FDA Patient-Focused Drug Development Guidances

Considerations for Trial Readiness in Rare Developmental and Epileptic Encephalopathies

Anne T. Berg, PhD, Natasha N. Ludwig, PhD, Mary Wojnaroski, PhD, Chere A.T. Chapman, MBA, MHSc, Rebecca Hommer, EdD, Gabrielle Conecker, MPH, JayEtta Z. Hecker, MS, and Jenny Downs, PhD

Correspondence
Dr. Berg
a-berg@northwestern.edu

Neurology® 2024;102:e207958. doi:10.1212/WNL.0000000000207958

Abstract

Developmental and epileptic encephalopathies (DEE) are rare, often monogenic neurodevelopmental conditions. Most affected individuals have refractory seizures. All have multiple severe impairments which can be as life-limiting as or more limiting than the seizures themselves. Mechanism- and gene-targeted therapies for these individually rare, genetic conditions hold hope for treatment, amelioration of disease expression, and even cure. The near absence of fit-for-purpose (FFP) clinical outcome assessments (COA) to establish the benefits for nonseizure outcomes of these new therapies in clinical trials poses significant challenges to drug development. The Food and Drug Administration Patient-Focused Drug Development guidance series provides direction for how to overcome these challenges and to ensure FFP measures are available for trials. The goal is to have measures that address outcomes of importance to patients and caregivers, reliably and accurately measure the outcome in the spectrum of abilities for the target disease, and are sensitive to meaningful change over time. The guidances identify 3 primary strategies: (1) directly adopting and implementing available outcome measures; (2) creating measures de novo; and (3) a middle path of adapting or modifying existing measures. Emphasized throughout the guidances is the indispensable and extensive role of the patient or caregiver to assuring the goal of having fit measures is achieved. This review specifically considers the difficulties of adopting available COAs in severely impaired patient groups and ways to adapt or modify existing COAs to be FFP as encouraged in the guidances. Adaptations include alternative scoring, use of assessments in out-of-intended age ranges, and modifications for individuals with sensory or motor impairments. Some additional considerations that may facilitate achieving adequate clinical outcome measures, especially for rare diseases, include use of personalized endpoints, merging of existing COAs, and developing a consortium of rare DEE advocates and researchers to ensure fitness of adapted COAs across multiple rare disease groups. The FDA guidances help ensure that clinical trials targeting nonseizure outcomes, especially in severely impaired populations, will have adequately valid and sensitive outcome measures. This in turn will strengthen the ability of trials to provide informative tests of whether treatments provide meaningful therapeutic efficacy.

Introduction

The 21st Century Cures Act[1] provides the impetus for robust inclusion of and reliance on patient experience data when evaluating clinical benefits of new therapies. The Federal Drug Administration (FDA) Patient-Focused Drug Development (PFDD) guidances were issued in response to the Act to aid researchers and drug developers in designing clinical outcome assessments (COA) that provide optimal tests of a therapy's benefits.

Notably, these guidances focus exclusively on clinical outcomes and emphasize the essential and extensive role of patient-caregiver involvement in most aspects of COA development.[2-5] The

From the Department of Neurology (A.T.B.), Northwestern-Feinberg School of Medicine, Chicago, IL; Decoding Developmental Epilepsies (A.T.B., G.C., J.Z.H.), Washington, DC; Department of Neuropsychology (N.N.L.), Kennedy Krieger Institute; Department of Psychiatry and Behavioral Sciences (N.N.L.), The Johns Hopkins School of Medicine, Baltimore, MD; Department of Psychology (M.W.), Nationwide Children's Hospital; Department of Pediatrics (M.W.), The Ohio State University, Columbus; Ardea Outcomes (C.A.T.C.), Halifax, Nova Scotia, Canada; Connections Beyond Sight and Sound Maryland & DC Deaf-Blind Project (R.H.), University of Maryland, College Park; The Inchstone Project (J.Z.H.); Telethon Kids Institute (J.D.), The University of Western Australia; and Curtin School of Allied Health (J.D.), Curtin University, Perth, Western Australia.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

The Article Processing Charge was funded by the authors.

Glossary

COA = clinical outcome assessments; **COI** = concept of interest; **CVI** = cerebral visual impairment; **DEE** = Developmental and Epileptic Encephalopathies; **FFP** = fit-for-purpose; **FDA** = Federal Drug Administration; **GAS** = goal attainment scaling; **GSV** = growth scale values; **NDD** = neurodevelopmental disorders; **PFDD** = patient-focused drug development; **SMA** = spinal muscular atrophy.

first 2 guidances address methods for obtaining representative stakeholder input[2] and identifying patient priorities.[4] The third guidance, "Selecting, Developing or Modifying Fit-for-Purpose Clinical Outcome Assessments,"[3] provides a roadmap for ensuring a COA is appropriate. The fourth guidance addresses trial endpoints developed from COAs, the outcomes to be tested.[6] Table 1 provides (A) a summary of the guidances and (B) the primary types of COAs in common use and concepts and terms used in the guidances. Here we consider specific challenges to identifying fit-for-purpose (FFP or fit) COAs faced by many recently recognized rare, genetic neurodevelopmental disorders (NDD) and potential solutions to address these challenges and facilitate robust assessment of the clinical benefits of new therapies in line with the FDA guidances.

Precision Medicine and Rare Diseases

In the context of rapid gene discovery and intense interest in precision therapies that target underlying molecular mechanisms, the PFDD guidances ground drug development in the human condition they are intended to improve. In theory, precision medicines should not simply suppress individual disease symptoms (e.g., seizures) but should reregulate underlying pathophysiology, halt disease progression, and potentially reverse salient disease manifestations in a manner that measurably improves patients' lives. Approval from a regulatory body such as the FDA generally requires demonstration of therapeutic efficacy: A therapy must produce better clinical outcomes than control therapies or placebo. Although FDA approval has been granted based on robust surrogate endpoints (biomarkers directly reflecting the well-established pathophysiology of the disease) in the absence of primary clinical outcomes demonstrating an effect,[e1] this is exceptional. Evidence of clinical efficacy is typically required.

For rare diseases, determining meaningful improvement presents particular difficulties.[7,8] These challenges have been faced by several rare neurometabolic and neuromuscular diseases now being treated with gene/mechanism-targeted therapies such as antisense oligonucleotide ("ASO") treatments for amyotrophic lateral sclerosis[e1] and spinal muscular atrophy (SMA),[e2] exon-skipping therapies for Duchenne muscular dystrophy,[e3] gene replacement for adrenoleukodystrophy,[e4] and enzyme replacement for neuronal ceroid lipofuscinosis.[e5] These diseases were recognized as distinct clinical syndromes long before their genetic underpinnings were identified, and there were extensive natural history data often defined by the same or similar COAs ultimately used in clinical trials.

Owing to recent advances in gene discovery, numerous NDDs, which previously were not clinically distinct from each other, are now being recognized as individually rare, monogenic diseases. This represents a shift from clinical syndromes driving gene discovery to gene discovery driving clinical phenotyping. Many NDDs have epilepsy as a prominent component, and the umbrella term "Developmental and Epileptic Encephalopathy" (DEE) is often used for those. Because DEEs are traditionally conceptualized as "epilepsies," seizures have been the primary outcome targeted in clinical trials (e.g., fenfluramine,[9] ganaxalone[10]). Although extremely disruptive, frightening, and dangerous, seizures are not necessarily the only or even primary concern for patients and caregivers. DEEs are also associated with severe to profound cognitive and behavioral morbidities, other neurologic and extraneurologic manifestations, high mortality,[11-13] and life-long dependence on caregivers and often on assistive equipment.[14,15] As genetic diagnostic indications expand, the clinical spectrum associated with any given "DEE gene" now includes individuals who never have epilepsy. This raises special challenges for drug development: Which nonseizure outcomes should be targeted? How will they be measured? What constitutes "better" to patients and caregivers?

Barriers to clinical trial readiness are especially challenging for NDDs and for DEEs in particular. Unlike some rare neurometabolic and neuromuscular diseases that have been recognized and studied for decades, most DEEs are relatively recently recognized through genetic diagnosis and are heterogeneous in expression. There are few if any natural history data to guide selection of nonseizure trial outcomes and no reliable biomarkers. For rare diseases, such data can take years to amass. Furthermore, available COAs often have significant content and psychometric limitations for DEEs. Currently, few if any therapies are approved for nonseizure indications for any DEE. Failure to show clinically meaningful differences for a primary indication could spell the end for a novel therapy. If that happens, it is essential that the lack of measured therapeutic effect be due to a true lack of treatment effect and not to a poor choice of outcome or its measurement.[16]

PFDD guidance #3 is particularly pertinent in this regard for nonseizure outcomes. Here, we (1) highlight key challenges with many COAs when used to study DEEs, (2) discuss strategies for arriving at a fit COA, (3) provide additional suggestions specific to rare DEEs, and (4) consider the role of COAs in defining trial endpoints.

Table 1 Patient-Focused Drug Development Guidances, Key Concepts, and Terms: Text Is Directly From the PFDD Guidances[2-4]

A. Patient experience data and purposes of the patient-focused drug development (PFDD) guidances	
Patient experience data	The Cures Act defines the term "patient experience data" to include data that: (1) are collected by any persons (including patients, family members and caregivers of patients, patient advocacy organizations, disease research foundations, researchers and drug manufacturers); and (2) are intended to provide information about patients' experiences with a disease or condition, including (A) the 'impact (including physical and psychosocial impacts) of such disease or condition or a related therapy or clinical investigation; and (B) patient preferences with respect to treatment of the disease or condition. This expansive definition of "patient experience data" includes a wide range of opportunities for the collection of information that might be used to inform and provide a greater patient focus in medical product development The range of patient experience data that would fit within the Cures Act statutory definition includes: patient registry data, natural history study data, patient focus group or meeting reports, patient survey data, clinical outcome assessment (COA) data collected during clinical trials, and elicited patient preference data.
Guidance 1 (G1)	From whom do you get input, and why? How do you collect the information? Guidance 1 discusses sampling methods that could be used when planning a study to collect patient input. It also provides a general overview of the relationship between potential research question(s) and method(s) when deciding from whom to get input (including defining the target population and development of the sampling strategy).
Guidance 2 (G2)	What do you ask, and why? How do you ask nonleading questions that are well-understood by a wide range of patients and other stakeholders? Guidance 2 [discusses] methods for eliciting information from individuals identified in Guidance 1, gathering information about what aspects of symptoms, impacts of their disease, and other issues are important to patients. It discusses best practices in conducting qualitative research and reference-related resources.
Guidance 3 (G3)	How do you decide what to measure in a clinical trial and select or develop fit-for-purpose COAs? Guidance 3 [addresses] refining the list of concepts of interest important to patients for measurement. Given that not everything identified as important by patients, caregivers, and clinicians can be addressed by an investigational treatment or be measured, this guidance addresses issues related to selecting what to measure in a medical product development program and identification or development of *fit-for-purpose* COAs to assess outcomes of importance to patients.
Guidance 4 (G4)	How do you incorporate a given COA tool or set of measures into a defined clinical study endpoint? How would you define a meaningful change in that endpoint? Guidance 4 [addresses] topics related to incorporating COAs into endpoints for regulatory decision making including COA-related endpoint development, defining meaningful within-patient score changes, and collection, analysis, interpretation, and submission of data.

B. Key concepts and terms in the PFDD guidances[a]	
Common acronym	**Term and definition**
COA	Clinical Outcome Assessment: • A measure that describes or reflects how a patient feels, functions, or survives
PRO	Patient-Reported Outcome: • Reports come directly from the patient • Useful for assessment of symptoms (e.g., pain intensity, shortness of breath), functioning, events, or other aspects of health from the patient's perspective o Examples include any of the PROMIS (Patient Reported Outcome Measure Information System) instruments and other measures of quality of life, pain, stress, fatigue, etc.
ObsRO	Observer-Reported Outcome: • Reports come from someone other than the patient or a health professional (e.g., a parent or caregiver) who has opportunity to observe the patient in everyday life • Useful when patients such as young children cannot reliably report for themselves, or to assess observable aspects related to patients' health (e.g., signs, events, or behaviors) o Examples would include the Vineland Adaptive Behavior Scales or the Aberrant Behavior Checklist, either completed by parents or teachers
ClinRO	Clinician-Reported Outcome: • Reports come from a trained health care professional using clinical judgment • Useful when reports of observable signs, behaviors, clinical events, or other manifestations related to a disease or condition benefit from clinical judgment o Examples include the Gross Motor Function Measure as adapted for trials of SMA or the results of a neurological exam
PerfO	Performance Outcome: • A measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions o Examples include an IQ test or a timed walk test
COI	Concept Of Interest: • The aspect of an individual's experience or clinical, biological, physical, or functional state that the assessment is intended to capture (reflect)
COU	Context Of Use: • Considerations for context of use include (i) how the COA will be used in a trial to support a COA-based endpoint; (ii) target population; (iii) study context; (iv) timing; and (v) implementation

Continued

Table 1 Patient-Focused Drug Development Guidances, Key Concepts, and Terms: Text Is Directly From the PFDD Guidances[2-4] *(continued)*

B. Key concepts and terms in the PFDD guidances[a]

Common acronym	Term and definition
FFP	Fit-For-Purpose: • The level of validation associated with a medical product development tool is sufficient to support its context of use as determined by the strength of the evidence in support of interpreting the COA scores as reflecting the concept of interest within the context of use
RS	Raw Scores: • The unaltered sum of points obtained from each item in an assessment
SS	Standardized Scores: • The conversion of a raw score to a score that allows direct comparisons between individuals and where they stand in the population for which the test was created. Standardized scores (e.g., intelligence quotient (IQ) score or developmental quotient (DQ)) are typically adjusted for age and sometimes other key factors
GSV	Growth Scale Values: • GSV are closely related to raw scores but differ in that each item is differentially weighted to reflect its difficulty based on Rasch analysis and Item Response Theory. In theory, GSVs provide interval scale measures where, e.g., a 10 point difference between 10 and 20 represents the same gain as a 10 point difference between 50 and 60[b]
TRT	Test-Retest Reliability: • The degree to which the same answers and same scores are obtained on repeated application of a measurement (of any kind) on the same individual over a brief period of time during which no change would reasonably be expected to occur. For example, the degree of cortical atrophy seen on 2 MRIs performed a week apart on the same healthy person using the same MRI equipment and protocol
IRR	Inter-Rater Reliability: • The degree to which 2 independent evaluator arrive at the same result when assessing the same set of data. For example, 2 EEG readers agree on the percentage of a tracing that contains spike wave
MCID	Minimal Clinically Important Difference: • The smallest change in a domain (e.g., mobility or expressive communication) that is considered important to patients or caregivers. This may not be the smallest detectable change but a change that has clear impact on function, quality of life or another criterion

[a] For a detailed explanation of these terms, please visit the COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments site at cosmin.nl/.
[b] For a detailed overview of GSVs, please see Daniels MH and Vannier L-C, Growth Scale Value (GSV): Theory, Development, and Applications. [GSV Technical Report 1], NCS Pearson, 2022.

Challenges for Studying DEEs With Available COAs

Concept of Interest May Be Inadequately Captured by a COA

For individuals with severe to profound impairments, fundamental concepts—communication, gross motor skills, sleep, quality of life—may be quite different from what is seen in the typical population. Mismatches between an instrument as used in its intended population and its applicability to a new target population can occur at multiple levels.

Defining the Concept of Interest (COI) in the Target Population

For some populations, a COI may have facets that are not present in the general population. Several pediatric sleep questionnaires were reviewed by parents of children with Dravet syndrome for interpretability, relevance, and completeness.[17] Parents identified that sleep disturbances due to seizures were not assessed. This was considered a significant deficiency of the instruments because frequent nocturnal seizures are highly disruptive and potentially dangerous. Questionnaires developed for use in typical children also consider certain sleep-related behaviors, such as cosleeping, to be "problematic." Parents of children with nocturnal seizures often sleep in the same bed or room with their children to monitor for nighttime seizures. The questions, interpretation, and scoring of such questionnaires require redesign for people with nocturnal seizures.

Understanding the Questions

Fitness also requires that respondents understand the questions and answer choices as intended by the developers (Table 2).[3,18,19] COAs designed for the general population frequently incorporate assumptions about available behavioral repertoire. For example, a question about whether a child "deliberately disobeys" assumes sufficient comprehension for deliberate disobedience to be a possible behavior and the capacity to perform activities that could constitute disobedience (running away) or compliance (cleaning up toys). Parents do their best to interpret the questions and response categories. Whether that translates into interpretable or reliable responses and scores is unclear. Ultimately, scores suggesting "normal" or nonproblematic behavior may be due to limited abilities to manifest the challenging behaviors as assessed in the COA.[20]

Relevance of Content

Many COAs are multidimensional and provide both summary and more specific scores. The composite Vineland

Table 2 Components of a Fit-For-Purpose COA: Text Is From Guidance 3, Table 1.[3,a]

Component
1. "The concept of interest should be assessed by *a specific COA* because [*investigators provide the rationale*]" • Example: Subjective symptoms such as pain or fatigue are best assessed with a patient-reported or occasionally proxy-reported assessment. By contrast, an assessment of gait mechanics is best conducted by expert exam, often assisted with computerized quantification of videos taken under specified conditions
2. "The COA measure selected captures all the important aspects of the concept of interest" • Example: Not including nocturnal seizures in a sleep measure for people with epilepsy may result in poor understanding of apparent sleep disorders
3. "Respondents understand the instructions and items/tasks of the measure as intended by the measure developer" • Example: Questionnaires may query whether a child completes homework without multiple reminders or displays disruptive behaviors such as hitting. If a child is not in an academic school setting and receives no "homework" or cannot purposefully manipulate her hands and arms to hit someone, it is unclear how to answer this essentially double-barreled question
4. "Scores of the COA are not overly influenced by processes/concepts that are not part of the concept of interest" • Example: Communication is typically assessed on the basis of verbal language. In individuals with NDD who are often nonverbal, failure of a COA to accommodate sign language, adapted and alternative communication devices, written communication, and other non-language-based or nonsymbolic forms of communication will result in inadequate or inaccurate assessment of communication ability. Cultural and translation/linguistic validation can be considered part of this general criterion. A question about use of western table utensils would have to be redesigned for cultures in which chop sticks are commonly used
5. "The method of scoring responses to the COA is appropriate for assessing the concept of interest" • Example: Yes-no responses force respondents into absolute categories. For example, "Does your child look at you when you call her name?" How does one respond if this has happened once or twice, or happens sometimes but not usually? Meaningful gradations in response options provides more information and makes the response easier to provide
6. "Scores from the COA correspond to the specific health experience(s) the patient has related to the concept of interest" • Example: A mobility question might query the highest level of function an individual has (from can take no steps at all to ascends and descends stairs unaided). The response on the questions should directly reflect what the individual can do and does in daily life. The gradations between the two ends should be distinctly different, easy to differentiate, and meaningful in terms of patient function
7. "Scores are sufficiently sensitive to reflect clinically meaningful changes within patients over time in the concept of interest within the context of use" • Example: An instrument such as the Functional Motor Scale at 5 yards has 6 discrete levels that range from walking independently without impediment at a typical speed on all terrains to requiring a wheeled device for mobility. Such a scale would be insensitive to changes in someone who is initially unable to sit independently or hold his head upright but who, over time, becomes able to do those things and even stand without support but not yet take steps. The FMS 5 yards score would remain unchanged
8. "Differences in COA scores can be interpreted and communicated clearly in terms of the expected impact on patients' experiences" • Example: COA scores should ultimately translate into whether someone has experienced a clinically meaningful improvement. While a 3-point change in the GSV on the Vineland for communication might not result in noticeable and meaningful improvement in a person's communicative ability, a 10-point gain might reflect noticeable improvement by caregivers in their ability to understand whether an individual is (e.g.) in pain versus is hungry versus desires something specific

Abbreviations: COA = clinical outcome assessment; GSV = growth scale value.
[a] Listed components are those that are likely but not necessarily needed in the rationale for a specific COA, concept of interest, and context of use. Each rationale can be tailored to the proposed interpretation and use. Each component should be accompanied by comprehensive supporting evidence and justification.

Adaptive Behavior Scales ("Vineland") score has domain scores for motor, communication, daily living skills, and socialization. Each contains 2 or 3 subdomains, for example, expressive, receptive, and written for communication.[21] Expressive and receptive communication are important for DEE-affected individuals and their families; written communication is less so as many affected individuals do not have the necessary cognitive or fine motor capacity. Because writing is combined with other scores to yield an overall communication score, changes in expressive or receptive communication may be obscured in the overall score by stagnant writing skills. Including subdomains such as written communication can also be discouraging to parents and increase protocol burden.

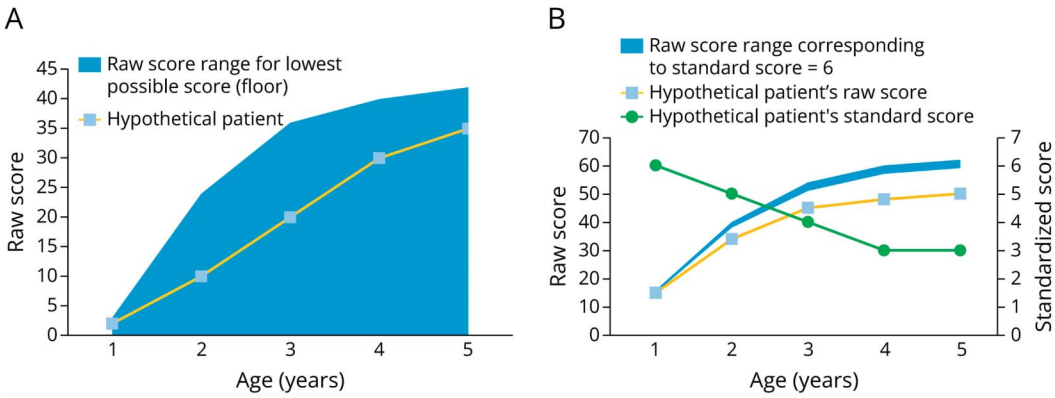
Mismatch Between the COA and the COI

Occasionally, an instrument designed for one COI is misinterpreted or used for another by the investigators. For example, adaptive behavior measures, such as the Vineland, assess skills and behaviors that are used to navigate one's environment in an effective, age-appropriate manner.[21] Adaptive behavior requires both acquiring a skill (development and ability) and using the skill appropriately. Consequently, an individual who can manipulate a fork and knife to cut food but is not allowed to have a knife because of rage attacks might receive credit on a fine motor ability measure but not on a daily living skill adaptive behavior assessment. Ensuring clarity about the purpose and intended use of the COA and re-evaluating it for the COI in the target population is essential to assessing its fitness.

Psychometric Properties of Available COAs Are Often Inadequate in Rare Disease Populations

Fit COAs must be reliable, sensitive to the relevant range of outcomes in the target population, distinguish between individuals with different levels of abilities, and detect changes within an individual over time (Table 1).[3]

Figure 1 Illustrations of Floor Effects and Declining Standardized Scores Despite Real Gains in Skills



(A) For a hypothetical COA, the lowest possible standardized score, 1, is assigned for raw scores of 0–3, 0–24, 0–36, 0–40, and 0–42 for children aged 1, 2, 3, 4, and 5 years (dark blue area). The hypothetical patient has made significant gains from year to year but never registers above the floor of the standardized score. (B) The raw scores required to maintain a given standardized score (SS) over time increase with chronological age. In this example, the dark blue area represents the range of raw scores that would receive a SS of 6 at ages 1 through 5 years. The yellow line represents a hypothetical patient whose raw score at 1 year corresponds to a SS = 6. Although the raw score increases each year, it does not keep up with what is needed to maintain a SS = 6 and the SS falls (green line).

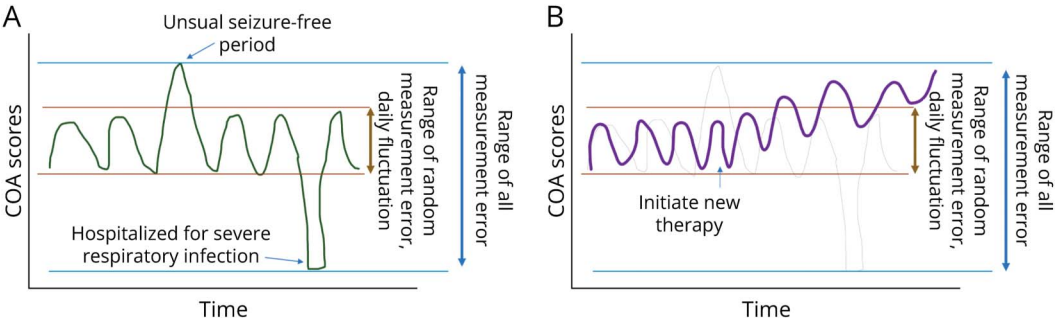
Floor Effects

DEE-affected individuals have moderate to profound impairments. Typically, more than half fall in a range represented by >3 SD below the norm-referenced mean for concepts such as development, IQ, and adaptive behavior[22-24]; only ~1–2/1,000 of the general population are expected to score in this range. Most measures are designed to distinguish above average from high average or low average from below average in the general population and do not have the granularity to make meaningful distinctions at the extremes of their score range. This lack of granularity renders the measure insensitive to real improvements or declines in severely affected individuals and for use in many disease groups (Figure 1A). It may also be a concern for therapies that halt disease progression but do not seem to reverse the disease state based on the same COAs used to assess disease progression.

Age Standardization

Many measures are standardized to chronological age-based normative data under the expectation that older children will have mastered more skills than younger children and therefore need higher raw scores at each age to maintain a given standardized score. In the DEEs, standardized scores decrease with increasing age.[22,25] This can reflect regression of skills over time but often reflects plateauing (failure to acquire new skills) or a slower rate of skill acquisition relative to same-aged

Figure 2 Evaluating Reliability When Systematic Changes in Background Health Status May Increase Measurement Error



(A) Scores on a COA naturally fluctuate within a small range from day to day because of random error or minor variations in the individual (e.g., blood pressure is rarely exactly the same each time it is taken but tends to vary within a small range). Such variation would generally not be considered clinically important. Systematic variation may occur in conjunction with significant events that produce a clear and important effect on the concept measured by the COA. These distort (increase) the true random error of measurement. (B) A new therapy is initiated and evaluated against the backdrop of random and systematic fluctuation. Scores continue to fluctuate but can clearly be shown to increase (depending on how many measures are made and when) above the levels expected for random fluctuation. By contrast, if variation due to systematic errors (significant events that create outliers) is not considered, the change in the COA scores does not measure above the range of measurement error. In psychometrics, the concept of reliability encompasses this concern and can be quantified with a series of measures called intraclass correlation coefficients or ICC.

peers (Figure 1B). Use of age-standardized scores renders many measures unable to detect real changes in severely impaired individuals.

Reliability of Measures vs Stability of Underlying Outcomes

Test-retest reliability is required to ensure a measure can demonstrate true change in the target outcome above and beyond random error (Figure 2). The occurrence of seizures, rescue medications, changes in medications, poor sleep, and illness all have systematic and sometimes large effects on the outcome but may be irrelevant to and interfere with assessment of treatment effect.[26,27] Empirically, we can see that adjustment for systematic perturbations that affect the outcome can improve instrument reliability and sensitivity.[27,28] Understanding these systematic influences can lead to modifications in trial design, analyses, and the COA itself. For caregiver or self-reported surveys, defined qualifiers such as "on a good day" or "typically" can make questions more meaningful and easier to answer. Although seemingly subjective, good questionnaire design helps ensure such terms can be clearly defined and used to elicit meaningful, consistent responses.

Modifying or Adapting COAs

The PFDD guidances present 3 approaches for arriving at a fit COA: (1) adopting an existing COA, (2) modifying or adapting existing instruments, or (3) creating a COA de novo. The previous examples illustrated challenges to wholesale adoption of existing, norm-referenced COAs. Creating new instruments is time-consuming, expensive, and results in a proliferation of redundant measures. Many available COAs are well-respected with extensive track records and rich item banks of relevant behaviors and skills. Modification of existing measures presents efficiencies and when feasible is encouraged.[3,7] Several strategies are already in use.

Alternative Scoring

Raw scores are simply the sum of points an individual achieves on a measure regardless of age. A change or difference in the number of points reflects a change or difference in the COI. Dramatically disparate assessments of difference and change were observed in a cohort of children with *SCN2A*-associated DEE when comparing the same set of Vineland data based on standardized vs raw scores.[22] Growth scale values (GSV) represent a more sophisticated alternative to raw scores[29,30] and are available for at least 3 well-respected neurodevelopmental instruments, the Vineland-3, the Bayley Scales of Infant and Toddler Development 4th edition (Bayley-4), and the Developmental Profile-4. Raw scores and GSVs do not rely on comparisons to same-aged peers. Randomized trials require measures that can identify change within an individual, and the raw and GSV scores seem preferable to standardized scores in this regard.[30]

Out of Age–Range COAs

Regardless of chronological age, most DEE-affected individuals function at levels comparable to infants and very young children. COAs used outside of their intended age range may provide useful measures. When used in older, severely impaired individuals, raw scores showed little floor or ceiling effects for COAs such as the Adaptive Behavior Assessment System (ABAS) 0–5 years[e6] or the Communication and Social Behavior Scalese[e7] developed for children younger than 18 months.[31] Furthermore, scores discriminated well between individuals with clearly different levels of ability and impairment. The Bayley-4 was successfully used out of age range in Angleman syndrome.[24] Although impaired adults are not equivalent to infants and toddlers, for narrowly defined basic communication, motor, and some behavioral functions, the content of COAs intended for infants and toddlers may be transferred with thoughtful modification to severely impaired, older children, adolescents, and even adults.

Modifying for Accessibility

Most individuals with severe cognitive and motor limitations also have significant sensory, motor, and communication impairments that interfere with test administration and hamper interpretation of test results. Cortical/cerebral visual impairment (CVI), a common finding in patients with DEE, may limit access to visual stimuli presented during performance-based testing, thereby compromising validity of results. Modifications of a hand-use COA in consultation with a CVI specialist resulted in adaptations that made fine motor tasks visually accessible for children with *CDKL5*-DEE.[32] Children who use alternative and adaptive communication devices are at a disadvantage with some COAs that inadequately capture ability to communicate other than with spoken language. Motor impairments may affect an individual's ability to respond for measures requiring pointing or manipulation of materials, even for tasks that are not primarily assessing motor skills. Efforts to adapt existing, established COAs include the Mullen for Rett syndrome[33] and the Bayley 3rd edition for children with severe impairments.[34]

Steps Toward COA Trial Readiness for the Rare Diseases and DEEs

The PFDD guidances provide clear steps needed to reach a fit COA. The following points build on those guidances and are intended to reflect needs specific to rare, heterogeneous diseases such as the DEEs.

Include Alternative Types of COAs

The fourth PFDD guidance introduces alternatives to traditional patient-observer–reported and clinician-reported and performance outcome measures (Table 1).[6,8] Personalized endpoints provide a feasible alternative to these traditional COAs, especially in rare diseases that have considerable heterogeneity in expression, severity, and progression.[3] Goal Attainment Scaling (GAS) is one such measure and assesses achievement of a patient's individual goals for a specific intervention. The goals and attainment levels are defined by the patient or caregiver and the clinician together and are therefore both clinically meaningful and patient-relevant. GAS avoids floor and ceiling effects and is increasingly used in randomized trials of, for example, cerebral palsy[35] and dementia.[36] This approach is especially valuable

when existing COAs have content or psychometric limitations. Given the substantial heterogeneity of disease severity and expression among DEE-affected individuals, even with shared etiology, personalized measures like GAS represent a valuable approach for assessing nonseizure outcomes.

Combine COAs to Cover the Outcome Spectrum

The 2019 Rare Disease draft guidance emphasizes the importance of a single COA to cover the full range of possible scores in the concept of interest for the disease.[7] This assures the same COA is used for all patients for a given outcome in a trial, thus optimizing statistical power and streamlining analyses. Most COAs are applicable for a restricted ability or age range. The original Hammersmith Function Motor Measure assessed nonambulatory patients with SMA types 2 and 3. To reduce floor effects, it was modified to incorporate aspects of the Children's Hospital of Philadelphia-Infant Test of NeuroDevelopment (CHOP-INTEND) developed for SMA type-1. To reduce ceiling effects, it incorporated parts of the NorthStar Ambulatory Assessment (NSAA) developed for Duchenne muscular dystrophy. This resulted in the "Revised Hammersmith Scale for SMA."[37]

As COAs move from the neonatal-infant period to childhood, adolescence, and adulthood, they shift from assessing (1) the most basic neurologic functions dependent on muscle tone and strength (e.g., grasping object placed in the hand) to (2) simple developmental skills or milestones (purposefully picking up an object) to (3) more complex skills (manipulating a fork to pick up food and place in the mouth) to (4) adaptive behavior (using utensils and eating appropriately at the table). Measures such as the Vineland (age 0–90 years) include each of these levels but are insufficiently granular at the youngest age range and are always framed as adaptive behaviors (i.e., *does* vs *can*). Supplementing and possibly combining COAs to ensure that the full range of the COI—from supportive neurologic functions to adaptive behavior— is meaningfully included could provide a significant resource for clinical trials.

Consider "Ability-Specific" Rather Than Disease-Specific COAs

COAs are often developed and validated for individual rare diseases.[32,38] Although some disease-distinctive features encourage this practice, many outcome domains that might be targeted in a trial are common across several rare diseases. Rather than pursuing separate COAs for each disease, measures that are ability- or impairment-specific and capture the full range of function across multiple diseases, especially for functioning >3 SD below the mean, could provide great efficiencies for trial readiness in the rare DEE space. For example, the Bayley-3, a well-accepted, frequently used instrument, was modified for individuals with severe impairments.[34] With these modifications, it may be used across disease groups characterized by severe impairments that were not well-assessed by the standard Bayley-3. Validation of the CHOP-INTEND, developed for SMA type 1, in children with X-linked myotubular myopathy and subsequent adaptation for trial purposes exemplifies the efficiencies of using existing impairment-specific instruments when appropriate.[39]

Process of Modifying a COA

While starting with an existing COA provides efficiencies, it still involves iterative steps that require designing, testing, evaluating, redesigning, and retesting until necessary criteria for fitness are met. As exemplified in the SMA literature,[37] achieving a fit COA requires close collaboration among content experts (clinicians, patients, and caregivers), mixed-methods researchers, and psychometricians. If the purpose is for use in an FDA-approved trial, the FDA is the ultimate judge of fitness.

Trial End Points

COAs are the basis for defining trial end points. An essential criterion for a COA is sensitivity to meaningful change in the outcome domain, as defined by parents and caregivers (Guidance-4).[4] Although establishing a trial end point is not a modification or adaptation of a COA, the end points inform COA selection and modification.

Three types of change might be considered: (1) A reliably detectable change is a function of a measure's granularity and reliability (test-retest and inter-rater). (2) "Minimal Clinically Important Difference" (MCID) is the smallest change that has meaningful consequences to the individual or family.[40] MCID is defined at the individual level (e.g., proportion of patients with >50% seizure reduction) rather than group level (average reduction in seizures for a group).[4] The MCID is not a property of the COA but derived from an average estimate of meaningful improvement from patients and caregivers. (3) A worthwhile change reflects a balancing of risks and costs vs potential benefits. This may vary depending on the type of therapy or requirements of the trial, for example, an approved drug repurposed for a new indication vs a novel gene replacement therapy with uncertain risks. A worthwhile change begins with the MCID but may be greater in different circumstances.

Summary/Conclusions

In the context of precision therapy initiatives for the multitude of newly recognized, rare, genetic disorders, the fitness of outcome measures to test efficacy of novel therapies is receiving heightened attention. The DEEs and NDDs more broadly represent clusters of rare diseases that share specific challenges for outcome assessment. Many currently available COAs are not fit for detecting changes in nonseizure clinical outcomes for severely affected individuals. Novel "precision" therapies for these disorders will likely not receive regulatory approval if they cannot reliably demonstrate real, meaningful improvement in the clinical condition. A well-structured process for confirming the importance of a specific COI for a new disease and establishing the appropriateness of a COA with respect to patients' levels of ability could efficiently provide the bulk of the evidence needed to determine suitability of a given COA for that disease. Partnering with the

many advocacy organizations to engage caregivers as an integral part of the COA development process could greatly shorten the path to trial readiness for these rare diseases. The PFDD guidances offer support for development and modification of COAs to ensure their fitness. The flexibility expressed regarding an evidence-based rationale, especially for rare diseases, is encouraging.

Study Funding

The Inchstone Project, Washington, DC.

Disclosure

A.T. Berg serves on the speaker bureau for Biomarin. She reports consulting fees from Encoded, Biohaven, and Biogen. N.L. Ludwig reports no disclosures relevant to the manuscript. M. Wojarnoski reports no disclosures relevant to the manuscript. C.A.T. Chapman reports owning equity in Ardea Outcomes, a privately held contract research organization that receives funding from pharmaceutical, biotech companies, and nonprofit organizations. R. Hommer reports no disclosures relevant to the manuscript. G. Conecker reports no disclosures relevant to the manuscript. J. Hecker reports no disclosures relevant to the manuscript. J. Downs reports consulting fees from Marinus, Acadia, Orion, Taysha, Neurogene, and AveXis. She reports funding for clinical trials with Newron, Anavex. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* May 11, 2023. Accepted in final form September 11, 2023. Submitted and externally peer reviewed. The handling editor was Associate Editor Courtney Wusthoff, MD, MS.

Appendix Authors

Name	Location	Contribution
Anne T. Berg, PhD	Department of Neurology, Northwestern-Feinberg School of Medicine; Chicago, IL and Decoding Developmental Epilepsies, Washington, DC	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
Natasha N. Ludwig, PhD	Department of Neuropsychology, Kennedy Krieger Institute; Department of Psychiatry and Behavioral Sciences, The Johns Hopkins School of Medicine, Baltimore, MD	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
Mary Wojnaroski, PhD	Department of Psychology, Nationwide Children's Hospital; Department of Pediatrics, The Ohio State University, Columbus	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
Chere A.T. Chapman, MBA, MHSc	Ardea Outcomes, Halifax, Nova Scotia, Canada	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
Rebecca Hommer, EdD	Connections Beyond Sight and Sound Maryland & DC Deaf-blind Project, University of Maryland, College Park, MD	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
Gabrielle Conecker, MPH	Decoding Developmental Epilepsies, Washington, DC	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
JayEtta Z. Hecker, MS	Decoding Developmental Epilepsies, Washington, DC	Drafting/revision of the manuscript for content including medical writing for content; study concept or design
Jenny Downs, PhD	Telethon Kids Institute, The University of Western Australia; Curtin School of Allied Health, Curtin University, Perth, Western Australia, Australia	Drafting/revision of the manuscript for content including medical writing for content; study concept or design

References

1. United States Government. *21st Century Cures Act. H.R. 34. 114th Congress*; 2016. Accessed March, 2023. gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf.
2. Federal Drug Administration. *Patient-Focused Drug Development: Collecting Comprehensive and Representative Input Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders*. US Department of Health and Human Services; 2020. Accessed March, 2023. fda.gov/media/139088/download.
3. Food and Drug Administration. *Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-forPurpose Clinical Outcome Assessments Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders (Draft Guidance)*. US Department of Health and Human Services; 2022. Accessed March, 2023. fda.gov/media/159500/download.
4. Food and Drug Administration. *Patient-Focused Drug Development: Methods to Identify What Is Important to Patients. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders*. US Department of Health and Human Services; 2022. Accessed March 17, 2023. fda.gov/media/131230/download.
5. Food and Drug Administration. *Precision Medicine [online]*. Accessed March 17, 2023. fda.gov/medical-devices/in-vitro-diagnostics/precision-medicine.
6. Food and Drug Administration. *Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making, Draft Guidance*. US Department of Health and Human Services; 2023. Accessed April 30, 2023. fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints-regulatory.
7. Food and Drug Administration. *Rare Diseases: Common Issues in Drug Development Guidance for Industry (Draft Guidance)*. Services UDoHaH; 2019.
8. Food and Drug Administration. *Rare Disease Endpoint Advancement Pilot Program [online]*. Accessed June 14, 2023. fda.gov/drugs/development-resources/rare-disease-endpoint-advancement-pilot-program.
9. Lagae L, Sullivan J, Knupp K, et al. Fenfluramine hydrochloride for the treatment of seizures in Dravet syndrome: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2019;394(10216):2243-2254. doi:10.1016/S0140-6736(19)32500-0
10. Knight EMP, Amin S, Bahi-Buisson N, et al. Safety and efficacy of ganaxolone in patients with CDKL5 deficiency disorder: results from the double-blind phase of a randomised, placebo-controlled, phase 3 trial. *Lancet Neurol*. 2022;21(5):417-427. doi:10.1016/S1474-4422(22)00077-1
11. Berg AT, Tarquinio D, Koh S. Early life epilepsies are a comorbidity of developmental brain disorders. *Semin Pediatr Neurol*. 2017;24(4):251-263. doi:10.1016/j.spen.2017.10.008
12. Leonard H, Downs J, Benke TA, Swanson L, Olson H, Demarest S. CDKL5 deficiency disorder: clinical features, diagnosis, and management. *Lancet Neurol*. 2022;21(6):563-576. doi:10.1016/S1474-4422(22)00035-7
13. Talwar D, Hammer MF. SCN8A epilepsy, developmental encephalopathy, and related disorders. *Pediatr Neurol*. 2021;122:76-83. doi:10.1016/j.pediatrneurol.2021.06.011
14. Camfield C, Camfield P. Twenty years after childhood-onset symptomatic generalized epilepsy the social outcome is usually dependency or death: a population-based study. *Dev Med Child Neurol*. 2008;50(11):859-863. doi:10.1111/j.1469-8749.2008.03165.x
15. Gibson PA. Lennox-Gastaut syndrome: impact on the caregivers and families of patients. *J Multidiscip Healthc*. 2014;7:441-448. doi:10.2147/JMDH.S69300
16. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol*. 2007;6(12):1094-1105. doi:10.1016/S1474-4422(07)70290-9

17. Van Nuland A, Ivanenko A, Meskis MA, Villas N, Knupp KG, Berg AT. Sleep in Dravet syndrome: a parent-driven survey. *Seizure*. 2021;85:102-110. doi:10.1016/j.seizure.2020.12.021
18. Downs J, Jacoby P, Leonard H, et al. Psychometric properties of the Quality of Life Inventory-Disability (QI-Disability) measure. *Qual Life Res.* 2019;28(3):783-794. doi:10.1007/s11136-018-2057-3
19. Epstein A, Williams K, Reddihough D, et al. Content validation of the Quality of Life Inventory-Disability. *Child*. 2019;45(5):654-659. doi:10.1111/cch.12691
20. Kaat AJ, Zelko F, Wilkening G, Berg AT. Evaluation of the aberrant behavior checklist for developmental and epileptic encephalopathies. *Epilepsy Behav.* 2021;119:107958. doi:10.1016/j.yebeh.2021.107958
21. Sparrow SS, Cicchetti DV, Saulnier CA. *Vineland-3, Vineland Adaptive Behavior Scales, Third Edition, Manual*. Pearson Psychorp; 2016.
22. Berg AT, Palac H, Wilkening G, Zelko F, Schust Meyer L. SCN2A-developmental and epileptic encephalopathies: challenges to trial-readiness for non-seizure outcomes. *Epilepsia*. 2021;62(1):258-268. doi:10.1111/epi.16750
23. de Lange IM, Gunning B, Sonsma ACM, et al. Influence of contraindicated medication use on cognitive outcome in Dravet syndrome and age at first afebrile seizure as a clinical predictor in SCN1A-related seizure phenotypes. *Epilepsia*. 2018;59(6):1154-1165. doi:10.1111/epi.14191
24. Keute M, Miller MT, Krishnan ML, et al. Angelman syndrome genotypes manifest varying degrees of clinical severity and developmental impairment. *Mol Psychiatry*. 2021;26(7):3625-3633. doi:10.1038/s41380-020-0858-6
25. Nabbout R, Chemaly N, Chipaux M, et al. Encephalopathy in children with Dravet syndrome is not a pure consequence of epilepsy. *Orphanet J rare Dis.* 2013;8:176. doi:10.1186/1750-1172-8-176
26. Berg AT, Kaat AJ, Gaebler-Spira D. Measuring the inch stones for progress: gross motor function in the developmental and epileptic encephalopathies. *Epilepsy Behav.* 2022;137(Pt A):108953. doi:10.1016/j.yebeh.2022.108953
27. Jacoby P, Epstein A, Kim R, et al. Reliability of the Quality of Life Inventory-Disability measure in children with intellectual disability. *J Dev Behav Pediatr.*2020;41(7):534-539. doi:10.1097/DBP.0000000000000815
28. Mensch SM, Echteld MA, Evenhuis HM, Rameckers EAA. Construct validity and responsiveness of Movakic: an instrument for the evaluation of motor abilities in children with severe multiple disabilities. *Res Dev Disabil.* 2016;59:194-201. doi:10.1016/j.ridd.2016.08.012
29. Farmer CA, Kaat AJ, Thurm A, et al. Person ability scores as an alternative to norm-referenced scores as outcome measures in studies of neurodevelopmental disorders. *Am J Intellect Dev Disabil.* 2020;125(6):475-480. doi:10.1352/1944-7558-125.6.475
30. Farmer C, Thurm A, Troy JD, Kaat AJ. Comparing ability and norm-referenced scores as clinical trial outcomes for neurodevelopmental disabilities: a simulation study. *J Neurodev Disord.* 2023;15(1):4. doi:10.1186/s11689-022-09474-6
31. Berg AT, Kaat AJ, Zelko F, Wilkening G. Rare diseases - rare outcomes: assessing communication abilities for the developmental and epileptic encephalopathies. *Epilepsy Behav.* 2022;128:108586. doi:10.1016/j.yebeh.2022.108586
32. Saldaris J, Leonard H, Jacoby P, et al. Initial validation and reliability of the CDKL5 deficiency disorder hand function scale (CDD-Hand). *J Child Neurol.* 2022;37(6):541-547. doi:10.1177/08830738221091044
33. Clarkson T, LeBlanc J, DeGregorio G, et al. Adapting the Mullen scales of early learning for a standardized measure of development in children with Rett syndrome. *Intellect Dev Disabil.* 2017;55(6):419-431. doi:10.1352/1934-9556-55.6.419
34. Visser L, Ruiter SA, van der Meulen BF, Ruijssenaars WA, Timmerman ME. Accommodating the Bayley-III for motor and/or visual impairment: a comparative pilot study. *Pediatr Phys Ther.* 2014;26(1):57-67. doi:10.1097/PEP.0000000000000004
35. Logan B, Jegatheesan D, Viecelli A, Pascoe E, Hubbard R. Goal attainment scaling as an outcome measure for randomised controlled trials: a scoping review. *BMJ Open.*2022;12(7):e063061. doi:10.1136/bmjopen-2022-063061
36. Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *J Clin Epidemiol.* 2003;56(8):736-743. doi:10.1016/s0895-4356(03)00132-x
37. Ramsey D, Scoto M, Mayhew A, et al. Revised Hammersmith Scale for spinal muscular atrophy: a SMA specific clinical outcome assessment tool. *PLoS One*. 2017;12(2):e0172346. doi:10.1371/journal.pone.0172346
38. Hou W, Bhattacharya U, Pradana WA, Tarquinio DC. Assessment of a clinical trial metric for Rett syndrome: critical analysis of the Rett syndrome behavioural questionnaire. *Pediatr Neurol.* 2020;107:48-56. doi:10.1016/j.pediatrneurol.2020.01.009
39. Duong T, Harding G, Mannix S, et al. Use of the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND) in X-linked myotubular myopathy: content validity and psychometric performance. *J Neuromuscul Dis.* 2021;8(1):63-77. doi:10.3233/JND-200479
40. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res.* 2018;27(1):33-40. doi:10.1007/s11136-017-1616-3

eReferences are listed at links.lww.com/WNL/D246.