

Why Language Matters in Digital Endpoint Development: Harmonized Terminology as a Key Prerequisite for Evidence Generation

Lada Leyens^a Carrie A. Northcott^b Lesley Maloney^c Marie McCarthy^d
Nona Dokuzova^e Thomas Pfister^f
on behalf of the EFPIA digital endpoint joint sub-group
with contributions from Aude Clement

^aProduct Development Regulatory, Hoffmann-La Roche, Basel, Switzerland; ^bDigital Sciences and Translational Imaging, Pfizer, Inc., Cambridge, MA, USA; ^cProduct Development Regulatory, Genentech, Washington, DC, USA; ^dDECC, Novartis Ireland Ltd, Dublin, Ireland; ^eResearch and Development Quality, Boudry, Switzerland; ^fJohnson and Johnson Innovative Medicine, Allschwil, Switzerland

Keywords

Digital endpoints · Digital health technologies · Clinical trials · Drug development

Abstract

Background: Developments in the field of digital measures and digitally derived endpoints demand greater attention on globally aligned approaches to enhance digital measure acceptance by regulatory authorities and health technology assessment (HTA) bodies for decision-making. In order to maximize the value of digital measures in global drug development programs and to ensure study teams and regulators are referring to the same items, greater alignment of concepts, definitions, and terminology is required. This is a fast-moving complex field; every day brings new technologies, algorithms, and possibilities. A common language is particularly important when working in multifunctional teams to ensure that there is a clear understanding of what is meant and understood. **Summary:** In the paper, the EFPIA digital endpoint joint subgroup reviews the challenges facing teams working to advance digital endpoints, where different terms

are used to describe the same things, where common terms such as “monitoring” have significantly different meaning for different regulatory agencies, where the preface “e” to denote electronic is still used in some contexts, but the term “digital” is used in other, and where there is significant confusion as to what is understood by “raw” when it comes to data derived from digital health technologies. **Key Message:** The EFPIA subgroup is calling for an aligned lexicon. Alignment provides a more predictable path for development, validation, and use of the tools and measures used to collect digital endpoints supporting standardization and consistency in this new field of research, with the goal of increasing regulatory and payer harmonization and acceptance.

© 2024 The Author(s).
Published by S. Karger AG, Basel

Introduction

In the past decade, there has been an increase in our understanding of the value created by digital measures in drug development programs and seen a growth in integrating digitally derived endpoints in clinical trials

[1]. These developments demand greater attention to globally aligned approaches to enhance acceptance of digital measures by regulatory authorities and health technology assessment (HTA) bodies for decision-making.

To support global drug development programs, it is critical to align concepts and address the differences in the definitions to ensure study teams and regulators are referring to the same items. For example, the European Medicines Agency (EMA) defines the terms “digital measure” and “digital biomarker” [2], while the United States Food and Drug Administration (FDA) does not. Alignment provides a more predictable path for development and validation of the tools and measures used to derive digital endpoints. This consolidation will support standardization and consistency in this new field of research aiming at increased regulatory and payer harmonization and acceptance.

The European Federation of Pharmaceutical Industries and Associations (EFPIA) digital endpoint subgroup is proposing the establishment of a common lexicon that builds on established terminology and regulatory frameworks: Biomarkers, Endpoints, and other Tools (BEST) Glossary [3], EMA Q&A Guidance [2], FDA Remote Data Acquisition Draft Guidance [4], Critical Path Institute (C-Path) Lexicon [5]. The authors’ goal was to raise a call for action for alignment among all stakeholders, including the scientific community, industry, and regulatory authorities (e.g., EMA, FDA), as well as patients, clinicians, HTAs, and payers. In addition, we highlight the impact the classification of a digital measure as a biomarker or clinical outcome assessment (COA) has on its evidence requirements and use as well as the use of different concepts related to

- the digital health technology (DHT) that is used to collect and derive digital measures (the tool) and
- the digital measure used in clinical trials to derive endpoints (the measure).

The impact that classification and the use of different concepts have on evidentiary requirements is highlighted throughout the document; however, an in-depth description of the additional evidentiary requirements is out of scope of this paper and may be described in a future manuscript. Moreover, while this paper focuses primarily on regulatory frameworks in Europe and the USA, as they are regions with specific regulatory pathways and guidance (as described elsewhere in this paper), however, the principles can also apply to other regulators.

The Value of Digitally Derived Endpoints

Traditional clinical assessments performed at a research site or a clinician’s office, providing a snapshot of the assessed condition, and may be affected by patient-specific circumstances, such as travel, clinic environment, timing, and patient anxiety or motivation. Interpretation of symptoms and assessments may also differ from patient to patient and physician to physician, adding to the variability. Therefore, the resulting outcomes may not adequately reflect a patient’s overall health status, functioning, or quality of life in a precise, consistent, longitudinal manner. Such gaps in understanding may result in disease areas with “unmet measurement needs.”

Compared to data collected during a clinic visit, DHTs unlock the potential to measure relevant aspects of disease more frequently, sometimes continuously, in real-world settings and provide more granular insight into health status over time. DHTs are “systems that use computing platforms, connectivity, software, and/or sensors for healthcare and related uses [4]. They include technologies intended for use as a medical product, in a medical product, or as an adjunct to other medical products (devices, drugs, and biologics). They may also be used to develop or study medical products.” [3]. “DHT” has emerged to become an umbrella term for a digital device, sensor/wearable, software as a medical device; therefore, DHT is used throughout this document.

DHTs allow for quantitative and objective measurements of meaningful aspects of health, including symptoms of disease, that were not able to be measured previously, resulting in novel endpoints. Furthermore, DHTs can be designed to interfere minimally with patients’ daily lives and enable remote data acquisition outside the traditional clinical setting. Measurements obtained by DHTs have the potential to assess relevant aspects of disease in a less biased, patient-centric manner with ecological validity [6]. Furthermore, potentially, these novel digital endpoints can identify early efficiency in clinical trials. The use of DHTs can facilitate decentralization of clinical trials, enabling clinical sites to recruit patients independent of location, creating more patient-centric trials, boost recruitment, especially in rare disease trials, as well as promote diversity and equity in clinical.

For example, Duchenne muscular dystrophy (DMD) is a genetic disorder, typically affecting young males, that causes progressive muscle weakness and degeneration. There is no cure and limited treatment options [7]. A traditional endpoint used in DMD trials is the 6-Minute Walk Test [8], which captures episodic assessments related to ambulation directly observed by clinicians. This is

Table 1. Learning from the past: BP measurements

The measurement of blood pressure (BP) originated in 1733 when Sir Stephen Hales introduced a brass pipe connected to a glass tube into a horse's leg artery and observed the rise of the blood column to "8 feet and 3 inches above the level of the left ventricle" [11]

Since then, there have been many advancements to this measure, from using inflatable cuffs with a stethoscope to now "cuffless blood pressure" measures [12]. The many refinements of the tool, regarding optimal cuff sizing, rate of cuff deflation pressure, accuracy of the Korotkoff sound detection by automated digital methods, and importantly, smaller wearable devices for the estimation of ambulatory BP (i.e., daytime, nighttime, and 24-h average), have improved detection and management of cardiovascular conditions [13]

While we have now advanced to digital methods to collect this important measurement, nowhere do we read that this is a digital BP measure; it is just called the systolic and diastolic BP. To this end, while the field of measuring physiological functions using DHTs is "new" (it really is not that new), eventually the word "digital" should and most likely will be removed from being in front of the word "measure" (always with a word of caution that this should happen once experience increases, use cases and the differences with "traditional" COAs and biomarkers are better understood and gaps in understanding narrow, as noted in the previous section). Just as we do not say that it is a "digital BP measure" when we speak of blood pressure, it will be just accepted as a measure of what it is, an endpoint. It is important to note that as the BP measure advanced, the tools that provided the measures were always validated to ensure the accuracy of the endpoint

Although the technology of BP measurement has advanced from analog scales on pipes of mercury to nowadays digital BP cuffs and photoplethysmography, "digital BP measurement" is not a term commonly used.

contrast with a new endpoint – Stride Velocity 95th Centile (SV95C) – that uses a DHT to enable increased frequency of measurement, a quantitative and passive measure that is a potentially more accurate representation of the lived experience of a patient with DMD. This COA is validated and qualified in Europe as a primary endpoint in pivotal or exploratory drug therapeutic studies with ambulant DMD patients 4 years of age and above, with its recently published qualification opinion by EMA [9, 10].

Regulatory authorities should and will apply the same standards for all endpoints, regardless of how the data are acquired.¹ As with traditional clinical endpoints, these digitally derived endpoints require the measure to be meaningful to patients (COA), disease-relevant, and interpretable for objectively assessing changes in patient's health status, feeling, functioning, or survival. Introducing DHTs does not change the concepts of COA, biomarker, and endpoint but only adds new ways of acquiring them. Digital biomarkers, digital COAs, and resultant digital endpoints are simply biomarkers, COAs, and endpoints derived from data collected by a DHT. Thus, we envisage the use of "digital" preceding "established" terms as a short-term solution. As the use of digitally derived measures matures, we envision DHTs to become one more tool in the toolbox for conducting clinical trials and the use of the term "digital," denoting a rather unspecific characteristic on how the measure is collected, to become obsolete as occurred with blood pressure (BP) measurement (Table 1).

¹"Collected" and "acquired" are used interchangeably in this paper.

Lost in Translation: The Drive to Develop Consistent Terminology

Ensuring the optimal and successful adoption of aligned terminology in any new and fast-evolving field is critical. In Europe, there are significant ongoing efforts to generate high-quality health data [14] with a recognition that it is essential to have a commonality in understanding the tools and associated methodologies by stakeholders to avoid miscommunication and data being "lost in translation."

This is equally important related to the use of DHTs in drug development, where agreement on common terminology and lexicon sets an essential foundation to ensure multi-stakeholder alignment, comparison across global programs and trials, and realization of the potential for digital endpoints to deliver a promising future for patients.

The use of specific terms to define a digital drug development tool (DDT) can have a significant effect on how it is regulated. The convergence of various stakeholder with diverse expertise from differing backgrounds; drug development, medical device, and technology sectors, as well as data scientists, patient-centered outcomes researchers, clinicians, and patients, adds further complexity in the quest for a globally aligned lexicon.

The measure and the tool are separate but linked elements necessary to derive a digital endpoint. The digital measure must be validated and accepted as an endpoint based on its context of use (COU), and the DHT (i.e., tool) must also be validated and accepted by drug and device regulators depending on its intended purpose.

The Importance of Defining the Measure

Aligned terminology is essential as novel measures can be used in clinical trials as (digital) biomarkers or (digital) COAs, and subsequently used to construct endpoints, and the evidentiary requirements for each differ. Determination of the “classification” of digital measures into a biomarker or COA will depend on the concept of interest (COI) and the COU. The COI refers to the aspect of an individual’s clinical, biological, physical, or functional state or experience that the assessment is intended to capture (or reflect). The COU is defined formally as “a statement that fully and clearly describes the way the medical product development tool is to be used and the medical product development-related purpose of the use.” [3]. How a measure is classified has implications regarding its

1. Validation requirements. For example, COAs must have proven content validity, i.e., how meaningful they are to the patient. Biomarkers need to have a proven link to the pathophysiology of the disease.
2. Use in a clinical trial. Classification affects how the measure can be used to establish the treatment effect.

Although the terms digital measure, digital biomarker, digital COA, and digital endpoint may be used interchangeably by some, each term has its meaning linked to established drug development terms. Furthermore, as described above, the validation requirements and regulatory acceptance pathways are distinct and different for each (Fig. 1) [4, 15, 16].

The Importance of Defining the Tool

The regulatory classification of DHTs (i.e., medical device or not) and its intended use will impact the development, validation, regulatory pathway, and other potential regulatory requirements of the derived digital endpoint. To add to the complexity, the DHT’s regulatory classification may differ regionally depending on the local medical device regulations, and the identity may change as different features are added to the tool. An important point to note is that if the DHT is not used for a purpose that fits within the legal definition of medical devices and is only used to acquire data and derive a measure that will be used as an endpoint in the trial, it most likely will not be regulated under medical device regulations.

Proposed Aligned Terminology

The terminology below represents our recommendations for how regulatory authorities could consider updating or expanding their current lexicon and how the

drug development and scientific communities should align to drive greater consistency and enhanced understanding of DHTs and digital measures in clinical trials.

Methodology

We reviewed current guidance, regulations, and glossaries utilized primarily by European Union and US regulatory authorities, i.e., EMA, FDA, International Medical Device Regulators Forum (IMDRF) as well as multi-stakeholder consortia, for example, DIME, Critical Path Institute, Clinical Trials Transformation Initiative (CTTI), Pharmaceutical Research and Manufacturers of America® (PhRMA), to identify key terms for inclusion in the EFPIA lexicon proposal for global regulatory authorities to adopt or address.

The terms proposed below reflect EFPIA member company agreement on definitions that best represent our understanding. Where possible, we utilized an existing definition and attempted to highlight differences between established outcome measures and the newly emerging digital outcome measures. We acknowledge there are a variety of similar terms and meanings that stakeholders can utilize, and where appropriate, we discussed why we selected specific terms and identified where additional discussion and alignment may be needed.

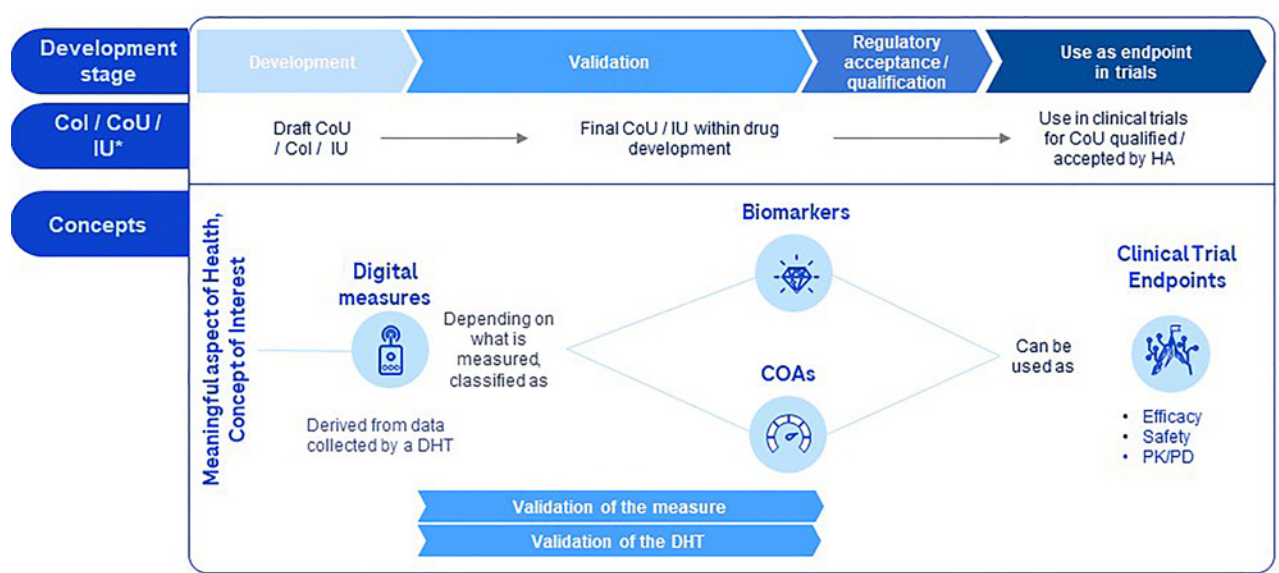
Measurement, Biomarkers, COAs, and Endpoints: Definitions Considering Digital Data Collection

Below, we present the established terms commonly used in drug development, followed by proposed digital terms. Any new proposed language or addition to the established term is marked in *italics*. The established term is inserted in brackets “[]” under the column “proposed digital term and definition” to depict the fact that the introduction of DHTs does not change the concepts of COA, biomarker, and endpoints but only adds new ways of acquiring them. The rationale for changes in established terms is provided under the table for each set of terms. The authors recognize that changing definitions or creation of new terms can have a significant effect on the regulatory landscape and application of its authorities, regulations, guidance, and recommendations for regulators; therefore, we aim to start a dialog to evaluate what changes are needed, and we start that dialog with the proposals below.

In the following table (Table 2), we outline the existing terminology and propose an alternative digital nomenclature for the most commonly used terms: measurement, biomarker, clinical outcome assessment, and endpoint. We outline a rationale as to why we believe this new term will provide a consensus approach.

Digital measures with the potential of becoming digital endpoints

Endpoints inferred from digital health technologies (DHT)



COA = Clinical Outcome Assessment; Col = Concept of Interest; CoU = Context of Use; IU = Intend of Use, DHT = Digital Health Technology; PK/PD = pharmacokinetic and pharmacodynamic

Fig. 1. This diagram represents the different components and development stages of a digital endpoint in a clinical trial. COA, clinical outcome assessment; COI, concept of interest; COU, context of use; IU, intended use; DHT, digital health technology; PK/PD, pharmacokinetic and pharmacodynamics (Simplification of CTTi Flowchart of Steps for Novel Endpoint Development [16]).

Terms Requiring Clarity

The following terms may require additional review and clarification. While we do not offer a solution or an alternative definition, we recommend a dialog in the digital measure community as we recognize that clarity around these terms and concepts is critical.

Active Task or Testing (PerfO) versus Passive Data Acquisition

We initially discussed including definitions for both active testing and passive data acquisition; however, no consensus could be reached with respect to passive acquisition, making it clear that more discussion is needed and an alternative term required for “passive data acquisition.” Passive data acquisition has the potential to become more relevant in the future as DHT developments advance and to bring more value to patients as data are collected with minimal additional burden and little interaction from patients. To initiate the dialog around this important term, we propose the following definitions and differentiation with active task or testing:

- Active task or testing (PerfO): a standardized task actively undertaken by a patient that can be captured with DHT. For example, an active task or test may require the patient to follow specific instructions in a mobile application, such as tapping, drawing, or recalling specific objects.
- Passive data acquisition: the recording of human behavior, functioning, or physiology with a data acquisition tool or DHT that does not require a patient’s active interaction with the tool nor requires the patient to perform a standardized task. Passive data acquisition is not performed according to instructions on a specific task nor standardized (however, there will be detailed instructions for the general use of the DHT), for example, a wrist-worn actigraphy device that the patient wears for the duration of the test period to assess sleep and physical activity.

The term “active task” or “active testing” is a term used by the digital measure community to differentiate activities measured with a DHT that are a specific, standardized, scheduled, or instructed from real-life activities measured with a DHT on an ongoing manner and that do

Table 2. Existing terminology and proposed alternative digital nomenclature for measurement, biomarker, clinical outcome assessment, and endpoint

Established term and definition	Proposed digital term and definition	Rationale
<p>Measurement: the obtained value using a test, tool, or instrument. (BEST Glossary) [3]</p>	<p>Digital measure: an [obtained value using a test, tool, or instrument] <i>derived from data captured by a digital health technology</i> (DHT) A digital measure can be used to derive a variable that can be used as an endpoint in the context of a clinical trial</p>	<p>Measurement: we support using the BEST Glossary [3] definition for measurement (i.e., measure) but prefacing it with the term digital Digital measure: we appreciate EMAs inclusion of the term “digital measure” (objective, quantifiable measure of physiology and/or behavior collected and measured through digital tools) in its 2020 Question and Answer (Q&A) guidance for digital technology-based methodologies [2]. As the field has evolved, we recommend simplifying the definition to be consistent with the proposed approaches for other “digital” definitions. Importantly, we recommend the removal of the term “objective” in the EMA definition, which conflicts with the proposal that digital measures can be developed into biomarkers or COAs, including patient-reported outcome assessments (PROs) and observer-reported outcomes (ObsROs), which introduce some level of subjectivity</p>
<p>Biomarker: a defined characteristic or set of characteristics that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention. A biomarker is not an assessment of how a patient feels, functions, or survives. Biomarker categories can be diagnostic, monitoring, prognostic, pharmacodynamic/response, and susceptibility/risk. (BEST Glossary) [3]</p>	<p>Digital biomarker: [a defined characteristic or set of characteristics that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention. A biomarker is not an assessment of how a patient feels, functions, or survives. Biomarker categories can be diagnostic, monitoring, predictive, prognostic, pharmacodynamic/response, safety, and susceptibility/risk] <i>derived from data captured by a digital health technology</i> (DHT)</p>	<p>Biomarker: we support the use of the BEST Glossary definition for biomarker, which broadens the scope beyond “a biological molecule” to a “defined characteristic.” Furthermore, we favor an expansion of the definition to consider a [set of characteristics] as proposed in a recent publication [17] since examples of biomarkers can include molecular, histologic, radiographic, or physiologic characteristics, such as cardiac troponin Digital biomarker: we encourage using the proposed definition above, which is intended to be consistent with the established term for biomarker and acknowledge the only difference is its collection method. We also encourage the EMA to consider eliminating the following language in their definition [2]: biomarkers being a measure of “behavior” and “The clinical meaning is established by a reliable relationship to an existing, validated endpoint.” These changes should add more clarity, as the inclusion of “behavior” within the definition (an objective, quantifiable measure of physiology and/or behavior used as an indicator of biological, pathological process or response to an exposure or an intervention that is derived from a digital measure. The clinical meaning is established by a reliable relationship to an existing, validated endpoint [2] can lead to confusion since it can also be considered as a “function,” which fits within the definition of a COA. In addition, the requirement for how to establish clinical meaning (e.g., EMAs definition requires it to be established from existing measures) should not be specified as part of the definition of a digital biomarker but, instead, be included as part of the evidence requirements</p>

Table 2 (continued)

Established term and definition	Proposed digital term and definition	Rationale
<p>Clinical outcome assessment: assessments of a clinical outcome defined by how a patient feels, functions, or survives. COAs can be made through a report by a clinician, a patient, or a non-clinician observer or through a performance-based assessment. Types of COAs include</p> <ul style="list-style-type: none"> • Patient-reported outcome (PRO) measures • Clinician-reported outcome (ClimRO) measures • Observer-reported outcome (ObsRO) measures • Performance outcome (PerFO) measures [3] 	<p>Digital clinical outcome assessment: [assessments of a clinical outcome defined by how a patient feels, functions, or survives] <i>derived from data captured by a digital health technology</i> (DHT)</p>	<p>Clinical outcome assessment: the authors recommend stakeholders align with the BEST Glossary definition [3], which is widely accepted</p> <p>Digital clinical outcome assessment: we recommend using the term digital COA to point stakeholders back to the conventional definition of COA. Although we understand the use of the EMA terminology “eCOA,” (a quantifiable measure used as a measure of how patients feel, function or survive that is derived from a digital measure. The clinical meaning is established de novo. Clinical outcomes can be assessed through a report by a clinician, a patient, a non-clinician observer or through an active performance-based assessment or passive monitoring of patient behavior or performance) [2]. We believe this term has historical context in the way it is used in the ecosystem currently, which links to the migration of PROs from paper to digital means (i.e., electronic patient-reported outcomes or observer-reported outcomes collected through a general computing system). Similarly, Critical Path Institute defines eCOA as “a clinical outcome assessment that has been implemented on an electronic data collection platform (e.g., smartphone or tablet).” [5]. To streamline and align language globally, we recommend no longer using eCOA as an umbrella term for COAs acquired digitally and referring to digital COAs, as defined above. In addition, regarding the definition provided by EMA [2], the specification of how the clinical meaning is established should not be part of the definition but of the evidence requirements</p>
<p>Endpoint: precisely defined variables intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. (BEST Glossary) [3]</p>	<p>Digital endpoint: [precisely defined variables intended to reflect an outcome of interest that is statistically analyzed to address a particular research question] <i>derived from data captured by a DHT</i></p>	<p>Endpoint: the authors recommend stakeholders align with the BEST Glossary definition, which is widely accepted</p> <p>Digital endpoint: while we appreciate that EMA defined a digital endpoint and agree with the definition provided, we recommend EMA simplify its terminology and point stakeholders back to the conventional endpoint definition. The authors propose removing the references to clinical relevance and reliable relationship from the definition, as noted above, for biomarkers and COA, as this is related to the validation needs and evidence requirements</p>

not require a structured patient participation and referred to as passive data acquisition or remote monitoring). FDA refers in the DDT library [18] on some occasions to COAs based on passive data acquisition as “DHT-Passive Monitoring COA.”

- Active standardized tasks are typically categorized as performance outcome assessments (PerfOs). PerfOs are defined by the FDA in its draft guidance [4] as “a measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions.” While DHTs enable quantification of previously inaccessible aspects, including functionally relevant behavior characteristics, such as gait parameters and acoustic features of speech, it is unclear if such elements should be deemed measures of patient function (i.e., COAs) or biomarkers. In addition, it is unclear if a PerfO can include both active testing and passive data acquisition, as PerfOs are standardized tasks and performed according to instructions, while passive measures are not. This lack of clarity indicates a need for either an update of the PerfO definition to account for passive data acquisition or the creation of a 5th COA category.
- In addition to clarity on the terms for developers to establish the validation plan, determining these tests as a COA or biomarker must be understood as it drives the level of evidence needed to establish the clinical relevance of a digital measure. More importantly, regulators should converge on classification as either biomarker or COA for the same digital measure in the same COU. We have intentionally not used the term “passive monitoring” (see “Monitoring discussion”).

Data Acquisition Tool

Some DHTs are referred to in the literature as “remote patient monitoring” or “digital data acquisition” tools. The purpose of these tools in clinical drug development studies is not to perform a medical device function, i.e., not intended to treat or diagnose a disease or support clinical decision-making, but to collect patient-generated health data deemed necessary to derive a digital measure and resulting digital endpoint. EFPIA members discussed whether a specific term is needed to describe the use of DHTs as a DAT with no medical purpose rather than identifying them as “patient monitoring” tools, especially since “monitoring” has a regulatory meaning under the European Union’s Medical Device Regulation (MDR) [19] or in vitro diagnostic medical device regulations (IVDR) [20] (see additional discussion in “Monitoring” below). The definition of DAT is clarified with the third revision (R3) of the International Conference on Harmonization (ICH) E6 on Good Clinical Practice [21], and

we recommend stakeholders align with the definition that is included in ICH E6 (R3). However, EFPIA members agreed further discussion is needed to understand regulatory requirements that might apply to DHTs used as DATs solely to acquire data in a clinical trial for drug development (e.g., for digitally derived endpoints).

Monitoring (in the Context of a Medical Device)

The EU MDR includes “monitoring of disease or disability” in its definition of the medical purposes that qualify an instrument or software as a medical device. Yet there is significant confusion around what monitoring means among stakeholders, including differences in interpretation between regulatory authorities and when such use meets the definition of a medical device. For example, a DAT/DHT described as a “remote patient monitoring tool” used in a drug trial to gather voice data over 24 h for purposes of developing a digitally derived endpoint may be viewed as a medical device by some regulatory authorities, even if the data are not intended for use in clinical decision-making. This interpretation can lead to significant time spent with regulatory authorities discussing if the DHT has a medical purpose under the MDR or IVDR. To provide clarity, we propose a more specific definition for a DHT that monitors and has a medical purpose (i.e., is regulated as a medical device) to include

- Tool is used to actively monitor a disease, injury, or condition of an individual patient and
- Data are analyzed or interpreted for clinical decision-making related to the individual patient.

Multicomponent versus Multimodal

EFPIA members discussed differences in meaning between multicomponent and multimodal measures, including whether the terms were interchangeable. As this topic is broader than DHTs and is currently being discussed by the FDA in the context of multicomponent biomarkers [22], the EFPIA team decided not to put forth specific definitions. Initial thoughts from the team included this definition for multicomponent measures derived from DHTs

- Multicomponent measure: the combined value of two or more measurements that are given equal or different weighting, for example, the combination of measures of fine motor speed, accuracy, and coordination, as measured by an active test, into a value reflecting fine motor dexterity.

Raw Data, Source Data and Parameters

EFPIA members discussed the need to define data at various levels of digital endpoint development, including the varying terms used to describe data at each level of

granularity and processing and what level of granularity and processing qualifies as source data. Although there was no agreement on single terms or whether these should even be considered “terms of art,” it was felt it to be necessary to share some of the questions and considerations raised. Terminology currently used by stakeholders includes raw data, epoch data, endpoint data, which can then be processed and/or combined with other parameters or variables to develop a digital measure. These terms are provided as an example in Figure 2 below. For instance, some consider raw data to refer to the accelerometer preprocessed data; for others, it refers to data that have been processed but not analyzed by algorithms.

We discussed that it is essential to understand and define the granularity, especially in relation to the level at which data are considered “source data,” as there are precise legal requirements for source data. As described in ICH E6 (R2) 6.10; 4.9.0; 5.1.2; 5.15.1; 6.4.9 [23], the source needs to be predefined and agreed with by the investigator; the investigator needs to have continuous access to it; it needs to be interpretable, human-readable; and finally, it needs to be stored for the legally required time for clinical trial data. Furthermore, monitors, auditors, and inspectors must be able to access and read it upon request when conducting inspections.

For these reasons, it was agreed that there is value in broader discussions around the required granularity of “source data” for data acquired by a DHT. As raw data from sensors are not necessarily human-readable and interpretable, the authors propose that the source data be that at which the data are “interpretable,” as demonstrated in Figure 2. Guidance such as the new computer system guidance from EMA [24] addresses this question for other types of systems, such as electronic data capture or computer systems supporting patient-reported outcome collection. However, the guidance does not provide further guidance for other technologies increasingly used in clinical trials, such as sensor-based DATs. The authors welcome opening a dialog with regulators on source data concerning data collected by DHTs. Additional key terms for navigating the digital tool and measure landscape are detailed in online supplementary Table 1 (for all online suppl. material, see <https://doi.org/10.1159/000534954>).

Why Aligning on Terminology Is Essential

The most widely used digital endpoints in clinical trials are those used to assess sleep and physical activity [1], which can also cause them to be subject to the greatest

“misuse” of terminology and language. While appearing to be simple concepts, there are many ways of measuring these common physiological functions, as many different parameters and stages can be measured. The case study presented illustrates the complexity and importance of the language used when selecting a DDT for a specific use case to ensure it is appropriate and fit for purpose. This case study should complement the case study by Vasudevan et al. [17] and showcase the need for a clear classification of DDTs.

Sleep: A Case Study

Sleep is a complex concept, and there are many parameters and stages that one could measure that are relevant to drug development. Similarly, there are many ways of measuring these parameters, and depending on the DHT and how the data are being used, the measure may vary in precision and accuracy in any given COU. When using sleep as an endpoint, it is essential that health authorities can rely on the data derived from the DHT and related algorithms, and it is also essential that the output is understandable and accurately reflects the treatment effect.

There are a variety of tools and methods available to assess sleep, including objective methods such as in-clinic polysomnography (PSG) or tools such as actigraphy where sleep patterns can be assessed longitudinally over weeks and months, to subjective methods such as sleep diaries and patient questionnaires.

- Subjective tools measure how a patient perceives they slept.
- PSG uses electrodes to measure brain activity and other physiological parameters to determine the stages of sleep objectively.
- Actigraphy uses established algorithms to convert motor movement into sleep and activity patterns (typically, the algorithms are validated to PSG).

All three measures provide different information and details to investigators regarding the sleep being measured. Despite being measured in different ways and capturing various aspects or parameters, the same terminology is often used to describe the measures: number of awakenings, sleep onset latency, wake after sleep onset, total sleep time, etc. [25].

In addition to different tools used to assess sleep, the resultant measures in any clinical trial in a specific patient population can be endpoints considered either biomarkers or COAs, based on the intended use of that endpoint [26]. For example, if you measure sleep endpoints in a condition or disease in which sleep is an aspect of the condition, such as in a sleep disorder (sleep apnea),

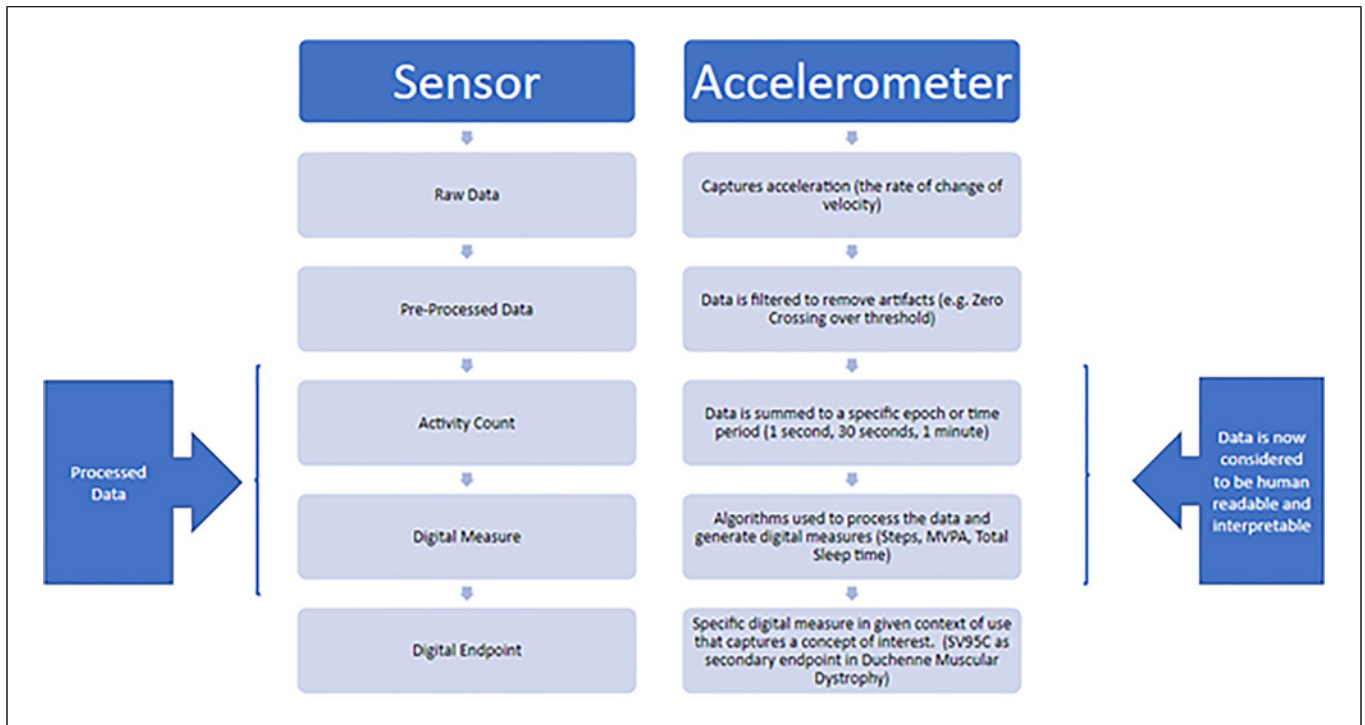


Fig. 2. When are data derived from a sensor deemed to be human-readable and interpretable?

you are directly measuring an aspect of the disease. Since it is a defined measure of the pathogenic processes, the digital measure derived from an objective tool, such as PSG, could be considered a biomarker [27]. In contrast, if you measure sleep parameters in a condition, such as dermatomyositis, and it is an indication of the health-related quality of life, the measure of sleep is not necessarily a “part” of the physiology of the condition. If it is demonstrated that the sleep measurements have a good relationship to how the patient feels the next day while awake (i.e., meaningful to the patient),” then, in this case, sleep is being evaluated as a measure of how the patient “feels, functions, or survives” due to the condition. The specific sleep measure would be considered a COA.

Conclusion

In this paper, EFPIA member companies who are part of the digital endpoint subgroup propose aligned terminology for concepts related to digital measures. Where there was consensus, we proposed a term; where there was misalignment, we recommend further discussion amongst the community. We agreed with the transient nature of the term “digital” and that an endpoint should not depend on

how it is collected but on *what* it measures. At this stage, “digital” can still be used to delineate how data are acquired, but once the field has matured, we should revert to the established terms “biomarker,” “COA,” and “endpoint.”

An aligned lexicon is essential when working in multifunctional teams to ensure a common language and a clear understanding of what is meant and understood. This gains even more importance when different stakeholders are brought into the discussion, such as in regulatory interactions. Furthermore, since a DHT can generate several distinct digital measures and scores, defining the specific digital endpoints derived from these scores in a given COU requires a common vocabulary to enable unified understanding and standards. For instance, a common error in published materials is that the term indicating the measurement methodology is also used to describe the endpoint, for example, actigraphy.

The language matters, and the choice of a specific endpoint matters as it needs to measure a particular COI in a specific COU for a given patient population and needs to be fit for informing regulatory decision-making. Furthermore, we envision a future where measures may be based on different modalities and composite scores may be created out of measures that are collected in many ways (e.g., in the combination of a biomarker and a

patient-reported outcome and a PerfO) or by “multiple sensors” to capture complex multifactorial measures. To enable that level of complexity, we need to set the correct bases, align at these earlier stages, and continue aligning as the complexity increases.

Next Steps

In this position paper, we present a call for

1. alignment in language related to digital measures and digitally derived endpoints;
2. convergence on fewer relevant terms rather than establishment of new terms;
3. reuse of common existing terms as much as possible;
4. importance should not lie on how we measure but on what we measure;
5. regulatory authorities to utilize common terminology to better enable global drug development programs and trials; and
6. an aligned classification of measures within COA or biomarkers.

As noted above, we envision a future where digital endpoints are more established and routinely used for regulatory decision-making. In the end, the method used to collect the endpoint will not need to be mentioned; therefore, we can revert to the terms as biomarker, COA, endpoint, and measure.

We call for companies and regulators to reflect on this lexicon within their organization as they are advancing their knowledge and experience in this field, to align on terminology, and to quickly move to the more critical questions on what evidence is needed in the validation of these measures, given what category they fall under.

Following the initial discussion on the lexicon, we plan to explore in more detail in future publications these evidence needs concerning digital measures, especially where there are open questions for developers. This should accelerate the identification of areas of uncertainty and contribute to discussions with regulators to increase our understanding of expectations.

References

- 1 Digital Medicine Society (DIME). DiMe’s library of digital endpoints. 2023. Available from: <https://dimesociety.org/get-involved/library-of-digital-endpoints/>.
- 2 European Medicines Agency (EMA). Human Medicines Division. Questions and answers: qualification of digital technology-based methodologies to support approval of medicinal products; 2020. Available from: https://www.ema.europa.eu/en/documents/other/questions-answers-qualification-digital-technology-based-methodologies-support-approval-medicinal_en.pdf.
- 3 FDA-NIH Biomarker Working Group. BEST (biomarkers, EndpointS, and other tools) resource. Food and Drug Administration (US). Co-published by National Institutes of Health (US); 2016-. Available

from: <https://www.ncbi.nlm.nih.gov/books/NBK326791/>.

- 4 Food and Drug Administration (FDA). Digital health technologies for remote data acquisition in clinical investigations; guidance for industry, investigators, and other stakeholders; draft guidance. FDA-2021-D-1128-0002; 2021. Available from: <https://www.regulations.gov/document/FDA-2021-D-1128-0002>.

Acknowledgments

We want to thank Aude Clement, Jonas Santiago, Thorsten Vetter, and Francesca Cerreta for their review and comments on the manuscript.

Statement of Ethics

An ethics statement is not applicable because this study is based exclusively on published literature.

Conflict of Interest Statement

Nona Dokuzova is an employee of Bristol-Myers Squibb and owns company stock. Lada Leyens is an employee of Hoffmann-La Roche and owns company stock. Marie McCarthy is an employee of Novartis Ireland Ltd. and may own company stock. Lesley Maloney is an employee of Genentech and owns company stock. Carrie Northcott is an employee of Pfizer, Inc. and owns company stock. Thomas Pfister was an employee of Janssen Pharmaceuticals and owns company stock, he is an employee of Vectura Fertin Pharma, Basel, Switzerland and may own company stock.

Funding Sources

Nona Dokuzova, Lada Leyens, Marie McCarthy, Lesley Maloney, Carrie Northcott, and Thomas Pfister did not receive funding for the preparation of this manuscript.

Author Contributions

Nona Dokuzova, Lada Leyens, Marie McCarthy, Lesley Maloney, Carrie Northcott, and Thomas Pfister substantially contributed to the conception of this manuscript, played a role in drafting and revising the article, and approved the final version for publication.

Data Availability Statement

All data are available in the tables and supplementary material.

- 5 Critical Path Institute (C-Path). eCOA: getting better together initiative. eCOA Lexicon v1.0. https://c-path.org/wp-content/uploads/2021/02/GettingBetterTogetherInitiative_eCOALexicon_v1.pdf [Accessed 15th March 2023].
- 6 Kruizinga MD, Stuurman FE, Exadaktylos V, Doll RJ, Stephenson DT, Groeneveld GJ, et al. Development of novel, value-based, digital endpoints for clinical trials: a structured approach toward fit-for-purpose validation. *Pharmacol Rev.* 2020;72(4):899–909.
- 7 Duan D, Goemans N, Takeda S, Mercuri E, Aartsma-Rus A. Duchenne muscular dystrophy. *Nat Rev Dis Primers.* 2021;7(1):13.
- 8 Haberkamp M, Moseley J, Athanasiou D, de Andres-Trelles F, Elferink A, Rosa MM, et al. European regulators' views on wearable-derived performance measurement of ambulation for Duchenne muscular dystrophy regulatory trials. *Neuromuscul Disord.* 2019;29(7):514–6.
- 9 Qualification Opinion for Stride velocity 95th centile as primary endpoint in studies in ambulatory Duchenne Muscular Dystrophy studies. European Medicines Agency. 2023 EMADOC-1700519818-1127132.
- 10 Servais L, Camino E, Clement A, McDonald CM, Lukawy J, Lowes LP, et al. First regulatory qualification of a novel digital endpoint in Duchenne muscular dystrophy: a multi-stakeholder perspective on the impact for patients and for drug development in neuromuscular diseases. *Digit Biomark.* 2021;5(2):183–90.
- 11 Booth J. A short history of blood pressure measurement. *Proc R Soc Med.* 1977;70(11):793–9.
- 12 Rader F, Victor RG. The slow evolution of blood pressure monitoring: but wait, not so fast. *JACC Basic Transl Sci.* 2017;2(6):643–5.
- 13 Watanabe N, Bando YK, Kawachi T, Yamakita H, Futatsuyama K, Honda Y, et al. Development and validation of a novel cuff-less blood pressure monitoring device. *JACC Basic Transl Sci.* 2017;2(6):631–42.
- 14 European Commission, Public Health. Proposal for a regulation - the European health data space; 2022. Document 52022PC0197. Available from: https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_en.
- 15 European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP). Qualification of novel methodologies for drug development: guidance to applicants. EMA/CHMP/SAWP/72894/2008. 2020; Revision 4. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf.
- 16 Food and Drug Administration (FDA). Center for drug evaluation and research (CDER), center for biologics evaluation and research (CBER). Qualification process for drug development tools guidance for industry and FDA staff. Final level 1 guidance. FDA-2010-D-0529; November 2020. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/qualification-process-drug-development-tools-guidance-industry-and-fda-staff>.
- 17 Vasudevan S, Saha A, Tarver ME, Patel B. Digital biomarkers: convergence of digital health technologies and biomarkers. *NPJ Digit Med.* 2022;5(1):36.
- 18 Food and Drug Administration (FDA). CDER & CBER drug development tool qualification project search. Available from: <https://force-dsc.my.site.com/ddt/s/> [Accessed on 16th March 2023].
- 19 European Parliament and Council. The medical device regulation (MDR); 2017. Regulation (EU) 2017/745. Available from: <https://de-mdr-ivdr.tuvsud.com/Medical-Device-Regulation-MDR.html>.
- 20 European Parliament and Council. Vitro diagnostic medical devices directive (IVDMD); 2017. Regulation (EU) 2017/746. Available from: <https://de-mdr-ivdr.tuvsud.com/In-Vitro-Diagnostic-Medical-Devices-Directive-IVDR.html>.
- 21 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH E6 Guideline for good clinical practice (GCP) - update on progress public web conference reports May 18 & 19; 2021. Available from: <https://www.ich.org/page/efficacy-guidelines> [Accessed 16th March 2023].
- 22 Food and Drug Administration (FDA). Identification of Concepts and Terminology for Multi-Component Biomarkers. 2022. Available from: <https://www.fda.gov/media/155485/download>.
- 23 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Integrated addendum to ICH E6(R1): guideline for good clinical practice E6(R2); 2016. Available from: https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf.
- 24 European Medicines Agency. Guideline on computerised systems and electronic data in clinical trials. 9 March 2023. EMA/INS/GCP/112288/2023. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-electronic-data-clinical-trials_en.pdf.
- 25 Bawany F, Northcott CA, Beck LA, Pigeon WR. Sleep disturbances and atopic dermatitis: relationships, methods for assessment, and therapies. *J Allergy Clin Immunol Pract.* 2021;9(4):1488–500.
- 26 Lucas J, White C, Koo B, Milham M, Klein A. Assessing actimeters for inclusion in the healthy brain network. *BioRxiv preprint.* 2017.
- 27 Wichniak A, Wierzbicka A, Jernajczyk W. Sleep as a biomarker for depression. *Int Rev Psychiatry.* 2013;25(5):632–45.