

Outcome Measures for Clinical Trials with Children with Fragile X Syndrome
November 5 and 6, 2009
Bethesda Marriott Pooks Hill
Bethesda, MD

Executive Summary
Day 1: November 5, 2009

This meeting was sponsored by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), the Office of Rare Disease Research (ORDR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS).

Welcome

Stephen Groft, Pharm.D., Director, ORDR, NIH

Dr. Groft welcomed the participants and gave an overview of the activities of the ORDR. He described the Therapeutics for Rare and Neglected Diseases program, an effort to stimulate drug development even without an early pharmaceutical partner, and noted the success of ORDR's Scientific Conferences program, which has sponsored more than 1,000 conferences. Dr. Groft outlined an approach to coordinating efforts for successful product development and rare disease research that utilizes disease-specific steering committees to bring together a full range of potential partners in government, academia, and the private sector. He noted that products for rare diseases are finally in the pipeline and offered the help and resources of the ORDR.

Introduction and Goals for the Meeting

Tiina K. Urv, Ph.D., Health Scientist Administrator, Intellectual and Developmental Disabilities Branch, Center for Developmental Biology and Perinatal Medicine, NICHD, NIH, HHS

Dr. Urv welcomed the participants and introduced Dr. Linda Brady to discuss the meeting goals.

Fragile X Syndrome

Linda Brady, Ph.D., Director, Division of Neuroscience and Basic Behavioral Science, NIMH, NIH

Dr. Brady reviewed a number of ongoing fragile X syndrome (FXS) clinical trials and noted that with the growing interest in components there is a need to develop clinical endpoints.

Dr. Brady asked the participants to consider whether sufficient core measures and instruments exist to test FXS treatments effectively or whether new measures must be developed. She outlined the charge to the work groups:

1. Develop a conceptual framework to identify the concepts and domains to be measured, the possible applications, and the population.

2. Identify available population-appropriate assessments that correspond with the domains to be measured and can be realistically measured across a variety of clinical settings.
3. Assess the measurements for reliability, validity, and effectiveness.

Dr. Brady stated that the goal for the meeting was to achieve a consensus on endpoints for FXS clinical trials. An end product would be publishing a resource article on the proceedings.

Overview of Measurement Issues—FDA Perspective

Laurie Burke, R.Ph., M.P.H., Food and Drug Administration

Ms. Burke presented an overview of measurement issues from the U.S. Federal Drug Administration (FDA) perspective. She outlined the various types of endpoints in clinical trials, including objective tests (lab or device measurements or performance-based), patient-reported outcome (PRO) measures, caregiver-reported outcomes measures, clinician-reported outcome (ClinRO) measures, or composite endpoints.

Ms. Burke presented an outline of the framework for FDA review of PRO instruments to support labeling claims. She explained that the adequacy of a PRO instrument as a measure to support medical product labeling claims depends on its documented measurement properties that demonstrate that the instrument measures the targeted concept in the targeted context. Data from non-patient observers, such as caregivers, are not PROs because the patient is not the respondent.

Ms. Burke explained the process of FDA instrument review. She emphasized that the FDA can only evaluate an instrument in the context of its intended use (i.e., specific clinical trial or desired labeling claim), so there is no such thing as an instrument that is validated for all uses. The initial and most critical consideration is whether content validity has been established—does the instrument measure the thing that supports the claim? Other measurement properties cannot make up for problems with content validity.

Ms. Burke discussed three FDA review tools:

1. The endpoint model, which is a diagram of the hierarchy of relationships among all endpoints, both PRO and non-PRO, that corresponds to overall study objectives, study design, and the data analysis plan for the clinical trial. It is specific to and defined by the population, the disease, the treatment, and the endpoints chosen.
2. The conceptual framework of a PRO instrument, which is a diagram of instrument items that explains how each item contributes to each concept represented by each score produced by the instrument. It provides a clear description of the relationship among concepts, domains, and items.
3. The PRO instrument dossier, which is submitted with the clinical study protocol and statistical analysis plan (SAP). It includes an extensive array of information, such as instrument and user manuals, targeted claims, endpoint model, conceptual framework, and content validity documentation.

Ms. Burke also gave guidance on health-related quality of life (HRQOL), a regulatory definition. Claiming a statistical and meaningful improvement in HRQOL implies that: (1) all HRQOL

domains important to interpreting change in how the clinical trial population feels or functions as a result of the targeted disease and its treatment were measured, (2) an overall improvement was demonstrated, and (3) no decrement was demonstrated in any domain. Ms. Burke cautioned that HRQOL must be differentiated from quality of life, which would never be a claim.

Discussion: Ms. Burke acknowledged the participants' concern that FXS is a complex disease, which makes it difficult to identify a single measure. However, she emphasized that it is critical to state what the study will measure and ensure that it does measure what is important to the patient population. The key issue is whether the objective test is measuring the right thing.

Working Group Reports—Round One

Cognitive Working Group Presentation

Leonard Abbeduto, Professor, Educational Psychology, University of Wisconsin-Madison Waisman Center, Paul Wang, M.D. (update employment?), David Posey, M.D., M.S., Associate Professor of Psychiatry, Indiana University-Purdue University

Drs. Abbeduto, Wang, and Posey reported that the work group had identified the following cognitive domains and measures for FXS subjects across a wide age range:

- **Language**—includes prelinguistic foundations, speech, vocabulary, syntax, and social language). Measures: parent/informant report, standardized tests, observational, and laboratory-based.
- **Executive function**—includes attention modulation, impulsivity/inhibition, problems managing transitions, perseveration, sequencing problems, and working memory. Measures: parent/teacher report, standardized batteries, laboratory, and clinical.
- **Memory and learning**—includes both explicit and implicit memory. Measures: standardized tests and experimental tasks for assessment.
- **Social cognition**—includes theory of mind and emotion perception. Measures: parent/teacher report (developed for autism, not for FXS) commercial neuropsychological tests, tests reported in the literature for theory of mind and emotion recognition.
- **Visual processing**
- **Motor**
- **Academic achievement**—includes achievement in mathematics, reading, and writing. Measures: standardized tests and tests of functional academics.
- **Adaptive behavior**

Behavior Working Group Presentation

David Hessel, Ph.D., Assistant Professor, Department of Psychiatry and Behavioral Sciences, M.I.N.D. Institute, University of California, Davis

Dr. Hessel reported that the work group identified the following domains and measures for clinical trials for a population of males and females with FXS full mutation, ages 5 years through adulthood with functional levels from severe intellectual disability through normal IQ.

- **Adaptive behavior**

- **Aggression/irritability**
- **Anxiety**—Measures: ADAMS, PARS, CASI, BASC, CBCL
- **Attention**—SNAP-IV, parent rating
- **Hyperactivity/impulsivity**—Measures: SNAP-IV, parent rating, ABC
- **Repetitive/compulsive behavior**
- **Self-injurious behavior**
- **Sleep** (shared with medical/physical work group)
- **Social avoidance**
- **Social reciprocity**

Dr. Hessel presented an informal “top 10” aberrant behavior FXS checklist ($N = 274$) of behaviors that were rated as most severe by parents in his lab. He suggested that many of the researchers present have data sets that could be reviewed to mine similar data and share information:

1. Easily distractible
2. Impulsive acts without thinking
3. Repetitive speech
4. Restless, unable to sit still
5. Repeats a word or phrase over and over
6. Has temper outbursts or tantrums (1)
7. Does not pay attention to
8. Temper tantrums outbursts (2)
9. Mood changes quickly
10. Talks excessively

Medical/Physical Working Group Presentation

Allan Reiss, M.D., Professor of Psychiatry, Director, Center for University School of Medicine for Interdisciplinary Brain Sciences Research, Department of Psychiatry and Behavioral Studies, Stanford

Dr. Reiss reported that the work group identified the following domains and measures:

- **Neuroimaging**—MRI, DTI, MRS, fMRI, and PET
- **Peripheral measures**—EEG
- **Psychophysiology**—Measures: PPI, eye blink conditioning, vagal tone, and electrodermal response
- **Physical** (side effects and motor skills)

Potential biomarkers in humans for FXS could include extracellular-signal-regulated kinase (ERK) phosphorylation, amyloid precursor protein (APP), and salivary cortisol.

Discussion on work group reports: Participants discussed outcomes measures, noting that few performance measures are validated for FXS. It is also challenging to get this population to perform tests. Participants debated the relative value of the informant report and concluded that it is an important measure but must be corroborated by other sources of information. It was also suggested that perhaps the number of measures should be streamlined because multiple measures

are a high burden in clinical trials with the FXS population. It might be more efficient to create a scale of three or four questions with well-defined anchors that deal with issues that parents see over and over, which could be administered in a half hour instead of 2 hours.

Participants were very interested in Dr. Hessler's "top 10" aberrant behaviors, and some urged researchers to go back to their files and do some data mining to see if there are indications that any of these behaviors are improved with treatment.

Dr. Reiss expanded on the use of MRI in a multisite study, saying investigators must vigilantly address calibration and quality assurance issues over the life of the study. He pointed to the potential predictive capability of imaging, noting that imaging in stroke victims has predicted disease progress or the likelihood of response to therapy. Imaging findings also predate some changes in cognitive function, so there is potential for using imaging as a predictor of change. Moreover, approximately 20% of FXS patients have seizures, so EEG has great potential to show areas that could be improved by medication.

Working Lunch Discussion

Ms. Burke noted the three domains identified so far (cognitive, behavior, and medical/physical) and suggested that the work groups must consolidate these into an overarching definition of FXS and identify what can be confirmed.

Dr. Ni Khin, also from the FDA, agreed that it is necessary to define in detail the clinical target and come to a consensus on what that target is in this patient population. She suggested that the schizophrenia MATRICS workshop would be a helpful model for developing guidelines for an FXS clinical trial. It is also important to ensure that the research can translate and culturally adapt in other countries. In terms of clinical trial issues, Dr. Khin suggested considering comorbidities in this population and the drugs that may be taken for them as well as thinking about the duration of the trial in relation to how quickly results can be seen. She emphasized that drug development must relate to measures by identifying outcome measures that show enough response.

Participants discussed a possible misalignment between targeted domains in industry and what is heard from the community, suggesting that there is a need to know the concept of greatest value to the population and what to aim drug development toward. Dr. Khin agreed, saying drug development must be very specific with a primary outcome measure and some secondary outcome measures, but not 20 measures.

Participants enumerated hurdles for FXS clinical trials, including: the complex paradigm for FXS that does not represent the usual psychiatric clinical trial, the fact that the FDA demands that the sponsor identify a primary endpoint without much prior knowledge so if the wrong choice is made the trial fails, and the involved FDA process to endorse a biomarker for drug approval.

Participants made suggestions about clinical trial design issues, including: (1) Define outcome measures that are suitable for conducting drug trials for more general issues afflicting *adults* with FSX, (2) limit biomarkers to functional-oriented outcomes in smaller trials to see if they impact the system and phenotype, and (3) consider sleep in FXS patients as a prime area for behavioral intervention. Ms. Burke noted that biomarkers used as outcome measures in psychopharmacological studies had not been part of the FDA approval process.

The participants defined the mission of the meeting as achieving a consensus about specific outcome measures that capture this disorder and are ready for testing right now and the type of studies that would best utilize them. The group might also consider what further development is needed in all domains going forward.

Working Group Breakout Work Sessions—The three working groups met individually from 1:00 p.m. until 2:45 p.m. then reconvened in full session to report on their progress.

Working Group Reports—Round Two

Cognitive Working Group: Dr. Leonard Abbeduto

Dr. Abbeduto stated that the work group prioritized the most important indicators of functional outcomes in children and identified those measures that are ready now versus measures needing more data across age ranges:

- **Language**—Measures include standardized tools for language research. The procedures are well defined, but transcription makes them labor intensive.
- **Memory**—Measures are available for both auditory and visual memory.
- **Executive function**—There are some measures but more are needed.
- **Social cognition**—Measures included the social responsiveness scale (SRS) that parents can complete. Eye tracking is probably not a useful outcome in a very large trial.

Behavior Working Group: Dr. David Hessl

Dr. Hessl reported that the group had narrowed down the number of domains to:

- **Anxiety**
- **Hyperactivity/impulsivity**
- **Attention/inattention**
- **Irritability/tantrums/SIV**
- **Social avoidance**

In terms of specific measures for the domains, the pediatric anxiety rating scale holds promise but needs to be developed in FXS for validation. Parent report should be correlated with other objective ratings. An FXS-specific rating scale should be developed.

Medical/Physical Working Group: Dr. Allan Reiss

Dr. Reiss stated that behaviors like irritability and tantrums may not represent core features of FXS resulting from the disease pathway but rather represent the interaction of environmental factors and non-FMR1 genetic factors. Because many behaviors that present clinically in FXS

patients may represent secondary environmentally shaped behaviors, there is a need for more objective measures.

The group identified the following top five measures that could be employed in small trials and as add-ons in large trials:

1. **ERK** signaling
2. **Salivary cortisol** (easy to do)
3. **Prepulse inhibition (PPI)**

These three have been demonstrated across species.

4. **Vagal tone** (can be done across multiple centers)
5. **Imaging** (has potential for cross-species and behavioral correlates)

The group identified five other measures that are not as ready:

1. EEG
2. APP
3. Brain-derived neurotrophic factor (BDNF)
4. Eye blink
5. Eye track

The group discussed motor skills and side effects. Some felt that motor skills are not ready to be used as a measure, but it could be argued that they are an objective indicator of changes.

Discussion on working group reports: Participants asked about the length of time needed to administer the measures, suggesting a maximum of 40 minutes per session would be optimal for this population. The following time estimates were presented: language samples, 10 to 15 minutes; digit span, 2 to 3 minutes; SRS is completed by parents; and eye tracking requires setup time. Participants agreed that it is possible to see changes within a 4 to 16 week time span.

Participants noted a number of anxiety scales: the Anxiety, Depression, and Mood Scale (ADAMS), which was developed for people with intellectual disorders across a broad age range, PARS, which has been used in clinical trials, and the Childhood Anxiety Sensitivity Index (CASI), which is currently being used in an FXS trial. Ms. Burke opined that because anxiety is well defined in this population, the primary outcome of anxiety could probably be labeled.

Summary of Day 1

Elizabeth M. Berry-Kravis, M.D., Ph.D., Professor, Rush University Medical Center

Dr. Berry-Kravis cited a need for a common set of exploratory-level measures to be consistent across sites in early studies and a need to define what is available and what should be developed for pivotal clinical trials as well as what is needed for this development.

She noted that the work groups flagged some common problems:

- There is a very broad range of function/ability to target.
- Patients' problems are not simple deficits; the problems have complex overlap.

- There is a lack of information on response to interventions.
- There is a lack of information on how measures correlate to the most important clinical problems, such as quality of life.
- There is a need to differentiate tools that are useful to explore phenotypes and response profiles versus potential primary endpoints.

The individual work groups identified a number of target areas and potential measures. Overlapping domains targeted by the cognitive and behavioral work groups were (measures in parentheses): cognitive impairment/poor ability to learn, executive function/attention (KiTAP, contingency naming), memory (Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)), language (language sampling), social cognition/social withdrawal (SRS, eye tracking), hyperactivity, anxiety, and irritability/aggression.

The medical/physical group targeted translational pathway measures (ERK), systemic biomarkers (cortisol), biophysical measures (PPI, vagal tone), imaging (fMRI, PET) sleep and side effects/motor (Safety Monitoring Uniform Research Form (SMURF)).

Day 2: November 6, 2009

Drs. Urv and Beckel-Mitchener presented a series of questions to focus the discussion prior to the final working group breakout work sessions:

Diagnosis

1. Using the endpoint model presented by the FDA, what is the conceptual framework and content validity (evidence) that defines FXS as a unique phenotype(s)?
2. What are the unique features?
3. How do the domains identified by the working groups map on to the unique characteristics of the phenotype?

Clinical State and Symptom Severity

1. Can we isolate phenotypes/functional domains that are core features associated with the pathophysiology of FXS?
2. Is it possible to isolate specific temporal changes in symptom domains during the course of the disorder?

Paradigm

1. Do other diseases (such as autism?) provide a model that can inform development of outcome measures for FXS clinical trials? Or,
2. Does FXS need to modify existing measures or develop new endpoint measures/biomarkers to assess disease course and treatment intervention specific to FXS?

Working Group Breakout Work Sessions—The three working groups met individually and then reconvened in full session to make their final reports.

Summary of Working Group Final Reports:

Cognitive Working Group: Dr. Leonard Abbeduto

Dr. Abbeduto enumerated the final constructs that the working group targeted. Measures already exist for each of these constructs and are noted along with the approximate administration time:

1. **Language** (grammatical complexity, frequency/pace, and perseveration)
 - a. Measure for all of above: Standardized language sample (10–20 minutes)
2. **Executive function**
 - a. Problem solving and planning—Woodcock Johnson (W-J) planning test (5–10 minutes)
 - b. Inhibition—Contingency naming, KiTAP (20 minutes for all basic tests)
 - c. Processing speed—W-J rapid naming (5 minutes)
3. **Working memory** (auditory and visual)
 - a. Measures include W-J auditory working memory and digits reverse (10 minutes) and Corse block or equivalent (5 minutes)
4. **Social cognition**
 - a. Measures include SRS (by parent, 15 minutes) and eye tracking in emotion recognition (10 minutes)
5. **Learning**
 - a. Object memory—Lab-based, Cambridge Neuropsychological Automated Testing Battery (CANTAB) (5–10 minutes)
 - b. List learning—RBANS, substest Child Learning Competency Test (CLCT) (5–15 minutes)
 - c. Fast mapping—New word learning (10 minutes)
6. **Level of cognitive function**
 - a. Measures include short form of Stanford-Binet-V (SB-V) (10 minutes) and Vineland Adaptive Behavior Scale (interview or informant report)

Behavior Working Group: Dr. David Hessel

Dr. Hessel enumerated the domains and measures that the working group targeted:

1. **Hyperactivity/impulsivity**—Aberrant Behavior Checklist (ABC) Hyperactivity Scale (1st line tool)
2. **Aggression/irritability/self-injurious behavior (SIB)**—ABC Irritability Scale (1st line tool)
3. **Inattention and hyperactivity**—Swanson, Nolan and Pelham-IV (SNAP IV) (1st line tool)
4. **Anxiety**—PARS (a less clear measure)
5. **Social avoidance**—SRS, ABC Withdrawal Scale, Matson Messier, Clinical Global Impression (CGI) (less clear measures)

Dr. Hessel noted that some things are still needed, including: (1) better anchors of severity and frequency in terms of how well they track behavior and differ from parent reports, (2) correlations with actual behavior ratings to rule out rater bias, (3) more work on anxiety scale testing and validation, (4) evaluations of measures for sensitivity in FXS, and (5) more information on how much overlap there is between the constructs of hyperarousal and anxiety.

Dr. Hessler suggested two future projects:

1. Combine ABC data across sites to get a true ranking of the most common problems (at least those within this scale).
2. Develop a FXS-specific rating scale.

Medical/Physical Working Group: Dr. Allan Reiss

Dr. Reiss stated that the work group members agreed that side effects must be measured across studies. There was also consensus that sleep should be measured using standard brief questionnaires since sleep disruption is common in FXS and studying sleep would be both low-cost and low-burden. Ms. Burke noted that there are specific issues that must be considered in sleep assessment.

The work group addressed the uniqueness issue, asking what sets FXS apart. There are no validated biomarkers of biological measures of protein expression at this time. Validating ERK and PPI measures as outcomes responsive to drug treatment should be considered.

The work group agreed that of the core behavioral features of FXS, measuring hyperarousal with cortisol and vagal tone holds the best promise as tools that can be used right now, although as secondary measures. Both correlated with the construct of hyperarousal. PPI has not yet been shown to be correlated with hyperarousal, although it is highly correlated with IQ.

Discussion: Participants noted that the behavioral 1st line measures have a strong track record in many populations. Dr. Hessler noted that the work group did not discuss quality of life. Ms. Burke stated that it is very difficult in a regulatory setting to claim improvement in HRQOL.

The population for the clinical trials was discussed. Participants noted that thought would have to be given to whether adults in group homes would be eligible, because of the difficulty in finding qualified informants.

Ms. Burke asked the participants to keep the endpoint model concept in mind. She cautioned that although all the issues being discussed are important, the group must identify what endpoint used in the intervention will demonstrate treatment benefit for FDA approval. The focus must be on what to measure that will show that the drug works in this patient population, and this requires much more specificity. Participants felt that the domains that were identified and the behaviors within them had objective bases and were chosen because imaging studies support many of those constructs. It was noted that from the standpoint of uniqueness, it is possible to differentiate the FXS brain with MRI with over 95% sensitivity, so that might be one of the best measures for use right now. Participants wondered what percentage of FXS patients are able to tolerate the imaging session. Dr. Reiss said that if the patients are properly prepared with videos, listening to MRI sounds, and if possible, practicing in a mock scanner, about 60% can do it. Only about 2% of 1- to 3-year-olds can tolerate imaging, so they must be sedated.

Concept Framework and Content Validity of FXS Endpoint Model

Dr. Brady asked the group to look at the domains and consider how they map on to FXS and how the measures are validated for FXS.

Dr. Abbeduto felt that the cognitive domains all map on to the phenotype. Some measurements are not relevant for high-functioning women, and others cannot be completed by some portion of the FXS population. There is no one measurement per domain that will work for everyone.

Dr. Hessel stated that the behavioral domains map on very well based on information from clinicians, researchers, and parents. There are also increasing data to validate many measures. However, there are more domains than measures and the measures listed were not developed solely for FXS. There is a need for a new scale that hits the FXS mark more fully, particularly if the goal is to modify the FXS phenotype.

Ms. Burke noted that the FDA is working with the NIH and other consortia to qualify tools for use in clinical trial and will publish guidance on this within the next few months.

Final Discussion:

Ms. Burke was asked whether the FDA needs a separate endpoint for each factor that plays a part in the complex model of FXS. She stated that it would be unlikely that the FDA would approve, for example, irritability without knowing that each patient in the trial has FXS. She acknowledged that there is not clear guidance on this but emphasized the key to getting the model right is whether one can conclude that this patient is benefitting from this treatment.

A participant questioned whether a drug can impact a learning disorder. Noting that in autism the only treatment is early intensive behavioral intervention, the participant questioned the focus on drug therapy when there must be behavioral treatment at an early age. Other participants agreed and suggested working on a background form to determine what other interventions FXS patients are getting through family, school, or behavioral interventions so it is not mistakenly concluded that the drug is working when it's actually a great teacher. Another participant added that families have usually done everything they can with behavioral interventions but also need medications to put the brain in a better place for learning. It was noted that at this time no behavioral treatment is effective for FXS and participants encouraged NICHD to identify behavioral interventions that are most efficacious in other disorders (e.g., autism) and try them first with FXS patients.

In terms of developing an FXS-specific measure, the representative from the National Database for Autism Research (NDAR) encouraged the group to use NDAR as a resource.

Ms. Burke suggested that the FXS model should be three-dimensional because the impacts of different treatments must be considered, and this is something that the FDA considers.

Next Steps:

1. Publish a paper on the results of this meeting to be used as a resource.

2. Develop a steering committee comprised of the leadership of each work group to move forward to develop a more overarching FXS endpoint model.
3. Consider development of an FXS-specific measure.