Contemporary Clinical Trials 91 (2020) 105962

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Review

Considerations for development of an evidence dossier to support the use of mobile sensor technology for clinical outcome assessments in clinical trials

M.K. Walton[a], J.C. Cappelleri[b], B. Byrom[c,*], J.C. Goldsack[d], S. Eremenco[e], D. Harris[f], E. Potero[g], N. Patel[h], E. Flood[h], M. Daumer[i]

[a] Janssen Research and Development, Titusville, NJ, USA
[b] Pfizer Inc, Groton, CT, USA
[c] Signant Health, London, UK
[d] Digital Medicine Society (DiMe), Sarasota, FL, USA
[e] Critical Path Institute, Tucson, AZ, USA
[f] Takeda, Boston, MA, USA
[g] Biogen, Cambridge, MA, USA
[h] AstraZeneca, Gaithersburg, MD, USA
[i] Sylvia Lawry Centre for Multiple Sclerosis Research, Munich, Germany.

ARTICLE INFO

Keywords:
Mobile sensors
Wearables
Clinical endpoints
Clinical validation
Analytical validation
Medical product labelling
New drug applications
Evidence dossier
Clinical outcome assessments
Novel endpoints

ABSTRACT

Background: Mobile sensors offer enormous potential for the collection of informative clinical endpoints in clinical trials to support regulatory decision making and product labelling. There are currently no specific guidelines on the information needed to enable regulators to review and accept proposed endpoints derived from mobile sensors for use in drug development trials.

Objective: The purpose of this working group report is to recommend the structure and content of an evidence dossier intended to support whether a clinical endpoint derived from mobile sensor data is fit-for-purpose for use in regulatory submissions for drug approvals.

Evidence dossier: The structure and content of a dossier to provide evidence supporting the use of a sensor-derived clinical endpoint is described. Sections include clinical endpoint definition and positioning, the concept of interest, the context of use, clinical validation and interpretation, study implementation, and analytical validity with sensor performance verification in support of the selected sensor.

Conclusions: In the absence of definitive regulatory guidance, this report provides a considered approach to compiling a comprehensive body of evidence to justify acceptance of mobile sensors for support of new drug applications.

1. Introduction

Measuring patient function is integral to understanding intervention effects and disease progression in clinical trials. Measuring function outside routine clinic visits – for example, by assessing physical activity – has the potential to provide a more complete picture of treatment effects, and an understanding of functioning in real-world settings, both of which aid clinical and regulatory decision making. While patient-reported outcome measures continue to provide a vital insight into the effects of treatment from a patient perspective, they leave an important gap when it comes to a complete assessment of important aspects such as mobility in daily life. The rise in ubiquity of mobile sensors enables other, objective observations to be gained that add additional insights into the effects of disease and its treatment. The majority of innovation to date has focused on the consumer sector, in particular the area of personal health and wellness. The same mobile technology has great potential in clinical research to measure the effects of clinical interventions, and support labelling claims.

Despite this increased interest in using mobile sensors in clinical research, the biopharmaceutical industry has been slow to adopt new mobile sensor technologies, including wearables, to measure clinical outcomes in drug development programs. Reasons for slow uptake include: limited regulatory guidance and lack of precedence, the absence of standards and established approaches to implementation, lack of appropriate resources to manage the product validation, aversion to changing existing practice, complexities in how to extract meaningful

* Corresponding author at: Signant Health, Waterfront, Hammersmith Embankment, Manbre Road, London W6 9RH, UK.
E-mail address: bill.byrom@signanthealth.com (B. Byrom).

https://doi.org/10.1016/j.cct.2020.105962
Received 21 November 2019; Received in revised form 5 February 2020; Accepted 17 February 2020
Available online 20 February 2020

and valid parameters from the raw data, and uncertainty around how to define and analyze derived endpoints based on these parameters. The lack of a consensus approach to implementation was illustrated in a review of the use of mobile accelerometers in the assessment of activity in chronic obstructive pulmonary disease (COPD) clinical research studies [1]. Across the 76 studies examined there was considerable variation in wear location, wear time and the physical activity endpoints studied. Such operational variability will make it difficult or impossible to perform meaningful meta-analyses and comparisons between studies, and can increase uncertainty and inaccuracy in the interpretation of results.

To address concerns relating to regulatory acceptance of data collected using mobile sensor technology, a number of initiatives have begun to provide a useful framework, in particular recommendations developed by the Clinical Trials Transformation Initiative (CTTI) [2,3] and the Critical Path Institute's Electronic Patient-Reported Outcome (ePRO) Consortium [4]. Recognizing the value in expanding the work conducted by CTTI and the ePRO Consortium, a working group was convened among members of the Drug Information Association's (DIA) Study Endpoint Community, CTTI, the ePRO Consortium and the Digital Medicine Society (DiMe) including experts from biopharmaceutical companies, eClinical technology providers, clinical research organizations (CROs), academia and non-profit organizations. The working group seeks to provide further recommendations and guidelines to facilitate the adoption of mobile sensor technology in clinical trials, allowing for the objective demonstration of treatment benefit in a real-world setting.

This paper proposes a structure for an evidence dossier suitable for submission to regulatory bodies to support the use of mobile sensor technology in regulatory decision-making. The recommended evidence dossier structure is similar in concept to the patient-reported outcome (PRO) dossier structure, which identifies specific information important to regulatory authorities in their review of PRO measures (PROMs) used in approval and labelling decisions [5], and identifies the evidence needed to support both the endpoint derived from the mobile sensor and the mobile sensor technology itself. However, mobile sensor derived endpoints that are the focus of this dossier, have important differences to PROMs, and this is reflected in differences in the proposed dossier structure and content. As the use of mobile sensor technology within clinical trials is still developing, sponsors are encouraged to have early discussion with regulators about their intended use. The proposed evidence dossier provides a structure that may be helpful in guiding this discussion and highlighting areas of specific attention recommended during regulatory consultation. The dossier may also provide a structure to support a submission through FDA and EMA's respective Drug Development Tool Qualification Programs.

1.1. Scope

In this manuscript we refer to mobile sensor technology, or mobile sensors, to describe a multitude of types of devices that can be used to provide measurements related to human health status (Table 1). A sensor, or microsensor, “detects and measures physical or chemical information from a surrounding physical environment and translates this into an electrical output signal” [4]. Such measurements may include light, heat, acceleration, rotation (motion), moisture, pressure, chemical content, or others. While PRO data can be collected using mobile technologies, this is not within the scope of this report. Firmware within the sensor device and/or software operating externally to the sensor device can filter the raw sensor data, translate the raw data into understandable physical units, and in some cases algorithmically extract parameters from the data to provide measurements that will be used to form the data of the outcome measure.

While some mobile sensors are able to generate measurements for diagnostic purposes, the focus of this report relates to the use of mobile sensors to generate data that are used to compose clinical endpoints, both efficacy and safety, to support the regulatory submissions and labelling associated with new interventions and treatments. Development of a robust approach is needed in order to provide evidence in a form that is suitable for regulatory reviewers.

1.2. Dossier context and definitions

The proposed evidence dossier contains a detailed description of the mobile sensor technology(ies) intended to be used for measuring the endpoint, how the technology and resulting data will be used, and the evidence to support acceptance of the endpoint. This dossier content and structure is chiefly targeted for new endpoints that are intended to be used in therapy development and be key evaluations of the clinical benefit accepted by regulatory agencies for making decisions on treatment approval and labelling. For endpoints developed as exploratory for regulatory purposes, these topics are also important to consider but submission of supportive evidence to regulatory agencies will generally not be needed.

For acceptance as a descriptor of clinical benefit, an endpoint must directly or indirectly reflect changes in a meaningful aspect of a patient's health [6]. For endpoints based on mobile sensor data, the meaningful aspect of health will be a particular aspect of how a person functions in ‘daily’ life. As part of the dossier, the validity of the relationship between concept of interest ([6], see Section 2.3) of the clinical trial(s) and the meaningful aspect of health will need to be demonstrated. In some cases, the concept of interest may itself be the meaningful aspect of health, such as with measuring certain aspects of daily walking. Use of mobile sensors may be considered as an approach to measure the concept of interest, especially when the intended meaningful aspect of health may be difficult to measure easily, reliably or with the required precision using other means. By the time of completing this dossier for submission, developers will have considered how this clinical endpoint would be used in the developing estimand framework [7] when this endpoint is used in confirmatory trials. Features of the clinical endpoint development, such a context of use, will influence the estimand description in aspects such as population for

Table 1
Categories of mobile sensor technologies of interest for use in endpoint data generation for clinical drug development.

Category	Definition	Examples
Wearable sensor	Mobile sensors incorporated into clothing or accessories that can be worn on the body such as adhesive patches, wristbands, belts, headbands, contact lenses, and glasses.	• Wrist-worn accelerometer • Patch worn to measure heart rate • EEG headband
External mobile sensor	Devices that the user can interact with but are not worn, implanted or ingested.	• A portable spirometer • A motion-sensing camera
Implantable sensor	Devices that are inserted into the body.	• Cardiac arrhythmia monitor • Brain fluid pressure sensor
Ingestible sensor	Sensors that are swallowed by the user.	• Core body temperature sensor • Medication ingestion tags [29] such as those incorporated in the first FDA approved digital pill for Abilify MyCite [30]

study, intercurrent events that may be planned for, and the population-level summary.

The overall evaluation of fit-for-purpose related to a sensor technology comprises three key components: 'verification', 'analytical validity' and 'clinical validation'. By 'verification' we refer to demonstration that the sensor technology provides raw data that has adequate technical performance characteristics such as accuracy, reliability, precision, consistency over time, uniformity across mobile sensor generations and/or technologies, and across different environmental conditions. Raw data refers to the data provided by some sensors before processing by firmware or software, for example the magnitude and timing of accelerations measured in g-force detected on each axis of a tri-axial accelerometer.

'Analytical validity' establishes the outcomes data generated by the mobile sensor technology firmware and any associated software, have adequate technical performance characteristics such as their accuracy, reliability, precision, consistency over time, uniformity across mobile sensor generations and/or technologies, and across different environment conditions that might be encountered. For example, this may refer to demonstration that a sensor technology designed to count steps is able to adequately translate the g-forces detected on each axis of a tri-axial accelerometer into a step count using its associated firmware and software.

By 'Clinical validation' we mean the process of establishing that the processed sensor data as used to comprise the endpoint acceptably identifies, measures, and/or predicts the intended meaningful aspect of a person's functioning. There are several distinguishable types of validity that together comprise the totality of 'clinical validation' and should be addressed in the dossier.

Evidence of 'content validity' aims to demonstrate that treatment effects observed on the endpoint will reflect meaningful treatment effects on the aspect of health of particular interest. When the concept of interest and the meaningful aspect of health are the same, meaningfulness may be clear and content validity is more straightforward to demonstrate. Evidence of 'construct validity' demonstrates that the filtered, transformed, or otherwise processed raw sensor data are accurately measuring the concept of interest that is intended to be directly measured. 'Concurrent validity,' demonstrating that the new endpoint is shown to be related to other measurements that are accepted as measuring the aspect of health when measured concurrently, is one approach to support construct validity.

The 'reliability' of sensor data is important to achieve reliable clinical endpoints derived from them. Reliability demonstrates that repeated measurements 1) using the same technology, 2) within the same patient (within-unit or within-subject agreement), 3) during a stable period between different units of the specific mobile sensor technology, or 4) using different units and different patients with the same clinical status (between-unit or between-subject agreement), yield adequately similar values.

'Sensitivity' is an assessment of how small a difference in a measurement of clinical state, either over time in the same patient or between patients, can be observed as a difference in the measured endpoint value.

2. Evidence dossier structure

In efforts to provide researchers with a comprehensive approach to the provision of information to support the use of mobile sensors in clinical endpoint measurement intended for regulatory review and acceptance, we have formulated a generic structure for an evidence dossier. This structure and content may be adapted to fit the needs of the patient population, type of mobile sensor technology being employed, strategy for treatment, and/or transformation of the raw data, and intended interpretation of the endpoint.

The proposed evidence dossier structure is presented in Fig. 1. This proposed structure takes into account that reviewers from different areas within a regulatory agency would be expected to participate in the evidence review, each focusing on different aspects of the evidence provided. Within FDA for example, Section 2.9, which focuses on the verification of the technology itself, will often be reviewed by staff with the expertise usually found in the FDA's Center for Devices and Radiological Health (CDRH), with the clinical use and interpretation information found in the remainder of the dossier reviewed by staff in the Center for Drug Evaluation and Research (CDER) or the Center for Biologics Evaluation and Research (CBER) when the intended medical treatments will be drugs (including therapeutic biologics). While the dossier is intended to be submitted in full, its structure facilities easy dissemination of the different sections by the agency.

The evidence provided in the first part of the dossier (sections 2,2.3,2.4) describe what the endpoint is believed to provide, the circumstances for appropriate use of the endpoint, and how to use the mobile sensor to provide the data to generate the endpoint. Sections 2.5 to 2.8 provide the justification for using and interpreting the endpoint as intended and further details of practical implementation to ensure consistent and correct use of the mobile sensor to measure that endpoint. Section 2.9 provides information on the specific sensor(s) to be used, analytical validity and verification evidence and practicalities of using the mobile sensor. Below we describe the content and considerations for each section of the proposed evidence dossier.

2.1. Section 1. Executive Summary

The evidence dossier should begin with an executive summary describing the content of the dossier, with brief statements describing the endpoint definition, the associated concept of interest and meaningful aspect of health it describes; a brief description of the mobile sensor technology and its fit-for-purpose assessment; the context of use; and evidence to support the clinical endpoint.

2.2. Section 2. Intended Goal

This section of the dossier provides a clear and concise description of the intended goal for the mobile sensor technology in the context of the clinical trial, including the specific endpoint definition(s), where the endpoint will be positioned in the overall endpoint hierarchy, the meaningful aspect of health intended as the interpretation of the endpoint, and the targeted labelling claim or other intended regulatory use of the endpoint based on the mobile sensor.

2.2.1. Section 2.1 Endpoint Definition

The specific endpoint based on the mobile sensor technology is defined in this section. The endpoint definition is a statement describing the clinical data comprising the endpoint and its calculation. Endpoint descriptions should include information defining what the values mean, how and when they are measured, and how they are analyzed. For example, an endpoint for physical activity measured using a mobile sensor could be the difference in mean number of walking episodes per day of at least 2 min duration with a cadence of at least 60 steps per minute at week 12 compared to baseline in a 12-week treatment trial, calculated only for patients providing at least 4 days of 10 h or more wear time per day within each assessment week.

2.2.2. Section 2.2 Endpoint Positioning

The endpoint positioning for the clinical trial should be presented in this section in an endpoint model to show the position of the mobile sensor-related endpoint relative to the other relevant trial endpoints. The endpoint model should include all primary and secondary endpoints that are related to the usage of the proposed endpoint. Endpoint positioning examples can be found in the FDA PRO Guidance Appendix [5].

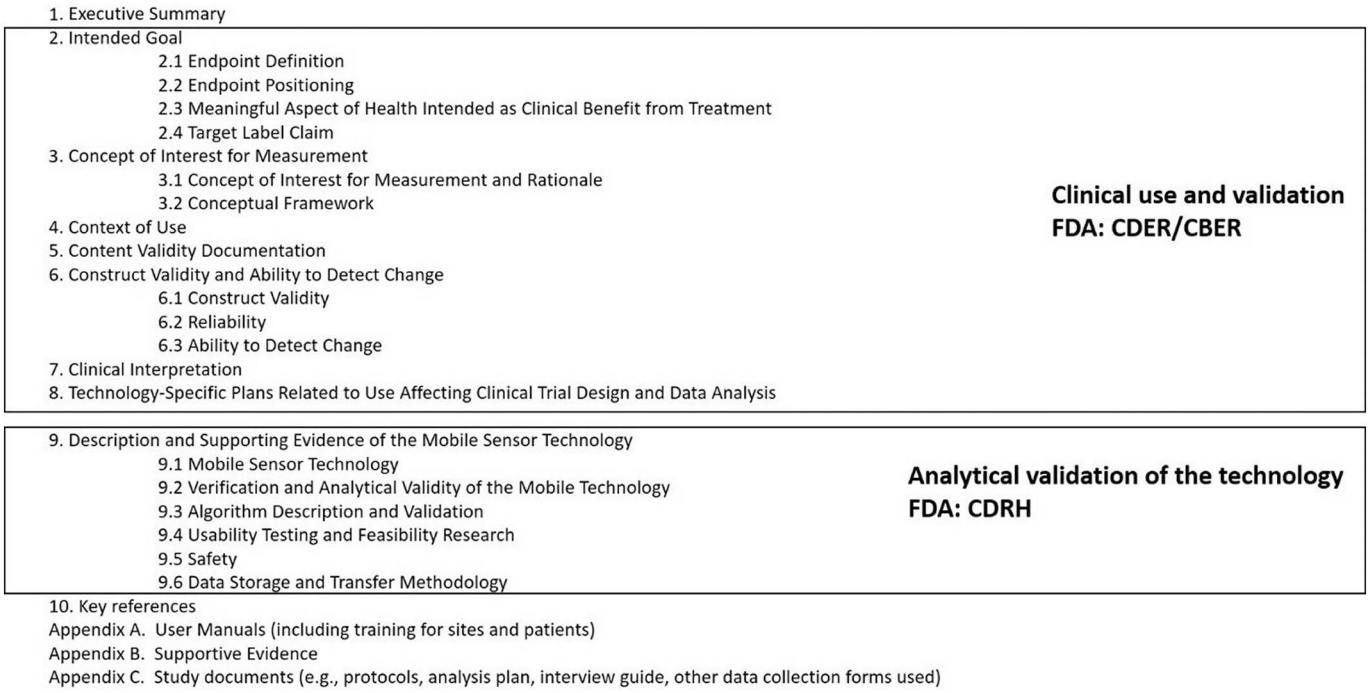

Fig. 1. Structure of evidence dossier to support the use of a mobile sensor to provide data to derive clinical endpoints to support regulatory decision making.

2.2.3. Section 2.3 Meaningful Aspect of Health Intended as Clinical Benefit from Treatment

The aspect(s) of patient health that are meaningful to patients with the relevant disease/condition/disorder(s), and that the endpoint is intended to measure, should be clearly described. These aspects of health should be potentially benefitted by treatment and this reflected by measured effects on the endpoint. These meaningful aspects of health will be the ones shown related to the endpoint as part of content validity evidence (Section 2.5).

2.2.4. Section 2.4 Target Label Claim

This section describes target labelling claims related to the proposed mobile sensor-based endpoint in the context of use (e.g., disease or condition, target population, trial design). For example, if a specific phrase stating a claim of benefit is intended, that proposed phrase should be provided. Iff the endpoint is intended for inclusion within the clinical studies section of the product labelling without further explanation, that intended regulatory use should be stated.

2.3. Section 3. Concept of Interest for Measurement

2.3.1. Section 3.1 Concept of Interest for Measurement and Rationale

The intended clinical benefit will have been stated in Section 2 (Intended Goal). In the regulatory perspective this is “an effect on a clearly identified aspect of how a patient feels or functions. This aspect must have importance to the patient and be part of the patient's typical life” [6].

The sensor-based endpoint may directly measure a meaningful benefit or may measure a more basic impairment caused by the disease/disorder from which the treatment's effects on the meaningful aspect of health of the patient can be inferred (i.e., indirect measurement of meaningful function).

This section will identify the concept that is intended to be directly measured. It is important to clearly state whether the concept being directly measured is the meaningful aspect of health or a concept that will be measured and shown to have a good relationship to the meaningful aspect of health.

For example, patients who have had a stroke, have multiple sclerosis, or have Parkinson's disease may have impaired abilities (in different ways) to use their hand or arm (or both) in performing meaningful daily activities (aspects of health) such as self-care or self-feeding. Improvement in performing those daily activities could be an important treatment benefit for these patients and a valid treatment goal. Additionally, in disorders where treatment is intended to prevent further decline in function, the meaningful change may represent prevention of worsening as opposed to improvement. Direct measurement of these activities usually relies on a patient-reported or observer-reported outcome (ObsRO) questionnaire. Improvements on the reliability or sensitivity of these tools could be useful and this provides an opportunity for a sensor-based endpoint to be a worthwhile advance. Adequacy or ease of performance of self-care activities cannot be directly evaluated by mobile sensors, but the more basic actions of motion of the hand and arm could be measured. With appropriate filtering and transformation of the raw sensor data, the resulting endpoint variable might enable inferring changes in the meaningful functions. For example, measurement of Parkinson's tremor, ataxic movements in multiple sclerosis, or range of motion in stroke might be shown to have validity for inferring effects on some selected daily activities. In this example the selected hand and arm motion, reported by the sensor as the endpoint variable, is the concept of interest for measurement.

When the endpoint (defined in Section 2.1) is directly measuring the meaningful function of patients, the relationship between the concept of interest for measurement and the meaningful aspect of health will usually require little to no explanation, but can be stated directly.

In contrast, if the meaningful aspect of health is not being directly measured, the concept that will be measured should be clearly stated and accompanied by the rationale for why treatment effects on the endpoint are expected to support inferring a treatment benefit on the meaningful functions. Thus, an explanation of the relationship between the concept of interest for measurement and the intended treatment benefit should be provided. For example, for use of a wearable accelerometer to measure specific aspects of movement (physical activity), explaining the relationship of the specific movement measured by the mobile sensor to the larger intended interpretation of functioning in daily life is needed.

Table 2
Examples of intended treatment benefit and concepts of interest that could be measured using mobile sensor technology.

Disease indication	Intended treatment benefit	Potential Concept(s) of interest
Chronic Obstructive Pulmonary Disease (COPD)	Improved distance and pace walked without resting	• Length of episodes of walking with defined criteria of speed/ cadence and minimum duration.
Parkinson's disease (PD)	Improved performance of upper-extremity-related activities of daily living	• Hand and forearm tremor • Functional reaching volume • Dexterity
Sarcopenia	Reduction in falls and fall risk	• Stride length variability • Gait symmetry

Irrespective of whether the endpoint is directly or indirectly measuring the intended treatment benefit, evidence of content validity to support the relationship between the endpoint measurements from the mobile sensor and how a patient functions (and thus treatment benefit) will be presented in Section 2.5 to demonstrate that the expectation is adequately justified.

Examples of treatment benefits and associated concepts of interest in the context of data collected using a mobile sensor technology are provided in Table 2.

The endpoint definition should describe how the measurements reported by the mobile sensor are used as the endpoint variable. When measurements are transformed from the sensor's reported measurement to provide and define the endpoint variable, the process should be fully described as part of the concept of interest. For example, measurements may be transformed into a score within a defined scale; several measured actions might be combined into a composite score that becomes the endpoint variable; or other transformations might be developed. The rationale for any specific scoring or transformation of the clinical measurements into the endpoint variable should be described.

The relationship between meaningful aspects of health, the concept of interest, and mobile sensor measurements is presented in Fig. 2.

2.3.2. Section 3.2 Conceptual Framework

Often there is value in providing a pictorial depiction of the conceptual framework for the endpoint derived from the mobile sensor in terms of meaningful benefit. This may be particularly useful when the composition of the concept of interest is more complex (e.g., contains more than one sensor- reported component).

The initial activity of endpoint developers often includes identifying the many options for what sensor measurements to obtain, what concepts of interest could be constructed from the sensor measurements, which meaningful aspects of health may be related to each of the concepts of interest. In the process of endpoint development decisions will be made to progressively narrow the options. Fig. 2 provides an example conceptual framework.

2.4. Section 4. Context of Use

Context of use is defined as “a statement that fully and clearly describes the way the medical product development tool [such as a wearable sensor] is to be used and the medical product development-related purpose of the use” [8]. In this section, the sponsor should clearly state context of use elements, such as (but not necessarily limited to) those listed below:

- Disease(s) and target population(s) within the full range of the disease (e.g., the major disease related inclusion and exclusion criteria for trials)
- Clinical trial general design (e.g., randomized, blinded)
- Specific guidelines for obtaining valid measurements with the particular sensor-based endpoint determined during the endpoint development process, such as:
  - o Frequency or study-timing of assessment over the course of the study (e.g., every 4 weeks, at baseline and study end only)
  - o Timing of assessment within a day (e.g., assessed in the morning) or relative to other activities during a day (e.g., not within 1 h of prior vigorous activity)
  - o Duration of assessment at each timepoint of measurement (e.g., for 1–3 h, for 7 consecutive days).

Some components of context of use will be described in other sections and can be briefly noted in this section with reference to the more complete description elsewhere. For example, endpoint positioning and how the endpoint results will be used in the regulatory setting are also part of context of use. The dossier may include some explanation of why the specific boundaries on use have been selected.

2.5. Section 5. Content Validity Documentation

Content validity for mobile sensor-based assessments means ensuring the sensor-derived endpoint, when measured within the stated context of use and analyzed according to the endpoint definition, is adequately related to the intended meaningful aspect of health of the patient. Specifically, this permits inferring treatment benefit effects based on the observed treatment-related change in the endpoint. This dossier section provides the full details of the evidence to adequately demonstrate that the clinical endpoint derived from the mobile sensor data has an adequately strong relationship to the intended meaningful aspect of health and is sufficiently comprehensive for the intended claim of benefit. The evidence for content validity should be relevant for the proposed context of use.

A precise relationship will often not be feasible to determine. The demonstrated relationship must nonetheless be sufficiently clear to justify use of the sensor-based endpoint to infer treatment benefit (see section 2.7). However, the less precise the relationship is, the greater the difficulty to quantify how changes in the endpoint relate to meaningful changes in the intended treatment benefit. Thus, it is useful to demonstrate as precise a relationship as feasible. Possible sources of content validity evidence, while not exhaustive, are provided in Table 3 for consideration.

Demonstrating content validity will need more extensive evidence when the endpoint is not directly measuring the intended meaningful aspect of health than when the endpoint is measuring this directly. However, even when the endpoint seems to be directly measuring the meaningful aspect of health, there may need to be evaluation on the comprehensiveness of the endpoint for the intended meaningful function across the range of patients within the context of use.

In some clinical contexts where a selected concept of interest for measurement may be already well-established and accepted, comprehensive new research may not be needed, and existing literature or other evidence can justify accepting the endpoint. This may be the case, for example, when considering pulmonary function tests in asthma or COPD administered using a portable spirometer. However, if the new sensor-based endpoint uses a different specific assessment method than the previously accepted methods, evidence to support accepting that the mobile sensor-based endpoint is providing essentially the same information as the established methods will be needed.

Although a previous assessment method has been accepted as

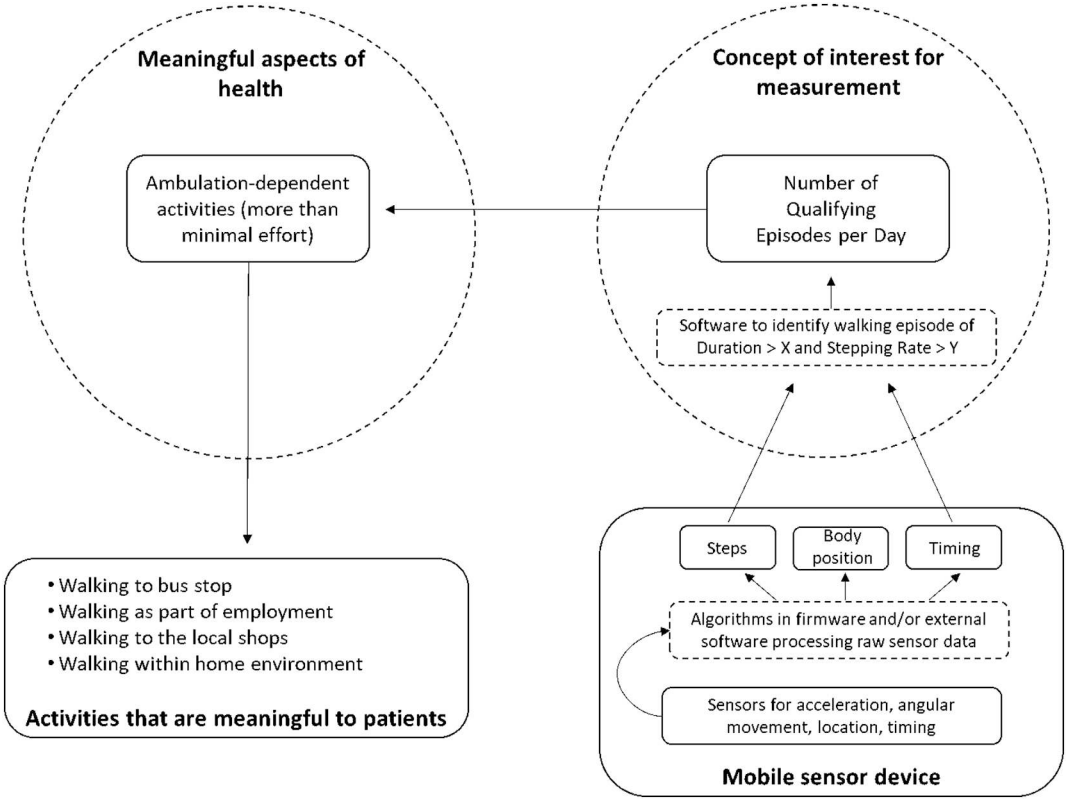

Fig. 2. Conceptual Framework for the Mobile Sensor-based Endpoint included in the Dossier.
Legend: A clear figure illustrating the important concepts and relationships for the proposed endpoint should be provided in the dossier, an example for a COPD activity-based endpoint relating to the disease-imposed limitations of mobility is shown. This figure will illustrate only the methods and concepts selected for the proposed endpoint from among the many options that would be considered at the start of sensor-based endpoint development. Endpoint developers selected ambulation activities of more than very minimal effort as the concept for a meaningful function in daily life that should among the benefits of a new treatment. Several specific examples among such activities are provided in the figure. There must exist a relationship between those activities and the new endpoint so that the meaningful benefit can be inferred from treatment effects on a new endpoint derived from mobile sensor measurements. From among several options for the concept of interest for measurement, selected as the endpoint variable is the daily number of walking episodes that meet criteria of a specific duration (X) and step rate (Y). This was hypothesized as relevant to the meaningful health aspect and viewed as feasible to measure. Endpoint developers also determined how raw sensor data is processed with algorithms that may reside on the sensor unit (i.e., firmware) and/or in external computer software after the data are uploaded from the sensor unit. Sensors and software in this example provide data indicating the occurrence of a step during walking and time-tagging for the physical movement data. Analytical validation would assure that the raw sensor measurements are reliably transformed into the information needed to extract step-events and their timing. Construct validation assures that the primary measurements of steps, timing, etc. are correctly transformed into the concept of interest (total daily number of qualifying walking episodes). Because the mobile sensor of this example can be evaluated only on the step occurrence output, validation of the step occurrence would assure both analytical validity and a portion of construct validity. Content validation provides assurance that there is an adequate relationship between the meaningful aspect of health and the endpoint variable (concept of interest) justifying inferring an impact on the meaningful aspect of health when an effect on the endpoint variable is observed in a clinical trial.

validly reflecting treatment benefit, a mobile-sensor based measurement may nonetheless have significant advantages. Advantages might include feasibility of more frequent or longer measurement providing a better or more accurate description of the patient's true functional status, or be more sensitive, be lower cost or less burdensome to measure (e.g., not requiring frequent clinic visits), among other advantages.

2.6. Section 6. Construct Validity and Ability to Detect Change

2.6.1. Section 6.1 Construct Validity

Evidence to show that the clinical values of the endpoint obtained from the mobile sensor, as filtered and transformed by the specified algorithms, can be interpreted as the intended concept of interest for measurement (i.e., construct validity) should be provided. For example,

Table 3
Sources of content validity evidence documentation to support the use of endpoints derived from mobile sensor data.

Item	Type	Description
Literature review	Qualitative and/or quantitative	Summary of key works in peer-reviewed literature illustrating the ability of the endpoint, or very similar endpoints, to reflect differences or changes in the meaningful function in the patient population or similar population(s)
Patient interview	Qualitative	Patient interviews to assess the patient's perspective of the mobile-sensor-derived endpoint and its relationship to the meaningful function
Expert consensus	Qualitative	Clinician/expert interviews to assess the relationship between the mobile-sensor-derived endpoint and the meaningful function

if a mobile sensor is intended to provide a count of steps taken by the patient during the period of assessment, construct validity evidence should demonstrate that the values provided by the device (and transformed by firmware and other associated software) are an accurate detection and count of steps. This may be accomplished by comparison to another measurement method such as a pressure-pad walkway or video capture. A mobile sensor intended to assess hand tremor as tremor amplitude and/or frequency will need to demonstrate that the reported data do indeed depict the amplitude or frequency (or both) of tremor. Construct validity will be important irrespective of whether the endpoint is directly or indirectly reflecting the meaningful aspect of health.

For some sensor-based endpoints, construct validity and analytical validity are so integrated that they are established simultaneously. For example, the construct validity of an accelerometer-based endpoint of total steps per day requires demonstration that the sensor device is able to adequately determine the occurrence and timing of steps, which is essentially analytical validity of the sensor technology.

Evidence for construct validity will often rely on a concurrent validity approach. Other assessments already accepted as evaluating the desired clinical variable will be measured at the same time as the sensor measurements and seen to be adequately similar, typically through correlational or regression analyses. Researchers should identify how closely associated a sensor measurement should be with the comparator method and defend their definition of acceptance criteria for demonstration of construct validity.

2.6.2. Section 6.2 Reliability

Evidence that a sensor-based endpoint measurement remains stable in the absence of any change in the measurement concept should be presented. This may be demonstrated by measurement of test retest reliability in individuals in which the measurement concept is stable, which might be demonstrated by the use of a concurrent measurement of known reliability that assesses the same concept of interest or during a short enough pre-treatment period where the health condition of patients remains unchanged. Reliability assessment should include the use of multiple units of the mobile sensor to account for inter-unit variability.

2.6.3. Section 6.3 Ability to Detect Change

Evidence that a sensor-based endpoint can identify differences in scores over time in individuals or groups (similar to those in the clinical trials) who have changed with respect to the measurement concept should be presented. A review of the ability to detect change within an individual patient includes evidence describing the size of improvements and deteriorations across the measurement range expected for the target patient population that can be reliably detected. This could most rigorously be demonstrated through controlled studies involving an intervention that is known to create a change in the concept of interest for measurement. Evaluation with concurrent measurement using other, reliable, assessments of the same concept that indicate that the patient's state has changed with respect to the concept of interest can enable identification of true change and understanding comparative responsiveness. Evidence should demonstrate sensitivity to gains and losses in the concept of interest and to change across the entire range expected for the target patient population.

The ability to detect between-patient differences, however, cannot be assumed equal to the ability to detect within-patient changes. Evaluation of sensitivity of each type of differences might be a component of exploratory studies in the target patient population (e.g., phase II), obtained in observational trials, or data on endpoint sensitivity may be available in the peer-reviewed literature for the target patient population or similar (representative) patient groups if the technology and proposed endpoint have been evaluated previously. Although they cannot be considered equal, a highly sensitive measure that detects between-patient differences will often also be highly responsive to detect within-patient changes. An unreliable measure, such as one with a low test-retest reliability, will often be unlikely to be responsive to detection of change.

2.7. Section 7. Clinical Interpretation

Clear interpretation of measurements returned by mobile sensor technologies is central to their use and necessary in designing clinical trials; evaluating interventions; informing health policy makers involved with regulatory, reimbursement, and advisory agencies; and educating consumers of the product label such as healthcare providers and patients. Without interpretation guidelines, the clinical meaningfulness (clinical significance) of statistically significant improvements in an endpoint derived from a mobile sensor cannot be ascertained confidently.

For interpreting group-level change, the detectable difference in sensor scores between two treatment groups can be obtained [9–13]. If the distance between the group means exceeds this detectable threshold (and statistical significance is achieved), then the treatment is considered to provide a clinically detectable effect at the group level. Meaningful treatment effects, however, occur or do not occur in individual patients. Thus, a group-level criterion of difference should be accompanied by within-patient criterion for meaningful difference. For example, consider a treatment effect on an endpoint ranging from 1 to 100, and assume that each individual patient in the control group does not change. A treatment effect of mean group-level difference of 2 points may have no importance if it is due to an uninterpretable 2-point change for each of the treated patients but may be clinically very meaningful if due to 20% of the patients having a 10-point change, with the remaining 80% having no change.

Although no method is without limitations, anchor-based methods have generally been preferred for the determination of clinically meaningful change. Distribution-based statistical methods primarily describe the ability to reliably detect a group-level change or difference for the specific population evaluated. Examples include using thresholds such as 0.5 x standard deviation and 1 x standard error of the mean [13–16]. Distribution methods can serve as a useful adjunct to anchor-based methods, but do not establish whether that difference is clinically meaningful to the individual patients. It is important that the definition of meaningful change be determined in advance of data analysis intended to be evidence of treatment efficacy. Whenever feasible the determination should be separate from clinical studies in which the sensor-based endpoint is to be used to support labelling claims and other regulatory decision-making.

Clinical responders are individuals that achieve at least the within-patient change in the clinical endpoint that distinguishes a meaningful change from one that may be insufficient to be considered meaningful. In practice, for example, individuals would be identified as responders or non-responders by evaluating their change scores from pre-treatment to post-treatment against the clinically meaningful within-patient change threshold.

The clinically meaningful within-patient change threshold is estimated from data generated by anchor-based studies using, for example, descriptive methods or regression methodology [11–13,16–20].Such approaches should be supplemented with empirical cumulative distribution functions and probability density functions [5,17,20–21]. The results of these analyses should be included within the evidence dossier to support the proposed responder definition.

The clinically detectable difference in group means or the clinically meaningful change within patient may be estimable using results of Phase II or other studies designed to assess endpoint changes. Well-designed observational studies can also be a good setting to develop this information. In some cases this information may be obtained from the scientific literature. Not many examples of such estimation have been reported for endpoints derived from common mobile sensors. Demeyer [22] and Motl [23] report estimates for changes in total steps per day in

COPD patients (based on distribution methods) and multiple sclerosis (based on a cross-sectional association with two PROMs), respectively. More robust approaches using associations with within-patient changes in anchor measures should be employed for inclusion in an evidence dossier to support regulatory evaluation. While interpretation of group mean differences can be useful in certain situations, a greater emphasis from regulatory bodies is in the understanding and characterization of meaningful within-patient change (responders) [17].

Judging the overall clinical importance of a treatment effect observed in phase III studies using the endpoint can be aided if there are guidelines for the amount of change that can be viewed as minimally, moderately, and highly clinically important to patients. Although not essential to primary acceptance of a sensor-based endpoint as a valid efficacy endpoint, reaching benefit-risk judgments of future therapies can be aided with such guidelines when well founded.

2.8. Section 8. Technology-Specific Plans Related to Use Affecting Clinical Trial Design and Data Analysis

Endpoints derived from mobile sensor technology data should be developed with careful consideration of how the mobile sensor will likely be used in future clinical research. In this section, it is important to provide in detail critical information describing how patient assessments using the mobile sensor will be implemented in a consistent and reliable manner within a clinical trial protocol, and to provide detail on data management and statistical analysis. This promotes standardization that supports relying on the endpoint in future clinical studies for regulatory decision-making. This standardization also aids interpreting between-study comparisons (systematic reviews) and pooling of data from multiple studies (individual patient data [IPD] meta-analyses). In principle it would be desirable to have sufficiently open access to the raw data and algorithms used to extract the parameters. Information to consider outlining in this section is described below, and an example provided in Table 4.

Clinical trial design considerations

- Usage protocol
  - o Data collection location (e.g., during site visit or remotely at home).
  - o Technology wear location and attachment method for sensors to be worn by the patient; and usage instructions for non-wearable sensors.
  - o Wear/usage period(s) – e.g., worn continuously for the treatment period, or worn for X days at baseline and continued at specific X-day periods at points during treatment. This may include requirements for weekend and weekday wear/use where outcomes may be different at different times of the week – e.g., activity patterns among working adults.
  - o Time periods or timepoints when the mobile sensor technology should be used during each day within the usage period.
  - o Technology configuration settings, where appropriate.
  - o Technology removal and replacement times – e.g., before going to bed and after getting out of bed, if relevant.
  - o Patient instructions in the event of adverse effects (e.g., skin irritation) or sensor malfunction.
  - o Whether data collected are blinded to subject and/or investigator, or (partially) shared with the subject/investigator at specific points in the study.
- Usage monitoring
  - o Compliance requirements, alerts, and reporting. For example, minimum usage/wear requirements and methodologies for monitoring and prompting sensor usage.
  - o Compliance monitoring (protocol should specify who is responsible for monitoring compliance and ensuring that subjects are using the sensor consistently, and follow up measures for non-compliance).
- Data oversight
  - o Ability to track data flow.

Table 4
Example content for technology-specific plans related to use affecting clinical trial design and data analysis for a wrist worn accelerometer to calculate average steps per day.

Section	Item	Description
Clinical trial design considerations		
Usage protocol	Data collection location	Remote, home-based use
	Wear location	Wrist-worn, non-dominant hand
	Wear period	Worn for 7-day intervals at baseline and following visits 4, 8 and 12.
	Wear time	The sensor unit should be worn from the time of getting up until bedtime. In the case of removal, the sensor unit should be worn for a minimum of 10 h during each wear day. Non-wear intervals are estimated by periods of 60-min of zero counts with up to two 1-min epochs of up to 100 counts per minute.
	Configuration settings	Data recorded to 15 s epochs. Sampling rate 100 Hz.
	Removal and replacement times	Sensor unit can be removed at bedtime, and replaced on getting up.
	Data blinding	All data are blinded to the subject. Investigator has web portal access to step counts collected during remote monitoring periods.
Usage monitoring	Compliance requirements	Four valid days out of seven within each collection period are required. A valid day is defined as a day with at least 10 h of wear time.
	Compliance monitoring	Investigator email alerts for patients with at least 2 consecutive non-wear days. Investigation follow up by telephone to encourage compliance is recommended.
Data oversight	Investigator data rights	Read access and ability to flag/comment against data.
Device firmware / software	Firmware version	Define device firmware version.
	Software version	Define device software version.
Data management and statistical analysis considerations		
Endpoint calculation	Sensor data derivation	Number of steps per day and daily wear time are derived directly from the device firmware and/or software.
	Endpoint calculation	Calculation of change from baseline in mean total steps per day at each post-baseline assessment interval.
	Missing data rules	Number of steps per day standardized to a 16-h wear interval. Days with less than 10 h wear time are discarded. 7-day intervals with less than 4 valid days are discarded.
	Implausible data rules	Outliers suspected to represent the sensor unit worn by another individual are detected by (state methodology) and discarded.
Statistical analysis detail	Analysis approach	Repeated measures longitudinal models that may include variables such as baseline disease state, treatment, time, treatment-by-time interaction and other covariates (e.g., sex and gender), depending on the objectives of the study.

- Ability of investigator to mark, modify or delete data collected by the mobile sensor.
- Eligibility assessment
  - Where the mobile sensor technology is used to determine eligibility to participate, or to define study groups/strata, details of the considerations and calculations made.
- Technology settings and firmware
  - Firmware version and any configuration settings that are pertinent should be described. Assurance that the relevant firmware/software that generates the processed data is accurate, precise, consistent, and uniform, in particular across version updates should be provided where appropriate.

Data management and statistical analysis considerations

- Endpoint calculation
  - How the raw sensor data are transformed in the calculation of the clinical measurement. Where available, this should include the definition of data filtering and algorithms for endpoint variable derivation from the sensor's raw data. In particular, this should include a detailed description of any and all transformations that have taken place once the data have been transferred from the device. The firmware version, and associated external software version if any, should also be identified.
  - Process to follow when encountering missing or implausible data due to technology malfunction.
- Statistical analysis detail
  - Covariates that have been observed to have value in endpoint analyses.
- Data Management Plan (DMP)
  - The study-specific data management plans relating to the sensor data.
- Data Transfer Specifications (DTS)

2.9. Section 9. Description and Supporting Evidence of the Mobile Sensor Technology

Communicating the performance characteristics (e.g., sensitivity, specificity, reproducibility, accuracy) and suitability of a mobile sensor to measure a meaningful aspect of the disease/condition or treatment is paramount. This section describes the mobile sensor development/selection background including performance characteristics and its suitability for the intended use. The selection of a specific mobile sensor technology for use in collection of clinical endpoints in clinical trials is not dependent upon its market clearance/approval regulatory status. In our view, both approved and non-approved devices are suitable for consideration and should satisfy the evidentiary considerations detailed in this section.

2.9.1. Section 9.1 Mobile Sensor Technology

This section of the dossier provides a high-level description of the mobile sensor technology and any associated companion software. Some considerations for content are listed below.

- Technology definition
  - Generic/Trade Name(s) & Model(s): Detail the make(s) and model (s) of the technology, including firmware version where appropriate.
  - Manufacturer information (e.g., name(s), contact information).
  - Marketing approval status and indication for use (if available/applicable).
- Technology specification
  - Measurement(s) provided and their units. For multi-sensor devices it may be possible to limit these to those measurements relevant to the implementation and endpoints(s) derived. For example, for a tri-axial accelerometer used to measure activity this may include raw accelerations in the three axes (x, y, z), counts (proprietary unit), steps (number per epoch), metabolic equivalent rates (MET) and other measures.
  - Sensor technology detail, outlining the specific sensor within the unit that provides the measurements – for example, a solid state 3-axis MEMS accelerometer.
  - Firmware version.
- Data security and transmission
  - High-level description of how the data are stored on the device including encryption, and securely transmitted/acquired for inspection by the investigator and provision to the sponsor. Greater detail is provided in section 9.6.
- Software details
  - Description of any companion software required to operate, initiate, and access the data from the sensor-unit along with appropriate evidence of compliance with standards for its use in clinical trials (e.g., compliance with 21 CFR part 11 [24,25]).
- Manufacturing quality
  - Description of quality manufacturing processes to assure consistency of product.

2.9.2. Section 9.2 Verification and Analytical Validity of the Mobile Technology

Sensor verification and analytical validity encompasses the evidence demonstrating that the technology output is a consistent and reliable measurement of what we claim it measures. It involves the physics/mechanics of the sensor component itself, the firmware in the sensor device, and may involve post-download software processing if the firmware does not complete the transformation from raw data to the required clinical outcome measures. This evidence provides assurance that the clinical outcome measures assessed can be used to derive a clinical endpoint as described in the earlier sections of the dossier.

Best practices for measuring and reporting the performance of mobile sensor technologies are well described by the Clinical Trials Transformation Initiative [2] and the ePRO Consortium [4] and should include the components described below.

2.9.2.1. Verification of the sensor technology. In many cases verification evidence may be proprietary to the manufacturer and undisclosed, but when available it provides valuable evidence. When available, verification evidence should include demonstration that the sensor technology provides raw data that has adequate technical performance characteristics such as accuracy, reliability, precision, consistency over time, uniformity across mobile sensor generations and/or technologies, and across different environmental conditions. As described above, raw data refers to the data provided by some sensors before processing by firmware or software, for example the magnitude and timing of accelerations measured in g detected on each axis of a tri-axial accelerometer. Verification is typically conducted by the sensor manufacturer and performed by artificial laboratory testing (e.g., a multi-axis shaking table for accelerometer testing [26]).

Documentation of verification should include both the technology performance characteristics and their limitations. For example, verified within a measurement range of x1 to x2 if calibrated each m months and used within a temperature range of t1 to t2 with battery changes every d days.

Sponsors should ensure that mobile sensor technologies are used within their engineering specifications when deployed for data capture, e.g., that the sensor is used within the manufacturer's recommended range of external temperature conditions.

In some cases, verification evidence may be disclosed as part of regulatory market certification/clearance documentation, where appropriate. In many cases, however, it is acknowledged that verification evidence may be proprietary to the sensor manufacturer and undisclosed, and that raw data may not be made available in addition to the processed outcomes measures provided by the sensor technology. In

this case, this should be stated in this section, and focus should be confined to demonstration of analytical validation. In such cases where verification evidence is not disclosed, we consider analytical validation evidence associated with a representative patient population and being appropriate to the context of use in the clinical trial to be sufficient evidence.

2.9.2.2. Analytical validation. Analytical validation evidence should demonstrate that the outcomes data generated by the mobile sensor technology firmware and any associated software, have adequate technical performance characteristics such as their accuracy, reliability, precision, consistency over time, uniformity across mobile sensor generations and/or technologies, and across different environment conditions that might be encountered. For example, this may refer to demonstration that a sensor technology designed to count steps is able to do so adequately using its associated firmware and software.

Reliability data specific to the measures provided by the mobile sensor may be provided by the technology manufacturer, published (ideally by independent groups) in the literature, or provided in evaluations made by the endpoint developer. Reliability assessment should include intra- and inter-unit reliability assessments. As detailed in section 2.9.1, reliability data should be supplemented with assurance that manufacturing processes follow quality standards to ensure ongoing equivalence of mobile sensors manufactured over time and across batches.

Careful consideration of the applicability of validation evidence across patient populations should be applied. For example, validation evidence of an accelerometer's capability to measure steps obtained in study of healthy volunteers may apply appropriately to many patient populations, but additional data may be needed to support the validation of step counts generated when applied to populations with gait abnormalities (e.g., the shuffling gait associated with Parkinson's disease).

2.9.3. Algorithm description and validation

In some cases, algorithms may be a component of the mobile sensor technology system and as such the precise details may be proprietary and undisclosed. In this case, this should be stated along with the firmware/software versions utilized. The analytical validation evidence provided in Section 2.9.2 will have included the actions of the firmware and associated external software and support the conclusion that the processed data provide reliable and valid measures of the concept of interest. A high-level description of the algorithm should nonetheless be provided where available.

Where an algorithm is used to filter or derive clinical outcome measures for the mobile sensor data, the algorithm and its supporting validation evidence should be described in this section. Aigner et al. [27] provide an example of an approach to validate walking speed measures in a semi-controlled environment. Validation should occur in a patient group that is considered representative of the participant population of interest [3], and algorithm validation evidence may be contained in (and reference) the earlier description of analytical validity.

2.9.4. Section 9.4 Usability Testing and Feasibility Research

This section summarizes any usability testing or feasibility research conducted to support the use of the mobile sensor in the target patient population and within the proposed clinical trial study design. It should provide evidence to support the suitability of the mobile sensor technology for use in this patient population and in this study design.

Usability testing is the assessment of the patient's ability to use the mobile sensor in question and may take place in a clinic or artificial setting [28]. Feasibility research is the assessment of the use of a sensor in the context of a specific study design and assesses how well the patient is able to use the sensor in that context, particularly outside the clinic where he or she has to integrate it into daily life [3,28]. Feasibility research often comprises a separate observational stand-alone study conducted before implementation in a trial to reduce risk [3], but Phase II trials can also be an opportunity to assess feasibility. It is important to evaluate the usability and feasibility of mobile sensor technology in target or representative patient populations for which existing evidence is not available to ensure that the target population is able and willing to use the sensor in a future clinical trial setting.

Usability assessment should include determination of whether the mobile sensor technology can be used effectively by the target patient population. This may include consideration of form factor and wear location / usage requirements. For example, assessment can be made to determine that wrist straps or belts associated with wearables are long enough for obese groups, or short enough for comfortable use in frail older adults. Consideration of ease of use may include operation instructions, the complexity of removing and replacing a wearable sensor unit, how easily patients are able to charge or replace batteries for the mobile sensor, ease of understanding and interaction with the mobile sensor user interface (where appropriate), and patients' ability to transmit data including pairing with other devices such as smartphones where necessary. Regulatory guidance and certain ISO standards provide a guide to other areas that may be important to consider for specific device types.

Feasibility research may assess three areas: (i) the site experience with implementing the device in the study, (ii) the patient experience with use of the device including usage compliance, and (iii) the data quality in terms of missing or erroneous data [3]. While recommended, a feasibility study may not always be needed – especially if similar studies using the mobile sensor have already been conducted.

2.9.5. Section 9.5 Safety

Determination that a mobile sensor is safe for use by patients may be available via the technology manufacturer. The evidence needed is dependent upon the device and its use in clinical studies (i.e., its context of use). The intended manner of use in clinical trials may be different to the manufacturer's marketed use and in this case, it is possible that the existing manufacturer's safety data may be insufficient. The manufacturer may provide evidence of testing in a number of areas including, as applicable:

- Mechanical, electrical, and biological engineering performance, e.g., fatigue, wear, tensile strength, and compression
- Electrical safety and electromagnetic compatibility
- Sterility
- Stability and shelf-life
- Instructions for safe preparation, cleaning and re-use (where appropriate)
- Ability to wear for the required time period, where the mobile sensor is a wearable device. For example, if the device is to be worn in contact with the skin, the materials and metals used should be hypoallergenic and fit-for-purpose, and shown to not result in adverse effects (e.g., skin abrasion or tissue inflammation) when worn for the periods of time required by the study.

2.9.6. Section 9.6 Data Storage and Transfer Methodology

This section should describe the way data are stored and transmitted to the central study database. Data storage and transmission must be secure and encrypted, and this should be described at each step in the data acquisition process: on the mobile sensor, in transmission and within the trial database and other databases (such as a vendor cloud solution) as appropriate.

Data transfer may be accomplished in a number of approaches:

- Data upload during site clinic visits, for example using a USB connection between the mobile sensor and a PC. In this case, the data storage on the PC and method of secure transmission to the study

database or vendor cloud should be described. The PC software used to communicate with the mobile sensor should be described along with its conformance with appropriate regulations such as 21 CRF Part 11 [24,25].

- Data transmission from mobile sensor to a vendor cloud solution, for example using a 3G, 4G or wireless internet connection, or by Bluetooth or near-field communication with a hub device such as a smartphone or home hub. In this case, data storage in each location should be described along with the method of secure encryption used in data transport. The method of data acquisition from vendor cloud to study database should also be described.
- Connected solutions with direct connection of the mobile sensor to the clinical trial database, for example by using a study app running on a smartphone that pulls sensor data by interacting with the mobile sensor Application Programming Interface (API) via a Bluetooth or near-field communication connection, and transmits data to the trial database via 3G, 4G or wireless internet connection. In this case, data storage in each location should be described along with the method of secure encryption used in data transport.

2.10. Section 10. Key references

In this section, the sponsor should list the key literature and data sources referred to in the dossier.

2.11. Appendix A. User Manuals (including training for sites and patients)

This section should contain patient and site user manuals and training materials provided and used in the clinical studies.

2.12. Appendix B. Supportive Evidence

Appendix B should contain patient interview transcripts associated with primary qualitative research conducted to support components of the dossier, literature review findings, expert interview transcripts and detailed reports on qualitative/quantitative work, as appropriate.

2.13. Appendix C. Study documents (e.g., protocols, analysis plan, interview guide, other data collection forms used)

This appendix should contain documentation associated with primary research studies conducted to provide input to the dossier. This may include study protocols and analysis plans, interview guides for qualitative studies etc.

Conclusions

Today advancement in mobile sensor technologies offer innovative ways of measuring patient outcomes and relevant clinical endpoints in clinical trials to support regulatory decision making and product labelling. There are currently no specific guidelines on the information needed to enable regulators to review and accept endpoints derived from mobile sensors. This paper describing a proposed evidence dossier is a first attempt to support regulatory review and acceptance of endpoints derived from mobile sensor data to assess patient functioning as a measure of clinical benefit and the basis for product labelling. There are, however, multiple other stakeholders whose judgment about acceptance and application of study results observed with the endpoint are also important (e.g., healthcare providers, payers). While not the focus of this dossier, the information it contains will be useful to other stakeholders in their decision-making. The working group recommends engaging relevant stakeholders as appropriate throughout the development of mobile sensor based clinical endpoint, and encourages researchers to share the evidentiary data collected to reduce rework and accelerate the acceptance and adoption of clinical endpoints derived from mobile sensor data.

Acknowledgments

The authors thank Takeda Pharmaceuticals International Co. for paying the open access fee associated with this article.

References

[1] B. Byrom, D.A. Rowe, Measuring free-living physical activity in COPD patients: deriving methodology standards for clinical trials through a review of research studies, Contemp Clin Trials 47 (2016) 172–184.
[2] Clinical Trials Transformation Initiative (CTTI), Recommendations on Developing Technology Derived Novel Endpoints, https://www.ctti-clinicaltrials.org/briefing-room/recommendations/developing-novel-endpoints-generated-mobile-technology-use-clinical, (2017) Accessed August 25, 2019.
[3] Clinical Trials Transformation Initiative (CTTI), Recommendations for Using Mobile Technologies in Clinical Research, https://www.ctti-clinicaltrials.org/projects/mobile-technologies, (2018) Accessed August 25, 2019.
[4] B. Byrom, C. Watson, H. Doll, et al., Selection of and evidentiary considerations for wearable devices and their measurements for use in regulatory decision making: recommendations from the ePRO consortium, Value Health 21 (2018) 631–639.
[5] Food and Drug Administration, Guidance for industry on patient-reported outcome measures: use in medical product development to support labelling claims, Fed. Regist. 74 (235) (2009) 65132–65133.
[6] M.K. Walton, J.A. Powers, J. Hobart, et al., Clinical outcome assessments: a conceptual foundation – report of the ISPOR clinical outcomes assessment emerging good practices task force, Value Health 18 (2015) 741–752.
[7] International Council for Harmonisation (ICH), ICH Harmonised Guideline: Addendum on Estimands and Sensitivity in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1), 20 November https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf, (2019) Accessed February 1, 2020.
[8] FDA-NIH Biomarker Working Group, BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet], Food and Drug Administration (US), Silver Spring (MD), 2016 Glossary. 2016 Jan 28 [Updated 2018 May 2]. Co-published by National Institutes of Health (US), Bethesda (MD) https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-C Accessed August 25, 2019.
[9] J.C. Cappelleri, K.H. Zou, A.G. Bushmakin, et al., Patient-Reported Outcomes: Measurement, Implementation and Interpretation, Chapman & Hall/CRC Press, Boca Raton, Florida, 2013.
[10] J.C. Cappelleri, A.G. Bushmakin, Interpretation of patient-reported outcomes, Stat. Methods Med. Res. 23 (2014) 460–483.
[11] H.C.W. de Vet, C.B. Terwee, L.B. Mokkink, et al., Measurement in Medicine: A Practical Guide, Cambridge University Press, New York, NY, 2011.
[12] F.M. Fayers, D. Machin, Quality of Life: The Assessment, Analysis and Reporting of Patient-Reported Outcomes, Second edition, John Wiley & Sons Ltd., Chichester, United Kingdom, 2016.
[13] D. Revicki, H. Rd, D. Cella, et al., Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes, J. Clin. Epidemiol. 61 (2018) 102–109.
[14] D. Cella, D.T. Eton, J.-S. Lai, et al., Combining anchor and distribution-based methods to derive minimal clinically important differences on the functional assessment of cancer therapy (FACT) anemia and fatigue scales, J. Pain Symptom Manag. 24 (6) (2002) 547–561.
[15] G.R. Norman, J.A. Sloan, K.W. Wyrwich, Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation, Med. Care 41 (5) (2003) 582–592.
[16] L.D. McLeod, J.C. Cappelleri, R.D. Hays, Best (but of forgotten) practices: expressing and interpreting meaning and effect sizes in clinical outcome assessments, Am. J. Clin. Nutr. 103 (2016) 685–693 (with erratum).
[17] Food and Drug Administration, Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 3: Select, Develop or Modify Fit-for-Purpose Clinical Outcome Assessments, https://www.fda.gov/media/116277/download, (2018) Accessed August 25, 2019.
[18] C.D. Coon, J.C. Cappelleri, Interpreting change in scores on patient-reported outcome instruments, Ther. Innov. Regul. Sci. 50 (2016) 22–29.
[19] C.D. Coon, K.F. Cook, Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores, Qual. Life Res. 27 (2018) 33–40.
[20] Food and Drug Administration, Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 4: Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making, https://www.fda.gov/media/132505/download, (2019) Accessed February 2, 2020.
[21] L.D. McLeod, C.D. Coon, S.A. Martin, S.E. Fehnel, R.D. Hays, Interpreting patient-reported outcome results: US FDA guidance and emerging methods, Exp. Rev. Pharmacoecon. Outcomes Res. 11 (2011) 163–169.
[22] H. Demeyer, C. Burtin, M. Hornikx, et al., The minimal important difference in physical activity in patients with COPD, PLoSOne 11 (2016) e0154587.
[23] R.W. Motl, L.A. Pilutti, Y.C. Learmouth, M.D. Goldman, T. Brown, Clinical importance of steps taken per day among persons with multiple sclerosis, PLoS One 8 (9) (2013) e73247, , https://doi.org/10.1371/journal.one.0073247.
[24] Food and Drug Administration, CFR - Code of Federal Regulations Title 21, Part 11: Electronic Records, Electronic Signatures, https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11, (1997) Accessed August 25, 2019.
[25] Food and Drug Administration, Draft Guidance for Industry: Use of Electronic

Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 – Questions and Answers, https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm563785.pdf, (2017) Accessed August 25, 2019.

[26] D. Eslinger, A.V. Rowlands, T.L. Hurt, et al., Validation of the GENEA accelerometer, Med. Sci. Sports Exerc. 43 (2011) 1085–1093.

[27] G. Aigner, B. Grimm, C. Lederer, et al., Method to collect ground truth data for walking speed in real-world environments: description and validation, PeerJ Preprints 7 (2019) e27558v1 https://peerj.com/preprints/27558/ , Accessed date: 25 August 2019.

[28] S. Eremenco, S.J. Coons, J. Paty, et al., PRO data collection in clinical trials using mixed modes: report of the ISPOR PRO mixed modes good research practices task force, Value Health 17 (2014) 501–516.

[29] A. Mullard, FDA approves first digital pill, Nat. Rev. Drug Discov. 16 (12) (2017) 818.

[30] ABILIFY MYCITE (Aripiprazole Tablets with Sensor), for Oral Use; Initial U.S. Approval: 2002 https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/207202lbl.pdf. Accessed August 25, 2019.