

Commonalities and Challenges in the Development of Clinical Trial Measures in Neurology

Jesse M. Cedarbaum · Diane Stephenson · Richard Rudick · Maria C. Carrillo · Glenn Stebbins · Douglas Kerr · Jill Heemskerk · Wendy R. Galpern · Petra Kaufmann · David Cella · Maria Isaac · Marc K. Walton

Published online: 11 November 2014

© The American Society for Experimental NeuroTherapeutics, Inc. 2014

Abstract As neurologists and neuroscientists, we are trained to evaluate disorders of the nervous system by thinking systematically. Clinically, we think in terms of cognition, behavior, motor function, sensation, balance and co-ordination, and autonomic system function. But when we assess symptoms of neurological disorders for the purpose of drug development, we tend to create disease-specific outcome measures, often using a variety of methods to assess the same types of dysfunction in overlapping, related disorders. To begin to explore the potential to simplify and harmonize the assessment of dysfunction across neurological disorders, a

symposium, entitled, "Commonalities in the Development of Outcome Measures in Neurology" was held at the 16th annual meeting of the American Society for Experimental NeuroTherapeutics (ASENT), in February 2014. This paper summarizes the presentations at the symposium. The authors hope that readers will begin to view Clinical Outcome Assessment (COA) development in a new light. We hope that in presenting this material, we will stimulate discussions and collaborations across disease areas to develop common concepts of neurological COA development and construction.

Summary of a Symposium Presented at the 16th Annual Meeting of the American Society for Experimental NeuroTherapeutics (ASENT), Bethesda, MD, February 28, 2014

J. M. Cedarbaum (✉)
Neurology Early Clinical Development, Biogen Idec, 14 Cambridge Center, Cambridge, MA 02142, USA
e-mail: jesse.cedarbaum@biogenidec.com

D. Stephenson
Coalition Against Major Diseases, Critical Path Institute, Tucson, AZ, USA

R. Rudick
Value Based Medicine, Biogen Idec, Cambridge, MA, USA

M. C. Carrillo
Medical & Scientific Relations, Alzheimer's Association, Chicago, IL, USA

G. Stebbins
Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

D. Kerr
Clinical and Safety Sciences, Biogen Idec, Cambridge, MA, USA

J. Heemskerk
Division of Adult Translational Research, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

W. R. Galpern
Office of Clinical Research, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA

P. Kaufmann
Division of Clinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA

D. Cella
Northwestern University Feinberg School of Medicine, Chicago, IL, USA

M. Isaac
European Medicines Agency, London, UK

M. K. Walton
Office of Translational Sciences Center for Drug Evaluation and Research Food and Drug Administration, Silver Spring, MD, USA

Keywords Clinical trials · Clinical outcome assessments

Introduction

Jesse M. Cedarbaum MD

As neurologists and neuroscientists, we are trained to evaluate disorders of the nervous system by thinking systematically. Clinically, we think in terms of cognition, behavior, motor function, sensation, balance and coordination, and autonomic system function. But when we assess symptoms of neurological disorders for the purpose of drug development, we tend to create disease-specific outcome measures, often using a variety of methods to assess the same types of dysfunction in overlapping, related disorders. We now have a series of legacy clinical trial outcome measures, the clinical relevance of which is being challenged, or the precision and sensitivity of which appear less than adequate to demonstrate efficacy in chronic, progressive neurological disorders.

In clinical practice, we take a neurological history and perform a neurological examination. We modify the history we take or examination we perform based on the clinical data that emerge before us. And in doing so, we organize our history taking and examining according to domains of neurological function. Our psychiatric colleagues are advancing, in the Research Domain Criteria initiative described below, the concept of assessing *domains* of dysfunction and disability instead of adhering to a rigid nosology of disease [1]. Can we do the same in neurology?

To begin to explore the potential to simplify and harmonize the assessment of dysfunction across neurological disorders, a symposium entitled, “Commonalities in the Development of Outcome Measures in Neurology” was held at the 16th annual meeting of the American Society for Experimental NeuroTherapeutics, in February 2014. This paper summarizes the presentations at the symposium. The authors and contributors, who were drawn from the academic community, pharmaceutical industry, patient advocacy groups, the US National Institutes of Health and regulatory bodies, hope that readers will begin to view clinical outcome assessment development in a new light.

The goals of this summary are 3-fold: 1) to define the concepts and the regulatory context for the development of clinical outcome measures; 2) to illustrate using specific examples challenges identified and approaches being taken by investigators to meet these challenges in a number of areas of clinical neurology; and 3) to call attention to new tools and frameworks that may enable us to begin speaking with a common vocabulary as we evaluate outcomes across indications in neurotherapeutics development. This symposium represents only a beginning of this effort.

We hope that, through the diverse vignettes summarized herein, we will stimulate discussions and collaborations across disease areas to develop common concepts of neurological clinical outcome assessment development and construction.

Conceptual and Regulatory Background

Keynote: An Overview of Clinical Outcome Assessments

Mark K. Walton MD, PhD [Note: the views presented in this article are those of the author and do not necessarily reflect policies of the Food and Drug Administration (FDA).]

Planning a clinical trial should include careful selection of clinical outcome assessments (COAs) because of the critical role that COAs play in the success, or failure, of a study’s ability to evaluate the treatment effect. For a variety of reasons, there are many diseases for which the COA of choice is not well established. For example, there may be no COA known to be sensitive to the clinical manifestations of the disease, or to a particular aspect of disease impact that is of interest. Experience with existing COAs in prior clinical trials may have revealed weaknesses of a COA, such as with reliability, acceptability of the assessment procedure by patients, or sensitivity to change within a reasonable study duration. How a COA is used in the endpoint, as defined by the study design and data analysis, can ameliorate some problems. Nonetheless, the intrinsic properties of the COA will have a large influence on the usefulness of the endpoint.

A general process for developing new COAs has been described by the FDA [2]. If the COA is to be developed and qualified in advance of any particular drug development program’s use of the COA, the first step is to clearly understand the disorder, including features such as the phenotypes or other subtypes of the disease, the breadth of manifestations, the clinical course of the disease, and the range of severity, along with the variability in these. An important component of developing a good COA is to ensure that the “patient’s voice” has been heard in selecting what aspects of the disease to focus on. Which manifestations of the disease are most important to patients (and if that varies at different stages of the disease), and how the manifestations affect the patients in their typical daily lives need to be understood. The expected effects of a treatment can then be considered to select which aspects of patients’ functioning or feelings should be targeted as the meaningful benefit from treatment, and will need to be measured by the COA.

Because there is often a range of phenotypes within a disease, a range of stages of severity, as well as other differences among patients with the same disease diagnosis, a specific COA may be suitable for studying only a portion of the patients with the disease. Thus, the specific population of patients that is intended for study needs to be identified. The

intended population is one of a wide variety of elements [collectively called a “context of use” (COU)] that precisely defines how a COA can be validly used. In addition to the specific disease patient population, elements such as the study setting, geographic region, or concomitant care of patients may also need to be specified.

In some cases the intended health benefit on a patient’s daily life can be readily measured directly (e.g., as with a patient-reported outcome questionnaire), but often that is difficult to do. In such cases, study designers might find it better to infer the true health benefit based on a less direct measurement such as a physician evaluation in the clinic, or performance of an in-clinic procedure (e.g., timed walking, muscle strength testing). In this case, a concept of interest (COI) for measurement is defined, which is thought to have a useful relationship to the meaningful effects of treatment on patients within their typical daily life. Measurement in this manner, while a step removed from the true, meaningful, benefit of the treatment, might have reliability and/or sensitivity. When the measurement method is well defined, and it is understood how to interpret the COA (i.e., how to translate changes in measurements into changes in patient’s daily lives), this can be a very powerful approach.

After the COI is selected, a survey of existing COAs that measure the COI can be reviewed and considered to assess if any existing tools appear well suited, or closely enough to warrant the effort to modify the COA tool. Alternatively, there may be no existing COAs suited to the selected COI within the desired COU, and a new COA will need to be developed. Whether an existing COA is being used or modified, or a new COA is being developed, evaluating the COA properties in light of the desired COU should be done to establish that the COA is acceptable for the intended use. These properties include the content validity of the COA, reliability, sensitivity to change, and other related properties. The clinical meaningfulness of the COA is critical to evaluate: What do the measurements and changes in measurements mean for the patients in their typical daily lives? This will be easier to show for COA procedures that closely resemble patients’ activities in their daily lives than for measurement procedures that are only loosely related to daily life, but it is important to elucidate for both.

Developing new COAs is not a trivial task, but it can be very important to the success of therapeutic development programs. Because many disorders have similar impacts on patients’ lives as other disorders, there is substantial interest in whether, or how, COAs established for one COU can be transported to another COU, such as a different disease that has similar clinical manifestations. Because there would be much already known about a well-developed and evaluated COA, this may be an efficient approach to obtaining a COA for a disorder for which there has not been any, or any good, assessments available. Automatic applicability, however,

cannot be assumed. When the context of use changes, the measurement properties relied upon for developing new COAs should be reevaluated in the new COU. Even if the measurement properties of the original COA are not entirely suitable for the new COU, modification of the original COA for the new COU might be more quickly accomplished than creating an entirely new COA, and yield a COA with adequate properties in the COU.

All the issues described for the development of new COAs are relevant to the re-use in a new COU, and the COA properties within the new COU will need to be considered. Without careful consideration, weaknesses in an existing COA for a new COU might be unrecognized and negatively impact therapy development programs. Content validity, for example, might be different in two disorders. Although at a high level the functional domain of interest is the same in the two disorders, the detailed description of how the disorder’s impairments affect the patients’ lives may be different, and a COA comprehensive for the domain in one disorder may not be in another. Alternatively, if two disorders (or two different populations within a single disorder) have different ranges of severity, then the COA’s range of measurement may be suitable for one disorder but not for the other (e.g., ceiling or floor measurement effects may damage the usefulness of the COA).

Because many disorders have similar impacts on patients’ lives as other disorders, transportability of COAs from one disorder to another is of great interest. However, development of COAs can be complex. Development of a new COA explicitly with multiple COUs in mind might provide greater efficiency in achieving availability of good COAs for multiple diseases. Similarly, starting with a well-developed COA in one disease may be an efficient path to have a good tool for a similar disease, but will warrant some re-evaluation of the COA in the new COU to avoid setbacks in therapy development.

The Role of Regulatory Qualification of New Outcome Measures: European Regulatory Perspective

Maria Isaac MD, PhD, MFPM [Note: the views expressed in this article are the personal views of the author and may not be understood nor quoted as being made on behalf of or reflecting the position of the European Medicines Agency (EMA) or one of its committees or working parties or any of the national agencies.]

Clinical assessment tools are continuously evolving, and predictive accuracy needs to be validated using clear regulatory guidance to be accepted by regulators. For example, the development of new outcomes measures to be employed in clinical trials of drugs intended to treat predementia stages of Alzheimer’s disease (AD) requires involving regulators in discussions early in the development of each clinical trial program. Previous EMA guidelines require a co-primary outcome measure, which involves an assessment of cognition by

a clinician and a measure of disability such as scales that assess activities of daily living (ADLs). However, discussion about development of predementia has caused an update of the Committee for Medicinal Products for Human Use (CHMP) guidelines in dementia [3], following the first qualification procedures for novel clinical trial methods for AD studies published in 2011 [4]. The current scientific advice and qualifications procedures at the CHMP incorporate recommendations for how to measure cognition and disability earlier in the predementia stages of AD.

Scientific assessment of the potential for use of new methodologies (clinical outcomes, biomarkers, new statistical methods) in clinical trials can be advanced in a structured fashion through the process of *qualification*, which was recently introduced by regulatory agencies including the EMA and the US Food and Drug Administration. Regulatory qualification of a new method or tool for a defined context of use provides scientifically robust assurances to sponsors and regulators that accelerate appropriate adoption of new methods into drug development and clinical practice. Such assurance saves time and money by removing the burden of proof on each individual sponsor to provide data to regulatory agencies on performance and validation.

In Europe, the Qualification of Novel Methodologies procedure (<http://www.emea.europa.eu/pdfs/human/biomarkers/7289408en.pdf>), established by the EMA in 2008, provides the regulatory pathway for qualification of new outcome measures. The aim of the procedure is to facilitate the review of new methodologies employed in pivotal trials at the time of the Marketing Authorization Application (MAA), by providing prior assurance to EMA of their scientific and clinical validity. Regulatory qualification of new clinical assessment tools enables drug development teams that are using new clinical endpoints for targeting new indications to provide regulators with an understanding of their clinical meaningfulness. As part of this process, robust regulatory discussion that includes contributions from patient organizations, academic experts, and the pharmaceutical industry in the qualification process offers the possibility of presenting a common regulatory understanding in the development of new outcome measures and assessment tools.

Although the qualification process is entirely voluntary, sponsors are urged to contact the Scientific Advice Working Party as early as possible to receive qualification *advice*; they are allowed to approach the EMA with a dossier in support of a direct qualification *opinion*. Publication of draft qualification opinions opens new proposed tools and methods to scientific scrutiny and public comment to ensure that adopted opinions are broadly accepted within the community. The qualification process thus can play a valuable part in establishing the concept of interest and context of use for novel Clinical Outcome Assessments in the development of novel neurological therapies.

Session 1: Summaries of Efforts to Update and Obtain Regulatory Qualification of Selected Disease-specific Outcome Measures

Critical Path Institute: Accelerating Drug Development Through Regulatory Science

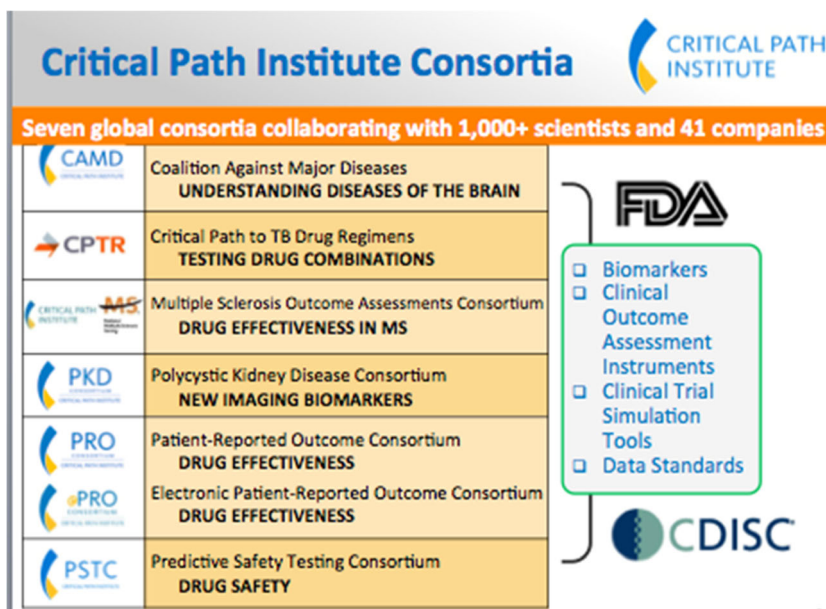
Diane Stephenson PhD

The Critical Path Institute (C-Path) is a nonprofit, public-private partnership that the US Food and Drug Administration (FDA) created under the auspices of its Critical Path Initiatives program in 2005. C-Path's aim is to accelerate the pace and reduce the costs of medical product development through the creation of new data standards, measurement standards, and methods standards that aid in the scientific evaluation of the efficacy and safety of new therapies. These precompetitive standards and approaches have been termed "drug development tools" (DDTs) by the FDA, which established a process for official review and confirmation of their validity for a given context of use. C-Path orchestrates the development of DDTs through an innovative, consortia-driven approach to the sharing of data and expertise and consensus building among participating scientists from industry and academia with FDA participation and iterative feedback.

C-Path is organized by program areas that focus on either a specific aspect or stage of medical product development or on a specific disease type. Currently C-Path research is divided into the following programs: Coalition Against Major Diseases (CAMD), Patient-Reported Outcome (PRO) Consortium and Electronic PRO Consortium, Critical Path to TB Drug Regimens, Polycystic Kidney Disease Consortium, and the Predictive Safety Testing Consortium. In addition, C-Path has recently entered into a partnership agreement with the Clinical Data Interchange Standards Consortium (CDISC) to create the Coalition for Accelerating Standards and Therapies, which will develop clinical research data standards in additional therapeutic areas identified as high priority by the FDA and industry (Fig. 1).

CAMD was formed in 2009 with the realization of the exceptional challenges and need for collaboration in successful therapeutic development for age-related neurodegenerative diseases [5]. The mission of CAMD is to accelerate the development of therapies for Alzheimer's disease (AD) and Parkinson's disease by generating methods and tools for evaluating drug efficacy, expediting clinical trials, and streamlining review by regulatory agencies [3]. CAMD's accomplishments to date include: 1) development, in collaboration with the CDISC, of the first AD and PD (with the involvement of National Institute of Neurological Disorders and Stroke) data standards for use in clinical trials; 2) the first biomarker qualification from the European Medicines Agency (EMA) for the use of magnetic resonance imaging to select

Fig. 1 Critical Path Institute consortia. ADNI = Alzheimer’s Disease Neuroimaging Initiative; PPSB = Private Partner Scientific Board; MCI = Mild Cognitive Impairment; AD = Alzheimer’s disease; CAMD = Coalition Against Major Diseases; pCOA = prodromal Clinical Outcome Assessment; FDA = Food and Drugs Administration; EMA = European Medicines Agency; EP = endpoint



patients with early stages of cognitive impairment for AD clinical trials; 3) development of a transformational quantitative drug–disease trial modeling and simulation tool for mild and moderate AD, representing the first disease model advanced for a regulatory decision; 4) establishment of a unified clinical database for AD from the placebo arm of 24 clinical trials, with patient-level data from 6500 patients; the CAMD CODR [Critical Path Institute (C-Path) Online Data Repository (CODR)] database is available to qualified researchers and is facilitating addressing a variety of unaddressed research questions.

Established instruments to measure cognition lack sensitivity and responsiveness in early stages of the AD process (e.g., Raghavan et al. [6]). CAMD’s most recent project aims to address this problem by advancing a composite clinical outcome assessment tool built from elements of existing scales through a formal regulatory path for qualification as a primary outcome measure for prodementia AD trials. This project originated as an alliance between CAMD and another public–private partnership, the Alzheimer’s Disease Neuroimaging Initiative, when it became clear that duplicative efforts were taking place by individual pharmaceutical companies. Individual industry and academic efforts have proposed more sensitive and responsive instruments in early stages of AD. Leading proposals include items from existing measures (ADASCog, Clinical Dementia Rating Scale Sum of Boxes, Mini-Mental Status Examination) as a composite clinical endpoint emphasizing both cognitive and functional measures of performance.

This project, as outlined in Fig. 2, aligns with the FDA’s 2013 draft guidance on early AD (Guidance for Industry: Alzheimer’s Disease: Developing Drugs for the Treatment of

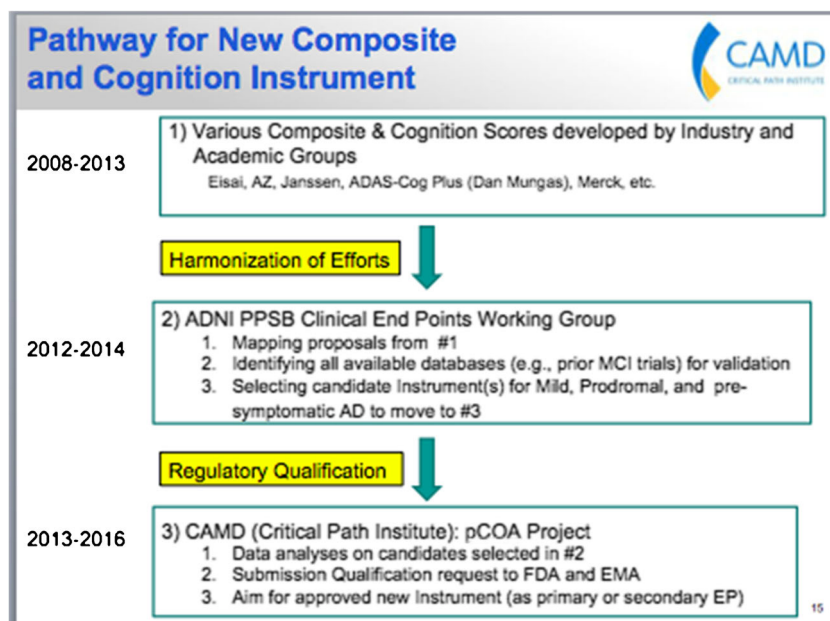
Early Stage Disease, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338287.pdf>), as well as the EMA’s 2013 Concept Paper on Need for Revision of the Guideline on Medicinal Products for the Treatment of Alzheimer’s Disease and other Dementias (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/10/WC500153464.pdf). CAMD’s prodementia clinical outcome assessment tool project received feedback from both the FDA and EMA to a formal letter of intent, including the intention to align between regulatory agencies with parallel reviews, as well as the importance of addressing clinical meaningfulness with any proposed composite measurement tool. CAMD has aligned with numerous public–private partnerships, including C-Path’s PRO consortium to address this gap [7]. Sharing learning across brain diseases is also underway, as highlighted by the American Society for Experimental Neurotherapeutics session, and promises to benefit multiple stakeholders.

The Multiple Sclerosis Outcome Assessments Consortium

Richard Rudick MD

The Multiple Sclerosis Outcome Assessments Consortium (MSOAC) is one of several consortia established by the Critical Path Institute to create drug development tools to accelerate development of therapeutics. MSOAC was established through a grant from the National Multiple Sclerosis Society (NMSS) to the Critical Path Institute, with the purpose of developing an improved clinical outcome measure for disability progression in multiple sclerosis (MS). MSOAC is a 43-member organization, with

Fig. 2 Proposed pathway for new composite and cognition instrument. TB = tuberculosis; MS = multiple sclerosis



representation from the academic community, pharmaceutical companies, patient advocacy organizations, and with liaisons to both the Food and Drugs Administration (FDA) and the European Medicines Agency (EMA). The mission of MSOAC is to develop and support adoption throughout the MS community (patients, clinical investigators, pharmaceutical industry, regulatory agencies, and advocacy groups) of a clinical outcome assessment (COA) tool for future MS clinical trials. The purpose of this COA will be to reflect the impact of an intervention on disability due to MS. MSOAC will seek regulatory qualification of the COA for registration trials. Therefore, the COA must be useful for demonstrating clinical change due to MS.

The MSOAC members conceptualized the concept of interest (COI) as indicated in Fig. 3. The target population is people with MS, and the COI is MS-related disability. Examples of activities of daily living related to MS disability were identified, and bodily activities connected to those activities highlighted. Subcomponents of the bodily activities were defined, and potential measurement tools were listed. In this way, the potential measurement instruments can be conceptually tied to the underlying concept of interest—disability.

MSOAC members adopted a focus on neuroperformance measures that relate to aspects of the MS disease process, largely because these measures have favorable psychometric properties compared with clinical ratings. The members identified 7 characteristics of candidate measures. They should be:

1. Objectively measurable by trained observers
2. Acceptable to patients
3. Practical
4. Cost-effective.

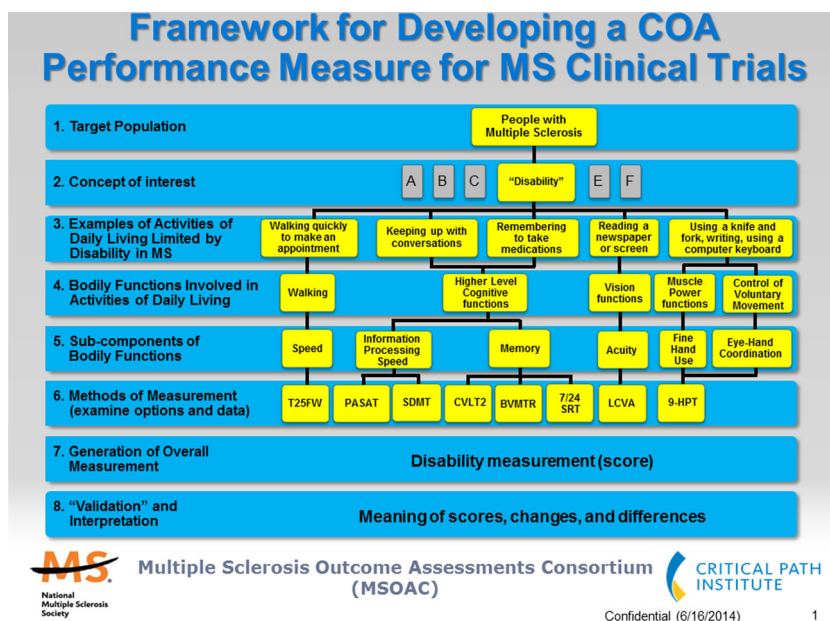
In addition, candidate measures should:

5. Have *strong psychometric properties* of reliability and validity
6. *Reflect impairments and disabilities that are important to patients*
7. *Change over time* to allow demonstration of a therapeutic effect.

Measures of pain, fatigue, sexual dysfunction, depression, and bowel and bladder dysfunction, among other symptoms, are common in MS, and are extremely important, but will not be included directly in the COA because these symptoms are not directly measurable using performance measure approaches.

Because of known limitations of the traditional MS outcome measures—relapses and the Kurtzke Expanded Disability Status Scale—the NMSS, in the mid-1990s, convened a clinical outcomes task force, which recommended a 3-part neuroperformance measure—the Multiple Sclerosis Functional Composite (MSFC)—consisting of a timed 25-foot walk, the 9-hole peg test, and the 3-second version of the Paced Auditory Serial Addition Test. The NMSS published an operations manual, describing exact methods for administering the MSFC. The MSFC never achieved acceptance as a primary outcome measure for MS clinical trials for several reasons. The recommended approach of converting individual component scores into individual *z*-scores based on reference populations, then averaging the *z*-scores, resulted in trial results that were very difficult to interpret. Further, the Paced Auditory Serial Addition Test was viewed as undesirable by both investigators and patients. Finally, the MSFC did

Fig. 3 Framework For performance measure for multiple sclerosis clinical trials



not contain a measure of vision, even though visual disturbance is common in MS, and easily measured objectively. However, because the NMSS task force recommended that the MSFC be included in prospective clinical trials in 1996, numerous research sponsors and academic investigators prospectively included MSFC (and other neuroperformance measures) in clinical trials beginning in the late 1990s. Consequently, a large number of completed trial datasets are potentially available for analysis, containing longitudinal, prospectively collected data on more than 20,000 patients with MS.

Because of the promise of using neuroperformance measures, the MSOAC members elected to seek clinical trial datasets from completed trials (that contain neuroperformance measures, relapse data and Expanded Disability Status Scale, and patient-reported outcomes), map the data elements to a common Clinical Data Interchange Standards Consortium (CDISC) standard to create a pooled dataset, and analyze the behavior of neuroperformance measures in the pooled dataset. Neuroperformance measures and their change will be compared with patient self-report to determine whether these measures reflect change that is perceived as important to the patient.

Thus, the MSOAC project entails the steps shown in Table 1.

The MSOAC members have identified 22 studies with nearly 17,000 patients. These studies were selected based on prospectively defined criteria established to be certain that the datasets contain information valuable for the purpose of the MSOAC project. Discussions are ongoing with the sponsors and investigators who own these datasets, and data sharing agreements are being discussed and finalized. At the same

time, a literature review is being formulated and discussed with the FDA and EMA. The literature review will be used to augment the data contained within these legacy datasets, and to identify more contemporary datasets of value for the project. A data standards working group is overseeing this aspect of the project. A statistical analysis working group, with members from academia and industry, is currently developing an analysis plan to carry this work forward. Once the analysis is completed, it is the intention of MSOAC to prepare a qualification package for submission and review by FDA and EMA. If the data are considered adequate, a new clinical outcome measure for disability in MS can be qualified.

The implications of this project are significant:

1. The project will result in a CDISC data standard for MS. This will improve the ability to compare data across

Table 1 Steps in the Multiple Sclerosis Outcome Assessments Consortium project

1. Create Clinical Data Interchange Standards Consortium Standard for multiple sclerosis (MS), based initially on elements from the National Institute of Neurological Disorders and Stroke MS common data element project.
2. Create a pooled data set from completed clinical trials containing traditional clinical outcomes (relapses and Expanded Disability Status Scale), neurological performance measures (e.g., Multiple Sclerosis Functional Composite), and patient-reported outcomes
3. Analyze pooled dataset to determine components of a neuroperformance-based composite outcome measure, and recommended approach to using the composite outcome.
4. Seek regulatory approval of a new MS outcome measure.

- studies, to interpret findings, and to analyze pooled datasets
2. A database of pooled, de-identified clinical trial data will be mapped to the CDISC standard. This will allow studies of clinical outcomes assessment tools, disease modeling, and potentially imaging or biomarker studies
 3. A new COA consisting of neurological performance measures will be developed. This could lead to an FDA- and EMA-approved disability endpoint in future MS clinical trials
 4. The project will be guided by FDA and EMA outcome measure qualification pathways. This could provide an example for other groups interested in pursuing better outcome measures through regulatory sciences
 5. The project depends on transdisciplinary collaboration and data sharing. This aspect of the project is considered very important, as solving the major problems in MS will require collaboration among academia, industry, regulatory bodies, and patient groups.

Initiatives in the Development of Clinical and Biomarker Measures for Alzheimer's Disease

Maria C. Carrillo PhD

Several recent Federal Agency Guidelines and decisions have necessitated an increased awareness for the need to synergize activities across the drug development ecosystem for Alzheimer's disease (AD). The Alzheimer's Association is participating in several initiatives in response to recent federal statements and decisions. Examples of such activities, which are driving the need for consensus in clinical trials for AD and other neurodegenerative diseases, include:

- National Institute on Aging Alzheimer's Association (NIA-AA) and International Working Group criteria for diagnosis in AD
- Food and Drugs Administration (FDA) Draft Guidance for Early Stage Alzheimer's Assessment in Drug Development, and the recent European Medicines Agency (EMA) concept paper on the need for revision of their current AD guidance (www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/10/WC5001534).

Efforts to support these needs have come from many places, including the Alzheimer's Association Research Roundtable (AARR), where issues are discussed amongst industry and academic leaders in the field, as well as regulators from FDA and EMA, biomarker consortia, and from our collaborations with the Coalition Against Major Diseases, as previously discussed.

The urgency to develop treatments for AD, especially in its earliest stages, is huge. More than 5 million people are living with AD, including 200,000 under the age of 65 years with

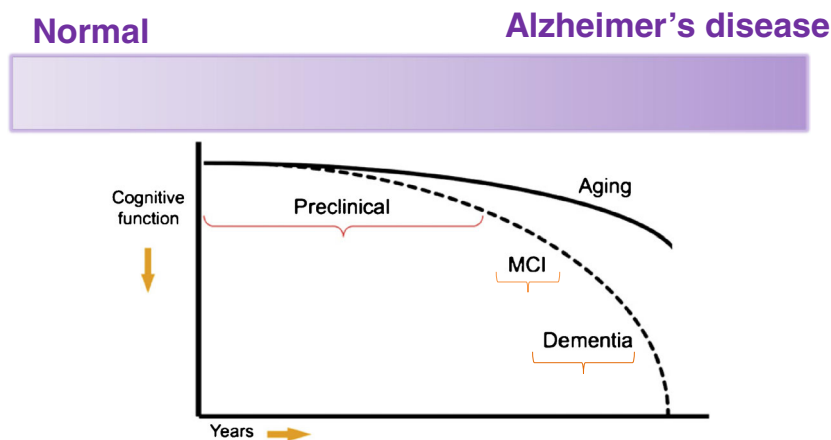
younger-onset disease. Every 68 s, someone in the USA develops AD; by 2050, it will be one every 33 s. AD is the sixth leading cause of death across all ages, and is the fifth leading cause of death for those aged 65 years and older. It is the only cause of death among the top 10 in the USA without a way to prevent, cure, or even slow its progression.

Our understanding of the course of the disease is also evolving. In 2009, the AARR got together to decide whether the time was right to revise our diagnostic criteria for AD. The result was a collaborative effort with the National Institute of Aging (our partner in developing the original National Institute of Neurological and communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria), comprised of 3 working groups. The results of these efforts were published in a series of papers in 2011 that outline recommendations for diagnostic criteria for 3 stages of the disease: dementia [8], mild cognitive impairment (MCI) due to AD [9], and a new “preclinical” stage encompassing persons with biomarker evidence of, but no or minimal, disease signs and symptoms [10].

One of the key goals in AD drug development is to understand how we can begin treatment in the earliest, even clinically silent, phases of the disease (Fig. 4). The FDA has been an active and engaged partner in this endeavor and has recently issued a draft guidance to inform drug development in this arena. The AARR, comprised of industry leaders working in collaboration with the FDA, the EMA and now the Centers for Medicare & Medicaid Services (CMS), has played a seminal role as an open forum for collaboration and conversations in this process. In October of 2013, the AARR hosted a session on functional outcome measures for clinical trials in early stages of AD. Approaches considered included development of new tools, modification of or retrofitting of existing outcome measures and even reviving little-used ones. Again, enabling communication and collaboration in the precompetitive space was seen as key to making sure that all stakeholders stay informed and are not undertaking duplicative efforts.

An example of this sort of consortium approach is the Collaboration for Alzheimer's Prevention (CAP) Initiative, facilitated by the Fidelity Foundation and the AA. Under this initiative, Fidelity and the Association meet with the leadership of 4 ongoing prevention trials and trial networks to help the tools and technologies, including biomarkers, imaging methods, and outcome measures across studies. The goal of this effort is to enable, as much as possible given the differing natures of the study populations, sharing of information, resources and, ultimately, comparisons across study results across these and other related clinical trials. Furthermore, it should be noted that all these studies have committed to making both their baseline and final data available to the public through open access, specifically on the Global Alzheimer's Association Interactive Network (www.GAAIN.org). Table 2 provides a snapshot of the outcome assessments being used in these

Fig. 4 The clinical course of Alzheimer’s disease (adapted from Sperling et al. [10]). MCI = Mild Cognitive Impairment



studies. All of these instruments probe similar areas of cognition and function, combined in novel ways in order to try to capture the earliest detectable changes in populations of interest. The FDA, especially the Division of Neurology Products (DNP), has been and continues to be very active and engaged in these discussions. It is important for them to be engaged so that we know they are supportive of the measures being developed as exploratory aims in these studies.

Cognitive biomarkers are important, but as we all know biological fluid and imaging markers are important parts of the equation not only in mild-to-moderate AD, but also especially in prevention trials that are looking much earlier, including in preclinical stages of disease. ADNI is an excellent example of such collaborations, and a hallmark success story for the field. Worldwide, ADNI allows comparison of data from across countries. Representatives from these countries meet yearly and on conference calls to ensure standardization of measures across studies that are actually starting in multiple places. Openness of information and data sharing and, of course, open access is important to creating populations in which these types of standardized measures can be validated externally.

Table 2 Outcome measures being employed in prevention trials underway in preclinical Alzheimer’s disease (AD)

A4
Computerized cognitive composite
• CogState card playing tasks on iPad
Face-name association and pattern separation
API
Composite cognitive score based on religious orders study/DIAN data
TOMORROW
Specific cognitive instruments were selected and decision rules developed to capture reliably the relevant domains affected in cognitive decline and in mild cognitive impairment AD
DIAN
Cogstate-based testing
Additional neuropsychological tests

A working group is focusing on hippocampal volumetry, finding a way to harmonize the way in which we segment the hippocampus in clinical trials. Another important standardization effort, the Cerebrospinal Fluid Quality Control (CSF QC) Program, has now graduated to an official program of the Institute for Research Methods and Materials (IRMM), which itself is closely aligned with the Federation of Clinical Chemistry. The goal here is to create a common standardized matrix that cerebrospinal fluid samples can be measured against. If this is successful, the IRMM would actually be willing to make this material for distribution across the globe at cost. This would allow all companies to use the same matrix material because what we are discovering is that the largest barrier is not how to measure amyloid in cerebrospinal fluid, even though that is a challenge, but it is rather the matrix material that we are having issues with. This is a 3-year project because one has to demonstrate to IRMM that one has a stable matrix and stable material.

Another initiative that is really critical comes from the fact that neurological diseases like AD do not exist in a vacuum. The Biomarkers Across Neurological Diseases (BAND) initiative recognizes that amyloid may not be the only cause of cognitive disorders but that there may be other underlying neuropathological factors that are perhaps additive or also synergistic. An example is Parkinson’s disease. This program, sponsored by the AA in collaboration with the Michael J. Fox Foundation for Parkinson’s Research and the Weston Brain Institute, will leverage resources from ADNI and Parkinson Progression Markers Initiative (PPMI). This works because both projects are so closely aligned. If we can mine these datasets and figure out ways in which these two diseases intersect, finding common markers, maybe we could find a way to accelerate progress in both.

Lastly, is the Accelerating Medicines Partnership (AMP) is a collaboration with participation from pharmaceutical companies the NIH and non-profit groups that, represents an opportunity to leverage existing resources and projects such as the Dominantly Inherited Alzheimer’s Network (DIAN)

and the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) study to be able to mine their databases for additional information, and to actually embed a common suite of biomarkers across studies. Hopefully, this will ensure that we are not always going to be wondering what to do across the board in clinical trials with respect to biomarkers and to determine their clinical utility.

In summary, there is a great deal of urgency to develop new consensus clinical and biomarker outcome measures for AD. Collaboration and the ability to have conversations among different groups is one of the ways by which we can accelerate progress and also leverage our pooled resources.

Parkinson's Disease: Updating the Unified Parkinson's Disease Rating Scale

Glenn Stebbins PhD

The Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) was initiated following an International Parkinson and Movement Disorder Society (IPMDS) Task Force critique of the UPDRS [11]. The UPDRS was originally published by Fahn and Elton [12], and became one of the most commonly used rating scales in Parkinson's disease (PD). The scale encompassed 4 parts: part I—mood, mentation and behavior; part II—activities of daily living; part III—motor examination; and, part IV—complications of therapy. Most of the items were based on a 5-point Likert-type scale, although some items were binary or descriptive only.

The MDS Task Force critique identified numerous strengths of the UPDRS, including its widespread use in clinical and research settings, the comprehensive assessment of motor aspects of PD, and the acceptable levels of reliability and validity for parts II and III. The task force also identified specific weaknesses of the UPDRS, including ambiguities in some instructions to the raters, poor inter-rater reliability on some items, and incomplete assessment of nonmotor symptoms of PD. An additional concern was that the overall scaling was weighted to the severe end of the spectrum and was not sensitive to milder manifestations of PD. The overall recommendation of the task force was to modify the UPDRS to address these weaknesses.

From this critique, a new task force was commissioned with the task of developing the MDS-sponsored revision of the UPDRS (the MDS-UPDRS). The process for revising the UPDRS followed traditional scale development steps, including assigning a Delphi panel, or panel of experts, to review the required domains of interest for assessment, select the scaling metric, and either modify existing items or generate new items to be included in the scale. Once the items were finalized, cognitive pretesting was conducted on each item assessing patient and rater comprehension and comfort with the questions and response items [13]. Multiple rounds of cognitive

pretesting and subsequent modification of items were required. The final scale retained the 4-part structure of the original UPDRS but refocused each part so that part I focused on nonmotor experiences of daily living, part II on motor experiences of daily living, part III remained the motor examination, and part IV motor complications. The scaling metric was modified (Table 3) so that all items followed a 0–4 Likert scale with 0 = no impairment, 1 = slight (instead of mild in the UPDRS), 2 = mild (moderate in the UPDRS), 3 = moderate (severe in the UPDRS), and 4 = severe (marked in the UPDRS). The scaling metric was also modified so that both frequency or intensity of impairment *and* impact of impairment was assessed.

Upon successful cognitive pretesting, the MDS-UPDRS was subjected to field testing in a sample of 877 English-speaking patients with PD, spanning the gamut from very mild to very severe disease, in a multicenter study. The purpose of this field testing was to assess the basic clinimetrics of the new scale, specifically testing floor and ceiling effects, internal consistency, item-to-total correlations, and both exploratory and confirmatory factor analyses (Table 4). The results of these assessments were very encouraging [14], with no floor or ceiling effects noted, strong internal consistency for each part of the scale, and a valid factor structure found for each part. It was determined that there would be no overall MDS-UPDRS score, but rather scoring would be selective for each part individually.

In addition to representing a new scale for the unified assessment of PD, the MDS-UPDRS is part of a web-based rater training and certificate program [15] sponsored by the IPMDS (www.movementdisorders.org) [15]. Additional studies of the MDS-UPDRS have included independent validation of the entire scale, validation of part I and part II [16–19], development of a postural instability gait disorder/tremor dominant formula [20], and a calibration method for converting UPDRS scores to MDS-UPDRS scores [21]. Further, there is an active translation program aimed at

Table 3 Revised Unified Parkinson's Disease Rating Scale scaling metric

Old scoring	New Scoring Metric
Mild (1) Moderate (2) Severe (3) Marked (4)	Slight (1) = low frequency or intensity causing no impact on function Mild (2) = frequency or intensity sufficient to cause modest impact on function Moderate (3) = frequency or intensity causing considerable impact, but do not prevent function Severe (4) = prevent function

Table 4 Assessing the clinimetric properties of the Movement Disorder Society–Unified Parkinson’s Disease Rating Scale

Clinimetric property	Statistic
Internal consistency	Cronbach’s α
Concurrent validity	Spearman correlation coefficient ρ , Lin’s CCC
Exploratory and confirmatory factor analyses of parts	CFI (≥ 0.90)
DIF by race, age, and sex	IRT
Sensitivity to change	Minimal clinically important change

DIF = Differential Item Functioning; CCC = Concordance Correlation Coefficient; IRT = Item Response Theory

generating validated versions of the MDS-UPDRS in different languages. As of June 2014, there were 9 official non-English translations with an additional 11 translations in process.

The MDS-UPDRS is copyrighted by the IPMDS and is available on its website (www.movementdisorders.org).

Motor Neuron Diseases: Clinical Trial Measures in Amyotrophic Lateral Sclerosis

Douglas Kerr MD, PhD

Amyotrophic lateral sclerosis (ALS) is a heterogeneous and relentlessly progressive disorder that affects multiple aspects of motor function, including respiratory function, upper and lower extremity strength, and bulbar functioning. The ALS Functional Rating Scale-Revised (ALSFRS-R [22]) has been the most widely used composite measure of function in ALS over the last 15 years. It was created and revised before current psychometric approaches were widely utilized to evaluate the consistency and appropriateness of the measure. Applying psychometric principles to the ALSFRS-R reveals challenges with this measure. Despite showing good internal consistency and a strong correlation with survival in early validation studies, the ALSFRS-R suffers from certain inadequate metric qualities [23]. Specifically, there is evidence of multidimensionality, which argues against the use of a single summed score to represent disease status and progression. In addition, there is poor rating category functioning on some items, which suggests that there might be problems with question wording and/or numbers of response choices. Of note, several of the items on the ALSFRS-R show a mismatch between the person (disease severity) and item (Fig. 5). This means that the item “misses” capturing disease in some patients and that some of the item responses do not correspond to the disease severity of any patient in the clinical trial. Category response curves show that many items do not occupy a unique and linear spot on the disease severity scale, but rather have disordered or subsumed positions to other responses. Therefore, these items and responses need to be altered to make the instrument stronger.

Additionally, a problematic feature of the ALSFRS-R is a reduced sensitivity to changes occurring in the extremes of the severity continuum. This is typical of a measure in which raw scores are summed from ordinal response, but is also partially derived from the types of items available for patients to endorse. For example, patients with higher functioning may not be able to endorse enough items, meaning that there are likely other domains of functioning that are important and not measured (i.e., cognitive function, pain). This is a problem in clinical trials in which the measurement tool should equally assess function consistently across time and across the severity spectrum.

We therefore undertook a Rasch analysis of the ALSFRS-R using data from a recently completed large, multicenter clinical trial. The results of this analysis underscore the importance of modifying the assessment of function in ALS [22]. Further work is underway to improve the assessment of function in ALS in the following ways:

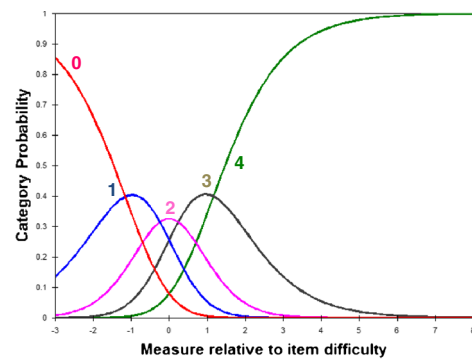
- using fewer response categories on some questions;
- evidence of multidimensionality suggests multiple instruments that can be used for composite scoring;
- consider developing new unidimensional measures;
- bottom-up approach to capture meaningful concepts in ALS functioning and progression;
- address ceiling effects by capturing health domains important to low severity patients;
- create meaningful category responses for each question;
- measure activities of daily living per the standard (outside of this measure).

Another challenge in assessing function in ALS is how to measure function accurately in the context of a significant number of patients dying during the trial. Survival and objective pulmonary functioning are widely considered to be important outcome measures in ALS. What if a drug caused enhanced mortality of late-stage ALS, leaving the healthier patients to measure? An inaccurate conclusion of the study might be that the drug improves functioning in ALS. Conversely, what if a drug prolonged survival, leaving patients with low function to be measured? An inaccurate conclusion of the study might be that the drug worsens function in ALS. One potential solution is to impute function from patients who died during the conduct of the clinical trial. However, imputation rules such as last observation carried forward are inappropriate in a progressive disease. And imputing a low score, such as “0”, for the functional status at the time of death is too punitive, as patients with ALS almost invariably die with ALSFRS-R scores >10 . Therefore, imputation does not solve this problem in ALS.

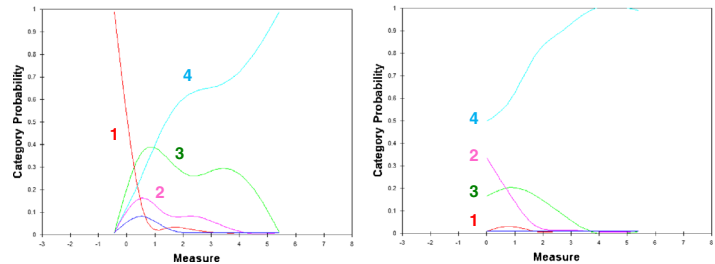
A novel solution to this quandary, the Combined Assessment of Function and Survival (CAFS [24]) (Fig. 6), has recently been utilized in a phase 2 and then a large phase 3 study in ALS [25, 26]. The CAFS is a joint ranking procedure

Fig. 5 Item Mismatch in the Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R)

Example of a “well-behaved” ALSFRS-R item



Examples of “poorly behaved” ALSFRS-R items



in which every patient in the trial is compared with every other in the trial based on mortality or functional decline. CAFS ranks each patient according to their outcome, with the worst outcome assigned to the patient who dies first in the study and the best outcome assigned to the patient who survives with the least functional decline. After establishing the ranking, the treatment groups are unblinded and an assessment is made of the mean rank in each group, with higher being better.

The CAFS analysis is nonparametric and does not rely on statistical assumptions required for many of the standard techniques such as linearity or data imputation. It further combines relevant measures of disease progression in ALS (survival and motor function) and it appropriately accounts for death while measuring function. Ongoing and future studies will further define the utility of this measure in defining ALS disease progression.

Session 2: New Tools to Enable Cross-disease Instrument Development

National Institute of Mental Health Research Domain Criteria Initiative: A Framework for Psychopathology Research

Jill Heemskerk PhD

Current diagnostic schemes [Diagnostic and Statistical Manual (DSM) and International Classification of Diseases (ICD)] are based on clinical symptoms. Clinical symptoms alone (e.g., fever, headache) cannot identify underlying mechanisms to guide treatment development and selection. However, DSM/ICD remain the default standards for disease

classification for research grants, journal publications, clinical trials, and regulatory approval use [1].

On average, a marketed psychiatric drug is efficacious in approximately half of the patients who take it. One reason for this low response rate is the artificial grouping of heterogeneous syndromes with different pathophysiological mechanisms into one disorder. The National Institute of Mental Health Strategic Plan therefore includes the following goal, with its attendant subgoals:

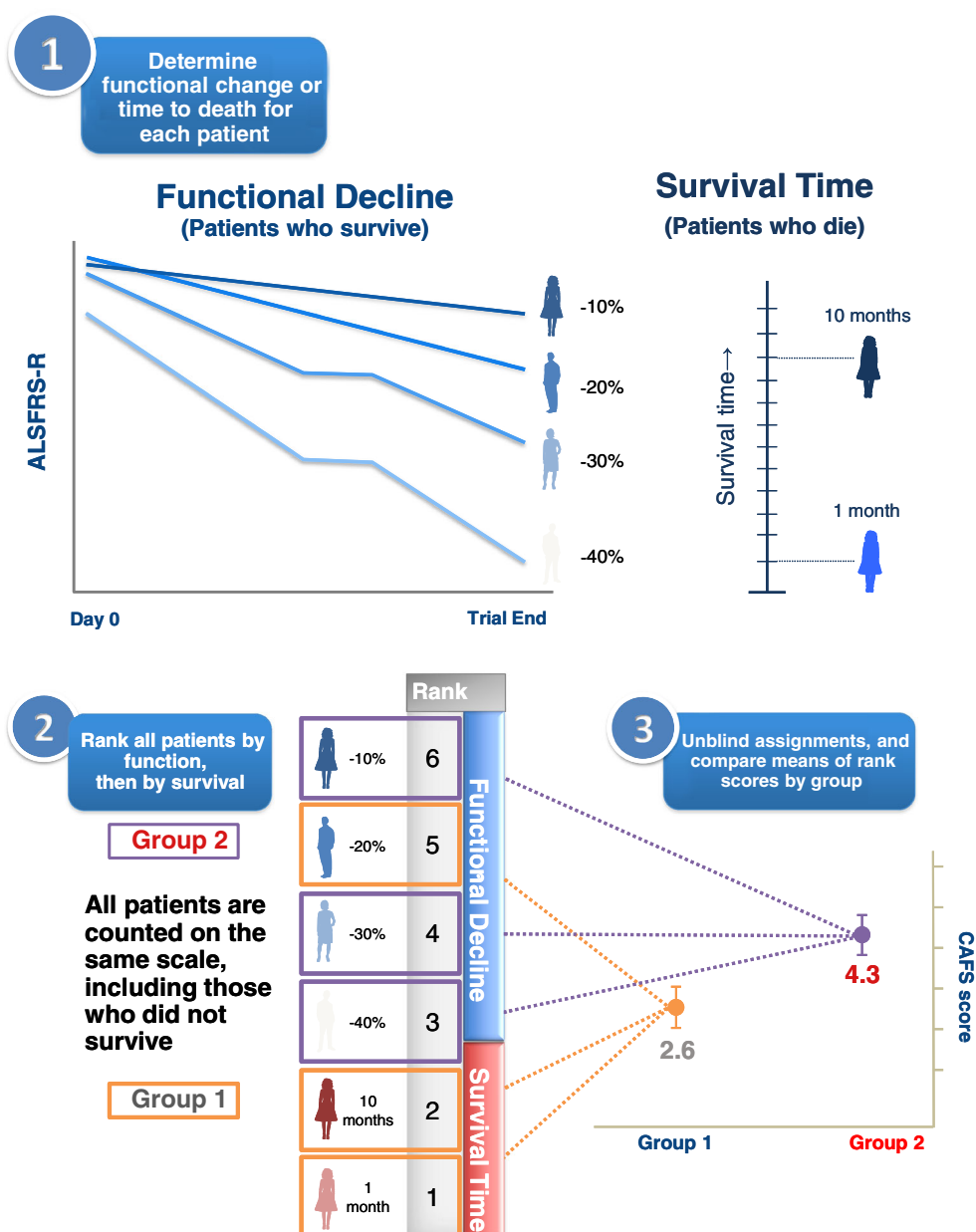
Goal: To develop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures;

1. Identify fundamental components that may span multiple disorders (e.g., executive function, affect regulation)
2. Determine the full range of variation, from normal to abnormal
3. Integrate genetic, neurobiological, behavioral, environmental, and experiential components
4. Develop reliable and valid measures of these fundamental components for use in basic and clinical studies.

The goal was to understand psychiatric dysfunction in terms of biological and behavioral underpinnings. In order to accomplish this aim, 5 workshops were held, focusing on the following themes:

1. Negative valence
2. Positive valence
3. Cognitive systems
4. Systems for social processes
5. Arousal/modulatory systems.

Fig. 6 Combined Assessment of Function and Survival analysis. ALSFRS-R = Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised



The resulting framework matrix for assessment of neuropsychiatric dysfunction domains is illustrated in Fig. 7.

The implications of the research domain criteria (RDoC) framework for clinical trials include, first, the recognition that use of heterogeneous DSM categories confound studies of mechanism in clinical trials [1]. Thus, no 1:1 relationship is expected to exist between a DSM diagnosis and a particular disease mechanism (e.g., not all patients with schizophrenia have cognitive deficits). And, second, a single feature can appear in multiple DSM disorders (e.g., cognitive deficits are seen in some patients with schizophrenia, bipolar disorder, and depression).

In particular, with respect to early-phase clinical trials, the RDoC framework leads to the following considerations:

1. Focus on a novel mechanism relevant to a clinical problem regardless of DSM diagnosis (e.g., anhedonia, working memory)
2. Enroll patients based on deficits in the mechanism, not DSM diagnosis
3. Trial outcomes should reflect the changes in the target mechanism

Finally, it should be recognized that the matrix is evolving; new mechanisms can be proposed for study as new knowledge is acquired and as concepts evolve.

Thus, the RDoC matrix constitutes a truly “translational” approach: disorders are viewed in terms of dysregulation in basic mechanisms rather than as symptom clusters. It therefore

Fig. 7 The research domain criteria matrix

Domain ▶ Construct ▶ Subconstruct	----- Units of Analysis -----							
	Genes	Molecules	Cells	Circuits	Physiology	Behavior	Self-Reports	Paradigms
Negative Valence Systems								
Acute threat ("fear")								
Potential threat ("anxiety")								
Sustained threat								
Loss								
Frustrative nonreward								

represents a framework to study mechanisms that cut across traditional disorder boundaries. Finally, the developers of RDoC believe that its use will inform, not compete with, future versions of DSM and ICD, and it is hoped that the RDoC framework will lead the development of psychiatric therapeutics toward engagement of personalized medicine in psychiatry, consistent with other areas of medicine. Further information on RDoC can be found on the National Institute of Mental Health's website (<http://www.nimh.nih.gov/research-priorities/rdoc/index.shtml>).

The National Institute of Neurological Disorders and Stroke's Common Data Elements Project

Wendy Galpern MD, PhD

The complexities and costs associated with clinical research highlight the need to introduce efficiencies into the research process. However, at present, there are many redundancies and inefficiencies that may impact research progress, as well as cost. Specifically, individual study case report forms (CRFs) are often created and variable definitions may be employed for the same items between different studies. Consequently, study start-up is often delayed, costs are greater than may be necessary, and comparisons between studies can be hampered. In efforts to address such issues, the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health initiated the NINDS Common Data Elements (CDE) Project, which has been a collective effort of hundreds of disease experts around the world aimed at facilitating clinical research across neurological disorders.

The objectives of the NINDS CDE Project are to develop publically accessible standards for data collection and documentation, as well as template CRFs for use in clinical studies. It is anticipated that these efforts will decrease study start-up time and cost, as well as facilitate data sharing and comparisons between studies. The project was initiated in 2007 and, to date, disease-specific CDEs have been developed for a variety

of neurological disorders along with "general" CDEs that are broadly applicable across all studies (Table 5). CDEs are also under development for additional disorders as noted.

The terminology "CDE" refers to a logical unit of data pertaining to one kind of information. Each element has a name, a definition, and a value code, if applicable. The CDEs are classified according to their anticipated use in studies: "General Core CDEs" are those that will be used by most, if not all studies. Elements that would be used by all studies in a particular disease are classified as "Disease-specific Core

Table 5 Available and planned Common Data Elements

Epilepsy
Headache
Mitochondrial disorders*
Movement disorders
• Parkinson's disease
• Huntington's disease
Multiple sclerosis
Spinal cord injury [†]
Stroke
Neuromuscular disorders
• Amyotrophic lateral sclerosis
• Friedreich's ataxia
• Muscular dystrophies
- Becker, congenital*,
-Duchenne, facioscapulohumeral muscular*
-Myotonic dystrophy*
• Myasthenia gravis
Duchenne, facioscapulohumeral muscular
• Spinal muscular atrophy
• Traumatic brain injury

Current disease areas in the NINDS CDE Project. All are publically available unless otherwise indicated

*Planned; [†] Under development

CDEs” whereas those that are commonly, but not always, collected are “Disease-specific Supplemental CDEs”. In order to capture novel elements that are of interest but perhaps not yet validated, a fourth category is also included, “Disease-specific Exploratory Elements”.

The development of the disease-specific CDEs was conducted by a working group of international experts in each disease area. The working groups typically met over a 1-year period to review the elements commonly collected in the disease area, develop a series of recommendations for standardized data collection, and to incorporate feedback. Public comment was solicited prior to finalization of the CDEs, and the CDEs were subsequently posted for public use. The available products include a listing of standardized instruments with explanations regarding recommended use, template CRFs, and data dictionaries that define CDE names, definitions, and permissible values. Importantly, the development is an iterative process. For each disease area, the recommendations are reviewed on a periodic basis to ensure they encompass current research practice.

Efforts have also been undertaken to harmonize the NINDS CDE Project with existing projects and international data standards. While the CDEs will be utilized in NIH-funded clinical research studies, it is anticipated there will be broader applications across the research community. For example, the standards organization, Clinical Data Interchange Standards Consortium, has used our CDEs to develop the Clinical Data Interchange Standards Consortium standard for Parkinson’s disease. It is hoped that this publically accessible resource will increase the efficiency of clinical research by harmonizing data collection across the research community.

All of the products associated with the CDE Project are publically available for use and can be accessed via the NINDS CDE website (www.commondataelements.ninds.nih.gov). For copyrighted instruments, information is provided as to how to obtain the instrument.

Outcome Measures for Clinical Trials: The National Institutes of Health Toolbox

Petra Kaufmann MD, MSc

For many neurological diseases, there is no suitable outcome measure for therapeutics development. In some cases, this is owing to insufficient robustness of the methodology used or to a lack of sufficient datasets on the use of the measurement in the indication of interest. To address these limitations, researchers, at times, resort to the development of new outcome measures. However, creating several new outcome measures for the same indication can create a new set of challenges. First, the number of outcome measures used in a given indication may become large. This can lead to lack of comparability between datasets so that meta-analyses are complicated. Also, using different outcome measures limits opportunities to combine datasets for better trial planning.

This can be particularly problematic in rare diseases. Second, creating outcome measures de novo requires more resources and time than building on existing options.

Therefore, it would be advantageous if clinical researchers built on existing components when seeking improved outcome measures. If these components were methodologically robust, the result could be better outcomes, increased comparability between datasets, and less development time and cost so that progress could be accelerated.

To offer such tools to clinical investigators, the National Institutes of Health (NIH) have promoted a number of initiatives to develop improved and publicly available outcome measures for the assessment of neurological and behavioral function, called the NIH Toolbox. The Toolbox is not a patient-reported outcome measure, but a brief, unified set of measures that allow a clinical researcher to assess function in a testing or examination setting. The Toolbox is meant to support epidemiological studies and clinical trials. Many of the Toolbox measures can be used in pediatric and adult populations because they utilize the same constructs over the lifespan to the extent possible. The Toolbox aims to provide objective measures rather than self-reporting.

The NIH Toolbox has 4 principal domains:

1. Cognition
2. Emotion
3. Motor function
4. Sensation.

For each of these domains, there are subdomains, which are again divided into specific functions. For the Cognitive Domain, for example, the Toolbox includes assessments of Executive Function (Inhibitory Control, Working Memory, Cognitive Flexibility), Episodic Memory (Visual and Auditory), Language (Vocabulary, Comprehension, and Reading Decoding), Processing Speed, and Attention. For the Emotion Domain, the Toolbox assesses Positive Affect (Happiness, Life Satisfaction, and Well-Being), Negative Affect (Sadness, Fear, Anger, General Distress and Apathy), Stress and Coping (Perceived Stress, Coping Strategy, and Coping Self-efficacy), and Social Relationships (Social Support, Social Network, Integration, and Loneliness). For the Motor Domain, the Toolbox includes measures of Endurance, Locomotion, Strength, Dexterity and Balance (Nonvestibular). For the Sensation Domain the Toolbox includes assessments of Olfaction, Taste, Audition, Somatosensation, Vision, and Vestibular Balance.

This structure results in a set of domain-level test batteries that each takes no more than 30 min so that the entire Toolbox can be administered in less than 2 h. English and Spanish versions are available. To encourage innovation, the Toolbox also offers a set of additional, more explorative instruments referred to as “Tool Shed”.

The NIH Toolbox is publicly available online free of charge at www.nihtoolbox.org. The website offers a training manual, administration manual, technical manual, scoring, and interpretation guide, as well as quick reference to accessibility and accommodating special needs. If successful, the use of the NIH Toolbox could serve as a “common currency” of functional assessments across a broad range of datasets with the potential to accelerate clinical research.

Including the Patient’s Voice in Neurotherapeutics Research

David Cella PhD

Increasingly, clinical neurotherapeutics research and, in particular, clinical trial research require reliable, interpretable, and meaningful data obtained directly from patients regarding the efficacy and safety of new treatments. These data have come to be referred to as patient-reported outcomes (PROs). The “source” for PRO data is the patient, and the “score” is typically derived by summing or otherwise summarizing patient responses to sets of standardized health status questions into a defined range of scale values. This is in contrast to clinician-reported outcomes, which are based on the health provider’s assessment of patient status on a range of domains ranging from depression and anxiety to functional status. In addition to these “reported” outcomes are performance measures drawn from patient performance on standardized tests or tasks, and clinical assessments conducted by laboratory, radiographic, or other clinical or diagnostic evaluation. Table 6 places PROs in this larger context of health measures that aid in clinical diagnosis and outcome evaluation.

In order to ensure that the most relevant and important questions are being asked in a PRO questionnaire, one typically embarks upon a rather extensive qualitative research effort. This research can involve focus groups or individual interviews, aimed to elicit the most salient content, as

described by patients with the condition. In the case of general concepts such as pain, fatigue, depression, or independent functioning, the relevant content can usually be elicited from any people with experience in that particular life domain.

Once content has been elicited, and the investigator is confident that the range of relevant content has been saturated, actual questions or questionnaire statements are constructed, typically with multiple choice response options (e.g., from “no problem” to “very much” of a problem, or from “never” to “always” a problem), a second round of qualitative research is typically conducted to ensure that people understand the question to be asking what was originally intended. Often, this is done with the “think aloud” technique, which asks people to verbalize their thought process as they answer the question. By listening to the thought process, the interviewer can determine if the question was read by the patient with the intended meaning.

After constructing the questionnaire, the next step is to determine whether a set of questions can be “scaled” to comprise a single summary score, and to validate the questionnaire for use in clinical research. Recently, and because of its useful measurement properties and flexibility of options, many have used item response theory techniques to scale and validate modern PROs. Such was the case with the Patient Reported Outcomes Measurement Information System and the Neurology Quality of Life initiatives led by our group with funding from the National Institutes of Health. Patient Reported Outcomes Measurement Information System (www.nihpromis.org) and Neurology Quality of Life (www.neuroqol.org) are related PRO measurement systems that together provide more than 80 item banks and scales to measure a wide range of symptoms and functional concerns of people being treated for neurological disorders. In addition, the NIH Toolbox provides a range of performance measures evaluating sensory, cognitive and motor function, as well

Table 6 Patient-reported outcomes (PROs) in the larger context of clinical trial outcomes

Source: patient	Source: clinician/laboratory
PROs	Clinician-reported outcomes
Standardized and scored assessment based on patient self-report, completed alone or with minimal assistance (Examples: SF-36; PROMIS; Neuro-QoL; Toolbox Emotion; MSQLI; PDQ-39)	Assessment based on provider interpretation of patient status (Examples: Ham-D, Karnofsky; EDSS; Rankin/mRS)
Performance-based assessment	Clinical assessment
Objective test measuring patient performance (Examples: reaction time, 6MWT, FEV ₁ ; Toolbox Motor, Sensory and Cognition)	Based on clinical examination and diagnostic evaluation (Examples: disease stage; histology; pathology; EEG; radiography)

SF-36 = ??; PROMIS = Patient Reported Outcomes Measurement Information System; Neuro-QoL = Neurology Quality of Life; MSQLI = Multiple Sclerosis Quality of Life Inventory; PDQ-39 = Parkinson’s Disease Questionnaire - 39 item; 6MWT = 6 Minute Walk Test; FEV₁ = Forced Expiratory Volume (first second); Ham-D = Hamilton Depression Rating Scale; EDSS = Expanded Disability Status Scale; mRS = Modified Rankin Scale; EEG = electroencephalography

as emotional health (www.nihtoolbox.org). These modern patient-based measures, both self-report and performance-based, are available for neuro-therapeutic research going forward. Their use will help build a knowledge base about the patient's direct perspective on the value of neurological treatments being studied in clinical trials.

A Not-So-Final Word

Jesse M. Cedarbaum MD

In his keynote address, Dr. Walton identified the need to establish a Concept of Interest (COI) supporting the development of an outcome measure. The COI should bear a demonstrable relationship to the health benefit intended to be conferred by the therapeutic intervention under study. In some cases existing measures may be adapted to use in new indications or disease areas; in other instances, entirely novel tools will need to be developed. Rigor and a thorough understanding of the intended Context of Use (COU) must be applied in the application of new instruments in clinical trials. Dr. Isaac discussed the regulatory qualification process that has been developed in the European Union. The aim of this procedure is to help assure acceptance of assessment tools by reviewers by describing a framework to the evidence to be provided in support of their use. The US Food and Drug Administration (FDA) has a similar guidance on qualification of outcome measures, and both agencies often attempt to cooperate in their qualification efforts. Diane Stephenson continued the theme of regulatory qualification by describing the activities of the Critical Path Institute, a precompetitive, public-private partnership that was created in 2005 as part of the FDA's Critical Path Initiative. The aim of this organization is to leverage a consortium approach to accelerate the pace and reduce costs of therapeutics development through creation of new data, measurement and methods to aid in evaluation of new therapies. In particular, and with relevance to neurotherapeutics development, the Coalition Against Major Diseases is pursuing development of disease progression models, and regulatory qualification of biomarker and clinical outcome measures for Alzheimer's and Parkinson's diseases in accord with relevant US FDA and European Medicines Agency guidance.

In Session 1, researchers working in different disease areas described a variety of issues common in the development of new Clinical Outcome Assessments (COAs). Richard Rudick kicked off the session on individual neurological diseases by describing the approach being taken to develop a new outcome measure for progressive multiple sclerosis clinical trials by MSOAC, one of the Coalition Against Major Diseases-sponsored consortia. This consortium is leveraging data from 22 studies, performed by multiple pharmaceutical companies and academic investigators, encompassing data from nearly 17,00 patients. The project will develop a new COA, based on

the principles of understanding COI and Context of Use, a corresponding Clinical Data Interchange Standards Consortium data standard utilizing metrics developed within Critical Path. In the end, the goal is to pursue regulatory qualification of the new measure so that its utility is understood by regulators and payors worldwide. Maria Carrillo described similar efforts in Alzheimer's disease, including the challenges of moving the focus of therapy development to early stages of the disease, and in coordinating worldwide efforts to understand disease evolution. She spoke of the variety of instruments currently in development, indicating the lack of consensus in the field, and also of international collaborative efforts to harmonize standards for biomarker collection and analysis by creating a standard biological reference for cerebrospinal fluid biomarker assays to ensure comparability of results between studies and laboratories. Glenn Stebbins described the methodology employed by the Movement Disorder Society to update and validate, as well as to standardize, administration of the Unified Parkinson Disease Rating Scale. In the process of revising and updating the scale, "backwards compatibility" was an important concern. Finally, Doug Kerr presented his team's recent efforts to apply modern clinimetric tools of Rasch and item response theory analysis to understand and address limitations that have arisen in the use of the Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised. He ended by presenting the "Combined Assessment of Function and Survival" analysis, a novel method for incorporating both survival and functional outcomes into a single assessment that minimizes the impact of missing data caused by deaths on study while incorporating functional change in outcome assessment of a uniformly fatal disease.

Session 2 focused on new concepts and tools available to the neuroscience research community. Jill Heemskerk summarized the Research Domain Criteria (RDoC) Initiative, which seeks to replace disease-based nosology with a system of classification based on behavioral symptoms and their underlying neuropsychological and biological domains. Fundamental to the Research Domain Criteria concept is the idea that our current disease constructs lead to failed clinical trials because they are artificial and phenomenologically based. It is hoped that treatments that focus on domains of dysfunction rather than application to nosologically defined disease entities will enhance the success of our therapeutics development efforts by moving us closer to the application of personalized medicine in psychiatry. Wendy Galpern next outlined the National Institute of Neurological Disorders and Stroke's Common Data Elements (CDE) Project. Common Data Elements are logical data units that pertain to information common across studies. These may either be general (e.g., demographics) or disease-specific. Template case report forms and data dictionaries have been developed for a variety of neurological disorders, and are available on the National

Institute of Neurological Disorders and Stroke's website; others are planned or under development. Use of standardized data collection methods should facilitate collaborative work such as that described by Stephenson, Rudick and Carrillo, already employed in Parkinson's disease, and much needed in the development of new instruments for amyotrophic lateral sclerosis. Petra Kaufmann presented the National Institutes of Health Toolbox, which is a unified set of brief outcome measures that allow clinical researchers to assess domains of function, including cognition, emotion, motor function and sensation, using well understood and accepted clinical tests. Training administration and technical manuals, as well as scoring and interpretation guides, have been developed. Use of the NIH Toolbox could serve as a "common currency" of functional assessments across a broad range of datasets with the potential to accelerate clinical research across neurological disorders. The development of the Toolbox represents the first systematic attempt to address the question posed at the beginning of the symposium. The key challenge is having it gain acceptance and for research teams to apply its tools in real-world situations to test the validity of the approach.

As Marc Walton indicated in his keynote speech, identifying the relationship of an outcome measure to clinically meaningful benefit to be gained from a new therapeutic is our ultimate goal in COA development. It was therefore fitting that David Cella closed the symposium with a discussion of "Including the Patient's Voice in Neurotherapeutics Research". Clinician-reported outcomes, which were discussed by all the speakers in Session 1, are but one tool we use in addition to performance-based assessments, diagnostic evaluations and patient-reported outcomes. Methods used for patient-reported outcomes development can and should also be applied to development of novel clinician-reported outcomes. These methods include elicitation of a range of relevant assessment items by qualitative patient/caregiver interviews, determining concept saturation (have we asked all relevant questions?) and defining severity/frequency responses. Item response theory techniques are then applied to ensure appropriate scaling of the measure, as applied to the Amyotrophic Lateral Sclerosis Functional Rating Scale in a post hoc way by Doug Kerr in Session 1. These methods have been used in the development of the Patient Reported Outcomes Measurement Information System and the Neurology Quality of Life measures. These scales are now available online to be applied in neurological clinical trials.

The need for novel effective therapies for neurological and psychiatric disorders grows daily as the burden of neurological disease around the world continues to increase. New tools and scientific understandings emerge almost daily, offering the promise and opportunity for the field of neurotherapeutics to deliver meaningful benefit to patients. At the same time, our growing and evolving perspectives on the clinical and biological manifestations of disease require us to think in new ways

about how we assess the impact of new potential treatments being tested in the clinic, thus creating the need for new, and hopefully more precise and patient-relevant, outcome measures. As we do so, we must leverage systems approaches to understanding how neurological disease manifests. At the same time, we all recognize that our collective resources are limited, that time is of the essence, and that the challenges our patients face in their daily lives share much in common across diseases, regardless of their diagnosis and its ICDM code. Hopefully, the presentations summarized in this paper will stimulate thought and collaboration amongst interested groups that can result in a systematic, harmonized approach to outcomes assessment in neurotherapeutics.

Required Author Forms Disclosure forms provided by the authors are available with the online version of this article.

References

1. Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry* 2014;13:28-35
2. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. U.S. Department of Health and Human Services, Food and Drug Administration, 2009
3. Concept paper on need for revision of the guideline on medicinal products for the treatment of Alzheimer's disease and other dementias EMA/CHMP/617734/2013.
4. Isaac M, Vamvakas S, Abadie E, Jonsson B, Gispen C, Pani L. Qualification opinion of novel methodologies in the predementia stage of Alzheimer's disease: cerebro-spinal-fluid related biomarkers for drugs affecting amyloid burden—regulatory considerations by European Medicines Agency focusing in improving benefit/risk in regulatory trials. *Eur Neuropsychopharmacol.* 2011;21:781-788.
5. Stephenson DT, Avilas A, Bain LJ, et al. Coalition Against Major Diseases: Precompetitive collaborations and regulatory paths to accelerating drug development for neurodegenerative diseases. *Ther Innov Regulat Sci* 2013;47:632-638.
6. Raghavan N, Samtani MN, Famum M, et al.; Alzheimer's Disease Neuroimaging Initiative. The ADAS-Cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement* 2013;9(1 Suppl.):S21-S31.
7. Ropacki M, Hannesdottir K, Hendrix S, et al., on behalf of the Critical Path Institute's Coalition Against Major Diseases and the Patient-Reported Outcome (PRO) Consortium. Consortia driven approach to addressing clinical meaningfulness in early AD. AAIC, Copenhagen, July 2014.
8. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:263-269
9. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association

- workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:270-279
10. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:280-292.
 11. The Movement Disorder Society Task Force on Ratings Scales for Parkinson's Disease. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. *Mov Disord* 2003;18:738-750.
 12. Fahn S, Elton RL, UPDRS program members. Unified Parkinsons Disease Rating Scale. In: Fahn S, Marsden CD, Goldstein M, Calne DB (eds) *Recent developments in Parkinsons disease*, vol 2. Macmillan Healthcare Information, Florham Park, NJ, 1987, pp. 153-163.
 13. Tilley BT, La Pelle NR, Goetz CG, Stebbins GT. Using cognitive pretesting in scale development for Parkinson's disease: the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) example. *J Parkinsons Dis* 2014;1:395-404.
 14. Goetz CG, Tilley BC, Shaftman SR, et al., Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008;23:2129-2170.
 15. Goetz CG, Stebbins GT, Chmura TA, Fahn S, Poewe W, Tanner CM. Teaching program for the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale: (MDS-UPDRS). *Mov Disord* 2010;25:1190-1194.
 16. Martinez-Martin P, Rodriguez-Blazquez C, Alvarez-Sanchez M, et al. Expanded and independent validation of the Movement Disorder Society-Unifed Parkinson's Disease Rating Scale (MDS-UPDRS). *J Neurol* 2013;260:228-236.
 17. Gallagher DA, Goetz CG, Stebbins GT, Lees AJ, Schrag A. Validation of the MDS-UPDRS Part I for nonmotor symptoms in Parkinson's disease. *Mov Disord* 2012;27:79-83.
 18. Martinez-Martin P, Chaudhuri KR, Roho-Abuin JM, et al. Assessing the non-motor symptoms of Parkinson's disease: MDS-UPDRS and MNS Scale. *Eur J Neurol* 2013 Apr 22 [Epub ahead of print].
 19. Rodriguez-Blazquez C, Rojo-Abuin JM, Alvarez-Sanchez M, et al. The MDS-UPDRS Part II (Motor Experiences of Daily Living) resulted useful for assessment of disability in Parkinson's disease. *Parkinsonism Relat Disord* 2013;19:889-893.
 20. Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify Tremor Dominant and Postural Instability Gait Difficulty groups with the MDS-UPDRS: comparison with the UPDRS. *Mov Disord* 2013;28:668-670.
 21. Goetz CG, Stebbins GT, Tilley BC. Calibration of Unified Parkinson's Disease Rating Scale scores to Movement Disorder Society-Unifed Parkinson's Disease Rating Scale scores. *Mov Disord* 2012;27:1239-1242.
 22. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A, BDNF ALS Study Group (Phase III). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *J Neurol Sci* 1999;169:13-21.
 23. Franchignoni, Mora, Giordano et al. Evidence of multidimensionality in the ALSFRS-R Scale: a critical appraisal on its measurement properties using Rasch analysis. *J Neurol Neurosurg Psychiatry* 2013;84:1340-1345.
 24. Berry JD, Miller R, Moore DH, et al. The Combined Assessment of Function and Survival (CAFS): a new endpoint for ALS clinical trials. *Amyotroph Lateral Scler Frontotemporal Degener* 2013:162-168
 25. Cudkovicz ME, van den Berg LH, Shefner JM, et al; EMPOWER investigators. Dexamipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. *Lancet Neurol* 2013;1059-1067.
 26. Cudkovicz M, Bozik ME, Ingersoll EW, et al. The effects of dexamipexole (KNS-760704) in individuals with amyotrophic lateral sclerosis. *Nat Med* 2011;17:1652-1656.