Mobile Devices in Clinical Trials for Neurological Diseases: CDISC Standards Development Workshop

March 10, 2017 (8:00 a.m. – 5:00 p.m.)

Pointe Hilton Tapatio Cliffs | Phoenix, AZ

ATTENDEES

Name	Affiliation
Jane Rhodes	Biogen
Jenny Barnett	Cambridge Cognition
Michael Ibara	CDISC
Barrie Nelson	CDISC
Jon Neville	Critical Path Institute
Stephen Arnerić	Critical Path Institute
Jennifer Ashley Downs	Critical Path Institute
Dan Hartley (*remote*)	Critical Path Institute
Volker Kern	Critical Path Institute
Diane Stephenson	Critical Path Institute
Lynn Hudson	Critical Path Institute
Derek Hill	IXICO
Adria Martig	Michael J. Fox Foundation for Parkinson's Research
Jeffrey Kaye	ORCATECH - OHSU
Daniel Karlin	Pfizer, Inc.
Farhan Hameed	Pfizer, Inc.
Jane Mohler	University of Arizona
Lee Ryan	University of Arizona
Rebecca Spencer	University of Massachusetts, Amherst
Ray Dorsey (*remote*)	University of Rochester

AGENDA

Continental Breakfast (7:30 – 8:00 am)	
8:00 – 8:15 a.m.	**Welcoming Remarks & Overarching Objectives for Concepts-of-Interest (COIs)** *Stephen P. Arnerić (Critical Path Institute)*
8:15 – 8:30 a.m.	**CDISC Standards & Digital Health** *Michael Ibara (CDISC)*
8:30 – 9:00 a.m.	**Data Flow: From COIs to Data Archiving** *Barrie Nelson (CDISC)*
9:00 – 9:15 a.m.	**Analogous Use Cases for Representing Device Data Using CDISC Standards** *Jon Neville (Critical Path Institute)*
Break (20 min)	
SESSION I: Sleep	
9:35 – 10:05 a.m.	**BMDs for Sleep** *Rebecca Spencer (University of Massachusetts) - 20 min.* *Derek Hill (IXICO) - 10 min.*
10:05 – 11:15 a.m.	**Capturing Key COIs & Data Flow** *Barrie Nelson & Rebecca Spencer*
Lunch (45 min)	
SESSION II: Mobility	
Noon – 12:30 p.m.	**BMDs for Mobility/Frailty** *Ray Dorsey (University of Rochester) - 20 min.* *Jane Mohler (University of Arizona) - 10 min.*
12:30 – 1:30 p.m.	**Capturing Key COIs & Data Flow** *Barrie Nelson & Lynn Hudson*
SESSION III: Cognition	
1:30 – 2:10 p.m.	**BMDs for Cognition** *Lee Ryan (University of Arizona) - 30 min.* *Jenny Barnett (Cambridge Cognition) - 10 min.*
2:10 – 3:30 p.m.	**Key COIs & Data Flow** *Barrie Nelson & Lee Ryan*
Break (20 min)	
3:50 – 4:50 p.m.	**Summary and Next Steps** *Barrie Nelson, Jon Neville, Stephen Arnerić*

ACTION ITEMS

ACTION ITEMS	LEAD/RESPONSIBLE	DEADLINE
1. Share ontology (developed by Dan Karlin, Pfizer) with CAMD.	**Dan Karlin**	**4/30/17**
2. Identify relevant questions to pose to the panel session with regulators during the follow-on *Biometric Monitoring Device Workshop* on May 9 - 10, 2017.	**All**	**3/31/17**
3. Invite EMA as well as PMDA representatives to the May workshop.	**CAMD**	**3/31/17**
4. Consider release of this workshop's presentations for posting on CAMD's website.	**All presenters**	**3/31/17**

HIGHLIGHTS AND SUMMARY

Advancing CDISC Standards for BMD Use in Clinical Development of Neurological Treatments, Dr. Stephen P. Arnerić, Critical Path Institute

Dr. Stephen Arnerić, PhD, Executive Director, Coalition Against Major Diseases (CAMD), Critical Path Institute, welcomed all attendees to the meeting on behalf of Critical Path Institute's Coalition Against Major Diseases. He requested all participants to introduced themselves and provide some background. Dr. Arnerić briefly reviewed the proposed vision for digital drug development tools with initial (years 1 through 3) focus on development, socialization, and implementation of a regulatory roadmap that will enable the advancement of regulatory science supporting the use of biometric monitoring devices (BMDs) in clinical trials. In parallel and slightly staggered approach (years 2 through 5), create a data repository of de-identified patient-level BMD data where quantitative disease-progress modeling and exploration of novel clinical outcome assessments can be examined at the earliest stages of disease to support specific contexts-of-use. Dr. Arnerić emphasized that various symptoms and signs that are impacting the patient's/care-giver's quality-of-life are prevalent across multiple disease areas; initial focus will be on sleep, mobility/frailty, and cognitive performance. Dr. Arnerić identified the envisioned deliverables for this meeting: To identify current gaps in data standards required to advance clinical Drug Development Tools that assess Physical Function/Frailty, Sleep and Cognition using Biometric Monitoring Devices (BMDs); and to fill these gaps to enable the use of BMDs in Registration Studies, and the creation of actionable databases of disease progression, and treatment responses across neurological diseases. Dr. Arnerić announced that a follow-on meeting is scheduled for May 9-10, 2017 in Bethesda, MD: "*Advancing CDISC standards for biosensors assessments in neurological clinical drug trials*". This follow-on meeting will attempt to understand the BMD landscape for concepts-of-interest, highlight regulatory considerations, and socialize a plan forward. Participants were requested to focus on the identification of the most important aspects in the disease areas of interest that will be applied to CDISC standards development.

CDISC Standards and Digital Health, Dr. Michael A. Ibara, CDISC

Dr. Michael Ibara, Head of Digital Healthcare, CDISC, provided a brief overview of CDISC, including the organization's mission to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. CDISC is a global standards development organization developing global consensus-based standards focusing on Clinical Research. Submission of all new drug applications in CDISC Standards is now required by U.S. FDA (since December 2016) and Japan's PMDA. CDISC's goals for standards include enabling of innovation, support of all types of research from protocol through analysis and reporting, streamlining of research processes and enabling of data sharing/ aggregation, linkage of healthcare delivery and clinical research through EHRs/eSource. Dr. Ibara explained CDISC's work with eSource in finding value in digital healthcare data. Mobile/wearable space innovation is occurring at several levels including hardware, software, clinical application, consumer/patient use, and regulatory acceptance. Currently, there is little consistency across the field of mobile/wearable devices and use of terminology and "metaphors" is confusing/can be misleading. Dr. Ibara suggested to "Begin with the end in mind" with specific attention to clinical usefulness, biomarker validity and regulatory endorsement. Assumptions need to be tested when innovation and immediate application are discussed. He suggested to seek an equilibrium amongst the best clinical concept vs. clear operational definition vs. representation in standards for regulatory use.

CAMD Mobile Devices, Dr. Barrie Nelson, CDISC

Dr. Barrie Nelson, Vice President, Standards, Terminology and Technical Services, CDISC, provided a brief overview of CDISC's activities and highlighted that CDISC supports community volunteers to develop open, freely available standards. CDISC provides global support and standards that have been downloaded in 90+ countries. CDISC standards drive prospective research efficiencies by allowing investigators to implement standards early in the process. In addition, CDISC standards may allow for identification of hidden discoveries in existing datasets (example: Autosomal dominant polycystic kidney disease, where the disease endpoint could be shifted to an earlier stage/earlier disease detection, *Critical Path Institute's Polycystic Kidney Disease Outcomes Consortium*). With FDA and PMDA requiring CDISC standards and the Chinese FDA and EMA recommending use of CDISC standards for data aggregation, integration, and analysis, implementing standards significantly shortens the time needed to develop a regulatory submission and leads to overall cycle time reduction. It is FDA's goal to reduce time to access and analyze data – leading to increased time available for review.

An entire spectrum of end-to-end standards and process automation exists that covers the areas of preclinical, protocol design, analysis plan, CRF design, data collection, analysis, and submission (**Figure 1**). Diverse data exchange scenarios exist. Dr. Nelson described the clinical research process in detail: CDISC standards include a multitude of models, including a Protocol Representation Model (PRM), Clinical Data Acquisition Standards Harmonization Lab Model (CDASH Lab), Study Data Tabulation Model (SDTM) for clinical studies - including medical devices, Analysis Data Model (ADaM), Operational Data Model (ODM), Standard for the Exchange of Non-Clinical Data (SEND), Controlled Terminology (CT), etc. For each part of the process, the CDISC standards provide standard answers to key questions: 1) How should the data be organized/grouped into datasets? 2) What individual concepts belong in each dataset? 3) What terminology should be used to represent questions and responses?

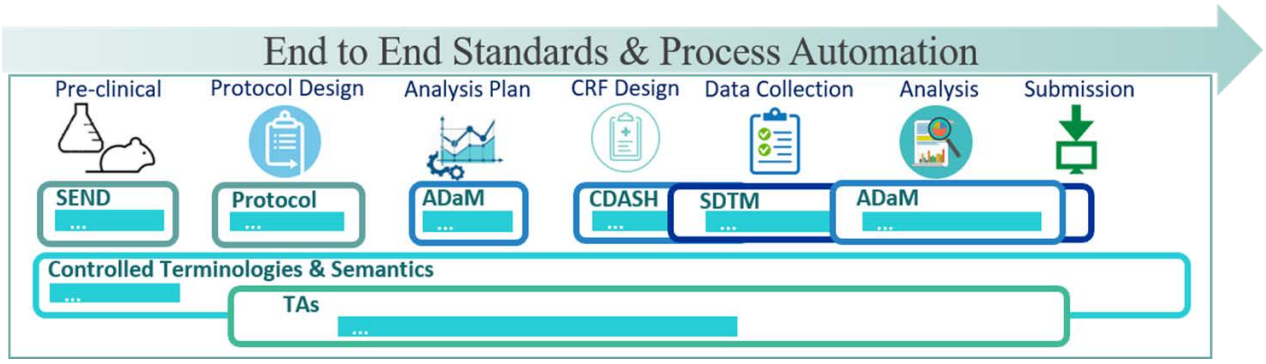

Figure 1: CDISC end-to-end standards and process automation

Dr. Nelson briefly described the benefits of the SHARE metadata repository, including deployment of the same metadata through multiple electronic systems, definition of common data via shared metadata in multiple systems and traceability, metadata synchronization automatically across multiple systems, version control, multiple language/dialect support, consistency and quality of data across systems, facilitated interoperability and data aggregation/reporting, up-front investment with downstream savings. Dr. Nelson also introduced concept mapping as a way to organize thoughts and share ideas between different disciplines, to enable clinicians and subject matter experts to communicate with technical standards developers, and to facilitate the translation, or mapping, of clinical concepts into CDISC standards. In closing, Dr. Nelson reviewed biomedical concepts and ontologies that can drive big data analyses. The transition from non-standard data via SDTM-conformant data to SDTM-conformant data with linked metadata is CDISC's ultimate goal.

Representing Device Data Using CDISC Standards, Jon Neville, Critical Path Institute

Jon Neville, Program Director Data Standards and Management, Critical Path Institute, introduced the CDISC Study Data Tabulation Model Implementation Guide for Medical Devices (SDTMIG-MD). Many existing Therapeutic Area Data Standards are available, including standards for Alzheimer disease, Parkinson disease, Multiple Sclerosis, Traumatic Brain Injury, etc., with additional standards under development (Huntington disease, etc.), developed in collaboration with the Critical Path Institute. Mr. Neville reviewed an associated concept map that was developed for imaging. A typical concept map summarizes the methods applied, devices used, other factors included, assessments, and linkages. Concept maps for biomarkers were highlighted as examples.

For Biometric Monitoring Devices, a real-world concept needs to be developed to answer the questions: 1) Where are we attempting to end up? 2) What's the most clinically-relevant measure? 3) How to capture device properties? For regulatory submissions, not all operational data need to be provided, but data has to be available upon FDA request. Fitting Biometric Monitoring Devices into a concept map framework may be challenging and ground rules need to be established to define: 1) What is the beginning-to-end life cycle of these data? 2) What happens to the raw data beyond interpretation by software? 3) Which parameters are important to capture? It will be important to log the processes during data collection and to fully understand them.

SESSION I: SLEEP

Measuring Sleep and Sleepiness with Mobile Devices, Dr. Rebecca Spencer, University of Massachusetts

Dr. Rebecca Spencer, Associate Professor, Department of Psychological and Brain Sciences, Center for Personal Health Monitoring, University of Massachusetts, Amherst, started her presentation by asking the question: "Why measure sleep?" Good sleep is important to sound memory function, decision making, stress management, cleaning of "brain waste", emotion regulation, and immune function; sleep quality and quantity is influenced by aging, social/economic factors, parenting, psychological disorders, adolescents, and neurodegenerative disorders. Sleep can be influenced/treated behaviorally with sleep hygiene being an important factor. Using devices may improve sleep without pharmaceutical interventions. Sleep is not homogenous. Dr. Spencer highlighted different stages of sleep including REM sleep (important for emotion regulation), non-REM sleep (nREM, transitional sleep stage), and Slow Wave Sleep (SWS, memory consolidation). All these stages serve a unique function; SWS may be particularly essential in neurodegenerative diseases given the role in glymphatic clearing. Key features of sleep that influence quality-of-life include sleep quantity (more sleep supporting all of these functions), sleep quality (less time wasted awake in bed improves function), and sleep sufficiency (i.e., does it meet the sleep need without resulting in sleep during the day).

Polysomnography includes electroencephalography (EEG), electrooculography (EOG), and electro-myography (EMG). Devices such as the *Actiwatch* are utilized for comparison to polysomnography. Dr. Spencer presented some example sleep recordings and introduced the audience to how to interpret results. Open questions include if the devices capture sleep accurately. Sleep start and stop logs/diaries (taken by the patients) help in interpretation. Polysomnography relies on EEG to identify sleep, while EMG/EGO are used to identify sleep stages.

Key features of sleep that impact quality-of-life include sleep quantity with total sleep time equaling sleep period minus wake time after sleep onset; and sleep quality with sleep efficiency equaling total sleep time divided by [sleep period x 100]; often referred to as the "sleep score". Actigraphy relies on BMDs such as the *ActiGraph* device or the *ActiWatch*. These devices utilize triaxial accelerometers and can provide an estimate of the sleep/wake cycle via movement (or absence of); based on many assumptions. These BMDs summarize the frequency of motions into epochs of specified time duration and they store the summary in their internal memory or through cloud-based storage. Dr. Spencer presented an example actigraph and highlighted interpretation features. Research-based actigraphy has many advantages (such as objectivity [compared to questionnaires, observation], they can be worn over multiple days/weeks, correlation between actigraphy and PSG-defined sleep estimates). Disadvantages include the fact that the accuracy is lower for some groups, they cannot score sleep architecture, scoring of data is tedious, and validated data requires simultaneous diary. When different commercially-available BMDs are compared, it becomes obvious that the measured outcomes (total sleep time, sleep efficiency/sleep score, deep vs. light sleep) differ significantly. Many challenges need to be overcome: Actigraphy-based sleep measures are *generally* reliable for total sleep time and sleep efficiency; however, most studies are limited to healthy young adults (additional work needs to be completed to characterize the elderly and obese population), focused on night-time (supine) sleep, inactivity vs. napping vs. sleepiness mid-day is indistinguishable. BMDs do not capture sleep stages accurately. Given that Slow Wave Sleep may be key and nREM may be of limited use, total sleep time measurement may not be enough. Furthermore,

opportunities to take the device off need to be limited, battery life and waterproofness need to be improved.

Operationalizing sleep from actigraphy in normal elderly and neurodegenerative disease, Dr. Derek Hill, IXICO

Dr. Derek Hill, Executive Director, IXICO, opened his presentation by providing an overview of the challenges to deploying BMDs for sleep in clinical trials of older people and those with neurodegenerative diseases. He highlighted the interconnectivity amongst core disease symptoms, and sleep and activity, as well as other factors such as mood disturbances and circadian rhythm disturbances. Dr. Hill described efforts related to actigraphy and polysomnography measurements and he described the related processes. When comparing actigraphy and polysomnography (Cole-Kripke algorithm (Cole *et al.*; LINK) and ESS algorithm [Borazio *et al.*; LINK]) poor correlation is apparent. In contrast, with deep-learning methodologies correlations are uncovered. Raw data are needed, especially from the older population.

Addressing IXICO's GCP (good clinical practice) solution closed system deployment and data management, Dr. Hill described the platform IXICO is employing, including collection of actigraphy data (algorithm), regulatory-compliant data handling, etc. IXICO is focused on operational delivery at the clinical site (distribution, training, support), the CRO (training and monitoring, data upload, device distribution), at IXICO (development of training material, data processing), and when working with sponsors (protocol development, compiling of study results).

Future work includes improving analytics validity within and between sensor device performance (test:re-test), the quest to collect more data to improve clinical contextualization, and to extend focus beyond sleep towards mobility, mood, etc. Collaborations across clinical trials need to be leveraged to improve clinical meaningfulness and ensure between-device standardization (sensors need to be interchangeable with the same analysis algorithm eliminating the need to repeat clinical studies). Moving from the exploratory stage to secondary endpoints will require more useful data and expanded collaboration. Data on reliability and reproducibility of BMDs will be important to communicate to FDA, especially CDRH.

Discussion (Session 1 – Sleep)

Dr. Arnerić opened the discussion by highlighting a draft high-level concept map with focus on sleep-related concepts to consider (**Figure 2**). He requested the participants to identify, if possible, the three top-tier/must-have/key specific outcomes based on their potential ability to:

- Impact patient/care-giver quality-of-life (QoL), and improve health;
- Address unmet needs;
- Achieve label claims;
- Use as pre-manifest disease outcome assessment.

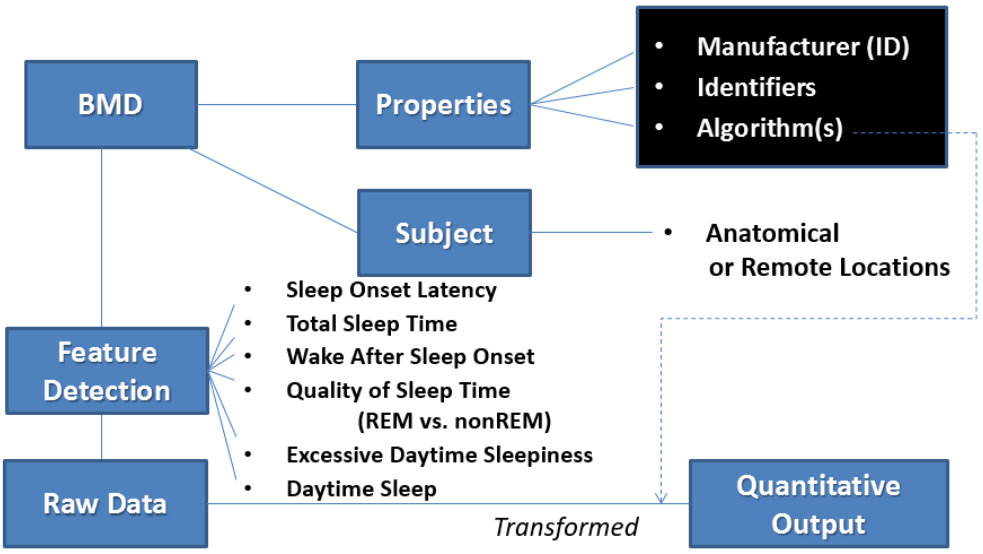

Figure 2: Draft concept map – Sleep

Pin-pointing prioritized specific outcomes/key Concepts of Interest (COIs), such as total sleep time or sleep efficiency, is dependent on the functional implications and this task was generally viewed as a significant challenge.

Participants fully agreed that "bridging" between devices and standardization is a necessity. Commercially available devices are considered "black boxes" and are of limited use; research-grade devices/a suite of technologies are needed; however, changing software versions and algorithms may make older datasets obsolete. Better technologies are needed that allow for rapid deployment and allow for easy data sharing. The ADNI dataset was discussed drawing parallels with what is already known regarding imaging data. While ADNI MRI data are comparable, the ADNI dataset represents a limited very selective cohort and generalizability to clinical trial subjects is problematic.

Participants discussed endpoint selection vs. data measurement capability; it was mutually agreed that the approach needs to be data-driven. Large datasets, close collaboration, and sharing of data are needed. Device and software standardization is needed; devices need to be able to measure the right parameters for a particular population. For example, actigraphy is approved (by the American Academy of Sleep Medicine) for measuring insomnia; actigraphy could be utilized as a baseline, or as a standard. Dr. Derek Hill suggested the development of a "standardized" robotic arm that can mimic movement; the data gathered could then be compared to other devices and be utilized as a standard. At this point of time, data are not always comparable across several devices; mathematical understanding of "what happens" to the data within the devices is required.

As a clinical trialist community, we need to understand what is of value to the patient, and to define a viable way to assess the concept. This needs to be compared to what is doable. Data sharing needs to be incentivized. A detailed concept map may be difficult to create; the action should be taken to draw a "user map" (what the end user needs) and compare it with a "device map" (what available devices can do). Focus needs to be on key topics amongst different populations. In this context of understanding what the

"user" needs, Dr. Arnerić briefly reviewed a survey "*The Voice of Those who Care*" that CAMD is preparing to launch shortly. Patients and care-givers will be asked eleven questions regarding quality-of-life domains in an attempt to identify and understand what is most valued by the patient and their care-givers regarding innovative treatments. This is an area of growing interest to regulatory agencies.

SESSION II: Mobility/Frailty

Future of Biometric Monitoring Devices in Alzheimer Disease, Dr. Ray Dorsey (remote), Professor of Neurology, Director of the Center for Human Experimental Therapeutics, University of Rochester

Dr. Ray Dorsey, Center for Human Experimental Therapeutics (CHET), University of Rochester opened his presentation by stating that a typical patient is only exposed to a medical professional for about 1% of their total lifetime and much of the available information is not being captured that could aide in providing better care and prospective therapies. Our present perspective of a patient's illness experience is episodic and limited to clinic visits (gathered by various rating scales). In addition, many of the current outcome measures are subjective and sub-optimal; as a direct result, the productivity of the drug development industry continues to decline (ref. new molecular entities per $1 billion in R&D (inflation adjusted), 1950-2010). The burden of neurologic diseases is increased in an aging population while current clinical trials for neurodegenerative disorders are "chasing poor signals". In this context, Dr. Dorsey briefly highlighted recent phase III trials that failed to replicate phase II findings. The way we measure disease needs to be changed. As new opportunities and new devices are becoming available, there is an identified need to broadly apply new technologies including digital biomarkers.

Smartphone applications for medical research are available, such as the *mPower* app for Parkinson disease. Applications offer survey components, structured cognitive tests/active tasks, and passive monitoring. These apps can detect responses from medications as demonstrated in changes of tapping frequency in individuals with Parkinson disease before and after medication. Pharmaceutical companies are incorporating such devices into their early stage development efforts (example: Roche). Dr. Dorsey briefly described the MC10 *BioStampRC* pilot study that is aimed to demonstrate the feasibility of data collection, to compare sensor data to standard clinical assessments, to develop algorithms to characterize abnormal movements, to assess response to medication, and to detect and quantify previously unmeasured symptoms (80 participants enrolled [20 with PD, 20 with HD, 20 with prodromal Huntington disease, 20 controls], 5 sensors on each participant). In this study, wearable sensors are able to capture data objectively and continuously in the clinic environment and at home with the focus on defining the proportion of the day individuals spend lying down, sitting, standing, and walking, opening new windows into measuring disease. Results of preliminary analyses showed healthy controls had distributions of 33% of time laying down, 1/3-time sitting and 1/3-time walking/standing. People diagnosed with PD spent 38% of time laying down. Unexpectedly, people with HD were found to show >50% of daytime laying down. Time walking was also much less for HD than for other participants. This illustrates the types of information that can be gathered with continuous monitoring that may translate to quality of life. Novel objective measures are generally accepted and are increasingly included into clinical trials, such as the use of accelerometers in a congestive heart failure trial (Redfield *et al.*, 2015 [LINK]).

Potential for Alzheimer disease: Current assessment measures for Alzheimer disease (such as ADAS-Cog) are considered subjective and categorically leading to long, large, and expensive trails. Cognition assessment in day-to-day life (such as the Cognition Kit used in collaboration with Apple Watch) allows for frequent, objective measurements. Smartphone technology to assess lifestyle and cognitive health is being leveraged, such as the Mind Share application developed by Digital Artefacts L.L.C. on Apple's *ResearchKit*. For example, recent results indicate strong correlation between gait and cognitive function with gait change as cognitive decline progresses in elderly population. In closing, Dr. Dorsey reiterated that health care needs a new class of measurements that can provide high frequency data, allowing for measurements per person per minute - not per day, week, or year. Dr. Dorsey invited CAMD and the participants of this workshop to consider publishing the proceedings in the new multidisciplinary journal "Digital Biomarkers".

Frailty Risk Adjustment – A Role in the Development of Neurological Tx?, Dr. Jane Mohler, Professor, University of Arizona Colleges of Medicine, Division of Geriatrics, General Internal Medicine and Palliative Medicine, Public Health, Nursing and Bioengineering GIDP Director, Healthy Brain Research Center, Associate Director (past), Arizona Center on Aging, University of Arizona

Dr. Jane Mohler, University of Arizona, opened her presentation by pointing out that that U.S. population ages 80 and older will nearly triple between 2010 and 2050; the number of people ages 90 and older will quadruple. The population age >65 account for the majority of health care spending. Aging is heterogeneous and age is not a precise indicator of health status, future care needs or health care costs. Risk stratification is needed with an approach to help us understand who is at increased risk for poor outcomes, how to risk stratify treatment and management strategies, how to more precisely estimate outcomes (adjusting for functional capacity/frailty), and to support optimization of therapies in the elderly with the availability of validated tools capable of predicting outcomes. Frailty equals "biological aging"; it is a hyper-inflammatory geriatric syndrome resulting from age-related cumulative declines across multiple physiologic systems, with impaired homeostatic reserve and a reduced capacity of the organism to resist stress. Frailty prevalence in the community is about equal to Alzheimer disease – approximately 11% of population >65 years (and higher percentage in older populations). Outcomes are poor quality-of-life, frequent hospitalization, treatment complications and adverse events, disability, institutionalization, and finally death. In populations diagnosed as frail, the 3-year survival estimates show a death rate of 81% (Fried *et al.*, 2001; LINK). Frailty is a more sensitive predictor of outcomes than is age: Assessment is important and inclusion of frail patients in clinical trials should be considered. A correlation of frailty and cognitive impairment has been demonstrated (Robertson *et al.*, 2013; LINK). Presently, there are few practical, objective instruments that can categorize age-related functional status. Frailty assessment methods include single markers (such as grip strength, walking speed), Phenotypic Frailty Indices (CHS [Fried] index, SOF index, FRAIL index), Multi-dimensional Indices (such as Rockwood, FI-CGA-10, MPI, SHERPA, HARP), and Functional Decline Instruments (such as ADL and CCI). For example, the CHS – Fried Frailty Index defines frailty as a clinical syndrome in which three or more of the following criteria are present: unintentional weight loss (10 lbs. in past year), self-reported exhaustion, low physical activity, weakness (grip strength), slow walking speed (Fried *et al.*, 2001; LINK).

Dr. Mohler presented examples of sensor-based frailty measures that are currently available. These are mostly based on inertial sensors (gyroscopes and accelerometers). For example, a novel test using upper-extremity motion (Toosizadeh *et al.*, 2015; LINK), UEF (upper-extremity function), allows measuring various kinematic parameters associated with slowness as measured by speed of movement, weakness as measured by power and moment on arm, exhaustion as measured by reduction in speed of movement and speed variability, flexibility as measured by elbow range of motion, and variability. The associated algorithm can be found online. Measuring aging function including the "frailty" status allows for adjustment for aging heterogeneity "biological age", supports predicting who will benefit, and identifies who will have poor outcomes.

Discussion (Session 2 – Mobility/Frailty)

Dr. Arnerić opened the discussion by highlighting a draft high-level concept map with focus on mobility/frailty-related concepts to consider (**Figure 3**). He requested the participants to identify, if possible, the three top-tier/must-have/key specific outcomes that consider the ability to impact patient/care-giver quality-of-life (QoL), improve health, the ability to address unmet needs, the ability to achieve label claim, and the ability to use as pre-manifest disease outcome assessment.

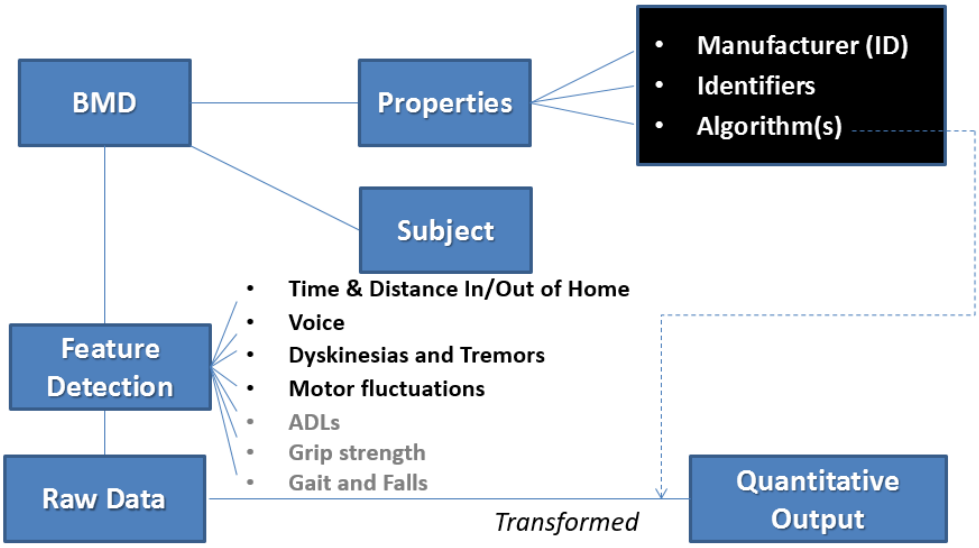

Figure 3: Draft concept map – Mobility/Frailty

Participants briefly reviewed types of motor function (including gait speed, grip strength, etc.), functional impact, frailty scale, pharmacoeconomic impact, etc. Changes in clinical meaningfulness (when applying mobile/sensor-based technology) need to be interpreted alongside traditional measures, allowing for cross-validation. Participants were in agreement that the standardization task needs to be defined. It was suggested to start with controlled standardized data collection with focus on a minimum dataset and blank CRFs. Mobility, physical activity, frailty, gait speed, other measures of activity are considered "low-hanging fruit" to map the ontology of data. Dr. Dan Karlin (Pfizer) offered to share with CAMD an ontology

that he developed at Pfizer as part of the BlueSky initiative collaboration with IBM Watson (**Action #1**). A conceptual framework needs to be proposed; CDISC is a partner that offers expert public review allowing for consensus and harmonization. Participants suggested to create a well-characterized accelerometer data-based study that follows a defined process. It will be important to focus on digital biomarkers that clearly demonstrate clinical meaningfulness with fundamental components that are "future-proof". Application of available accelerometer technology shows promise across multiple domains (sleep and movement/relationship between speed of gate and cognition). Devices, sensors, algorithms, etc. need to be comparable allowing for cross-dataset analyses; dataset collected side-by-side will allow for comparison of metadata and validation.

In summary: There is good familiarity with available devices; it was suggested to focus investment on generating longitudinal data (vs. the current standards). Based on the presentations of this session, frailty should be considered and frailty measures be included. Complex, decontextualized datasets, even if they are exploratory, would be of high value; however, defined protocols need to be established. "Teaching" computers to integrate related context (i.e., movement while asleep; movement in home; movement outside of the home) needs to be implemented in order to provide a valid interpretation.

SESSION III: Cognition

BMDs for Cognition, Dr. Lee Ryan, Professor and Department Head, Associate Director, Evelyn F. McKnight Brain Institute, Director, Cognition and Neuroimaging Laboratory, University of Arizona

Dr. Lee Ryan, University of Arizona, offered general comments on cognitive testing, including large-scale drug trials. A battery of standardized tests to investigate memory and executive function with high reliability and repeatability does exist for mild to moderate AD (such as the NIH Toolbox). Most tests have a long history and are non-specific and impacted by various environmental factors and practice effects (from performing the test multiple times); they are simple models focused on learning, storage, retention, and retrieval. These tests can detect impairment – but they cannot pinpoint small changes. Composite scores can be established by combining multiple tests – overcoming fluctuation and increasing power.

Targeted cognitive tests are considered to be unsuccessful as early predictors with limited clinical meaningfulness. New more sensitive tests need to be devised. However, at this point in time it is a challenge to answer what measures should be targeted. An important next step in cognition needs to be the introduction of targeted tests that are not widely known and where normative data are not available. In addition, more control samples and a real-world connection to what is happening in a patient's home vs. clinic are needed.

Cognitive Testing on Mobile and Wearable Devices, Dr. Jenny Barnett, Chief Scientific Officer, Cambridge Cognition

Dr. Jenny Barnett, Cambridge Cognition, opened her presentation by sharing her view on challenges for cognitive testing on mobile devices. Automated cognitive testing can work even in impaired populations with appropriate adjustments; e.g., technological literacy and fiddle factor, screen real estate for stimuli

and responses, intelligent responding and support for queries. However, day-to-day variability, time of day effects, pain, emotion, etc. have to be considered when determining the baseline/results. Cognitive tests must challenge participants to be valid. The goal is to maximize adherence to a valid cognitive assessment protocol. In order to achieve scientifically valid cognitive assessments on mobile devices, opportunities to increase validity (automated and dynamic, can measure not just what a patient does but how they do it, real world data, high frequency data) and threats to validity (unsupervised testing of cognitively impaired users, usability issues of the device, direct translation to digital may or may not be meaningful) need to be addressed. Dr. Barnett described a study with the goal to develop a high-frequency cognitive and mood assessment protocol that is acceptable to patients – based on a wearable device. As part of this effort, a proof-of-concept study was performed in healthy volunteers (Q2 2016; n=10; measuring heart rate, skin temperature, GSR; measuring continuously 9 a.m. to 7 p.m.) with cognitive and mood assessments scheduled hourly and four CANTAB tasks [battery of cognitive tasks developed by Cambridge Cognition] and a PRO administered daily. Results from 1-minute "microtests" demonstrated meaningfulness across-time variation, expected correlations with standard tests, expected effects of age, ease of recall varying across symbols, and some participants reaching ceiling. Goals for the next stage of development (Stage 2; Q3/4 2016) included changing symbol design and bank of symbols to reduce verbal strategies, titrating the difficulty level without increasing patient burden – role of symbol complexity, timing parameters and number of symbols, and change to Apple Watch. Cambridge Cognition recruited 944 patients aged 20-64, with 88 that completed two full-length CANTAB tests. Results from Stage 2 testing determined significant effect of stimulus presentation time on performance, significant impact of number of symbols in trial, symbol design (e.g., number of segments + arrangement) did not significantly impact performance. Overall, sensitivity to working memory performance was confirmed. Phase 3, the real-world testing in healthy participants, has the goals to confirm that the same parameters work on the watch as on the web, to ensure that participants tolerate the paradigm over an extended time/schedule, to ensure that the app is stable and fit for purpose, and to minimize barriers to patient adherence and understanding. In designing protocols for mobile and wearable studies, compromises on validity, usability or regulatory compliance are not acceptable. However, compromises can be considered when evaluating: 1) which device to use, 2) how much assessment will be applied and how frequently, 3) ways to motivate adherence (prompts, feedback, the right level of challenge), and 4) when to adjust comfort levels.

Discussion (Session 3 – Cognition)

Dr. Arnerić opened the discussion by highlighting a draft high-level concept map with focus on cognition-related concepts to consider (**Figure 4**). He requested the participants to identify, if possible, the three top-tier/must-have/key specific outcomes that consider the ability to impact patient/care-giver quality-of-life (QoL), improve health, the ability to address unmet needs, the ability to achieve label claim, and the ability to use as pre-manifest disease outcome assessment.

Participants reiterated that cognitive performance is being impacted by mood and pain. In addition, time-of-day is known to potentially vary a result by more than one standard deviation. Data standards, based on the combination of multiple measures, can serve to reduce variability. A core set of information from a feasible device is needed – in addition to controlled user input.

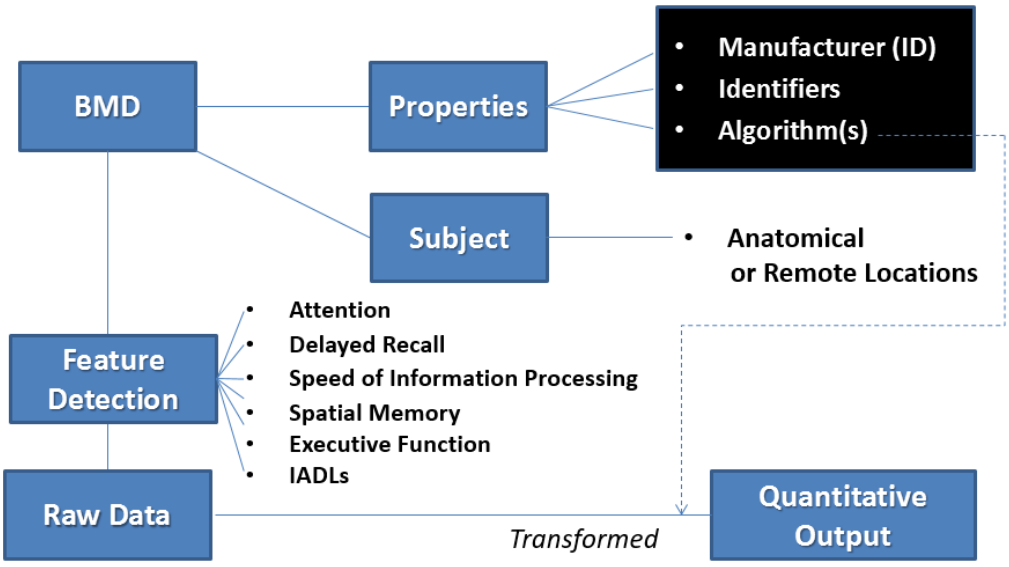

Figure 4: Draft concept map – Cognition

Overall Summary

Attendees reiterated the need to focus on an accelerometer data-based core set of elements (metadata) that can be combined with/compared to other, more traditional measures. That study needs definition and resulting data need to be mapped to SDTM (Study Data Tabulation Model), the standard structure for human clinical trial (study) data tabulations, enabling the creation of standardized data marts. Concept maps that reflect the user's perspective as well as device capabilities need to be devised; focus on sleep and mobility/frailty appears to be most feasible; a longitudinal dataset (to allow for additional analyses) would be most-effective; data sharing needs to be encouraged. Existing frameworks for integration of such novel measures include existing cohort studies such as PPMI and ADNI. Correlation with PROs will assure constant alignment with existing measures and allow the potential to assess clinical meaningfulness. For cognition measures, attributes of parameters need to be defined; available devices need to be vetted; the terminology/ontology needs better definition; initially, cognition measures could be included alongside other measures to provide greater perspective. Participants agreed that it would not be useful to develop a new strategy for each new concept - but to be able to reuse a defined standard/standards.

The potential issue of *pseudo-specificity* and clinical relevance across multiple diseases was raised. Changes in improvement of quality-of-life do not have to attach to a particular disease. A follow-on workshop is planned for May 9 and 10, 2017; in Bethesda, MD "*Biometric Monitoring Device Workshop*; *Advancing CDISC Standards for Biosensors Assessments in Clinical Drug Trials*". Device developers and attendees from regulatory agencies will be invited (FDA, EMA, etc. – to be determined). Panel discussions will facilitate opportunities to hear and probe the regulators points of view, to understand the BMD landscape for the identified concepts-of-interest, to highlight regulatory considerations, and to socialize a forward plan. Participants were requested to consider specific questions that could be raised to the regulators during the May workshop (**Action #2**).

In summary, the question regarding how to capture associated and sufficient contexts and what other inputs might be needed requires attention. Use of uniform terminology regarding sensor, device, and measurement needs to be a focus item. High priority should be given to understanding which functional outcomes patients and caregivers want. In addition, the payer's perspective should not be ignored. However, this could be a later priority when alignment with regulators has been achieved.

It was suggested to invite EMA as well as PMDA representatives to the May workshop (**Action #3**).

Based on the earlier invitation by Dr. Ray Dorsey, participants suggested to summarize the proceedings of this meeting in a manuscript with focus on the necessity to create CDISC standards, to be submitted to *Digital Biomarkers*.

Dr. Lynn Hudson (Chief Science Officer, Critical Path Institute) informed participants of a recent effort within C-path to leverage C-Path's individual consortium activities on mobile health/wearable devices/biosensor monitoring devices and coordinate/unify C-Path's approach and plans forward. The DRASTIC (Device Regulatory and Standards Team in C-Path) team, co-led by Dr. Hudson and Dr. Volker Kern, is composed of members of seven C-Path consortia with a focus on CDISC standard development and the creation, in collaboration with the regulatory agencies, of a more-defined regulatory pathway.

In closing, Dr. Arnerić thanked all participants for their invaluable contributions and their support of CAMD. Minutes/notes of this workshop will be compiled and will be posted on CAMD's website together with the presentations (Dr. Arnerić requested presenters to approve public posting of their contributions; **Action #4**). He encouraged all participants to consider registering for the follow-on May 9 and 10, 2017 workshop in Bethesda, MD.