Remote Cardiac Safety Monitoring through the Lens of the FDA Biomarker Qualification Evidentiary Criteria Framework: A Case Study Analysis

Elena S. Izmailova^a William A. Wood^b Qi Liu^c Vadim Zipunnikov^d
Daniel Bloomfield^e Jason Homsy^f Steven C. Hoffmann^g John A. Wagner^h
Joseph P. Menetski^g

^aKoneksa Health, New York, NY, USA; ^bLineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA; ^cOffice of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research Food and Drug Administration, Silver Spring, MD, USA; ^dJohns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ^eAnthos Therapeutics, Cambridge, MA, USA; ^fTakeda Pharmaceuticals International, Cambridge, MA, USA; ^gFoundation for the National Institutes of Health (NIH), North Bethesda, MD, USA; ^hCygnal Therapeutics, Cambridge, MA, USA

Keywords

Digital health · Cardiac · Remote monitoring · Biomarkers · Sensors · Evidentiary criteria

Abstract

Clinical safety findings remain one of the reasons for attrition of drug candidates during clinical development. Cardiovascular liabilities are not consistently detected in early-stage clinical trials and often become apparent when drugs are administered chronically for extended periods of time. Vital sign data collection outside of the clinic offers an opportunity for deeper physiological characterization of drug candidates and earlier safety signal detection. A working group representing expertise from biopharmaceutical and technology sectors, US Food and Drug Administration (FDA) public-private partnerships, academia, and regulators discussed and presented a remote cardiac monitoring case study at the FNIH Biomarkers Consortium Remote Digital Monitoring for Medical Product Development workshop to

examine applicability of the biomarker qualification evidentiary framework by the FDA. This use case examined the components of the framework, including the statement of need, the context of use, the state of the evidence, and the benefit/risk profile. Examination of results from 2 clinical trials deploying 510(k)-cleared devices for remote cardiac data collection demonstrated the need for analytical and clinical validity irrespective of the regulatory status of a device of interest, emphasizing the importance of data collection method assessment in the context of intended use. Additionally, collection of large amounts of ambulatory data also highlighted the need for new statistical methods and contextual information to enable data interpretation. A wider adoption of this approach for drug development purposes will require collaborations across industry, academia, and regulatory agencies to establish methodologies and supportive data sets to enable data interpretation and decision-making.

© 2021 The Author(s)
Published by S. Karger AG, Basel

FDA Biomarker Qualification Framework

The US Food and Drug Administration (FDA) has been working with external stakeholders to develop biomarkers as drug development tools to promote efficiencies and innovation. The FDA Center for Drug Evaluation and Research (CDER) has established a Biomarker Qualification Program to facilitate this mission. A list of qualified biomarkers can be found on the FDA website [1]. Biomarker qualification efforts started in 2008 and were carried out in collaboration with external stakeholders yielding qualifications for safety biomarkers, such as urinary nephrotoxicity biomarkers [2] and serum/plasma cardiotoxicity biomarkers [3]. Later on, diagnostic biomarkers [4], prognostic biomarkers [5], and monitoring biomarkers [6] were qualified.

In December 2018, the FDA CDER and Center for Biologics Evaluation and Research (CBER) jointly published a draft guidance titled *Biomarker Qualification: Evidentiary Framework* [7], providing recommendations on general considerations for developing and qualifying a biomarker as a drug development tool. This guidance drew upon several years of stakeholder meetings and consensus from the biomarker development community. Initial discussions started in August 2015 (symposium cosponsored by the University of Maryland CERSI/FDA/Critical Path Institute) and continued in additional meetings in 2016. An initial framework was suggested by Leptak et al. [8] with broad input from the international experts, including the EMA. To be qualified as a tool for clinical development, a biomarker of interest should address a specific medical need, be measured using reliable methods, and be interpreted in a context appropriate for this tool use and regulation. The evidentiary framework includes the following components: (1) describing the statement of need, which consists of a problem statement and a drug development need; (2) defining the context of use; (3) considering potential benefits and risks associated with the proposed use; and (4) summarizing the state of the evidence which describes existing data and knowledge. Context of use is defined as a statement that fully and clearly describes the way the medical product development tool is to be used and the regulated product development and review-related purpose of the use [9]. The evidence necessary for qualification consists of analytical validation (establishing detection method performance characteristics) and clinical validation (establishing a biomarker's relationship with an outcome of interest). These principles are agnostic to biomarker types and categories.

Remote monitoring, which is defined as data collection outside the clinic while study subjects are not being

supervised by a healthcare professional, has become a very attractive approach for use in drug development trials with high expectations for capturing subject information in real-life settings. This growing field of devices and potential applications has generated great enthusiasm and questions about how to develop these remote approaches for use within clinical trials. The FNIH Biomarkers Consortium Remote Digital Monitoring for Medical Product Development workshop was born from the need to provide a clear understanding on the amount and type of the evidence required in a remote monitoring context to make confident decisions. The workshop was designed to test the applicability of the FDA evidentiary criteria framework by relating it to real-world case studies. Workshop case studies were chosen to touch on multiple characteristics of potential measures that included sensor type, algorithm development, and clinical characteristics of the measure [10].

One of the case studies considered remote cardiac monitoring as a part of safety monitoring in drug development and it is summarized in this review. Clinical safety is defined here as any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment [11]. This case study was specifically chosen because it included known biomarkers, such as heart rate, respiratory rate, and body temperature, but was focused on remote data collection using 510(k)-cleared devices [12]. In addition, the studies had already been completed to assess the viability of this approach and draw a conclusion about devices of interest to be fit for purpose [13, 14]. The case study provided a good basis for determining fit-for-purpose characteristics in a clinical trial beyond FDA requirements for device clearance for a well-established concept measured digitally [15]. Moreover, the need for vital sign remote monitoring came into the spotlight during the COVID-19 pandemic and is discussed in this paper in addition to the workshop materials.

This article is organized according to the components of the biomarker qualification evidentiary framework followed by general conclusions about the applicability of this regulatory guidance to the cardiac monitoring use case.

Problem Statement

The primary objective of early-stage drug development clinical trials is assessment of the emerging safety/tolerability profile along with pharmacokinetic and phar-

Table 1. Comparison of vital sign monitoring during the conventional safety monitoring process and remote ambulatory monitoring

Feature	Conventional vital sign monitoring	Remote ambulatory vital sign monitoring (based on use cases)
Variables of interest	Heart rate, respiratory rate, body temperature, SPO ₂ , blood pressure, ECG-based cardiac abnormalities (yes/no)	Heart rate, respiratory rate, body temperature
Devices	Hospital/clinical pharmacology unit equipment	Technology of choice; a study assessing a sensor and data processing algorithm performance and usability is needed prior to deployment for safety monitoring
Data collection protocol	Resting and supine ^b	Ambulatory conditions which include various activities of daily living
Data collection venue	Hospital or clinical pharmacology unit	Hospital, clinical pharmacology unit, and free-living conditions
Duration of data collection	Predefined time points according to the study protocol ^{b,c}	Frequent dense data collection unless otherwise specified; the total duration of data collection is device and protocol dependent; the duration can range from several days to several weeks
Training requirements	Minimal, limited to site personnel training	Requires additional training for site personnel and study subjects; the level of training may be contingent on the technology literacy level of the intended recipients
Data analysis methodology	The methodology is well established; predominantly descriptive statistics	New statistical methodology is needed to leverage dense data; the data collected at spot checks can be analyzed similarly to conventional vital sign monitoring
Subject compliance	Very high as all procedures are done according to the schedule of assessments under the supervision of a healthcare professional while the study subjects are in the hospital or a clinical pharmacology unit	Can be variable, depending on device choice, device tolerance and usability by study subjects, the duration and frequency of data collection, the study population, and the training provided to sites and study subjects
Missing data handling	Missing data are rare – data are collected under controlled conditions	Data losses due to subject noncompliance, device connectivity, and filtering out invalid data can be substantial; a plan is needed to estimate potential data losses; handling missing data should be specified in the statistical analysis plan for primary and secondary endpoints
Metadata concerning the context of vital sign data collection	Usually not required because data are collected in a controlled environment	The context of data collection such as physical activity or other activities of daily living is critical for data interpretation; it can be collected by means of physical activity trackers and/or patient diaries
Data interpretation	Reference ranges and intervals are well established and used for data interpretation; device body placement is well established and aligned with reference ranges and intervals	Reference ranges and intervals are context dependent, e.g., intensity of physical activity, age, and medical conditions in a population of interest; device body replacement may require revisiting reference ranges and intervals.

^a The manual technique is conventionally used for respiratory rate data collection.

^b Does not apply to ambulatory Holter ECG and blood pressure monitoring.

^c Ambulatory Holter ECG and blood pressure monitoring are usually limited to 24 h or 48 h in a limited number of cases.

macodynamic properties of an investigational compound. Safety data collection includes vital signs, laboratory tests, and patient reports on signs and symptoms which may or may not be related to an investigational drug. Cardiovascular monitoring is an integral part of safety monitoring and routinely includes certain ECG parameters, heart rate, respiratory rate, body temperature, blood oxygenation (SpO₂), and blood pressure. Vital sign data are collected throughout the entire process of drug development, though the collection schedule is sparser in later-stage trials. In early-stage trials, vital sign data (multiple time points per day) are usually collected in first-in-

human and subsequent phase 1 studies. These data are collected before, during, and after administration of an experimental drug while study subjects are confined to a clinical pharmacology unit and during in-person follow-up visits. The confinement periods may range from several days to several weeks depending on the investigational compound properties, the study design, and the anticipated or emerging safety profile. Conventional 12-lead ECG data, SpO₂, blood pressure, heart rate, and body temperature data are usually collected at predefined time points by a clinical pharmacology unit healthcare professional using medical devices. Respiratory rate is usually

Fig. 1. Cardiac monitoring use case according to the FDA biomarker qualification evidentiary criteria framework.

collected manually by observation of chest falling and rising for a defined time. Holter monitoring (continuous multiple-lead ECG) is limited to one or several days during the study confinement for 24–48 h (Table 1).

Statement of Need

Clinical safety issues remain one of the reasons for high rates of attrition for experimental medicines during clinical development. Often, cardiovascular liabilities of the investigational drugs are not detected in early-stage clinical trials and become apparent when drugs are administered chronically for extended periods of time in a larger patient population [16]. Conventional data collection at predefined times during phase 1 studies can lead to missing safety signals as some of the signals can be transient in nature and no data are being collected when study subjects are asleep. Additionally, clinical pharmacology unit confinements for extended periods of time are inconvenient for study subjects and may not provide data reflective of normal day-to-day activity. Safety data collected after subject discharge from the unit are limited to a subject's memory recall, which can have biases, or infrequent follow-up during in person clinical pharmacology unit visits.

Remote ambulatory vital sign data collection may be able to capture the same parameters as conventional vital sign data collection in the unit while offering the oppor-

tunity for a deeper physiological characterization that permits earlier safety signal detection. However, outside of the clinic, adjunctive data, such as physical activity and patient-reported symptoms, are needed as an aid in interpretation of adverse event data collected without the supervision of a healthcare professional. The overall needs for remote vital sign data collection are multifold: (1) collect as much data as possible to better inform the emerging safety profile, (2) reduce the time of residential observation and the number of follow-up visits required while maintaining the quality of data, and (3) ensure that the collected data reflect the real-world patient experience (Fig. 1).

The need for remote vital sign data collection became even more pressing during the COVID-19 pandemic. Investigators and sponsors wished to continue the conduction of clinical trials while maintaining social distancing and limiting the participants' risk. Both US and EU regulators issued guidance for clinical trial conduction during the pandemic [17, 18], urging sponsors to conduct clinical trial assessments remotely whenever possible in order to protect the study subjects while giving access to experimental drugs and maintaining GCP compliance. This need also represented an opportunity – by working out appropriate considerations for outpatient vital sign monitoring, the field could accelerate the use of prolonged outpatient remote monitoring outside of the current phase 1 setting, leading in the long term to earlier safety signal detection and more efficient drug development

Table 2. Summary of the study design for use cases discussed by the cardiac monitoring working group

Study characteristics	Study 1 [10]	Study 2 [11]
Phase	Phase 1	Phase 1
Endpoint	Exploratory	Exploratory
Study population	Normal healthy controls	Normal healthy controls
ECG device/variables collected	HealthPatch by VitalConnect – Heart rate, respiratory rate, skin temperature – Step count (accelerometer)	BodyGuardian by Preventice – Heart rate, respiratory rate – Activity counts (accelerometer)
510(k) clearance status	510(k) cleared [16]	510(k) cleared [17]
Actigraphy device	Actiwatch Spectrum Pro by Philips – Physical activity – Sleep parameters	
510(k) clearance status	510(k) cleared [16]	
ECG tracer level data accessibility	Yes	Yes
Data processing algorithms	Proprietary	Proprietary
Data collection duration	10 days during clinical pharmacology unit confinement	2 days during clinical pharmacology unit confinement and 2 days at home
Data analysis and review	Retrospective	Retrospective

overall. However, important confounding considerations, such as the impact of SARS-CoV2 infection itself on key vital signs, quickly became apparent. Effective outpatient vital sign monitoring must be paired with a robust system of SARS-CoV2 testing and the ability to account for the expected effects of SARS-CoV2 infection on body temperature, heart rate, respiratory rate, physical activity, and other relevant physiologic parameters.

Context of Use

Vital sign elements of a safety profile, including cardiovascular parameters such as heart rate, respiratory rate, SpO₂, and body temperature, can be assessed comprehensively in early-stage clinical trials, during clinical pharmacology unit confinement, or after discharge from the unit by means of remote monitoring technologies and remotely if needed (Table 1). Many devices collect data passively, i.e., not requiring user input other than device placement on the body and recharging, thus improving the participant experience and potentially reducing bias (as data are collected passively and not influenced by the hospital environment, e.g., white coat hypertension). However, one cannot exclude the possibility of introducing a Hawthorne effect [19] by the fact of making study

subjects aware that they are being monitored and inadvertently modifying their behavior. In contrast to devices used for vital sign data collection in the clinical pharmacology unit, wearable sensors should have the following properties: (1) Bluetooth-enabled internet connectivity allowing data streaming into the cloud for in-study review if needed, (2) ability to collect data passively, and (3) sufficient memory and/or ability to stream data frequently to collect data for extended periods of time (e.g., days). There are substantial differences in vital sign monitoring during conventional safety monitoring processes and in remote ambulatory settings (Table 1), indicating the need for experiments to assess the impact of these differences on clinical study conduction and study data.

State of the Evidence

The authors discussed the advantages and shortcomings of remote cardiac monitoring using 2 publications describing evaluation of 2 single-lead ECG devices supplemented by a wrist-worn actigraphy device in phase 1 clinical trials to assess deployment and data collection feasibility and ascertain whether these devices were fit for purpose for vital sign data collection in clinical research [13, 14]. The device evaluation objectives were explor-

Table 3. Challenges of remote ambulatory vital sign data collection and potential solutions

Challenge	Solution
<i>Study design and operational</i>	
High variability in device technical characteristics (e.g., memory and battery life); form factors, body placement, sample level data, and derived variables	Maximizing the amount of information in the public domain for specific devices and device performance characteristics through publicly shared databases and peer review publications
Decreased patient compliance while unsupervised	Providing appropriate site and subject training, explaining to the study subject the importance of data collection, establishing study dashboards to enable communication with the study subject if the compliance drops beyond a certain threshold
<i>Data analysis</i>	
Lack of transparency for data processing algorithms	Selecting devices with sample level data availability and conducting device-specific evaluation or validation studies; results from independent device evaluation or validation studies can be very informative
Prolonged periods of monitoring requiring a review of large amounts of data	Novel statistical approaches that would account for diurnal patterns, state transitions (e.g., resting vs. active), establishing a baseline and detecting changes from baseline, and minimizing false positives (establishing the signal-to-noise ratio)
<i>Data interpretation</i>	
Reference ranges to support wearable data collection	Normative studies to be conducted to establish ambulatory reference ranges which also account for device body placement
Data artifacts requiring a manual review	Assessing device performance characteristics and establishing the face validity of the data prior to device deployment to support primary and secondary endpoints or study eligibility criteria
The need for metadata to enable results interpretation	Additional devices such as physical activity trackers or patient-reported outcomes/patient diaries may be required to gather contextual data to enable data interpretation. Devices or diaries should be defined in the context of use

atory; the data were placed in a “safe harbor” and they were not used for any clinical decision-making. All devices under evaluation had 510(k) clearance from the FDA [20–22] and were used continuously during both clinical pharmacology unit confinement and after the subject discharge (Table 2). Devices were evaluated by comparing the wearable device-derived data to vital sign measures collected as a part of conventional safety monitoring at matching time points, estimating the data loss and examining the face validity of the data, such as the proportion of values within a physiological range and diurnal change patterns. Overall, the evaluation experiments demonstrated the need for specifically designed experiments to establish the analytical accuracy of measures of interest. Specifically, the HealthPatch device was found to be problematic because of a large number of episodes with an elevated heart rate which were not corroborated after manual review of the selected ECG tracing [13]. In contrast, the BodyGuardian showed good accuracy in detecting the heart rate using the same evaluation approach [14]. Additionally, the use of the BodyGuardian device demonstrated the benefits of continuous heart rate data collection: an elevated heart rate

was detected during the periods of physical activity after treatment with amphetamine [14] and not detected during the periods of conventional vital sign data collection using a resting and supine protocol. Moreover, both studies indicated the need for monitoring physical activity as an adjunct measure to enable interpretation of certain vital sign measures such as heart and respiratory rate (Table 3). The general limitations included data losses due to subjects taking devices off for unspecified reasons and filtering out invalid data according to manufacturer’s recommendations (Table 3). The limitation specific to single-lead ECG devices is their inability to accurately detect QT interval prolongation [23], a side effect described for a number of experimental and approved medicines [24]. Additionally, ambulatory monitoring is prone to data artifacts introduced by physical activity and motion [25]. Skin temperature measured on the chest showed no correlation with body temperature measured in the oral cavity, indicating the need for establishing reference ranges for continuous data collection in a specific location which would guide data interpretation [13]. The use of these wearable devices was well received by both the study subjects and the site personnel, who

indicated the need for hands on experience prior to device deployment [13].

In the context of potential SARS-CoV2 infection, important vital signs of interest include body temperature, heart rate, respiratory rate, and SpO₂ [25]. An elevated temperature is a frequent presenting symptom of COVID-19, and continued fevers can indicate persistent infectivity. Individuals who have a clinically significant infection may develop an elevated heart rate from the infection and later may develop a baseline elevated respiratory rate secondary to pulmonary compromise. Continued respiratory involvement may lead to hypoxemia, detectable by pulse oximetry. Low blood oxygenation is linked to worse outcomes in COVID-19, whether through unmistakable clinical deterioration [26] or through more insidious “silent hypoxia” [27]. The majority of commercial pulse oximeters collect measures intermittently, but some of the models have been used successfully at home to collect data continuously at night in patients with COPD [28]. While evidence exists to support the use of outpatient monitoring for body temperature, heart rate, respiratory rate, and SpO₂ [25], the above considerations in COVID-19 illustrate that context is critical when interpreting outpatient vital signs data – whether vital signs are obtained in the setting of physical activity, in the setting of a COVID-19 infection, or otherwise. The issue of context is further elevated when clinical trials are specifically testing interventions for potential efficacy against SARS-CoV2. Understanding the anticipated effects of the virus on the normal human physiology as assessed by ambulatory remote monitoring is critical to assessing the therapeutic impact and/or safety signals from experimental treatments.

Risks and Benefits

All biomarker development is done in the context of the risks and benefits of the decisions made when using the biomarker. For example, a biomarker that is meant to detect lethal, irreversible, drug-induced organ damage will need to be very sensitive to a potential false-negative rate as the risk of getting the signal detection wrong will lead to death, though the benefit of getting it right is enormous. The risk will be directly related to the inherent risk in which the patient’s condition places him or her. A patient population is likely to accept more risk in making decisions with a biomarker if their condition is extremely debilitating and has no known therapy compared to a population with a less severe condition. Parameters that

describe the standard limits of the measurement need to be compared to the risk to the patient if misidentified and the risk that the drug developer is willing to accept for making development decisions. Like a traditional biomarker approach, the risk and benefit must be determined in the context for which the tool will be used and what decision will be made. In the case of measuring heart rate, respiratory rate, and body temperature, each as a separate variable constituting a safety signal in a drug development trial, the risk will be dependent on the study population and on the likelihood that the drug may cause a change in body physiology. If the study population is comprised of normal healthy volunteers and the drug is not suspected to generate any changes in heart rate, respiratory rate, or body temperature, the risk is low to the study subjects. However, if the population consists of severely ill patients and the drug has shown a potential for cardiac or respiratory functions or having a pyrogenic effect, the risk will be high. In this second case, the data accuracy will need to be extremely high and missing or spurious data will not be acceptable for decision making. The evidentiary framework provides a useful means to organize characteristics about a digital tool, establishing the purpose, validating tools for its intended use, and deploying digital measurements to get to interpretable results. More importantly, it provides a lattice to identify the risks, benefits, and key gaps that need to be filled before these tools can be used in clinical trials.

The potential benefits of continuous vital sign monitoring based on the use cases evaluated include early safety signal detection that can inform dose adjustments or lead to discontinuation of drug candidates with an unfavorable safety profile at an early stage. Additional benefits include a continuous pharmacodynamic assessment if relevant measurements, such as heart rate, are pertinent to the drug mechanism of action (Fig. 1). The limitation of this benefit is the detection of a potential safety signal restricted to arrhythmias. The risks include missing safety signals if a detection method has an unacceptably high false-negative rate. A high false-positive rate will lead to time-consuming data review and reporting. Additional risks, such as data losses and the need for metadata to interpret results, were identified (Fig. 1).

Changes that have come to the healthcare community due to SARS-CoV2 virus and COVID-19 have had a significant impact on many aspects of the risk-and-benefit equation. The impacts to ongoing trials compared to trials yet to launch can be very different. For ongoing trials, the risk of exposure to infection has changed the ability and willingness of trial participants to travel to clinical

sites, making in-person testing very difficult. In yet-to-start trials, the travel issues will affect enrollment and ultimately the likelihood that a trial will obtain enough subjects to generate decision-making data. Both situations offer opportunities for remote monitoring, reducing the risks of travel to clinical sites where exposure to infection is higher. In addition, trial sponsors may be willing to take more risk on established measurement methods [29] to allow some data collection and provide a trial result. If any trials are focused on COVID-19 therapies, the scales are even more in favor of adoption of the remote approach. Truly successful use of remote monitoring will await the results of what changes were made during this time, but the external environment, i.e., the pandemic, has moved the use of these approaches to the forefront of every clinical testing program.

Alignment with the Biomarker Evidentiary Criteria Framework

The case study showed that the determination of fit for purpose according to the FDA evidentiary framework for biomarker qualification could be applied for this remote monitoring approach. The level of evidence needed could be clearly derived by following a step-wise approach, defining the unmet need, the context of use, and the risk/benefit assessment of a biomarker's use (Fig. 1). However, the case studies revealed several challenges which are imposed by changes in the means of data collection. First, while the reference ranges for heart and respiratory rates are well established, these only apply to a resting and supine subject heart and respiratory rate data collection technique. The corresponding ambulatory reference ranges are highly context dependent; e.g., they vary depending on the intensity of physical exercise, age, and medical conditions a study subject might have, presenting a bigger challenge for data interpretation compared to the spot check data collection method. The data are available from Holter monitoring studies [30, 31], but this technique does not account for physical activity which has an impact on the range of measures. Moreover, skin temperature measured at the chest wall is different from the oral body temperature, being more impacted by environmental factors. Second, the requirements for device clearance vary with device type and do not necessarily include analytical validation. Additionally, 510(k) clearance summaries may contain only abbreviated information about device performance characteristics and are not sufficient to reconstitute experimental details and under-

stand the generalizability of the results. However, the requirement for ascertaining the quality of a measurement test exists for biomarker qualification according to the evidentiary framework. This case study confirmed the need for analytical validation for an established concept combined with digital measurement methods (Table 3). Third, there are certain limitations imposed by use of wearable devices. For instance, an important ECG-derived parameter, such as QT interval prolongation, can be problematic when measured by a single-lead ECG device [23]. This measure needs to be supplemented by conventional 12-lead ECG or multi-lead ECG wearable devices which became recently available [32]. In general, this case study showed that, for this application, the FDA evidentiary framework was completely compatible with development of the cardiac monitoring approach.

Statistical Considerations

Continuous vital sign monitoring collects a large volume of data that raises substantial challenges for extracting relevant signals but also provides unique opportunities for a deeper understanding of the patient status.

Given the current lack of unified standards for continuous ambulatory vital sign monitoring, the results of ongoing studies have to be interpreted through a prism of study-specific devices, raw data-processing pipelines (e.g., removing artefacts), and description data analytical methods at a high level. Before such standards are established and accepted by all key stakeholders, device manufacturers should publish evidence for validation of their methods. It is also highly desirable to have access to raw/sample level data to be able to perform retrospective analyses by study sponsors, i.e., to detect device malfunctioning, to identify values outside of the calibration range, or to deal with missing data.

Modern ambulatory cardiac monitors based on wearable device and accompanying apps can enable collection of important contextual information surrounding the time of cardiac events through: (1) patient-friendly reporting interfaces (collecting information about cardiac feeling such as heart palpitations and type of event) and (2) algorithm-based detection of sleep periods and identification of body posture using information from built-in accelerometers. Ambulatory cardiac monitoring should be used to establish accurate patient level normative values for heart rate rhythmicity or the cardiac burden including the frequency, duration, severity, and timing of adverse cardiac events and greatly facilitate detec-

tion of a divergence from the pretreatment baseline or flag a safety concern. Novel statistical approaches, such as time-frequency analysis and signal processing [33], functional data analysis (24-h diurnal patterns) [34], state transition time series analysis [35], and multi-modal multi-resolution analysis [36] efficiently leverage the temporal resolution of the collected data. These approaches provide a more sensitive quantification of cardiac function, particularly in the context of safety, efficacy, and estimating the main and side effects of a treatment.

Lessons Learned

We concluded that the approval/clearance of a biomarker test by the CDRH or by the CBER does not indicate qualification (or even fit-for-purpose use) of the biomarker for drug development. The examination of evidence in 2 publications [13, 14] demonstrated the need to conduct a validation or evaluation experiment to ascertain whether a biomarker and associated data collection methods are fit for purpose for use in a clinical trial. The data generated for the purposes of 510(k) clearance may be very different from the statement of need, context of use, and benefit/risk assessments which constitute the evidentiary criteria needed for qualification of the biomarker for use in a clinical trial. The information provided in a 510(k) summary, such as indications of use and intended use as well as device performance characteristics, can be informative for designing biomarker qualification experiments. However, the device performance (both hardware and software) and the benefit/ risk assessment need to be evaluated in the context of use (Table 3).

The evidentiary framework guidance is agnostic to biomarker type, class, and measurement method. The 2 specific use cases discussed here by the authors did not introduce novel biomarkers. Instead, they used the following well-qualified biomarkers for monitoring vital signs in clinical trials: heart rate, respiratory rate, and body temperature, but they changed the context of use, including the method of measurement. The changes in the context of use included new devices, single-lead ECG instead of 12-lead ECG, respiratory rate measured by a chest-worn device instead of manual counts, and skin temperature measured on the chest wall instead of in the oral cavity (change in body placement). The other parameter that changed is the duration of data collection; the devices were worn continuously for multiple days data instead of spot check data collection at predefined time

points. The examination of case studies revealed the number of challenges imposed by the context of use change (Table 3).

The results of the evaluation demonstrated the critical importance of the context of use and the need to design evaluation or validation experiments according to the context of use instead of extrapolating from the information used for device 510(k) clearance. Additionally, these studies highlighted the importance of assessing the accuracy of wearable devices in the context of use. Heart rate, respiratory rate, and skin temperature measures were evaluated by comparing the mobile measures to corresponding conventional safety measures obtained by an orthogonal method. The heart rate measures were additionally corroborated by a manual review of ECG tracing. Establishing the analytical validity of a digitally measured biomarker can be challenging because, in some cases, the data processing algorithms are proprietary and not available to the users. Nevertheless, this issue can be overcome by comparing a device readout to the readout done by an independent method and, in some cases, by evaluating sample level data [37].

The long duration of data collection also highlighted the need for new statistical methods that would leverage the richness of dense data. Moreover, the contextual information is needed to enable data interpretation (Table 3). In the context of conventional clinical trials, when the data are captured at predefined time points, this requirement does not exist.

This review does not consider aspects of technology evaluation outside of the FDA biomarker qualification evidentiary framework such as technology cost, security, data rights, and governance. These important aspects are considered elsewhere [38].

Future Directions

The cardiac monitoring use case confirmed the utility of the FDA biomarker qualification evidentiary framework for remote ambulatory vital sign monitoring in clinical trials. Specifically, the key elements of the framework were applicable to this use case. The alignment of the evidentiary framework with the remote cardiac monitoring approach highlighted the need for data collection method validation irrespective of the regulatory status of a device of interest. Currently, the information in the public domain, including the results of device clearance review by the FDA and data from independent evaluation/validation studies, is limited, indicating the need for more use

cases to be available for examination and knowledge generation in the scientific community. Additionally, the richness of the data afforded by remote monitoring devices will not be fully leveraged without an appropriate statistical methodology to analyze the data. Novel statistical approaches are needed to maximize the use of continuous data and provide novel insights. Moreover, remote monitoring for extended periods of time is feasible when supported by an appropriate device selection, data collection, and analysis methods. However, a wider adoption of this approach for drug development purposes will require collaborators across industry efforts to establish methodologies and supportive data sets to enable data interpretation and decision-making.

Conflict of Interest Statement

E.S.I. is an employee of Koneksa Health and may own company stock. B.W. is an advisor for Koneksa Health and Elektra Labs and received consulting fees from Teladoc and research funding from Pfizer. J.A.W. is an employee of Cygnal Therapeutics and may own company stock, and he is a member of the FNIH Biomarkers Con-

sortium Executive Committee. All of the other authors have no conflict of interests.

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or an implied endorsement of such products by the Department of Health and Human Services. This article reflects the views of the author and should not be construed to represent views or policies of the FDA.

Funding Sources

The FNIH Biomarkers Consortium, which hosted this workshop, is a major public-private biomedical research partnership supported by stakeholder membership including government, industry, academia and patient advocacy, and other not-for-profit organizations.

Author Contributions

All of the authors made substantial contributions to the conception and design of this work, data interpretation, and writing and revision of this paper and approved the content of this work.

References

- 1 FDA. List of Qualified Biomarkers. [cited 2020 September 14]; Available from: <https://www.fda.gov/drugs/cder-biomarker-qualification-program/list-qualified-biomarkers>.
- 2 FDA. FDA letter "Re: Review Submission of the Qualification of Seven Biomarkers of Drug-Induced Nephrotoxicity in Rats". [cited 2020 September 14]; Available from: <https://www.fda.gov/media/82532/download>.
- 3 FDA. FDA letter "Re: Biomarker Qualification Decision". [cited 2020 September 14]; Available from: <https://www.fda.gov/media/83152/download>.
- 4 FDA. Draft Guidance on Qualification of Biomarker — Galactomannan in studies of treatments of invasive aspergillosis. [cited 2020 September 14]; Available from: <https://www.fda.gov/media/89512/download>.
- 5 FDA. Qualification of Biomarker Plasma Fibrinogen in Studies Examining Exacerbations and/or All-Cause Mortality in Patients With Chronic Obstructive Pulmonary Disease. [cited 2020 September 14]; Available from: <https://www.fda.gov/media/92782/download>.
- 6 Qualification Decision Letter FD. (DDTB-MQ000044). [cited 2020 September 14]; Available from: <https://www.fda.gov/media/119374/download>.
- 7 FDA. Guidance for Industry and FDA Staff: Biomarker Qualification: Evidentiary Framework. [cited 2020 September 14]; Available from: <https://www.fda.gov/media/119271/download>.
- 8 Leptak C, Menetski JP, Wagner JA, Aubrecht J, Brady L, Brumfield M, et al. What evidence do we need for biomarker qualification? *Sci Transl Med*. 2017 Nov;9(417):eaal4599.
- 9 BEST. (Biomarkers, EndpointS, and other Tools) Resource. [cited 2020 November 30]; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK338448/>.
- 10 Izmailova ES, Wagner JA, Ammour N, Amondikar N, Bell-Vlasov A, Berman S, et al. Remote Digital Monitoring for Medical Product Development. *Clin Transl Sci*. 2020, Online ahead of print.
- 11 Clinical safety [cited 2020 November 30]; Available from: https://www.nidcr.nih.gov/sites/default/files/2018-03/safety-definitions-for-clinical-research_0.docx.
- 12 510 (k) clearance by the FDA. [cited 2020 December 2]; Available from: <https://www.fda.gov/medical-devices/device-approvals-denials-and-clearances/510k-clearances>.
- 13 Izmailova ES, McLean IL, Bhatia G, Hather G, Cantor M, Merberg D, et al. Evaluation of Wearable Digital Devices in a Phase I Clinical Trial. *Clin Transl Sci*. 2019 May;12(3):247–56.
- 14 Izmailova ES, McLean IL, Hather G, Merberg D, Homsy J, Cantor M, et al. Continuous Monitoring Using a Wearable Device Detects Activity-Induced Heart Rate Changes After Administration of Amphetamine. *Clin Transl Sci*. 2019 Nov;12(6):677–86.
- 15 Goldsack JC, Izmailova ES, Menetski JP, Hoffmann SC, Groenen PM, Wagner JA. Remote digital monitoring in clinical trials in the time of COVID-19. *Nat Rev Drug Discov*. 2020 Jun;19(6):378–9.
- 16 Lavery H, Benson C, Cartwright E, Cross M, Garland C, Hammond T, et al. How can we improve our understanding of cardiovascular safety liabilities to develop safer medicines? *Br J Pharmacol*. 2011 Jun;163(4):675–93.
- 17 FDA. Guidance for Industry. COVID-19: Developing Drugs and Biological Products for Treatment or Prevention [cited 2020 September 14]; Available from: <https://www.fda.gov/media/137926/download>.
- 18 EMA. Key changes from v2 (27-03-2020): distributor to trial participant IMP shipment, monitoring, remote source data verification and communication with authorities. [cited 2020 September 14]; Available from: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf.
- 19 Schwartz D, Fischhoff B, Krishnamurti T, Sowell F. The Hawthorne effect and energy awareness. *Proc Natl Acad Sci USA*. 2013 Sep;110(38):15242–6.
- 20 510(k) premarket notification for Health-Patch device. [cited 2020 September 14]; Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K152139>.

- 21 510(k) premarket notification for Body-Guardian device. [cited 2020 September 14]; Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K151188>.
- 22 510(k) summary for Actiwatch device. [cited 2020 September 14]; Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf/K983533.pdf.
- 23 Bekker CL, Noordergraaf F, Teerenstra S, Pop G, van den Bemt BJ. Diagnostic accuracy of a single-lead portable ECG device for measuring QTc prolongation. *Ann Noninvasive Electrocardiol*. 2020 Jan;25(1):e12683.
- 24 Nachimuthu S, Assar MD, Schussler JM. Drug-induced QT interval prolongation: mechanisms and clinical management. *Ther Adv Drug Saf*. 2012 Oct;3(5):241–53.
- 25 Manta, C., et al., An Evaluation of Biometric Monitoring Technologies for Vital Signs in the Era of Covid-19. *Clin Transl Sci*. 2020 Nov;13(6):1034–44.
- 26 Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. *Radiology*. 2020 Aug;296(2):E86–96.
- 27 Wilkerson RG, Adler JD, Shah NG, Brown R. Silent hypoxia: A harbinger of clinical deterioration in patients with COVID-19. *Am J Emerg Med*. 2020 Oct;38(10):2243.e5–2243.e6.
- 28 Buekers J, Theunis J, De Boever P, Vaes AW, Koopman M, Janssen EV, et al. Wearable Finger Pulse Oximetry for Continuous Oxygen Saturation Measurements During Daily Home Routines of Patients With Chronic Obstructive Pulmonary Disease (COPD) Over One Week: observational Study. *JMIR Mhealth Uhealth*. 2019 Jun;7(6):e12866.
- 29 Izmailova ES, Ellis R, Benko C; Remote Monitoring in Clinical Trials During the COVID-19 Pandemic. Remote Monitoring in Clinical Trials During the COVID-19 Pandemic. *Clin Transl Sci*. 2020 Sep;13(5):838–41.
- 30 Romano M, Clarizia M, Onofrio E, Caiazzo MR, Adinolfi L, Cutillo S, et al. Heart rate, PR, and QT intervals in normal children: a 24-hour Holter monitoring study. *Clin Cardiol*. 1988 Dec;11(12):839–42.
- 31 Paudel B, Paudel K. The diagnostic significance of the holter monitoring in the evaluation of palpitation. *J Clin Diagn Res*. 2013 Mar;7(3):480–3.
- 32 Multi lead ECG wearable device example. [cited 2020 September 14]; Available from: <https://www.preventivesolutions.com/hcp/body-guardian-mini>.
- 33 Kay S. Fundamentals of statistical signal processing. Prentice Hall PTR; 1993.
- 34 Ramsay JO, Silverman BW. Applied Functional Data Analysis: Methods and Case Studies. Springer; 2007.
- 35 West BJ. Where medicine went wrong: Rediscovering the path to complexity Vol. 11. 2006: World Scientific.
- 36 Addison PS. The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance. 2017: CRC press.
- 37 Godfrey A, Vandendriessche B, Bakker JP, Fitzer-Attas C, Gujar N, Hobbs M, et al. Fit-for-Purpose Biometric Monitoring Technologies: Leveraging the Laboratory Biomarker Experience. *Clin Transl Sci*. 2020, Online ahead of print.
- 38 Coravos A, Doerr M, Goldsack J, Manta C, Shervey M, Woods B, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. *NPJ Digit Med*. 2020 Mar;3:37.