Developing a Novel Measurement of Sleep in Rheumatoid Arthritis: Study Proposal for Approach and Considerations

Michelle Crouthamel^a Robert J. Mather^b Suraj Ramachandran^c Kai Bode^c
Godhuli Chatterjee^d Luis Garcia-Gancedo^e Joseph Kim^f Rinol Alaj^g
Matthew F. Wiperman^g Lada Leyens^h Henrik Sillenⁱ Tina Murphy^j
Michael Benecky^k Brandon Maggio^l Thomas Switzer^m

^aDigital Health & Innovation, Global Clinical Development, AbbVie Inc., North Chicago, IL, USA; ^bEarly Clinical Development, Pfizer Inc., Cambridge, MA, USA; ^cGlobal Regulatory Affairs and Clinical Safety (SR), MRL (KB), Merck & Co, Inc., Kenilworth, NJ, USA; ^dClinical Study Unit (India-South East Asia Cluster), Sanofi Healthcare India Private Limited, Mumbai, India; ^eValue Evidence and Outcomes, GSK, Stevenage, UK; ^fTranslational Technology and Innovation, Office of Digital Health, Eli Lilly and Company, Indianapolis, IN, USA; ^gClinical Outcomes Assessment and Patient Innovation, Global Study Strategy & Optimization (RA), Precision Medicine, Early Clinical Development & Experimental Sciences (MFW), Regeneron Pharmaceuticals Inc., Tarrytown, NY, USA; ^hProduct Development Regulatory, F. Hoffmann-La Roche Ltd, Basel, Switzerland; ⁱDigital Health R&D, AstraZeneca, Mölndal, Sweden; ^jRegulatory Affairs Innovation, Novartis Pharmaceuticals, East Hanover, NJ, USA; ^kGlobal Regulatory Affairs, UCB Biosciences, Inc., Raleigh, NC, USA; ^lDigital Trials – Global Clinical Operations, Boehringer-Ingelheim, Ridgefield, CT, USA; ^mEarly Clinical Development Informatics, Genentech, South San Francisco, CA, USA

Keywords

Novel digital endpoint · Validation framework · Hypothetical study

Abstract

The development of novel digital endpoints (NDEs) using digital health technologies (DHTs) may provide opportunities to transform drug development. It requires a multidisciplinary, multi-study approach with strategic planning and a regulatory-guided pathway to achieve regulatory and clinical acceptance. Many NDEs have been explored; however, success has been limited. To advance industry use of NDEs to support drug development, we outline a theoretical, methodological study as a use-case proposal to describe the process and considerations when developing and obtaining regulatory acceptance for an NDE to assess sleep in patients

with rheumatoid arthritis (RA). RA patients often suffer joint pain, fatigue, and sleep disturbances (SDs). Although many researchers have investigated the mobility of joint functions using wearable technologies, the research of SD in RA has been limited due to the availability of suitable technologies. We proposed measuring the improvement of sleep as the novel endpoint for an anti-TNF therapy and described the meaningfulness of the measure, considerations of tool selection, and the design of clinical validation. The recommendations from the FDA patient-focused drug development guidance, the Clinical Trials Transformation Initiative (CTTI) pathway for developing novel endpoints from DHTs, and the V3 framework developed by the Digital Medicine Society (DiMe) have been incorporated in the proposal. Regulatory strategy and engagement pathways are also discussed.

© 2021 TransCelerate BioPharma Inc.
Published by S. Karger AG, Basel

Introduction

Digital health technologies (DHTs) provide the following opportunities to transform drug development: “(1) existing endpoints that can be measured in new and possibly better ways, and (2) new endpoints that have not previously been possible to assess” [1]. These new measures have been shown to provide richer and more objective information on how patients function and detect disease activity faster than classical clinical outcome measures and can capture measures that are more relevant to patients and their lived experience [2, 3]. In addition, a novel digital endpoint (NDE), if used as the prespecified ranked endpoint in registrational studies, may enable promotional claims. Hence, many pharmaceutical companies are developing novel digital measures of health and disease to accelerate drug development, increase patient centricity, and support product differentiation. As of September 2020, over 100+ unique novel digital measures are used as clinical endpoints in industry-sponsored clinical trials [4]. One digital endpoint, moderate-to-vigorous physical activity measured by actigraphy, was accepted by the US Food and Drug Administration (FDA) as a primary endpoint for investigational product development in patients with pulmonary hypertension associated with interstitial lung disease [5]. Further, 95th percentile stride velocity measured with an accelerometer has received a positive opinion from the European Medi-

cines Agency (EMA) to be used as a secondary endpoint for evaluating the efficacy of drugs in Duchenne muscular dystrophy [6]. In addition, several public-private partnerships are identifying digital endpoints to assess fatigue, sleep, and activities of daily living in various indications (Table 1) [7–9].

The development of an NDE requires a multidisciplinary, multi-study approach with strategic planning and a regulatory-guided pathway to achieve regulatory and clinical acceptance. Early patient engagement is crucial to identifying the aspects of health that are meaningful to them and that reflect their experience and quality of life (QoL). To advance industry use of NDEs to support drug development and potential label enhancement, we present a use-case proposal to describe the process and considerations when developing and obtaining regulatory acceptance for an NDE to assess sleep in patients with rheumatoid arthritis (RA). We incorporated recommendations from the FDA patient-focused drug development (PFDD) guidance [10], the Clinical Trials Transformation Initiative (CTTI) pathway for developing novel endpoints from DHTs [1], the V3 framework for the evaluation of biometric monitoring technologies [11], and the structure and content of an evidence dossier for the use of a sensor-derived clinical endpoint proposed by Walton et al. [12]. Regulatory strategy and engagement pathways are also discussed.

Table 1. NDE initiatives led by PPPs

NDE initiatives	PPPs	Descriptions	Links
Fatigue and SD	IMI IDEA-FAST	Identifying novel digital measures for fatigue and SDs that will provide more objective, sensitive, and reliable measures of the severity and impact of these symptoms in the person's normal surrounding	https://idea-fast.eu/
PD	Critical path for PD (Watch-PD)	Developing a set of digital measurements in a prospective observational study using digital technologies with regulatory agency feedback to inform future PD clinical studies	https://c-path.org/
RWWS	IMI mobilize-D	Developing a system to analyze people's gait using wearable technologies in conditions, including COPD, PD, MS, hip fracture recovery, and CHF	https://www.imi.europa.eu/projects-results/project-factsheets/mobilise-d

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; IMI IDEA-FAST, Innovative Medicines Initiative, identifying digital endpoints to assess fatigue, sleep, and activities in daily living in neurodegenerative disorders and immune-mediated inflammatory diseases; MS, multiple sclerosis; NDE, novel digital endpoint; PD, Parkinson's disease; PPPs, public-private partnerships; RWWS, real-world walking speed; SD, sleep disturbance.

Measuring Sleep Disturbance in RA

Scientific Rationale

RA is a common chronic systemic inflammatory disease characterized by painful, stiff, and swollen joints, and patients with RA often have impaired physical capacity and sleep patterns [13]. RA patients often report joint pain, sleep disturbances (SDs), excessive daytime sleepiness (EDS), and fatigue, as well as reduced QoL [14–16]. It is known that joint pain affects physical activity and functional capacity. Many researchers have investigated functional assessment using wearable technologies [15, 16]. However, the research of SDs in RA has been limited due to the availability of suitable technologies. SDs are frequently reported in patients with RA, with a prevalence of 60–80% compared with 10–30% in the general population [17–21]. In these studies, patients with moderate and high disease activity reported fewer hours asleep than patients with low disease activity or those in remission. However, it is unclear whether SD is a direct result of RA disease activity or other secondary manifestations.

Regulation of sleep has been linked to the immune system and the release of cytokines [22–26]. Inflammatory cytokines, including interleukin-6 and tumor necrosis factor- α (TNF- α), are elevated in RA as well as disorders of EDS (i.e., sleep apnea and narcolepsy), suggesting that these cytokines might play a significant role in mediating sleepiness and fatigue in these patient populations [27]. There are observations that etanercept and infliximab, both of which block the effects of TNF- α (a mediator of inflammation), could reduce EDS in RA [24], and total sleep opportunity was increased after etanercept treatment when corrected for patient C-reactive protein (CRP) levels [28]. Together, these results suggest that sleep quantity and quality may be viable endpoints in anti-TNF- α therapy development. Assessment of sleep and the impact of pain and inflammation by clinician-reported outcomes may not accurately reflect a patient's daily life due to periodic evaluation. They may be influenced by recall bias because symptoms occur at night or may not be accurately remembered by the patient over time. Polysomnography (PSG), while highly accurate, requires visits to sleep labs, disrupting a patient's typical behaviors, and is not a scalable solution for more extensive clinical studies. As such, we present an opportunity for the development of NDEs utilizing DHTs for this case study given that sleep is impacted by RA disease severity, patients cite sleep disruption as a significant burden in their daily lives, and last, assessments like PSG and patient-reported outcomes (PROs)

suffer from scalability or sensitivity, respectively, needed for therapeutic development.

Clinical Meaningfulness

FDA's PFDD guidance [10] encourages sponsors to incorporate patients' experiences to justify the concept of interest (COI) chosen by the sponsors. Therefore, patient focus groups may be utilized early on to ascertain the COI measurement's meaningfulness and assess patient desires/wants or aspects of the disease that a patient does not wish to worsen, want to improve, or want to prevent [19, 29, 30]. Furthermore, a qualitative study can be used to elucidate these meaningful aspects of health. The goal becomes turning these patient desires into an endpoint or outcome that is feasibly measurable and relatable to patients' QoL. In the context of sleep and RA, patients may say, "Normally, I don't sleep very well. When my disease is active, I sleep even worse. I'm tired all the time, and it's hard to function. I wish there were something that would help me sleep better." Sponsors try to translate this into existing quantifiable measures – total sleep time (TST), sleep efficiency (SE), wake after sleep onset (WASO), number of nighttime awakenings, and sleep onset latency (SOL) – all well-established concepts in sleep medicine, which have been measured in RA patients in clinical sleep labs using PSG [20–34].

The Digital Advantage

After substantiating the justification for the measurement of sleep as a viable COI in RA, it is essential to articulate the value of digital measures beyond the existing assessment tools. In other words, why digital? By defining the COI (i.e., quantitative measurement of sleep in moderate-to-severe RA patients), we describe the conceptual framework for validation of the NDE. These requirements allow for selecting sensors and the analytical requirements needed to assess the accuracy, specificity, and sensitivity of the endpoint(s) required to meet the COI outlined in Figure 1 below.

Clinically, sleep is characterized by a behavioral and physiological state. Based on the electroencephalogram (EEG) assessment, sleep comprises 2 major phases: (1) nonrapid eye movement sleep and (2) rapid eye movement sleep, with the former being further divided into 3 stages (N1, N2, and N3) of increasing sleep depth. PSG, which records EEG, electromyography, blood oxygen level, heart rate, respiratory rate, and eye and leg movements, is regarded as the clinical "reference standard" for assessing sleep stage and quantity. However, PSG requires in-clinic sleep laboratory visits, which are influ-

Fig. 1. Conceptual framework for measuring sleep in RA. The COI for measuring sleep in RA patients will monitor treatment effect via the actigraphy device and algorithms. The measurement of the variables of sleep in a defined patient population provides information about a meaningful aspect of the patients' condition affected by RA and the effectiveness of treatment, thus defining the conceptual framework for validation of the NDE. TNF- α , tumor necrosis factor- α ; COI, concept of interest; RA, rheumatoid arthritis; NDE, novel digital endpoint.

enced by setting and discomfort and limit its scalability in global clinical trials. In addition, the manual scoring of the PSG results may increase the complexity and variability of the study results. Sleep quality may be assessed via a PRO questionnaire. However, these survey-based assessments are often just a "snapshot" in time. Using a PRO to assess daily or weekly sleep quality also introduces subjectivity and recall errors into the study. Due to the limitations of the existing tools, drug developers have not been able to conduct large-scale trials to study the impact of SD in RA despite its clinical importance to patients' QoL. With the advancement of DHTs in the marketplace today, the assessment of sleep quality and quantity may be feasible.

Digital technologies offer many advantages, as we have seen above. However, the patient's input remains crucial in developers' consideration for success. Sponsors benefit from considering patients' unique needs in living with their disease, including the invasiveness and added burden new technologies can bring to patients. Digital technologies may allow for a broader understanding of the diverse patient population's lived experience. Their introduction can also increase access inequity and lead to a bias

in the recruited patient population (e.g., tech-savvy patients and patients with cognitive impairment, impaired dexterity, or visual impairments). In addition to patient input on meaningful aspects of health related to their disease, it is also essential to consider their input early on. Components such as the user experience, the tool's design, and the measurement schedule support successful and compliant use in the clinical trial.

Selection of a Fit-for-Purpose Digital Sensor for Use Validation

Fit-for-Purpose Sensor Selection

Currently, based on the sensor type and application method, sleep monitoring sensors can be classified into 3 different categories: (1) in-bed and in-room sensors (e.g., radiofrequency), (2) EEG-based sensors, and (3) accelerometry-based sensors. Sleep parameters in RA patients have been evaluated using various methods, including PSG and actigraphy. Since RA patients have worse measurements for TST, wake after sleep onset, SOL, and SE than healthy controls [35], these sleep parameters can be

Fig. 2. Studies and objectives for validation. Based on our hypothetical NDE for sleep parameters in patients with RA. Δ , change; ClinRO, clinician-reported outcome; DHT, digital health technology; HV, healthy volunteer; PRO, patient-reported outcome; PSG, polysomnography; RA, rheumatoid arthritis; SE, sleep efficiency; TNF, tumor necrosis factor; TST, total sleep time; WASO, wake-after-sleep onset.

used as technical requirements for selecting fit-for-purpose sensor(s) and analytics.

During sensor selection, accuracy and reliability of the sleep parameters are vital criteria for consideration. The accuracy of a sensor and algorithm for a given location in the patient's environment or on the body should support the COI. In addition, a sleep assessment that allows for regular monitoring of sleep quantity at a patient's home over time is ideal. Form factor, user burden, battery life, mode of data transmission, Wi-Fi dependency, data privacy and security, and cost are also important considerations. Further, having a clear understanding of the capabilities of the algorithm/platform is critical to the success of the experimental design. There are significant benefits and trade-offs between open and closed platforms that should be considered during the selection of the mobile sensor device. Guiding this selection is the ability to validate the collection of data, the accessibility of the platform to develop and further innovate, and the ability of the sensor to provide a "complete" picture of the quantifiable biometric of interest.

In sleep measurements, collecting endpoints such as "total sleep" hours may be similar between devices. On the other hand, sleep staging will be evaluated differently between sensors as data are extrapolated based on algorithms and will vary significantly. It is essential to discuss with manufacturers how their analysis algorithms were

developed, validated, and benchmarked using 1 sensor over another and whether they were tested in the intended study population. If not, the sponsor may be required to revalidate for the intended clinical population.

According to the practice guideline from the American Academy of Sleep Medicine (AASM) [36], actigraphy can be a useful clinical tool for the evaluation of adult and pediatric patients with suspected sleep disorders. Hence, in this proposed study, we chose an actigraphy sensor to measure activity and quantify the sleep parameters. According to the V3 framework, besides the technical verification, the sleep sensor should be analytically validated against a PSG reference and its accuracy and reliability determined to measure sleep in this RA population. Suppose the sensor is only validated in the healthy population. In that case, additional analytical validation will need to be conducted to demonstrate the sensitivity and specificity of the measurements in the population of interest. PSG, the "reference standard" for measuring sleep parameters, will need to be used as the anchor for the validation, regardless of whether it is a consumer [37] or medical-grade sensor. Although a medical-grade sensor may be perceived to have a greater level of quality control, it is essential to select a fit-for-purpose sensor according to the FDA guidance on the qualification process for drug development tools (DDTs) and follow the regulation of clinical outcome assessments (COAs) [34] to measure

how a patient feels, functions, or survives. A medical device with 510k status (or CE mark) may be irrelevant during NDE development. It does not substitute analytical validation of the DHT or clinical validation of the novel measure in the context of novel endpoint development.

Define Validation Objectives

To develop and validate an NDE of sleep for RA drug development, ideally, an analytical validation study and a clinical validation study should be performed in a sequential manner [11]. However, it is worth noting that if a sensor already has significant analytical validation completed in the RA population, it is possible to move the sensor into clinical validation directly. Here, we propose a Phase 2 proof of concept study if standard safety and efficacy objectives for the investigational drug are in place. Additionally, we designate this as a clinical validation study in which we also include a sub-study of analytical validation for the digital measurement tool. The objectives of the validation study encompassing both analytical and clinical validation are summarized in Figure 2.

In this validation study, the eligibility criteria of the RA patients can be based on a published anti-TNF agent trial design (see online suppl. Material 1; for all online suppl. material, see www.karger.com/doi/10.1159/000518024) (NCT01185301 on <https://clinicaltrials.gov>). It is important to note that any comorbidity, such as pre-existing insomnia, obstructive sleep apnea, or restless leg syndrome, will be a confounder. Patients with these conditions may not be included in the study if the diagnosis is already known.

Clinical Validation Study

Outlining Study Objectives

According to the V3 framework, the DHT should be analytically validated in the targeted population first; however, analytical and clinical validation may be combined in 1 study to accelerate the development. We chose an actigraphy watch localized on the nondominant wrist that several independent research groups have analytically validated for assessing TST and was compared to the PSG reference standard and was found to have good agreement (97%) for TST [38] and overall agreement rates of 91–93% in adults [39, 40]. As such, it minimizes some analytical validation effort by the sponsor. However, each measure to be captured (e.g., SOL, TSD, and SE) must be validated against the reference standard, such as PSG, and in the population of interest (i.e., active RA

patients for the intended purpose such as being a surrogate endpoint) as defined by the DDT qualification process and the BEST (biomarkers, endpoints, and other tools) guidelines [41, 42]. Therefore, the first clinical goal of this study was to validate novel sleep measures using DHTs in an RA population without treatment intervention and compare DHT measures of sleep with reference standard measures of sleep (e.g., PSG).

The second clinical goal of the validation study was to demonstrate that the measurement can detect a clinically meaningful increase in sleep following the anti-TNF therapy in an RA population, ultimately supporting the drug efficacy claims. To support these claims, we will first evaluate prior hypotheses about the relationships of DHT parameters, clinical PRO parameters, biological parameters, and the reference standard for sleep (online suppl. Material 2). The first step required to answer these questions is to investigate the relationship between all parameters generated at screening, test a priori hypothesis about these relationships, and test what aspects of biology are being captured by novel DHT parameters. For example, is time on anti-TNF therapy inversely correlated with the Pittsburgh Sleep Quality Index (PSQI) [43], a PRO of sleep? Does the PRO of sleep, PSQI, correlate with the PSG reference standard? Understanding these relationships and the quality of each variable is critical to understanding how each variable may respond to treatment intervention. The variables and frequency of data collection for a possible validation study are detailed in a time-event table in online suppl. Material 3. Study design considerations for a possible validation study are outlined in Table 2.

Sample Size

When designing a clinical research study, it is critical to estimate sample size correctly. The validation study should be powered to detect the effect of sleep improvement using the reference standard PSG. Thresholds for desired effect sizes for any sample size calculation should consider covariate relevant factors (e.g., disease severity, age, or sex).

Two questions are relevant when considering sample size with wearable sensor data in this study context. The first is the number of patients needed to detect a treatment effect in a specific RA population for sleep improvement relative to placebo. Power calculations can be performed to determine the theoretical effect size between treated and untreated individuals. The estimated expected effect can be selected from the literature on PSG or other sleep PRO data. This process can be repeated for

Table 2. Study design considerations for development of a NDE

Properties	Questions	Analysis and considerations
Sample size How many patients are needed to answer key questions?	How many patients are needed to detect a treatment effect in an RA population for sleep improvement? Can we detect robust relationships between the reference standard and the wearable sensor parameters?	Power calculation to design study able to detect reference standard (PSG) effect Effect size estimates from the RA and the actigraphy literature can be used here Power calculation to design study able to detect correlation between 2 measures
Test-retest reliability Are parameters measuring values reliably?	Is the ICC for all parameters >0.5? (favored for clinical parameters)	Compute the CV and ICC of each parameter, adjusting for relevant clinical and demographic information Perform at baseline and post treatment
Construct validity: convergent	Variables expected to have positive or negative relationships a priori do in fact follow those expected trends	Correlate parameters with each other (e.g., Spearman correlation)
Construct validity: divergent	Is there a weak relationship between digital sensor and sleep diary scores against other measurements or endpoints that are hypothesized not to be associated with sleep?	Perform K means clustering to determine how many “groups” of variables there are
Construct validity: known groups	Do prespecified hypotheses about variable relationships follow expected trends?	
Ability to detect change over the course of treatment	Do changes from baseline values in the treatment arm significantly differ from those in the placebo arm? Are individuals posttreatment significantly different from their pretreatment state?	Standard clinical change from baseline analyses
Establishing meaningful change (e.g., responder/nonresponder)	Are there appropriate criteria to determine responders versus nonresponders?	Determine appropriate cutoffs for drug responders and nonresponders, and test hypotheses

CV, coefficient of variation; ICC, intraclass correlation; RA, rheumatoid arthritis; PSG, polysomnography.

actigraphy parameters, PRO, and PSG parameters. Population means and standard deviations can be estimated from the literature or public data, or may come directly from the manufacturer.

The second question about sample size relates to the need to detect robust relationships between the reference standard and the wearable sensor parameters, a component of analytical validation. For example, asking how many patients are needed to see a Spearman correlation coefficient of at least 0.3 between PSG and relevant DHT may be applicable. The necessary threshold set must be agreed upon by the regulatory authority and the study sponsor in advance. The lower the desired correlation coefficient, the larger the *N* in the study. In sum, understanding all relevant questions about study design before the study is conducted allows for appropriate power calculations and sample size estimations to be performed.

Define Test-Retest Reliability

Parameters measured for the same participant under the same biological setting should have the same (or

very similar) values if measured multiple times; this is called test-retest reliability. Test-retest reliability for any clinical measurement is an effort to quantitate how well the data for a particular measurement from the same patient are similar. This is typically done with the intraclass correlation coefficient (ICC). The ICC measures the consistency of data generated from the same patient and is particularly important when there are repeated independent measures within a patient of the same parameter as there are for many of the parameters in this study. ICC values range from 0 to 1, and values >0.5 are favored for clinical parameters [39, 40]. If there are no repeated measures, the coefficient of variation (CV) may be used. Both the CV and ICC can be computed from a mixed-effects model adjusted for covariates (e.g., sex, age, and BMI) and considered the random effect of patients.

Test-retest reliability can be computed overall (all 12 weeks of the proposed study [Schedule of Assessments in online suppl. Material 3]) and at the following binned one-week time points: screening/baseline and week 12. ICC for each parameter is computed similarly. One po-

tential outcome of this analysis is as follows: if the ICC of a particular parameter from a digital sensor (e.g., TST) is higher than the ICC for a PRO over the same period, then an argument can be made that the parameter from the wearable sensor has better test-retest reliability and may be a better clinical measurement. Repeated clinical measures are ideally appropriately independent, for example, measured over several days or weeks, rather than back-to-back in the same session.

Construct Validity

Construct validity is an effort to establish evidence that the sleep-related outcome scores derived from all metrics conform to prespecified, hypothesized constructs. Correlation between all parameters during a nontreatment screening phase allows for the determination of relationships between wearable sensor data, PROs, and reference standards of sleep at baseline in RA. Variables with significant (e.g., by p value) and strong correlation coefficients (e.g., Spearman's ρ) will be considered related. Correlation matrices of the relationships between parameters can be visualized to understand how the digital sensor outputs connect to the PRO data or any other relevant parameter. To determine how many aspects or unique types of data are captured, methods like k -means clustering can determine the optimal number of clusters of measurements that exist among all parameters. Collectively, these analyses provide an understanding of the overall relationship of all parameters, group parameters into related sets, and tell us which aspects of biology are being measured by PROs, biological, and/or digital sensor variables.

Internal validity of the study is demonstrated by showing that the variables expected to have positive or negative relationships a priori follow the expected trends. Divergent validity is demonstrated by showing a weak correlation between digital sensors and sleep diary scores against other measurements or endpoints hypothesized not to be associated with sleep. Finally, known-group validity tests can show if digital sensor data can detect a priori statistical hypotheses.

To demonstrate the internal validity of the digital sensor, sleep diary, and other outcomes of interest (e.g., RA PROs [PSQI and Medical Outcome Study Sleep Scale]), we can test for associations between the data generated from the wearable sensor and data from these other study measurements. For example, we can calculate the correlation coefficient between the continuous numerical outputs from the sleep diary (e.g., hours of sleep each night and quality of sleep) with continuous

data generated from the digital sensors and other measurements.

Ability to Detect Change throughout Treatment

This study design collects data from screening/baseline to posttreatment for a variety of parameters. For all parameters, change from baseline comparisons for treatment and placebo can be conducted to determine if each clinical parameter captures any clinical efficacy of the study drug. A prespecified window at each time point (e.g., a minimum of 7 continuous days of monitoring at each time point) may be used to determine the baseline and posttreatment data.

Although greater sensitivity to change than the reference standard is not a requirement, it is important that the NDE can detect a clinically relevant change within the anticipated timelines of the clinical study. The mean-to-standard deviation ratio of decline (i.e., the longitudinal change over the period during which the disease is expected to progress relative to population variability) is the preferred method to demonstrate sensitivity to change. Known clinically meaningful changes of reference standard measures combined with cross-sectional correlations can serve as an early indicator of the sensitivity to change of the new measurement. Still, these need validation in a longitudinal setting.

Computing change or percent change from baseline for each parameter over a period during which there is an expected effect is the standard method for determining differences between groups. This method has the statistical advantage of considering the random impact introduced by different patients. Suppose continuous data over the entire course of treatment are used in an exploratory manner and may also consider linear modeling of change over time with relevant covariates.

Finally, the critical concept of minimal clinically important difference, or the degree of improvement that is meaningful to patients, is typically reported as a change from baseline [44]. This is especially useful if there is a treatment intervention, and the changes can be compared between 2 groups (e.g., placebo vs. treatment at a particular time point vs. baseline). These scoring metrics can be applied to quantitative measures from the PSG reference standard, sleep PROs, and digital sensor parameters; both the significance (p value) and magnitude of the effect (effect size) can be determined and used to understand how these measures capture meaningful change. Minimal clinically important difference values for known PROs may be used as a reference to make a case for detecting significant meaningful change.

DHT-Specific Plans Related to Use Affecting Clinical Trial Design and Data Analysis

Planning and deployment of the chosen wearable within this proposed clinical trial should be reasonably straightforward. However, DHTs require sponsors to ensure that data collected via mobile technology or biometric monitoring technologies meet the requirements needed for audit and FDA, EMA, or other regulatory body inspection readiness through to the composite analysis. Sponsors must also consider strategies associated with data integrity, such as authentication of users and device placement, access, privacy, and security. Sponsors are giving a sensor to a patient and expecting it to measure something. This is not operationally different from an electronic PRO sensor (technology that has been well established). The sponsor defines the sensor requirements, sets the analytical validity of output from the sensor, the context in which it is used, and appropriate levels of access to the data. Additionally, sponsors must ensure that software and firmware design quality life cycle controls are in place for analytical platforms throughout the validation of the novel measurement. The sponsor must also develop and maintain data standards needed for reporting and qualification of an NDE.

As such, sponsors will need to describe endpoints, the rationale for inclusion, the technology selected, and the timing of data collection. Analytical plans can be included in the study protocol or the statistical analysis plan. Sponsors need to be mindful of completing formal clinical system validation activities to ensure data captured from the sensor are accurate, changes to study data are explicable, and adequate data security measures are in place.

Implementing digital technology into study milestones/phases along the drug development pipeline is preceded by a risk assessment to identify, analyze, and mitigate potential health risks to study participants. This assessment is focused on ensuring the integrity of the data because software/sensor malfunction may result in unforeseen compromises to clinical data management. Typically, digital safety is focused on the clinical risk because software/sensor malfunction may result in errors in clinical patient management. Digital security also includes patient data privacy and cybersecurity. In addition, the digital risk analysis should also have electrical safety, which provides for battery/charging cable failure, electromagnetic interferences (e.g., impact on implantable devices), and sensor material biocompatibility to mitigate risks from skin irritation from wearable sensors. Last, sponsors need to establish effective data surveillance measures to ensure that data are coming from the sleep sensor per pre-established plans.

Define Minimal Data Required for Data Analysis

Considering sleep patterns may be affected by many factors, it is unclear how many nights of data collection may be needed to gain insight into the night-to-night variability in sleep patterns. Thus, in our hypothetical study, continuous monitoring of sleep patterns from baseline to week 12 could be helpful to assess whether sleep parameters change from baseline following the anti-TNF treatment. Alternatively, shorter (1-week) assessments at prespecified time points (e.g., baseline [the week before starting study treatment], week 1 [first week of treatment], and week 12 [last week of treatment]) may be considered to reduce user (i.e., patient) burden. Although the primary goal is to collect nightly sleep data to minimize compliance issues (e.g., taking the sensor off during the day and putting it on at night), we propose that study participants wear the sensor continuously (24 h per day, 7 days per week). We also define a minimum wear time of 20 h per day to ensure all night-time sleep periods are captured; this time was chosen arbitrarily to allow the skin to rest or perform other activities that might require removing the watch.

Device manufacturer algorithms, either out of the box or developed as a collaboration between the manufacturer and the sponsor, are designed to detect the difference between an awake state and a sleep state. Wearable sensor devices that measure sleep employ various filters that can see differences between awake and sleep and are robust to missing data (e.g., if the sensor is removed for a brief period or if the sensor is worn improperly for a temporary period). In this context, missing data can be quantified, and determinations about whether the validity of analysis is affected by the amount of missing data can be made based on manufacturer recommendations for using the device properly.

Clinical Interpretation

Interpretation of the study results may include disease duration and scoring on the Disease Activity Score-28 for RA with CRP (DAS28-CRP) scale [45] and how the efficacy results of anti-TNF agent correlate with sleep improvement. Sleep quality can be assessed using the PSQI or any other appropriate scale. The patient's functional assessment can be performed using the Health Assessment Questionnaire Disability Index (HAQ-DI) [46] and the patient's assessment of sleep via a sleep diary. Objective sleep study results can be interpreted from the PSG result for each patient. Age and gender may also be stratified during the interpretation of the results.

Table 3. US/EU pathways depending on the regulatory goal of the interaction

Agencies	Early dialog	Acceptance for one trial	Qualification for specific COU (independent of a molecule)
US FDA EU EMA	CPIM ITF	IND path (e.g., type C meeting) Scientific advice	DDT program Qualification of novel methodologies

COU, context of use; CPIM, Critical Path Innovation Meeting; DDT, drug development tool; EU EMA, European Union European Medicines Agency; IND, Investigational New Drug; ITF, Innovation Task Force; US FDA, United States Food and Drug Administration.

Regulatory Engagement Pathways

The US FDA offers several pathways for sponsors to seek input on the development of NDEs. When an innovative methodology or technology is being conceptualized at the earliest stages of development, the Critical Path Innovation Meeting might be most appropriate, and feedback is nonbinding. Because Critical Path Innovation Meetings are outside the IND process, they offer an opportunity for sponsors to present novel tools for use in drug development to the Center for Drug Evaluation and Research (CDER) (and other centers if the sponsor requests their presence) and get conceptual feedback; however, this forum is not intended for the FDA to provide input on data [47].

Once sponsors have collected sufficient data to move forward with the qualification process, there are 2 options: the IND pathway using Type C Meetings [44] or the DDT pathway [34]. The IND pathway is used to get feedback and acceptance from FDA on the digital endpoints for use in a particular drug development program. Sponsors can request focused Type C Meetings on the COA/biomarker under development and invite specific experts such as COA staff members and the relevant review division for the sensor at the Center for Devices and Radiological Health (CDRH) [48]. Under the DDT pathway, the COA and Biomarker Qualification Program allows a broader qualification of the digital endpoint that can be used across multiple drug programs with the same context of use (COU). It should be noted that both the IND pathway and DDT pathway can be utilized in parallel, and it is recommended to pursue both avenues for the most comprehensive application of the digital endpoint in question.

In Europe, the EMA offers different pathways to discuss digital endpoints. Before confirmatory clinical data are collected at the conceptual stage, the sponsor can request nonbinding information through an Innovation Task Force meeting [49]. The Innovation Task Force

meeting is an opportunity to pressure test the plans, determine the relevance of the endpoint, and determine if the sponsor is going in the right direction.

Suppose the sponsor wants to start the process to get the endpoint accepted (i.e., qualified) by the EMA for use in drug development. In that case, there are 2 possible avenues, depending on the intent. If the goal is to use the endpoint in 1 single drug trial only, the developer can use the scientific advice process for this clinical trial; however, the agency may direct the sponsor to a dedicated qualification advice process to discuss the NDE. Suppose the aim is to qualify the endpoint for all trials within a specific COU. In that case, the “qualification of novel methodologies for medicine development” process should be considered, starting with the “qualification advice” process at the endpoint development stages or through the “qualification opinion” process when the endpoint has been validated by the sponsor (EMA published guidance [50] and Q&A [51]). Table 3 presents possible US/EU pathways depending on the regulatory goal of the interaction from an early nonbinding dialog at the conceptual stage (no clinical data/might have a clinical plan, proof of concept, literature, prototype, and conceptual framework) through interactions to get acceptance for the use of the endpoint in one specific trial and the pathways available to obtain health authority (HA) acceptance to use the endpoint broadly in all trials within 1 COU. As presented in Table 3, both the EMA and FDA offer routes for early interaction in endpoint development. They also allow interactions to obtain acceptance for single trials. Their processes differ for the whole qualification of the endpoints for a specific COU. FDA strives for transparency at all stages of the DDT program, publishing large parts of documents required for the 3-stage process: letter of intent, qualification plan, and full qualification plan submitted by sponsors [41], whereas the EMA only publishes the draft and final qualification opinion [50].

Furthermore, FDA has expected review timelines for each of the 3 stages, while the EMA does not require qual-

Fig. 3. HA interactions to discuss the development and validation of the NDE when developed in parallel to a molecule. These interactions are mapped against the following: phases of drug development, phases of digital endpoint development and validation, and the NDE qualification pathways available at the EMA and the US FDA. This figure illustrates how developers can present the evidence needed at each stage of the NDE qualification pathways through the interactions of the drug development program to obtain HA acceptance on the use of the endpoint in the pivotal clinical trial.

cal trial. COA, clinical outcome assessments; COI, concept of interest; COU, concept of use; CPIM, Critical Path Innovation Meeting; DHT, digital health technology; FQP, full qualification package; GxP, Good X Practices, where X refers to Laboratory, Manufacturing, Clinical, etc.; IND, Investigational New Drug; ITF, Innovation Task Force; LOI, letter of intent; QP, qualification plan; HA, health authority; US FDA, United States Food and Drug Administration; NDE, novel digital endpoint; EMA, European Medicines Agency.

ification advice before the opinion step. EMA does have a public consultation step at the opinion level and considers the comments from different stakeholders in their final opinion. Because of the 3 stages of the DDT process, the qualification timelines at the FDA can be longer than at the EMA.

It is important to note that even though agencies review the proof of analytical validation for the digital health technology tool (DHTT) used (actigraphy in our hypothetical case study), agencies do not qualify the DHTT used to infer the endpoint. They qualify the measure, COA, or biomarker and the endpoint itself. Consequently, it is essential to reflect the requirements of the DHTT in the qualification opinion to ensure that only fit-for-purpose DHTTs are used in clinical trials.

Regulatory Strategy for the Hypothetical Case Study Presented

As a sponsor developing an NDE for studying sleep in RA patients in parallel to the anti-TNF drug development program in which the endpoint will be used, the most appropriate regulatory pathway is the IND pathway at FDA (Fig. 3). Type C COA meetings could be requested before starting Phase 2 to demonstrate that the digital tool chosen is fit-for-purpose, discuss the endpoint development plan, and determine if the sleep endpoint discussed in the above sections is relevant to the target RA population of interest. At the end of Phase 2, the sponsor could request another Type C COA meeting to present the evidence supporting content validity and construct validity using the data gathered in the analytical and clinical studies discussed earlier in this study (Fig. 2). At the EMA, the qual-

ification advice and scientific advice processes could get the same confirmation as with the FDA. In our case study, we propose 2 meetings. Still, we acknowledge that sponsors may request additional meetings with the agencies depending on the progress of NDE development and any questions/issues during the drug development process. The time points for requesting such arrangements may occur at different times from that in our example. If the goal is to have the NDE accepted as a critical secondary or primary endpoint for use in a Phase 3 trial, we recommend that these meetings occur long before the start of the pivotal trial. In our example, we chose to use a digital device that is 510(k) cleared and CE marked; however, nonmedical devices may be used, and sponsors should consider that additional levels of validation will be needed to support fit-for-purpose, including proof of GxP compliance.

After these consultations with the agencies, if successful, the sponsor will have official acceptance from the EMA and FDA that the sleep endpoint can be used in the anti-TNF RA pivotal trial. In terms of timelines, the sponsor is targeting a global drug development program in RA and would submit the briefing packages and marketing applications to both authorities simultaneously.

Drug development is time-critical; therefore, the current endpoint qualification procedures may not be fit for purpose. Sponsors might not have the resources or the time to go through the DDT COA or qualification opinion procedure before the pivotal trial when the NDE development and drug development are happening in parallel. However, at any point in time afterward, the sponsor can take the strategic decision to qualify the endpoint officially through the DDT COA (FDA) and qualification of novel methodologies in drug development (EMA) pathways. This path would allow for broad use of the novel digital sleep endpoint in trials from other sponsors in the specific COU and ensure ubiquitous use of the endpoint. Sponsors are free to request these consultations when they deem most appropriate during their drug development process.

Requirements for Evidence Presentation

Since our NDE is a COA, we focus on the evidence requirements for COAs in this section. FDA provides detailed recommendations on the content to be included for each phase of the DDT COA submission [52]. It is suggested that sponsors follow this format to receive the best possible outcome from the FDA review. Figure 3 describes the evidence presented in briefing packages for the DDT procedure by stage of the IND procedure. For Type

C meetings under the IND pathway, this FDA guidance can also be consulted (Appendix 1 of the Discussion Guide of PFDD3: <https://www.fda.gov/media/116281/download>).

The EMA guidance on qualification of endpoints includes an annex with the high-level structure for briefing packages. The evidence structure is general for DDTs (e.g., including preclinical models) and not specific for digital endpoints [51]. The Q&A published in 2020 provides more clarity on submitting information to maximize the impact of HA interactions [47]. At the qualification advice stage, the sponsor can decide how much information to provide in each section of the briefing package depending on the stage of endpoint development and the sponsor's questions for the agency. For qualification opinion, the development plan and results have to be presented to support the qualification of the endpoint in the specific COU. An overview of the evidence needed to be provided to HAs to support the use of mobile sensor technology for COAs in clinical trials is presented in recently published articles from Goldsack et al. [11] and Walton et al. [12]. Even if the pathways for qualification of endpoints are different at the EMA and FDA, the evidence requirements for COA and biomarker qualification from both agencies do not differ, as can be seen in the guidance and Q&A published by the agencies [41, 50, 51]. We acknowledge that this field is still evolving. HAs are adjusting their positions and guidance as we learn more about the specificities and needs of NDE development.

HAs often urge sponsor companies to work together to develop NDEs, an approach we strongly endorse. Equally important is the collaboration between regulatory agencies to find new collaborative avenues to share the review and approval of the use of NDEs. We can only resolve the current unmet measurement needs and apply the potential of technology through true collaboration.

Conclusion

When planning to develop an NDE, sponsors must first ascertain 4 things: (1) scientific justification of what needs to be measured, (2) clinical meaningfulness to the patient, (3) reasons for not using existing tools (or *why digital*) before pursuing the technology, and (4) the accuracy and reliability of the new tool. Developing NDEs using a wearable sensor should follow a rigorous process, such as DDT as defined by FDA, CTTI, and the V3 framework. Verification evaluates the performance of DHT and confirms it is collecting accurate data within prespec-

ified thresholds set forth by the manufacturer. Verification is performed before collecting data from human subjects to ensure the collected data are consistently accurate and precise. Subsequent analytical validation entails the evaluation of a wearable sensor for generating meaningful data using an algorithm. This is performed by establishing a data capture protocol to collect verified raw data from a defined patient population and process it by a specific algorithm(s). The processed data must then be evaluated against an established reference to demonstrate a reproducible correlation to a health outcome. For example, sleep onset/wake can be validated against PSG. Through these processes, sponsors will explore improved and novel phenotypes that can be captured with wearables and sensors.

To illustrate the processes and considerations of developing, validating, qualifying, and using an NDE in the development program of a new drug, we (the TransCelerate NDEs team) embarked on the development of this hypothetical study proposal to validate the scientific hypothesis of an anti-TNF effect on sleep as well as the use of actigraphy as a sleep measure in the context of RA drug development. The proposed approach was a synthesis based on the collective knowledge from the FDA DDT qualification process, Foundation for the National Institutes of Health (FNIH), BEST, V3, and digital outcome assessment submission frameworks, as well as the CTTI Digital Health Initiative to provide a holistic consideration to support an NDE to measure sleep in RA patients. While this case study outlines many aspects of NDE development, sponsors should also consider the operational challenges and site-related issues with implementing technology in clinical trials, which have been outlined extensively in TransCelerate's publication titled "Accelerating Adoption of Patient-Facing Technologies in Clinical Trials: A Pharmaceutical Industry Perspective on Opportunities and Challenges" [53].

As many sponsors, clinicians, and patients are interested in the consistency of measurement solutions, it is important to strive for harmonization and alignment with stakeholders to avoid multiple measurement solutions for similar COIs to be developed. Industry consortia, such as Critical Path Parkinson and Innovative Medicine Initiatives, have been leading projects in the development of novel digital endpoints. At the current state, those efforts do not preclude companies and sponsors from developing endpoints based on the COI. In this case study, we are grounding our NDE in current clinical PSG terminology (e.g., TST). As such, we are bounded by the technical validation of the COI in com-

parison to the PSG standard. These topics are also covered under CTTI's guidance, which has been referenced within our study.

Whether it is an independent development or consortia approach, Walton et al. [12] have described a detailed evidentiary dossier framework. As for the clinical validation, due to SD having many potential confounders (i.e., fatigue, depression, and pain), this may be a challenging concept to validate as SD may be affected by external factors and not a modifiable anti-TNF agent. Hence, regulatory engagement pathways and strategic considerations are a critical part of this study proposal. We recommend gaining alignment with the regulatory agency early in the NDE development process and considering this to be an essential factor for success.

Acknowledgments

The authors gratefully acknowledge the support of TransCelerate Biopharma Inc., a nonprofit organization dedicated to improving the health of people around the world by accelerating and simplifying the research and development (R&D) of innovative new therapies. The organization's mission is to collaborate across the global biopharmaceutical R&D community to identify, prioritize, design, and facilitate the implementation of solutions to drive efficient, effective, and high-quality delivery of new medicines. The authors also gratefully acknowledge the support of the following TransCelerate working team members who contributed to the concepts described in this study: Amber Williams and Tyler Reynolds, for guidance, review, and project management support, and Frances Pu, Ph.D., for medical writing.

Statement of Ethics

An ethics statement was not required for this study proposal as no human or animal subjects or materials were used.

Conflict of Interest Statement

The authors declare no conflicts of interest. However, all authors are employees and/or stockholders of the companies with which they are affiliated.

Funding Sources

All authors are employees of the companies with which they are affiliated. All companies are members of TransCelerate Biopharma, Inc. and contribute both funds and workforce to it. Additionally, TransCelerate Biopharma, Inc. paid a nonauthor, Frances Pu, Ph.D., for medical writing support.

Author Contributions

All authors were involved in the writing, editing, and final approval of the article. Conceptualization was done by M.C., R.J.M., M.F.W., R.A., L.L., and T.M. All authors made substantial contributions to the conception or design of the work, participated in drafting and revisions, and provided final approval of the version to be published. The corresponding author attests that all listed authors meet the authorship criteria and that no others have been omitted.

Data Availability Statement

There are no data generated or analyzed in this study proposal.

References

- Clinical Trials Transformation Initiative. Novel endpoints project. 2017. Available from: <https://www.ctti-clinicaltrials.org/projects/novel-endpoints>.
- Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiol Meas*. 2018;39:05TR01.
- Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol*. 2017;15:e2001402.
- Digital Medicine Society. Library of digital endpoints. 2020. Available from: <https://www.dimesociety.org/index.php/knowledge-center/dime-generated-content/46-knowledge-center/digital-endpoints>.
- Nathan SD, Flaherty KR, Glassberg MK, Raghu G, Swigris J, Alvarez R, et al. A randomized, double-blind, placebo-controlled study of pulsed, inhaled nitric oxide in subjects at risk of pulmonary hypertension associated with pulmonary fibrosis. *Chest*. 2020;158:637–45.
- European Medicines Agency. Qualification opinion on stride velocity 95th centile as a secondary endpoint in Duchenne Muscular Dystrophy measured by a valid and suitable wearable device. 2019. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-secondary-endpoint-duchenne-muscular-dystrophy_en.pdf.
- Idea-Fast Consortium. Home page. 2019. Available from: <https://idea-fast.eu/>.
- Cedarbaum JM, Taylor K, Hill D, Alexander R, Luo Y, Rubens R, et al. Enabling efficient use of digital health technologies to support Parkinson's disease drug development through precompetitive collaboration. 2019. Available from: https://c-path.org/wp-content/uploads/2019/09/cpp_mds_2019_final_revised.pdf.
- Innovative Medicine Initiative. Mobilise-D. 2019. Available from: <https://www.imi.europa.eu/projects-results/project-factsheets/mobilise-d>.
- Food and Drug Administration. Patient-focused drug development: collecting comprehensive and representative input guidance for industry, Food and Drug Administration staff, and other stakeholders. 2020. Available from: <https://www.fda.gov/media/139088/download>.
- Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). *npj Digit Med*. 2020;3:55.
- Walton MK, Cappelleri JC, Byrom B, Goldsack JC, Eremenco S, Harris D, et al. Considerations for development of an evidence dossier to support the use of mobile sensor technology for clinical outcome assessments in clinical trials. *Contemp Clin Trials*. 2020;91:105962.
- Sparks JA. Rheumatoid arthritis. *Ann Intern Med*. 2019;170(1):ITC1–16.
- Barbour KE, Helmick CG, Boring M, Brady TJ. Vital signs: prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation, United States, 2013–2015. *MMWR Morb Mortal Wkly Rep*. 2017 Mar 10;66(9):246–53.
- Mckenna S, Tierney M, O'Neill A, Fraser A, Kennedy N. Sleep and physical activity: a cross sectional objective profile of people with rheumatoid arthritis. *Rheumatol Int*. 2018 May;38(5):845–53.
- Feehan LM, Lu N, Xie H, Li LC. Twenty-four hour activity and sleep profiles for adults living with arthritis: habits matter. *Arthritis Care Res*. 2020;72:1678.
- Abbasi M, Yazdi Z, Rezaie N. Sleep disturbances in patients with rheumatoid arthritis. *Niger J Med*. 2013;22:181–6.
- Goes ACJ, Reis LAB, Silva MBG, Kahlow BS, Skare TL. Rheumatoid arthritis and sleep quality. *Rev Bras Reumatol Engl Ed*. 2017;57:294–8.
- Gullick NJ, Ibrahim F, Scott IC, Vincent A, Cope AP, Garrood T, et al. Real world long-term impact of intensive treatment on disease activity, disability and health-related quality of life in rheumatoid arthritis. *BMC Rheumatol*. 2019;3:6.
- Hirsch M, Carlander B, Vergé M, Tafti M, Anaya JM, Billiard M, et al. Objective and subjective sleep disturbances in patients with rheumatoid arthritis. A reappraisal. *Arthritis Rheum*. 1994;37:41–9.
- Løppenthin K, Esbensen BA, Jennum P, Østergaard M, Tolver A, Thomsen T, et al. Sleep quality and correlates of poor sleep in patients with rheumatoid arthritis. *Clin Rheumatol*. 2015;34:2029–39.
- Irwin MR. Sleep and inflammation: partners in sickness and in health. *Nat Rev Immunol*. 2019;19:702–15.
- Krueger JM, Rector DM, Churchill L. Sleep and cytokines. *Sleep Med Clin*. 2007;2:161–9.
- Krueger JM. The role of cytokines in sleep regulation. *Curr Pharm Des*. 2008;14(32):3408–16.
- Opp MR. Cytokines and sleep. *Sleep Med Rev*. 2005;9(5):355–64.
- Lange T, Dimitrov S, Born J. Effects of sleep and circadian rhythm on the human immune system. *Ann N Y Acad Sci*. 2010;1193:48–59.
- Vgontzas AN, Papanicolaou DA, Bixler EO, Lotsikas A, Zachman K, Kales A, et al. Circadian interleukin-6 secretion and quantity and depth of sleep. *J Clin Endocrinol Metab*. 1999;84:2603–7.
- Wolfe F, Michaud K, Li T. Sleep disturbance in patients with rheumatoid arthritis: evaluation by medical outcomes study and visual analog sleep scales. *J Rheumatol*. 2006;33:1942–51.
- Digital Medicine Society. The playbook: digital clinical measures. 2020. Available from: <https://playbook.dimesociety.org/>.
- Coskun Benlidayi I. Sleep impairment: an obstacle to achieve optimal quality of life in rheumatoid arthritis. *Rheumatol Int*. 2018;38:2183–92.
- Straub RH, Detert J, Dziurla R, Fietze I, Loeschmann PA, Burmester GR, et al. Inflammation is an important covariate for the crosstalk of sleep and the HPA axis in rheumatoid arthritis. *Neuroimmunomodulation*. 2017;24:11–20.
- Taylor-Gjevne RM, Gjevne JA, Nair B, Skomro R, Lim HJ. Components of sleep quality and sleep fragmentation in rheumatoid arthritis and osteoarthritis. *Musculoskeletal Care*. 2011;9:152–9.

- 33 Luyster FS, Chasens ER, Wasko MC, Dunbar-Jacob J. Sleep quality and functional disability in patients with rheumatoid arthritis. *J Clin Sleep Med*. 2011;7:49–55.
- 34 Bourguignon C, Labyak SE, Taibi D. Investigating sleep disturbances in adults with rheumatoid arthritis. *Holist Nurs Pract*. 2003;17:241–9.
- 35 Grewal GS, Amlani A, Hazlewood G. 0915 Sleep in rheumatoid arthritis: a systematic review and meta-analysis of objective sleep parameters. *Sleep*. 2018;41(Suppl 1_1):A340.
- 36 Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med*. 2018;14:1231–7.
- 37 Khosla S, Deak MC, Gault D, Goldstein CA, Hwang D, Kwon Y, et al. Consumer sleep technology: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2018;14:877–80.
- 38 Jean-Louis G, von Gizycki H, Zizi F, Fookson J, Spielman A, Nunes J, et al. Determination of sleep and wakefulness with the actigraph data analysis software (ADAS). *Sleep*. 1996;19:739–43.
- 39 Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD. Sleep detection with an accelerometer actigraph: comparisons with polysomnography. *Physiol Behav*. 2001;72:21–8.
- 40 Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep*. 1994;17:201–7.
- 41 Food and Drug Administration. Qualification process for drug development tools guidance for industry and FDA staff. 2019. Available from: <https://www.fda.gov/media/133511/download>.
- 42 Food and Drug Administration. FDA-NIH biomarker working group. BEST (biomarkers, endpoints, and other tools) resource. 2016. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK326791/>.
- 43 Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The pittsburgh sleep quality index (PSQI): a new instrument for psychiatric research and practice. *Psychiatry Res*. 1989;28:193–213.
- 44 McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA*. 2014;312(13):1342–3.
- 45 Fransen J. Disease activity score-28 for rheumatoid arthritis with CRP (DAS28-CRP). Available from: <https://www.mdcalc.com/disease-activity-score-28-rheumatoid-arthritis-crp-das28-crp>.
- 46 Division of Immunology & Rheumatology, Stanford University School of Medicine. The health assessment questionnaire disability index (HAQ-DI). Available from: https://www.niehs.nih.gov/research/resources/assets/docs/haq_instructions_508.pdf.
- 47 Food and Drug Administration. Critical Path Innovation Meeting (CPIM). 2019. Available from: <https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/critical-path-innovation-meetings-cpim>.
- 48 Food and Drug Administration. Guidance for industry: formal meetings between the FDA and sponsors or applicants of PDUFA products. 2017. Available from: <https://www.fda.gov/media/109951/download>.
- 49 European Medicines Agency. Available from: https://www.ema.europa.eu/en/documents/leaflet/innovation-task-force_en.pdf.
- 50 European Medicines Agency. Qualification of novel methodologies for drug development: guidance to applicants. 2014. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf.
- 51 European Medicines Agency. 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.
- 52 Food and Drug Administration. Clinical outcome assessments (COA) qualification program resources. 2020. Available from: www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessments-coa-qualification-program-resources.
- 53 Polhemus AM, Kadhim H, Barnes S, Zebrowski SE, Simmonds A, Masand SN, et al. Accelerating adoption of patient-facing technologies in clinical trials: a pharmaceutical industry perspective on opportunities and challenges. *Ther Innov Regul Sci*. 2019 Jan;53(1):8–24. Available from: <https://journals.sagepub.com/doi/10.1177/2168479018801566>.