel only able to do a little.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
3	I feel very active.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
4	I feel like doing all sorts of nice things.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
5	I feel tired.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
6	I think I do a lot in a day.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
7	When I am doing something, I can keep my thoughts on it.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
8	Physically I can take on a lot.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
9	I dread having to do things.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
10	I think I do very little in a day.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
11	I can concentrate well.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
12	I am rested.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
13	It takes a lot of effort to concentrate on things.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
14	Physically I feel I am in a bad condition.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
15	I have a lot of plans.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
16	I tire easily.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
17	I get little done.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
18	I don't feel like doing anything.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
19	My thoughts easily wander.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true
20	Physically I feel I am in an excellent condition.	yes, that is true	❑1	❑2	❑3	❑4	❑5	no, that is not true

Fatigue Visual Analogue Scale (fVAS)

Mark on the line to represent the level of abnormal fatigue (tiredness) that you have experienced today:

Score = ☐☐☐ mm

C2: Sleep Assessment

Epworth Sleepiness Scale Can be filled out at home after study visit on day 1-3

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired?

This refers to your usual way of life in recent times.

Even if you haven't done some of these things recently try to work out how they would have affected you.

Use the following scale to choose the **most appropriate number** for each situation:

0 = would **never** doze
1 = **slight chance** of dozing
2 = **moderate chance** of dozing
3 = **high chance** of dozing

It is important that you answer each question as best you can.

Situation	**Chance of Dozing (0-3)**
Sitting and reading	
Watching TV	
Sitting, inactive in a public place (e.g. a theatre or a meeting)	
As a passenger in a car for an hour without a break	
Lying down to rest in the afternoon when circumstances permit	
Sitting and talking to someone	
Sitting quietly after a lunch without alcohol	
In a car, while stopped for a few minutes in the traffic	

Pittsburg Sleep Quality Index Can be filled out at home after study visit on day 1-3

Instructions: The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. **Please answer all questions.**

1. During the past month, what time have you usually gone to bed at night? _____
2. During the past month, how long (in minutes) has it usually taken you to fall asleep each night? _____
3. During the past month, what time have you usually gotten up in the morning? _____
4. During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.) _____

5. During the past month, how often have you had trouble sleeping because you…	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
a. Cannot get to sleep within 30 minutes				
b. Wake up in the middle of the night or early morning				
c. Have to get up to use the bathroom				
d. Cannot breathe comfortably				
e. Cough or snore loudly				
f. Feel too cold				
g. Feel too hot				
h. Have bad dreams				
i. Have pain				
j. Other reason(s), please describe:				
6. During the past month, how often have you taken medicine to help you sleep (prescribed or "over the counter")?				
7. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?				
	No problem at all	Only a very slight problem	Somewhat of a problem	A very big problem
8. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?				
	Very good	Fairly good	Fairly bad	Very bad
9. During the past month, how would you rate your sleep quality overall?				

	No bed partner or room mate	Partner/room mate in other room	Partner in same room but not same bed	Partner in same bed
10. Do you have a bed partner or room mate?				
	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
If you have a room mate or bed partner, ask him/her how often in the past month you have had:				
a. Loud snoring				
b. Long pauses between breaths while asleep				
c. Legs twitching or jerking while you sleep				
d. Episodes of disorientation or confusion during sleep				
e. Other restlessness while you sleep, please describe:				

MOS SS acute

For each of the following questions, please mark an ☒ in the one box that best describes your answer.

1. How long did it usually take for you to fall asleep during the pastweek?

0-15 minutes	16-30 minutes	31-45 minutes	46-60 minutes	More than 60 minutes
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

2. On the average, how many hours did you sleep each night during the past week?

Write in number of hours per night: ☐☐

3. How often during the past week did you…

		All of the time	Most of the time	Some of the time	A little of the time	None of the time
a	feel that your sleep was not quiet (moving restlessly, feeling tense, speaking, etc., while sleeping)?	☐	☐	☐	☐	☐
b	get enough sleep to feel rested upon waking in the morning?	☐	☐	☐	☐	☐
c	awaken short of breath or with a headache?	☐	☐	☐	☐	☐
d	feel drowsy or sleepy during the day?	☐	☐	☐	☐	☐
e	have trouble falling asleep?	☐	☐	☐	☐	☐

f	awaken during your sleeptime and have trouble falling asleep again?	☐	☐	☐	☐	☐
g	have trouble staying awake during the day?	☐	☐	☐	☐	☐
h	snore during your sleep?	☐	☐	☐	☐	☐
i	take naps (5 minutes or longer) during the day?	☐	☐	☐	☐	☐
j	get the amount of sleep you needed?	☐	☐	☐	☐	☐

MOS 12-Item Sleep Scale Acute – Revised 2010 United States (English)

Appendix 2 – Description of devices

AX6 Axivity

The AX3 supplied by Axivity is an accelerometer data logger. It features a MEMS 3-axis accelerometer, temperature sensor, light sensor, real time quartz clock and a flash based on-board memory for data storage. Sampling rate and range are configurable, which enable it to be optimized to different uses. In particular, the long battery life and high-resolution data make it ideal for collecting longitudinal movement data in clinical trials. There is a large repository of AX3 data (~100,000 subjects) within the UK biobank and several pharmaceutical companies, including Eli Lilly, are using it within clinical trials. There is also >20 peer reviewed publications on the AX3. The custom .cwa file format is optimized for storing and working with the data. Several libraries are available to access and manipulate the data in this form in addition to the OmGUI. In addition, several companies such as Ixico have developed algorithms for detecting sleep in patient populations (Parkinson's disease) that may be relevant to the consortia. In our experience collection up to 16 days is possible at 100Hz without charge. One major limitation is the requirement to return the device to the site in order to extract the data. In addition, high accuracy time locking between devices or events it is not possible. A manual activity (e.g. 'clap) is required, which has to then be detected as a sync pulse. Finally, there is lack of utility for the subject. The AX6 is broadly like the AX3 with placements and orientations of sensors the same between devices. The AX6 contains a 6-axis accelerometer and gyroscope that records data at a greater bit depth than the AX3 (16-bit).

McRoberts Movemonitor

Movemonitor by McRoberts is an inertial sensing unit that is designed for activity/sleep tracking, classification and energy expenditure assessment. The current version of the device is placed in the lower back of the user and the data is logged into a local memory card. It can log the acceleration values sampled at 100Hz up to 14 days. The activity classification algorithms have been validated in several scientific studies. Out of 17 reported studies, 7 were conducted with COPD patients, 2 with respiratory disorders, 1 with Parkinson disease patients, 1 with elderly, and the rest with healthy users. All together some ~600 subjects have been participating the studies in which the wearing time of the device varied from 15 minutes to 14 days. In activity classification studies, the person's movement was classified into e.g. walking, sitting, standing, and lying. The classification accuracy in the studies varied from 80% to 100%. Similar stationary movements such as standing and sitting tended to get confused with each other. In energy expenditure studies, a relative correlation of ~0.9 with reference was obtained. Movemonitor is technically relatively simple device, that provides reliable activity tracking information. In addition, its rather comfortable due to its small size and unnoticeable location and it can log up to 14 days in one go. On the other hand, the data must be manually extracted from the device.

VTT Bed Sensor

BedSensor is a force-sensitive piezo-electric film that is placed under mattress during sleep. The sensor is developed at VTT Technical Research Centre of Finland and is currently commercially available. The sensor is able to detect heart rate, breathing rate, breathing disturbances and movements of the person lying on the mattress. The sensor has been validated in number of studies (~10) at which the accuracy of heart rate, breathing rate, breathing disorders and sleep staging have been assessed. Heat rate during normal sleep is detected well (90%) and identified breathing disturbances (respiratory events) have been correlating well (r=0.93) with gold standard reference. In sleep staging, the bed

sensor has reached total accuracy of 77% when compared to clinical polysomnography-based classification. Bed sensor is an easy to use, unnoticeable solution for accurate sleep tracking. In addition, it has proved to be sensitive in revealing breathing disturbances such as sleep apnoea episodes.

Vital Connect Vital Patch

The Vital Patch, developed by Vital Connect, is a fully disposable adhesive device attached on the left chest for continuous monitoring of a patient's vital signs. It measures single lead ECG, heart rate (HR), heart rate variability, breathing rate (BR), skin temperature (ST), body posture including fall detection and activity. The Vista Solution platform is a patient monitoring system that consists of the Vital Patch for collection of vital sign data, relay devices like tablet or wall-mounted hub for bi-directional communications using encrypted BLE protocol, and a tablet or computer monitor for real-time viewing of data. The platform can be deployed in hospital/home care and allows for data capture, transfer and storage of encrypted de-identified patient data in cloud server.Vital Patch was validated in 57 healthy individuals in clinic. Participants wore the Vital Patch, an oronasal canula in the nostril connected to a portable Capnography monitor for BR reference, and an Actiheart device with 2 standard ECG electrodes on chest for HR reference. Overall HR accuracy compared to Actiheart device and BR accuracy compared to oronasal cannula were 3.8 ± 3 bpm and 3.1 ± 1.2 brpm respectively. ST accuracy from bench testing was 0.1C for a range of 15-50C; Absolute percent error of step count compared to manual counts was 4.7 ± 4.6%; Accuracy of postures compared to visual annotation were standing: 95.1 ± 5.9%, supine: 96.2 ± 3%, walking: 97.3 ± 7.8%. Specificity and sensitivity of fall detection (n = 20) were 100% and 93.8% respectively.Participants also wore Vital Patch for several days after to evaluate the wear duration. The average wear duration was 110.5+23.9 hours. There are 8 publications related to validation of Vital Patch. It is FDA cleared.

Everion Biovotion

Everion Biovotion is a medical wearable consisting of a multi-sensor platform and an elastic strap. It can be worn on the upper arm where it collects continuous real time data (1 Hz) that can be categorised into 22 features and parameters. It includes PPG (photoplethysmography). The vital sign parameters include heart rate, skin temperature, respiratory rate, blood oxygenation.

Dreem

The Dreem headband (DH) device is a wireless headband worn during sleep which records and automatically analyzes physiological data in real time without any connection (e.g., Bluetooth, Wi-Fi, etc.). After recording, the DH connects to a mobile device (e.g., smart phone, tablet) via Bluetooth to transfer aggregated metrics to a dedicated mobile application and via Wi-Fi to transfer raw data to the sponsor's servers. Five types of physiological signals are recorded via 3 types of sensors embedded in the device: brain cortical activity via 5 EEG dry electrodes; movements, position, and breathing frequency via a 3-D accelerometer located over the head; heart rate via a red-infrared pulse oximeter located in the frontal band. DH was tested on 31 healthy subjects, who wore a PSG and the DH simultaneously, in an overnight study at a sleep center. Pearson correlation for EEG frequencies between DH and PSG were high: alpha (r= 0.71 ± 0.13), beta (r= 0.71 ± 0.18), delta (r = 0.76 ± 0.14), and theta (r = 0.61 ± 0.12). Mean absolute error for heart rate, breathing frequency and respiration rate variability were 1.2 ± 0.5 bpm, 0.3 ± 0.2 cpm and 3.2 ± 0.6 % respectively. Automatic Sleep Staging accuracy was 83.5 ± 6.4% (F1 score: 83.8 ±6.3) for DH compared with an average of 86.4 ± 8.0% (F1 score: 86.3 ± 7.4) from 5 sleep experts. Heart rate variability could not be obtained for most records

due to insufficient resolution. The final dataset (of 25 participants) was small and homogenous in age and sleep profile. There are 2 publications related to validation of DH. The device is CE marked.

Byteflies Sensor Dots

Sensor Dots by Byteflies is a multimodal wearable device to continuously record physiological signals such as the brain (EEG), heart (ECG), muscle activity (EMG), eye movement (EOG) and activity (via a 3-axis accelerometer), in home or in the clinic. Sensor Dot has been validated in 2 studies (currently as 1 publication). The first study focussed on testing specific configuration of the Sensor Dot, i.e. 3 EEG channels recorded behind the ears (post auricular, PA) in a hospital setting. The second study validated the quality of EEG data in an ambulatory setting and compared it to a portable bulkier standard EEG device, the Medatec BrainWalker 3. A clinical review of this study is being prepared for publication. Sensor Dot was tested on 31 patients admitted for pre-surgical evaluation. Video EEG data for 3168 hours was collected, and 169 seizures were annotated on reference data. Epochs of 10 min containing seizures and 200 false-positive epochs of 10 min were selected. Each epoch was scored as a yes/no by 2 reviewers blinded to video EEG data and annotations. For the 2 reviewers, accuracy was 78% and 81% respectively, specificity was 97% and 78% respectively, and sensitivity was 49% and 85% respectively. For at-home validation, 8 people with refractory absence seizures were sent home with Sensor Dot (1 channel PA) and the reference device for 24 hours. 89.7% signal similarity properties of Sensor Dot vs. reference device was obtained. Unusable Data (RMS over 1 s > 500µV), No Signal (RMS over 1 s < 0.5µV), Minor Artifacts (350µV threshold), were low for both devices.A full clinical validation of Sensor Dot is currently pending. ISO 13485/CE Class IIa and FDA approval are also pending.

VTT Stress Monitor (Mobile Phone App)

The VTT Mobile Phone App is an Android mobile application for analysis of behaviour (originally in the context of detection of stress). It gathers mobile usage data and questionnaire data from a user, uses physiological signals from a wearable Android smartwatch (if available) and sends the pre-processed data to a cloud server.

The app was tested on 4 weeks of participants' daily activities collected with a Polar M600 smartwatch and a smartphone. The smartwatch data consisted of the heart rate, inter-beat interval and 3-axis acceleration. The smartphone data consisted of location, messaging, application usage, screen usage status. Additionally, a pop-up questionnaire was collected from user 3 times/day on participant's level of liveliness/sleepiness, calmness/nervousness, excitement/boredness, feeling of control and feeling of recovery in a 1-7 Likert-scale. Data was collected from smartwatch with ~10 min interval, which was then transferred to the phone through Bluetooth and then continuously forwarded to cloud-based server. All data were pseudonymized, and location and application data were anonymized by categorization. Physiological measurements were obtained from 35 participants at night-time (for a total of 441 nights) and phone activity data were obtained from 65 participants at day-time (for a total of 1008 days). Range of F1-scores for prediction of stress detection for personal, semi-personal and general models (at day/night-time/both) varied from 38% - 63%, with prediction scores being mostly above the baseline F1-score of 46.6%, 42.2%, 45.5% (at day/night-time/both respectively).

CANTAB Cognitive Assessment

The Cambridge Neuropsychological Test Automated Battery (CANTAB) provides cognitive tests via touch-screen devices like tablet computers or smartwatches. First developed at the Cambridge University, it is now a software product marketed for research by Cambridge Cognition Ltd. CANTAB

has been studied for more than three decades including research with respect to a broad selection of diseases and disorders, such as Alzheimer's disease, Huntington's disease, major depressive disorder, Parkinson's disease, and others. Overall, they state to have over two thousand peer-reviewed publications. The cognitive tests comprise memory, attention, psychomotor speed, emotion, social cognition, and executive function. The average completion time is approximately ten minutes. CANTAB also features an additional questionnaire for mood assessment according to the GDS-15 assessment for older adults, and another one for activities of daily living. CANTAB has been used both in clinical assessments and in longitudinal studies extending over several weeks. In the latter, the subjects would take the test independently, advised to do so several times a day. CANTAB is CE marked with FDA-clearance. It promises secure data integration into medical records systems; compliance with HIPAA and GDPR, ISO, and GCP standards; and data security through servers in private cloud and data encryption both at rest and during transfer using HTTPS. They also provide voiceover guidance in more than 20 languages and dedicated technical support.

ZKONE

ZKONE is a RF-based sensing device to monitor human respiration and heart rate in a contactless manner. The implementation is state of the art sensing systems based on UWB radar. The system is capable of sensing breathing patterns of a human from 3 meters distance when the person is stationary in sitting or laying position. The performance has been evaluated in an experiment consisting of one person. The reported mean error in breathing frequency detection was 0.24 and 0.25 breathing cycles per minute in laying and sitting positions respectively. In heart rate detection, the error varied from 6-9%. ZKONE has received a CE certificate.