

Outcome Measures for Clinical Trials in Individuals with Down Syndrome

April 23–24, 2015
Bethesda, Maryland

Welcome

Catherine Spong, M.D., NICHD

Dr. Spong welcomed the meeting attendees and noted that research related to Down syndrome (DS) and intellectual and development disabilities (IDDs) is essential to the NICHD mission.

Last fall, the National Institutes of Health (NIH) published a DS research plan, calling for new clinical trials. The attendees of this meeting would help define the outcome measures for the clinical trials and help shape future NIH research.

Dr. Spong said that she looked forward to their report.

Introduction and Goals for the Meeting

Tiina Urv, Ph.D., NICHD

Dr. Urv said that one of the problems in the DS field is the lack of outcome measures for clinical trials. There are a variety of tests for IDDs available, but are they appropriate for use in DS clinical trials?

The purpose of this meeting was to identify instruments that can assess DS clinical trial outcomes, either pharmaceutical or behavioral. The meeting participants had begun this work in February via teleconferences.

Dr. Urv outlined the steps the group was taking to identify outcome measures for DS clinical trials:

1. Establish working groups with members from a variety of backgrounds, including clinicians and clinical trialists, and people with varying expertise, including psychometrics, pharmaceuticals, and advocates.
2. Identify the concepts and domains to be measured.
3. Identify the assessments that are currently available.
4. Assess the potential instruments for reliability, validity, and ability to detect change in the domain to be measured.
5. Rate the existing measures. Identify measures that are ready to use, those that would need modification before they could be used, and those that should not be used.

The participants in this project already have begun working with the Food and Drug Administration (FDA) so that they understand FDA requirements, and the FDA understands what the group wants to achieve. The model for this work is the outcome measures effort that the NICHD conducted on Fragile X syndrome.

Overview of Measurement Issues—the FDA Perspective

Michelle Campbell, Ph.D., FDA

Dr. Campbell is a member of a study endpoints team in the Office of New Drugs, which provides consultation and advice on clinical outcome assessment (COA) development. Her office manages the COA qualification program for the FDA's Center for Drug Evaluation and Research.

Patient and parent or caregiver input is needed to ensure that the time, energy, and money spent on research is of interest to the patient population. The earlier the target population is involved, the better the measurement instrument is likely to be.

The FDA uses outcome assessments to determine whether or not a drug has been shown to provide benefit to patients. The four types of COAs are performance, clinician-reported, observer-reported, and patient-reported outcome measures. When a population can report outcomes for themselves, a patient-reported outcome assessment would be used. If clinical judgment is required to interpret an observation, a clinician-reported outcome assessment is chosen. If a population cannot report for themselves, an observer-reported outcome assessment is appropriate. In some cases, observing functional performance in the clinic is useful, and in this case a performance outcome assessment may be appropriate.

The FDA seeks to evaluate treatment benefit for patients, weighing the benefits with known risks of the product to make drug approval and labeling decisions. Direct evidence of treatment benefit is derived from studies with endpoints that measure survival or how patients feel and function in daily life. Indirect evidence of treatment benefit is derived from studies with endpoints that measure other things that are related to how patients survive, feel, or function.

For drug approval and labeling, it is necessary to consider the relationship between the outcome measures and the disease and treatment. Outcomes may be closely related to the condition and treatment, such as disease symptoms, or more distantly related to the condition, such as health-related quality of life, including psychological, physical, and social functioning. The more distant the outcome is from the treatment, the harder it is to connect the treatment to the outcome.

When distant concepts are used to assess clinical trial outcomes, the investigators must measure the variables that contribute to those outcomes. For example, to measure health-related quality of life, it is necessary to assess symptoms, adverse events and toxicities, and all the variables that could contribute to quality of life.

Measuring distant concepts is difficult in patients with multiple comorbidities that could affect quality of life. The FDA recommends choosing endpoints that are more closely related to the condition and treatment rather than broad concepts that are harder to measure, interpret, and show a treatment effect.

Drug developers must document evidence of treatment benefit using well-controlled clinical trials. The methods of assessment of a subject's response must be well-defined and reliable. "Well-defined and reliable" are the key criteria by which the FDA evaluates outcome assessments.

When the FDA evaluates whether a patient-reported assessment is well-defined and reliable, the agency first evaluates the tool's content validity. After content validity is established, the FDA considers construct validity, reliability, ability to detect change, and guidelines for interpreting meaningful change.

How might an investigator choose an acceptable instrument? Dr. Campbell showed a schematic roadmap to patient-focused outcome measurement in clinical trials and described it as a tool to help drug developers think through the issues when planning drug development programs.

The FDA recommends understanding the condition and conceptualizing the treatment benefit before selecting an outcome measure. The FDA encourages drug developers to discuss outcome assessments with FDA staff members early in development. The FDA can provide advice to help improve the likelihood of success.

When FDA staff members review an instrument's content validity, they find out what the instrument measures and determines whether that is the right thing to measure in that population. For example, when all the items from a questionnaire are combined into one score, what does that score represent?

Regulators emphasize content validity to ensure that the meaning of a change in score is clear and that the score change represents a meaningful treatment benefit. A very small change in score could be statistically significant but not clinically important or meaningful to the patient population.

It is the job of the FDA to weigh risks and benefits in order to make approval decisions. The FDA needs to know what patients view as meaningful change on various assessment tools and how they weigh the benefits and risks.

The FDA provides patient-reported guidance to help evaluate whether a measurement is well-defined and reliable. This guidance was developed for patient-reported outcome assessments, but many of the principles are appropriate to apply to any COA type.

Not all assessment instruments are appropriate for use as outcome assessments in clinical trials. An instrument may be useful in clinical practice with individual patients but not at the population level, making the instrument inappropriate for drug approval or labeling decisions. A patient-reported outcome questionnaire that has only one total score may not be appropriate because it may be unclear what is driving a score change. The FDA encourages investigators to use instruments that provide separate scores for specific domains.

Developing instruments specific to patients with a neurodevelopmental disability allows for more accurate measures to be used on the specific population. But developing instruments for a specific patient population, such as for DS, can be challenging. For example, it could be a challenge to develop an instrument to be used across multiple cognition levels. Would the instructions need to be modified for patients to understand?

Incorporating different perspectives is another measurement issue. Whose responses are the most important—clinicians', patients', or caregivers'? Would each group understand the questions in the same way?

There are often multiple research studies actively recruiting for participants. DS individuals and their parents and caregivers may be overburdened with the number of questions they are asked in a single research study and the number of research studies they are asked to participate in.

The FDA discourages the use of proxy-reported outcomes measures, such as parents reporting as if they were the child. It is best for observers to report observable behaviors when self-reporting is not possible. For example, a parent should not be asked to rate a child's pain. The parent should be asked about observable behavior such as crying or holding a body part.

When developing an instrument for a wide age range, use the lowest age range of understanding for the instrument.

It is important to train raters and provide instruction manuals to minimize inter-rater variability in clinical trials.

The FDA can work with stakeholders to develop clinical outcome assessments. The traditional way is within an individual drug development program. FDA encourages investigators to begin discussions with the FDA in the pre-Investigational New Drug stage. The FDA can also work with stakeholders within the Drug Development Tool (DDT) COA qualification process. The FDA can work with stakeholders to develop and quantify publicly available outcome assessment tools.

Through the DDT, the FDA has issued qualification guidance for industry that describes the qualification process for biomarkers, animal models, and COAs. These are not evidentiary standards but show that FDA agrees with the content and measurement properties of a particular instrument.

Dr. Campbell provided some links that could be helpful:

- The FDA's Patient-Reported Outcome Guidance for Industry
<http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071975.pdf>
- The DDT COA Qualification Program
<http://www.fda.gov/drugs/developmentapprovalprocess/drugdevelopmenttoolsqualificationprogram/ucm284077.htm>
- The FDA's DDT Qualification Program Guidance for Industry
<http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm230597.pdf>

Discussion

Paul Wang, M.D., of Autism Speaks, said that there will be endpoints that will be valuable to include in studies but may not meet the FDA endpoint criteria. Investigators use inputs such as patient self-rating their depression or parents rating the degree of attention deficit and hyperactivity. It would be good if DS investigators had a quantitative endpoint akin to the 6-minute walk that multiple sclerosis investigators have, but currently that is not the case.

Surrogate markers can predict a clinically meaningful outcome even though they are not meaningful by themselves. It is valuable to know a cholesterol level because it is a surrogate for the risk of cardiovascular disease. There may be surrogates that could be used in DS, but they would have to have a close link to clinical outcomes.

It is important to be able to use subgroups of the DS population. Some interventions may not be appropriate for the whole range of people with DS but may be appropriate for subgroups. There is precedent for approving a treatment or outcome for a subpopulation, such as drugs that are available to treat mild dementia.

Dr. Campbell agreed that it is important to allow for subpopulations. She also noted that the use of biomarkers often is of interest to investigators. The FDA could arrange a webinar on biomarkers if there were interest.

Dr. Wang also noted that information about the way a drug was tested can be included in the label. He gave the example of a drug to treat attention deficit hyperactivity disorder (ADHD) that was tested in a "fake" classroom giving the subjects an arithmetic test. This is a common technique in drug development and is often included in the approved label, but the test is not the sole basis of approval.

Dr. Campbell said that Dr. Wang raises the issue of the hierarchy of endpoints, which can be primary or secondary, exploratory or descriptive. The hierarchy is determined by the target population and what the investigator is trying to treat.

Michael Ropacki, Ph.D., of Janssen Research and Development, said that it is a lengthy process to develop norms for a new test and to have the test approved for use for DS. Which would be the faster route to FDA approval, using the DDT COA Qualification Program or the more traditional route of a drug development program?

Dr. Campbell said the DDT program involves three FDA divisions, so it does not necessarily go more quickly than the drug development approval process. But the DDT can help the FDA determine whether there is a wider range of uses for the test under review.

Dr. Ropacki said he wanted find out whether the DDT is a more efficient pathway to be able to use the tool for a trial, but it seems not. Is there any thought to giving the DDT program deadlines? That could help get reviews through more quickly.

Dr. Campbell said she was unable to answer the question. However, the FDA is looking at a system of compendiums, which may be able to speed the process. More information on the compendiums will be posted on the FDA website soon. The FDA seeks comments on the idea.

Roger Reeves, Ph.D., of Johns Hopkins University, asked whether the FDA will approve prenatal interventions. Dr. Campbell said that she could not answer the question.

Michael Harpold, Ph.D., of the LuMind Foundation, asked whether the decline in intellectual ability and the onset of Alzheimer's disease needs more attention. Dr. Campbell said that understanding the natural history of DS is very important, particularly in terms of the comorbidities. But she could not say whether the FDA would favor the approach over another. She would not want to sway investigators to go either way because the FDA is always evaluating their thinking on those issues.

Industry and Clinical Trial Overview

George Capone, M.D., Kennedy Krieger Institute

The field of pediatric cognitive enhancement (cognitive pharmacology) does not exist, but there is an unmet need for the field. As the neural mechanisms of learning, memory, emotion, behavior, attention, and executive control become better characterized, this leads to hope that more effective treatments can be developed.

Previous strategies for cognitive enhancement focused on challenging claims surrounding the use of dietary and nutritional supplements, including claims of nutritional deficiency. Some of the claims that a supplement would improve developmental or cognitive outcomes were advocated by parents, manufacturers, and providers, but the claims lack scientific backing.

Biomedical researchers had little interest in conducting clinical trials regarding the dietary and nutritional supplements for persons with trisomy 21 except to dispel exaggerated claims. There wasn't a lot of interest in working with the trisomy 21 population.

Early research efforts (1950s through the 1980s) looked at hormones and vitamins in relation to DS. Trials looked at pituitary extract, thyroid extract, vasopressin, and glutamic acid. Later studies looked at megavitamins and minerals, vitamin B₆ (pyridoxine), and tryptophan.

Clinical trials after 2000 included those investigating folate metabolism and antioxidant strategies. Recent clinical trials have looked at folic acid, L-acetyl carnitine, antioxidants, amino acids, and alpha lipoic acid and L-cysteine. This represents a change in rationale.

More recently, there has been interest in testing existing cognitive enhancement medications in the DS population. These medications have shown some benefit in non-DS adults with dementia. Some theorize that the mechanisms of action for dementia might be similar in people with DS and adults with dementia.

Small feasibility studies have begun on medications that have FDA approval for dementia that have a known safety and benefit profile and mechanism of action. They follow one of two hypotheses, the cholinergic hypothesis or the glutamatergic hypothesis. The rationale to these studies was questionable, but they are small feasibility studies. Some of the medications approved for use were memantine, piracetam, donepezil, and rivastigmine.

Studies using a mouse model found that memantine, pentylentetrazole, and Roche RG1622 were helpful. The cholinergic medication donepezil and the glutamatergic medication piracetam were not helpful.

Human pediatric trials using piracetam and rivastigmine were found not to be helpful. Donepezil was found not to be helpful in children and adults. Roche recently concluded a trial with RG1662 in young adults, but the results are still pending. A phase II trial of RG1662 is currently recruiting.

There are virtually no studies using psychotropic medications on behavior targets such as maladaptive behaviors or psychiatric disorders, such as ADHD, anxiety, sleep disorders, and repetitive compulsive behaviors. Treating some of these conditions with psychotropic medications could have a positive effect on cognitive functioning. This might be a more effective approach than the cognitive enhancement studies that are being conducted.

Jeannie Visootsak, M.D., Roche Innovation Center

The pharmaceutical industry has experience in trial design, but the experts in DS are the clinicians who treat it and the investigators who study it. None of the medications used in clinical trials for cognitive impairment in DS have been effective in humans, although piracetam may have an effect on NMDA glutamate receptors, which are involved in learning and memory.

Parents of DS children have varying attitudes toward clinical trials and may have limited knowledge about them. Some parents have come to accept their children as they are and in some cases worry that a clinical trial would affect their child's likable personality.

Other parents would like to see changes in their child. Of those who would like to see changes in their child, most want cognitive improvement. That contrasts with the parents of children with Fragile X, who are more likely to want behavioral improvement. Parents of children with DS also worry about independent living, being teased by peers, language problems, and self-care skills. Current treatments do not address these parental concerns.

Some parents may want investigators to guarantee that their child would suffer no adverse effects if they take part in a drug trial, but that cannot be done. There is a common misconception that the investigators are attempting to change the child in a fundamental way.

Dr. Visootsak described a phase I trial to determine safety that her group is running. Children who participate must be able to complete all tasks, including submitting to a blood draw and being able to sit still for at least 20 minutes. They must be verbal, and they must be able to swallow tablets. So the children who participate in the trial tend to be higher functioning. The study days are long, it is tiring, and

many children with DS have short attention spans. It can be difficult to ask questions in a way the child understands. An incorrect self-report from the child can affect the quality of the data.

Some children enjoy participating in the study so much that when they improve, the question arises whether the improvement is because of the medicine or because they like the staff and the activities.

There are also challenges to the caregivers to have the child participate in a study. The caregiver must drive to the clinic, do a full day of testing, and return home. Parents and participants get tired. One question Dr. Visootsak's team has wrestled with is how they could streamline the study day.

Ensuring quality data requires a number of steps, including having quality raters with experience with this patient population, ensuring there is no drift in assessments or scoring over the course of the study, and ensuring the accuracy of parents' self-reports by having a study coordinator sit with each parent during the administration to ensure the parent does not rush through it.

The group also provides the parent an outline of the tasks to be completed during the study day, and every attempt is made to streamline the day. Questionnaires that ask about issues that may bring up negative feelings are deferred to later in the day.

If possible, lab tests are done later in the day; however, fasting labs are done first thing. The same caregiver must attend each session, the same rater completes the interviews, and all questionnaires must be completed in the clinic.

A number of considerations go into evaluating the suitability of neurocognitive tests and functioning scales, including the number of children who complete the tests and the ceiling, floor, and practice effects.

Working Group Breakout Work Sessions

The participants next broke up into three working groups—cognitive, behavioral, and medical. The members of these working groups had been working together for a few months via teleconference. Their task during this session was to put together a presentation of the work they had done to identify outcome measures for DS clinical trials. After the morning working group session, each of the groups returned to present their work to the entire group.

Cognitive Working Group

Prior to the meeting, the Cognitive Working Group had produced 103 PowerPoint slides containing important cognitive outcomes in the categories of language. The group agreed to edit down the number of slides while maintaining the most important outcomes.

Working group members included Dr. Wang, Dr. Ropacki, Stephen Hooper, Ph.D., Carolyn Mervis, Ph.D., Fran Conners, Ph.D., Jamie Edgin, Ph.D., Leonard Abbeduto, Ph.D., Nancy Raitano Lee, Ph.D., Sharon Krinsky-McHale, Ph.D., Wayne Silverman, Ph.D., and Xavier Liogier d'Ardhuy, Ph.D.

The Cognitive Working Group was subdivided, so various members gave different parts of the presentation.

Dr. d'Ardhuy led off by showing a slide that summarized various problems that arise over the lifespan of a person with DS. These include developmental delays through life, infantile seizures, ADHD, oppositional defiant disorder, obsessive compulsive disorder, early-onset dementia, congenital heart disease, hypothyroidism, diabetes, and obesity. He then outlined some current treatments and said that the major unmet needs are for drugs to improve cognitive function and adaptive behavior.

The group next highlighted a schematic of DS cognitive and behavioral weaknesses. Roche Pharmaceutical worked with the FDA to develop the model, which connects all of the DS comorbidities. Roche is now trying to select their outcome scales based on these domains.

Some general issues that the working group considered are patient subpopulations and comorbidities. The group identified a need for tests in different languages and for different countries. Another issue the group considered was the overlap across cognitive and behavior domains and how those can combine to create a “task impurity” problem.

Dr. Krinsky-McHale gave the next part of the presentation, which examined attention. There is not a lot of data on attention for DS with the exception of a type of attention called cancellation. Other forms of attention include the following:

- Sustained attention, which is linked to the level of wakefulness or alertness
- Selective attention, which directs sensory and thought processes to a particular stimulus so that action can be taken
- Divided attention, which is the ability to attend to more than one stimulus at a time

Dr. Krinsky-McHale described several short tests that can be used with people with DS. Short tests are best because they keep the subject more engaged. Her group uses a paper-and-pencil test to measure cancellation. It involves asking the subject to cross out a particular line drawing among a group of line drawings. The subject then does the task a second time, crossing out a different line drawing. This task requires the individual to inhibit the response they gave to the first trial. This is a difficult task for those showing signs of dementia. It is a moderately sensitive test and takes 5 minutes to complete. It has been tested in adults with DS.

Another cancellation scale includes the Flanker Inhibitory Control and Attention Test in the NIH Toolbox. This test measures attention and inhibitory control and involves executive function. It has not been tested in the DS population. It can be completed in 3 minutes.

The Alzheimer’s Disease Neuroimaging Initiative Number Cancellation scale is another scale that measures attention control and sustained attention and also overlaps with executive function ability. It takes 3 minutes to complete. It has been used with adults in the general population.

The Brief Test of Attention, which measures divided attention, has not been tested in people with DS, but it has been used in clinical studies. The test is given orally and takes about 10 minutes.

The Ruff 2 and 7 Selective Attention Test measures sustained and selective attention. It has only been tested in the general population. This paper-and-pencil test takes about 5 minutes to complete.

The Test of Everyday Attention measures selective and sustained attention, attention control and switching, and divided attention. It has been tested in the general population and takes between 45 to 60 minutes to complete. There is also a children’s version.

Dr. Edgin discussed processing speed and executive function tests. There are several challenges to measuring executive functioning and processing speed. There is little or no retesting and practice-effect data available, very few tests have alternate forms, and the measures that do exist have not been tested in children younger than 6 years. Also, these tests have a “task impurity” problem in that the tasks may be assessing different domains of cognition.

One important need is to develop working memory tests that are understandable by children with DS and that focus on working memory as opposed to short-term memory.

The group identified concepts of interest for meaningful treatment benefit, including processing speed, which is linked to IQ and reading and math skills. Processing speed may mediate other cognitive processes such as working memory and inhibitory control.

Processing speed has not been studied systematically in DS. However, there are two reasons to research processing speed in DS individuals. First, slow processing speed is a hallmark of most forms of intellectual delay, and second, pharmaceuticals might be able to affect this domain.

The group produced a chart of tests for processing speed and classified them as ready to use, needs modification, needs to be developed, or is not appropriate for people with DS. The Cantab Simple Reaction Time test is ready. The Wechsler Preschool & Primary Scale of Intelligence (WPPSI) processing speed (PS) subtests, the Differential Ability Scales II (DAS-II) Rapid Naming test, and the Congruent Condition of Happy Sad Stroop Experimental Task could be used, but they would require modification for use with people with DS. A test for processing speed without a motor component needs to be developed. Among the tests not appropriate are the more complex processing speed measures from the Wechsler Intelligence Scale for Children (WISC) and Wechsler Adult Intelligence Scale (WAIS). There are test-retest (TRT) data on these instruments, although some data cover only the general population and not people with DS.

Executive function has three subdomains: working memory, inhibition, and shifting. Executive function is important to measure and treat because it is related to real world outcomes in typical children, including academic achievement and work behavior. It is strongly linked to fluid intelligence.

In DS, executive functioning deficits are seen starting in early childhood and into adulthood. In most studies, a DS individual has working memory, cognitive flexibility, and inhibition skills that are below their mental age. Decline in executive function may be an early marker of Alzheimer's disease in adulthood.

The Behavior Rating Inventory of Executive Function (BRIEF) and the Behavior Rating Inventory of Executive Function-Preschool (BRIEF-P) questionnaires have good TRT reliability, but they may not be suitable to use in pharmaceutical trials because they may not be sufficiently sensitive to changes.

The group included slides with further information on tests of the executive functioning domains of working memory, inhibition, and shifting, but the group did not discuss them.

Dr. Edgin noted that it is important to mention tests that are not appropriate for use in DS. Her research group recently found that a test they had been using had a TRT of 0.48. That was disappointing, but she also said that this TRT might not be unusual in the DS population.

She went through slides containing information about a variety of measures of inhibition they had evaluated. The slides contained information about the ages the test was appropriate for, how long it takes to complete, the mode of administration, TRT and practice effects, and additional comments. The NIH Toolbox Flanker has good TRT, but there may be practice effects, and there is a 17 percent noncompletion rate.

Shifting measures the child's ability to learn a rule and then transition to a new rule. The NIH Toolbox had good TRT and no practice effects, but it has not been tested in children with DS. In general, there are very few instruments available to measure outcomes for children younger than 6 years old.

With tests of executive function, there is poor TRT, so alternative forms are important in this domain. Otherwise, the children learn the strategies, and that can cause instability in the test when it is given again.

Within the working memory domain, there are some subtests from IQ tests that can be used, including the DAS-II Digit Span, the Wechsler Scales Digit Span, and the Wechsler Scales Spatial Span. There are also tests used with preschoolers, but they have to be tested within the DS population.

Dr. Hooper provided the next portion of the Cognition Working Group report. His area was related to memory and learning. Memory and new learning are challenges for the DS population. Explicit recall and implicit recall are both lower in people with DS, but explicit recall is disproportionately lower. For that reason, this might be the skill that investigators would want to look at more closely.

The Cantab and the Repeatable Battery for the Assessment of Neurological Status (RBANS) have been used with DS individuals. The Children's Memory Scale has also been used.

The Cantab is an automated battery of neuropsychological tasks that uses touch-screen technology. There are data available on that test. The Cantab has alternate forms and a number of subtasks. Each subtask takes 5 to 10 minutes. The test covers ages 4 to 90, and it is available in alternate forms. It includes tasks on visual and verbal memory, but it does not have backward number repetition, which can be very difficult for the DS population.

The RBANS is a paper-and-pencil test that has been used with individuals ages 12 to 89. It is not appropriate for people functioning below age 12. Tasks include list learning and story component. It has good TRT.

The Woodcock Johnson Test of Cognitive Ability has been used in clinical trials and has a long track record. It is primarily a paper-and-pencil test. The subtests take about 5 minutes. The age range is 2 to 90. This test has exceptional psychometric properties because item response theory was used to construct the test. In essence, that means that the raw score is standardized. The standardized raw score provides a good way to track outcomes over time. There are a range of tasks, including story recall and numbers recall, that measure explicit recall and learning.

The Stanford-Binet Intelligence Scales covers ages 2 to 85 and has subtests on a wide range of skills that can be used with the DS population. The test is well-standardized and produces a W score that is well suited for use for measuring outcomes in clinical trials.

The Leiter International Performance Scale is well-standardized, and the group may want to look at the NIH Toolbox.

It is important to measure the change in slope when measuring memory and learning.

This group also produced a chart showing possible tests and where they are categorized (ready to use, needs modification, etc.). Dr. Hooper suggested using one more category, "contraindicated," because some tests are not appropriate for the DS population. The group also identified some measures that have subtests that are ready to be used.

Dr. Abbeduto spoke next, noting that language onset is delayed in children with DS, but the deficit is greater in certain areas of language. Children with DS have more problems in expressive language, grammatical development is more delayed than vocabulary development, and receptive language may be more advanced than expressive language. This means that it is important to think about language acquisition in a more nuanced way rather than having a single language score.

Among the tests the group looked at were those that measured vocabulary (concrete and relational), grammar (morphology and syntax), communication (prelinguistic and nonverbal), speech (articulation and motor programming for speech), and language in context (intelligibility and fluency). It is important to measure domains that have real-world implications for the individual.

The group evaluated a variety of tests and subtests that measure language, rating them as other groups have done (ready for use, needs modification, needs to be developed). The tests that are ready have been used for individuals with DS. Tests that need modification would require a significant amount of work before they are ready for use, including developing norms for the DS population and determining the TRT value.

The Peabody Picture Vocabulary Test, a receptive language test, and the Expressive Vocabulary Test are ready to use. The psychometrics for these tests have been developed using a large population and have been used with people with DS. There is still more information needed on TRT. These tests also cover a broad age range, children to adults.

There are a variety of tests that the group rated as possibilities, but they would need modifications. They include the MacArthur-Bates Communicative Development Inventory (CDI), which has been validated against other measures and works well. The CDI does not need much modification, and it covers only young children.

There are no tests to measure grammar acquisition that are ready to use right now. Among the tests that are available but would need modification are the Children's Depression Inventory (CDI), Test of Reception of Grammar, version 2 (TROG-2), and Clinical Evaluation of Language Fundamentals (CELF). These tests have the advantage of having a long history and having good psychometrics for the general population. The psychometrics have not been developed for the DS population.

Speech articulation is a big concern expressed by parents of children with DS. They want their children to develop greater speech intelligibility, which could positively affect social interactions. The Goldman-Fristoe Test of Articulation and the Verbal Motor Production Assessment for Children have been around for a long time and could be useful, but they need modifications.

The Social Responsiveness Scale may be ready to use. It screens for autism spectrum disorders (ASDs) and has been used with DS.

The Wordless Picture Book is a test of expressive language, but it needs modification. There are studies currently using the test with both individuals with DS and individuals with Fragile X.

Dr. Abbeduto also summarized the group's work on social cognition. Social cognition involves processing social information in a social situation, recognizing what is important in social interactions, and storing and using the information to make decisions. It involves taking the perspective of another and is strongly associated with IQ, executive functioning, and language ability. There is some overlap in symptoms between people with DS and people with ASDs.

Social cognition is different from sociability. Sociability is a DS strength while social cognition is a DS weakness. Social cognition often falls below the DS individual's mental age.

The Social Responsiveness Scale may be a good instrument to measure this domain in DS individuals ages 2.5 years to adult. It is an informant report anchored in the most recent 6 months. It can be completed quickly and includes two subscales, social awareness and social cognition. A recent study by Channell et al. (2015) involving people with DS ages 10 to 21 found the test to be a good instrument for DS. The total score correlated with the Social Communication Questionnaire, nonverbal mental age, and receptive language scores.

The Emotion Judgment Test is directly administered to the person with DS. It uses video followed by easy multiple-choice questions, so there are few language demands on the participant.

The Social Resolution Task has been used in France, and there is an English version. It has high expressive language demands, but has been used with adults with DS and correlates well with other measures such as selective attention, inhibition, and receptive vocabulary.

There are commercial neuropsychological tests such as the Developmental Neuropsychological Assessment, second edition (NEPSY-II) that examine theory of mind, that is, assessing difficulty with comprehending the perspectives, experiences, and beliefs of others. They are widely available and have good psychometrics, but they are not always normed for DS.

Theory of mind has been an important concept used in the developmental literature. Tests that look at theory of mind contain hundreds of tasks that could be extended to use with people with DS.

Another measure often reported in the literature is emotion recognition. Tests that measure emotion recognition include the Children's Eyes Test, the Penn Emotional Recognition Battery, and eye tracking of dynamic or static social scenes. The Diagnostic Analysis of Nonverbal Behavior (DANVA) contains some DS data. Eye tracking has potential for use with DS, but the challenge is that eye tracking is difficult to do, and it can result in a large loss of data.

Dr. d'Arhuy said that the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-5), has changed how DS is characterized, moving away from the idea of IQ and placing more emphasis on ability areas. These include the conceptual (language, reading, math, etc.), social (empathy, social judgement, interpersonal communication, etc.), and practical (managing personal care, job responsibilities, money, etc.) domains.

The group identified four scales to measure adaptive behavior outcomes. The first is the Vineland Adaptive Behavior Scale II (VABS II), which has been used in DS clinical trials and is available in English and Spanish. There is also a French version under development. The scale covers birth to age 90, is standardized and normed, and has a survey interview form available. VABS II takes 45 to 60 minutes to complete and needs a well-trained rater. The test has good TRT and inter-rater reliability.

The Diagnostic Adaptive Behavior Scale (DABS) covers ages 4 to 21. The DABS takes 60 minutes to complete and is also available in a survey interview form. It is available only in English. It has good inter-rater reliability and TRT. It is standardized and normed and is adapted to the new DSM-5 definition of intellectual disability. The DABS is recommended by the American Association on Intellectual and Developmental Disabilities.

The Adaptive Behavior Assessment System (ABAS-II and -III) covers birth to age 89. It is in questionnaire form, is available in English and Spanish, and takes 20 minutes to complete. It is adapted to the DSM-5 definition of intellectual disability.

The Scales of Independent Behavior, Revised (SIB-R), covers birth to age 80, is available in survey interview form, takes 45 to 60 minutes to complete, and is available in English only. It is standardized and normed, has good inter-rater reliability, and has good TRT. There is a maladaptive domain within the test, too.

In summary, the group rated VABS II and SIB-R as ready for use in the DS population. However, VABS-II may overestimate the participant's abilities while SIB-R may underestimate them. ABAS-II needs modification because it is only in questionnaire form. The DABS age range is too narrow, so it would need further development before the test could be assessed for use with the DS population.

Dr. d'Arduy also presented information on measures that Roche has used in a non-drug study in teens and adults with DS. For IQ, Roche concluded that the Leiter scale was better than the Stanford-Binet scale because the Leiter scale has less of a floor effect. VABS-II was found to be suitable and stable, as was the BRIEF-P. They also found the Clinical Global Impressions-Severity (CGI-S) and the Clinical Global Impressions-Improvement (CGI-I) were suitable and complement the VABS. So far, Roche has not found a suitable tool to assess quality of life.

Behavior/Social/Emotional Working Group

Anna Esbensen, Ph.D., Cincinnati Children's Hospital Medical Center

Deborah Fidler, Ph.D., Colorado State University

Dr. Esbensen began the presentation by noting that the group kept the outline suggested by Dr. Urv, looking at their topic in terms of natural history, patient subpopulations, and the environment of the patients.

Dr. Esbensen said that people with DS have more social problems but fewer behavior problems than individuals with other types of developmental disabilities. Internalizing behaviors and moodiness emerge in adolescence, decline, and then re-emerge in later years with Alzheimer's disease.

Between 20 and 30 percent of DS individuals show some psychopathology, and the same percentage show maladaptive behaviors. Maladaptive behaviors are more common in younger children and with the onset of dementia.

Life changes such as transition out of school and to the workplace are related to an increase in mood and other mental health concerns. During these times, parents may describe their child as stubborn. Weaknesses in executive functioning, processing, and learning may underlie the maladaptive behaviors and noncompliance. There are more adjustment concerns after the caregiver transition or transition to independent living.

Dr. Fidler said that self-regulation is a challenge to people with DS, even compared to others of the same mental age. People with DS also demonstrate attenuated early exploratory behavior and difficulty with planning tasks. On the BRIEF-P, they score poorly in working memory and planning.

There is evidence in the laboratory of difficulty with shifting relative to mental age peers. However, difficulty with shifting does not show up in parent reports. This may be a weakness that appears with testing but is not a "real life" concern. DS adults also have difficulties with executive functioning and self-regulation.

Dr. Capone said adolescents and young adults may develop an intensification of pre-existing mental health conditions, or they may develop new mental health conditions. Among the conditions that can emerge or intensify are obsessive-compulsive behaviors, anxiety, depression, psychosis, mania, catatonia, declines in adaptive functioning, memory problems, changes in personality, and self-talk. With so many co-occurring symptoms, it is not easy to pinpoint the primary diagnosis. That makes it difficult to know how to proceed. These mental health conditions may be used as inclusion or exclusion criteria for clinical trials. It is important to exclude any medical problems before making a primary psychiatric diagnosis.

Elisabeth Dykens, Ph.D., said that the important thing to consider with these mental health presentations is how pronounced a change there is in the individual's own baseline. Self-talk is common among DS adolescents and is relatively benign, so that may not be a concern. However, if the self-talk takes on a difference in tone, there may be reason to be concerned.

Although providers have a sense of the mental health conditions that exist among individuals with DS who come to clinics, not as much is known about the mental health of the general DS population. When the DS phenotype is investigated in community samples, there appears to be a lower overall rate of these mental health presentations.

The diagnostic boundaries are blurry, so it is a challenge to define the phenotypes. These cases defy the tools that are available. Dr. Fidler agreed that there are large gaps in knowledge about mental health presentations among DS individuals because the data come from clinic samples. It is difficult to know how representative the clinic population is of the larger DS population.

There are subpopulations of DS individuals, including across the lifespan. Younger DS children are at risk for a variety of difficulties such as agitation and aggression, sleep problems, and inattention. In adulthood, there are problems with anxiety, depression, and obsessive-compulsive disorder. That means that the outcomes of interest will vary by age.

Comorbid mental health diagnoses include inattention, ASD, and dementia. These are distinct subpopulations. Individuals with both ASD and DS, for example, are very different from those who have DS only. The ASD/DS group may have more in common with individuals who have ASD and IDD.

There may be different genetic causes of DS, and those need to be considered.

When doing an assessment, it's important to get observations from a variety of environments, including home, school, clinic, and laboratory. Natural observations can be done in the home. Teacher ratings can be obtained from the school, where the child may manifest more problems because of school pressures. There can be observations in the clinic or the laboratory.

Parent and teacher ratings in school age individuals converge in some ways and diverge in other ways. Parents reported more trouble with control than teachers did. This might be an artifact in that parents might have their own ideas about behavior without the built-in reference group that teachers have.

Executive function skills as reported by the teacher were strongly related to school outcomes. Therefore, that might be important to measure in any studies. Emotional control and self-regulation could be outcomes to measure for treatment benefit.

One issue that is difficult is defining the phenotype that is specific to DS and different from IDDs. It would be helpful to determine which maladaptive behaviors DS individuals have in common with people with IDDs and which maladaptive behaviors they do not have in common.

It is important to establish good treatment outcome measures for mood disorders, including clinician reports and self-reports. It is important to get the right people in the study, avoiding those who have a medical condition that contributes to the mood disorder.

Investigations of sleep may include parent ratings or actigraphy.

Dementia is complex, difficult to study, and may require the use of multiple assessments. It is important to consider whether a change in the individual's environment might play a role in the mental confusion.

Potential COA types include patient-reported, clinician-reported, observer-reported and performance. Assessing emotional control and self-regulation could use all four of these assessment types. Assessments of maladaptive behavior would likely rely on observer reports and possibly performance. Dementia would rely on observer reports and performance, but not on patient reports. The group placed the greatest value on observer-reported outcome assessments.

The group developed a chart rating possible instruments to measure maladaptive behavior outcomes. They categorized the tests as in one of three categories:

- Measures the concept in the IDD population
- Needs modification
- Needs further development to assess its utility with DS individuals

The Autism Behavior Checklist (ABC) and the Child Behavior Checklist (CBCL) were among the instruments placed in the “measures the concept in the IDD population” category. The SIB-R and the VABS measures of problem behavior were among those categorized as needing modification.

Currently, there are no instruments to measure inflexibility, noncompliance, and coping with transition.

Among the measures being used in the IDD population for emotional control and self-regulation are the BRIEF and BRIEF-P.

Measures for self-report of mood being used in the IDD population are the Glasgow Anxiety Scale, the Glasgow Depression Scale, and the Social Research Database on Questionnaires (SRDQ). Informant reports of mood can draw on the Glasgow Depression Scale or the Hospital Anxiety and Depression Scale (HADS).

One thing to consider before deciding on which measure to use is to decide whether it is better to compare the experimental group with their same chronological age or developmental age peers. That can make a difference in which measure is used.

Psychiatric symptom outcome measures include the Psychiatric Assessment Schedule for Adults with Developmental Disabilities. Investigators should be aware that the diagnostic manual used to categorize patients, such as the DSM-5 or the International Classification of Diseases, version 10 (ICD-10), can change the outcomes.

Sleep outcomes can be measured by actigraphy, daily sleep log, or polysomnography.

There are many screening measures for dementia and aging, including the DS Mental Status Exam. A medication for dementia is needed, but it may be possible only to slow dementia rather than curing it. One big issue is that a better understanding of normal age-related declines will be needed before dementia can be fully understood. Another problem will be to better define typical dementia.

The working group also noted that a person with dementia may have fewer acting-out behaviors because the individual is withdrawing.

The group identified outcome measures that need to be developed for use with children, including compliance, self-regulation of behavior and emotion, physical activity, and sleep. Some psychophysiological measures that are needed include stress markers and biomarkers. There are some new apps that might be helpful in evaluating lower-functioning individuals.

Measures that must be developed for adolescents include compliance, specific psychopathologies, motivation, and functional and social adaptive decline.

Discussion

A participant said that she is working on a DS phenotype project involving 52 children with DS. They have found that both the Nysomer Scale and the school-age BRIEF have good TRT.

Dr. Ropacki asked whether anybody had considered creating a composite battery by drawing from multiple tests. Dr. Fidler said that is an option that they need to explore.

Ben Handen, Ph.D., asked whether the group had considered the Home Situations Questionnaire or the School Situations Questionnaire to measure noncompliance. Another option for noncompliance is the Eigler FB Scale.

Dr. Fidler said there is a measure of compliance on the School Function Assessment. Many of their discussions shifted to what caused the behavior, such as failure to understand direction because of language deficits.

Dr. Esbensen said that people should email any measures that they think would be good. Dr. Urv said that participants could also post suggestions on the website she has set up for the working groups.

Dr. Abbeduto said that it is best not to rely on parent reports as an outcome measure for a clinical trial. It is better to include third parties such as teachers.

Dr. Hooper said that this presentation has highlighted the complexity of the task of identifying DS outcomes for clinical trials. There are so many covariates and mediators that can moderate the outcomes, and they must be taken into account.

Medical/Physical Working Group

Kent McKelvey, Jr., M.D., University of Arkansas for Medical Sciences

The Medical/Physical Working Group broke down their examination of outcome measures by organ systems.

There is increasing longevity among people with DS. In 1910, the life expectancy of a person with DS was 9 years. A person with DS now has a life expectancy in the 60-year age range. The natural history of DS is not known, and as people live longer, new aspects of DS are emerging. Endpoints of aging might be different for DS, at least in some ways. For example, people with DS appear to get fewer solid tumors. People with DS have high rates of hearing loss and dental conditions. Some of the traits that appear as the individual with DS ages are part of normal aging. Some of the changes are more characteristic of people with DS.

Dr. McKelvey went through an Excel spreadsheet that listed outcomes by organ system. Each organ system included information on physical exam and clinical lab findings, biomarkers, imaging, clinical trials, and possible outcome studies (i.e., questions that need to be answered).

Dr. McKelvey went through the items to explain the group's approach to its work. The aim is to correlate some of the physical outcomes with the behavioral tests that are available. For example, the sheet lists biomarkers of oxidative stress in the neurological system. Would oxidative stress correlate with lower adaptive function? That is something that is not known but would be helpful to learn. Is Alzheimer's disease an inflammatory disease? DS individuals have higher rates of inflammation. Is there a behavioral test that would show that?

The working group would create Excel spreadsheets for various age ranges.

Dr. Reeves said that this exercise showed that the natural history of DS is not known, even though DS is a frequently occurring condition. Although investigators and clinicians gather much information on individuals with DS, the data is not aggregated to give a better overall understanding of the condition. That problem needs to be addressed.

A second issue is that biomarkers change over the lifespan and may also change during the course of treatment. Is the ideal to harmonize the DS biomarkers with biomarkers seen in the typical population?

A third issue is that there is a high degree of variability in this population, and there is a need to better understand the variability. DS is a perturbation of a set of genes that causes a series of changes and reduces the person's ability to maintain homeostasis. It is important to identify these changes as they appear.

Dr. Harpold said that DS-Connect™: The Down Syndrome Registry may provide a way to capture the natural history of DS. Clinicians would enter medical information about their patients into the registry. Over time, that could provide a rich source of information about DS by life stage. It also would give insight into the variability of the syndrome. This is a step that could be done sooner rather than later. There is already a lot of data collected; the problem is that it has not been collected into a central location, but DS-Connect™ could change that. Dr. Dykens agreed, saying that other groups have successfully used databases. It has worked well in rare diseases.

Gordon Worley, M.D., suggested examining the pathophysiology of DS. Areas of inquiry could include morphogenesis, free radical injury, inflammation, autoimmunity, abnormal energy production, metabolic function, and stem cell depletion in adulthood. The Medical/Physical Working Group would bring together what is already known about biomarkers with these pathologies. The working group is formulating a way to do this and then to relate them to things such as cognitive decline.

Melissa Parisi, M.D., Ph.D., said that DS-Connect™ could be a tool to collect natural history, if families will update their data yearly. This will not produce a quick resource for investigators and will take a while to bear fruit.

Dr. Parisi suggested that members of the DS Medical Interest Group could find out whether the interest group members have information that they could share with the Medical/Physical Working Group. That might also be a venue to promote DS-Connect™.

Dr. Visootsak said that health issues can affect cognitive and behavioral outcomes. Untreated sleep apnea has been connected to behavior problems and a decrease in IQ. Fifty percent of DS individuals have sleep apnea. Hearing loss is also common in DS. It can affect expressive language ability, which may further affect cognitive ability. Finally, it has been found that those who have a congenital heart defect have lower gross motor and social skills.

Dr. McKelvey said that an ill-fitting continuous positive airway pressure (CPAP) mask could affect inflammation and inflammatory biomarkers. So a better fitting mask could change the biomarkers. That's an instance where a social factor affects a biomarker, which affects health. These are all related.

Maddalena Adorno, Ph.D., said that there is a lot of variability in the DS population, so it is important to take into account the starting point of individual patients when examining interventions.

Andrea Videlefsky, M.D., said that the Medical/Physical Working Group had identified many biomarkers of interest. The challenge now is to make the connection between biomarkers and clinical outcome trials.

Large Group Discussion: How Do the Pieces Fit Together?

Moderators:

Robert Riddle, Ph.D., National Institute of Neurological Disorders and Stroke

Dr. Tiina Urv, NICHD

The discussion began with the question of whether to use new technologies to test outcomes or to continue with the paper-and-pencil tests that so many measures currently employ.

Dr. Harpold said that there are many new technologies that could be used in clinical trials. There is a new technology that can measure change in gait. This measure was used with Alzheimer's disease patients. Many valuable measurements could be obtained through current technology and in real time. Waking and sleep cycles, for example, can be measured by a watch or a mobile phone. New technology should be incorporated into their planning, for not only what can be done today but also what can be done in the future.

Dr. Ropacki said that they should add any new instruments that might be helpful. One problem in other fields is that investigators continued to use instruments that weren't really very good, simply because they had always been used. He suggested investigators begin using these new technologies. New technologies can gather a lot of valuable data with little burden to the investigator. They provide real-time monitoring and changes in skill or behavior could be quickly discovered. He gave the example of a smartphone measuring the speed at which an individual texts. That measure conceivably could be used to track mental decline.

Dr. Dykens said there are tools now available to bring big data into the field of IDDs. It would be good to use these technologies to see whether they correlate with existing instruments. But would the FDA accept this new technology for use in a drug trial? Dr. Ropacki speculated that the FDA would accept new technology that is shown to benefit patients.

A participant commented that when choosing new tools, they must be certain the instrument is measuring what they say it is measuring. For example, a test of processing speed should not require motor skills. Dr. Riddle asked whether they were talking about using existing technology or developing new technology.

Dr. Ropacki said that in some cases they would transfer an existing instrument to a new technology platform, such as creating a version of a paper-and-pencil test for the iPad. Technology offers the possibility of real-time measurement in the individual over time. The measurements would need to correlate with existing tests and real life changes in the individual. The idea is to move to the concept that function and cognition are related; an impact on cognition will have an impact on function.

Dr. Abbeduto said there is a temptation to think technology is the solution when it is not. For example, a cognitive test involving manipulation of objects is three-dimensional. On the iPad, it is two-dimensional and so may not measure the same thing. Using the iPad may make it a completely new task for the child. This is one example of how introduction of new technology adds another layer of complexity.

Dr. Handen said that technology gives multiple data points, which would help us understand variability. The iPhone can measure variables over 24 hours, providing numerous data points.

Dr. Abbeduto agreed that new technology carries many new possibilities, but investigators will need to develop all new psychometric data to validate these new tests.

Dr. Edgin said that children do like the computer. Her group alternates between paper-and-pencil and computer tests as a way to maintain interest.

Dr. Riddle asked how the NIH study sections would rate the development of new tools to assess DS outcomes. Would the study sections welcome those applications?

The NICHD has released a program announcement to encourage the development of new tools. But the applicant would need to make a compelling case that the instrument could be used more widely than within the DS population. Dr. Urv said that the NICHD is looking for assessment tools to use in clinical trials and is using a special emphasis panel to do so. Investigators should try to take advantage of the program announcement.

Dr. McKelvey noted that DS adults get osteoporosis and are prone to bone fractures. The osteoporosis drugs on the market do not work with DS adults. It would be good to study further, but it is not directly related to cognition. However, it can indirectly relate to cognition because of the fallout from the injury. Dr. Urv said that it might be best to take that study to the National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Dr. Riddle said that investigators who have a good idea for a study likely could find funding. The trick is to find the right Institute. Dr. Urv advised investigators to contact members of the NIH Down Syndrome Working Group for advice about funding. If the Working Group members cannot fund the project, members could help the investigator contact the appropriate person at another Institute.

Seth Keller, M.D., suggested that the amount of time that families spend in research laboratories could be reduced through the use of technology. For example, self-assessments and questionnaires could be completed in advance, using a secure site that meets Health Insurance Portability and Accountability Act requirements. Dr. Urv asked whether the pharmaceutical industry would agree to that. Dr. d'Ardhuy said that it might be a problem. Sometimes the forms must be filled out in the presence of the investigator to ensure the parent is engaged and understands the questions.

Dr. Visootsak said there is a period during which the questionnaire must be filled out. If the parent takes the form home, they might not complete it within the period. Dr. Keller said it would be online, on a secure site, and password protected. Dr. Campbell said that changing the mode of data collection would require additional psychometrics. Dr. Hooper agreed that it would require some preliminary psychometric work to ensure the responses between the two modes are equivalent. But it could be done. Dr. Wang said it could be done, but the tests would need to be standardized. The FDA has regulations about how to use electronic data storage and capture, and those regulations would need to be met.

Dr. Abbeduto said that many of the outcome measures the groups have identified so far are not normed for Spanish-speaking participants. Spanish-speaking people in the United States will be excluded from clinical trials because there is no outcome measure for them. That is a big issue and one that must be resolved soon.

Dr. Harpold said that one limitation of clinical trials is that they tend to include only high-functioning individuals. It might be possible to see more dramatic changes when the interventions are given to lower- or intermediate-functioning individuals. That is another issue that should be considered with the outcome measures. Dr. Dykens said it's more often the higher functioning and more motivated families (not individuals) that can volunteer for a clinical study.

Dr. Urv said she wanted to know what participants thought was the highest priority for the field. What is the issue that DS researchers most need to address? Each participant answered this question.

Dr. Dykens said she wants to have a DS phenotype measure for both the school and the home that captures the range of ways DS is expressed. It might be possible to create that measure by piecing together data that already exists. In addition, it would be ideal to be able to compare DS individuals with a particular phenotype to individuals who have other developmental disabilities but a similar phenotype. This approach might widen drug development research beyond DS to other IDD.

Dr. Capone said he was unsure how the pieces fit together. The cognition group had made the most progress in identifying outcomes. One thing that was clear was that investigators will need to consider the comorbidities of the people they study, parsing out people with AD, anxiety, depression, etc. This will create smaller experimental groups.

Dr. Campbell said that there is much overlap in the cognitive, behavior, and medical areas, and the participants in the meeting represent a range of disciplines. The challenge will be to get an overview of the topic, one that takes all of these areas into account.

Sigan Hartley, Ph.D., said that it is important to understand the natural history of DS. The age and developmental stage of the individual is important as they examine cognitive and functional outcomes. It is important to find out which outcomes would be most meaningful to DS families.

Dr. Abbeduto said they should develop a set of short-term and long-term priorities. He also said that there are problems with how clinical trials are run. Some trials begin with older children and never include younger children. Some trials have the wrong duration. Some drug trials look only at a drug and do not include a behavioral intervention. The whole model of how clinical trials are done may not work with the DS population.

Dr. Esbensen said that this meeting is only a start. More people are needed to work on this problem, but they should work in smaller groups to focus in on which measures have the most promise and how to take the next step. Dr. Esbensen will email to Dr. Urv the names of people who should be included in the next round of discussions.

Dr. Fidler said this is a complex field, and it is necessary to capture the complexity in linking treatment to outcome.

Dr. Lee said that one problem is that they have been identifying outcome measures without knowing whether the measures are for drug trials or for some other type of trial. In terms of drug trials, it is important to think ahead to what the short-term and long-term effects might be. It is possible that drug trials are being stopped too soon, before the drug is given a chance to show effectiveness.

Dr. Edgin said that there is very little information on early childhood. It would be good to have more investigations into behavioral outcomes for young children.

Dr. Mervis said the cognitive group has some tests ready to go. The next step may be to find out which drug trials are coming up and suggest which tests could be used as the outcome measure. It would also make sense to find out what drugs are in development and to suggest outcome measures for those future trials.

Huntington Potter, Ph.D., said the Cognitive Working Group showed that there is a consensus about which cognitive measures should be used in clinical trials. Another point is that DS is a lifetime disorder and the drugs to be investigated will treat different aspects of the syndrome. It is necessary to separate the different periods of life when thinking about which drugs are appropriate. And finally, the drugs that are tested in humans are selected based on preclinical research, which relies on mouse models. DS researchers do not have a good mouse model, and one should be developed.

Michelle Livingston said it is clear that more funding is needed to develop tests, run trials, etc. Ms. Livingston noted that it is difficult to get DS families to sign up for trials, but that could be because of the burden that participation imposes on the parents. Investigators should involve parent advisory groups when developing studies.

Dr. McKelvey said that DS is very complex and much about it remains unknown, including its natural history. He suggested that, since so much research is directed at behavior change, it would make sense to develop a master list of phenotypes. Also, people with DS have unique medical issues, and there already may be drugs available to help. It should be a priority to find those drugs and test them in the population.

Dr. Visootsak said that there are many misconceptions about clinical trials in the DS community. Parents believe that investigators want to fundamentally change their children and so refuse to participate. One priority is to develop better relationships with the community and do more education about the trials; otherwise, enrollment will continue to be difficult.

Dr. Wang said it is important to define the functions, behaviors, and cognitions that investigators expect of a drug. It is difficult to predict what effect a drug will have, so perhaps investigators should first do the drug trials in healthy volunteers. Also, there is a need to use clinically meaningful tests.

Dr. Ropacki noted that investigators find it difficult to get DS individuals and families to enroll in studies. He suggested that a DS registry that includes DS patients across the lifespan would be helpful. It would be possible to follow them longitudinally to answer some of the medical questions. The data would have cognitive and behavioral endpoints that could be associated with biomarkers. Dr. Ropacki is the principal investigator in the Cognitive Health in Aging Registry: Investigational, Observational, and Trial (CHARIOT) study of dementia. CHARIOT has blood and saliva samples, PET imaging of amyloid, patient consents, and other materials. This database has proven to be a valuable collaborative tool. CHARIOT has obtained the use of resources at no cost in return for sharing data or coauthoring a study. It has even been used to generate income. At the request of Dr. Urv, Dr. Ropacki said he would share more information about CHARIOT following the meeting.

Dr. Harpold said he sees three priorities. The first priority is to obtain a natural history. The second priority is to find a way to compare the outcomes of clinical studies that are ongoing right now. The third priority is to ensure that clinical trials are measuring the correct outcomes and that they measure what they say they measure.

Dr. d'Ardhuy suggested that electroencephalogram recordings should be used as a biomarker.

Dr. Reeves defended the DS mouse model, saying it has led to the only two clinical trials to improve cognition. Dr. Reeves also argued for doing gene sequencing. Many people resist having their DNA sequence in a government database. DNA data would be valuable in family studies. It costs up to four times as much to recruit and have a full medical work-up, but the sequencing is the geneticist's best biomarker.

Dr. Krinsky-McHale said that this workshop has concentrated on drug studies, but there are other types of interventions to consider. She also said that it is important to maintain the big picture of what outcomes they are trying to achieve. Another issue is that measures for preverbal groups are needed. It is possible that the greatest improvements will come from interventions at the younger ages.

Jeffrey Reznik, M.D., said that families may not be able to afford new interventions that are found to be effective. He advocated that investigators find lower-cost interventions whenever possible.

Dr. Videlefsky said that it is important to continue the interdisciplinary collaborations that were begun with this meeting. Things that have been discussed are very promising but would be impossible for one investigator or group to carry to fruition.

Dr. Adorno said that it is important not to reinvent the wheel. Many clinical trials fail. Why did they fail? Was it the wrong target, wrong endpoint, timing of treatment? Prenatal and infant studies are also needed. DS patterns are established early in life, so research should target that early age.

Dr. Keller said his focus is on aging, including aging parents. He suggested that clinical trial outcomes should not only include the individual with DS but also look at outcomes that include caregiver support and caregiver outcomes.

Dr. Worley said that the Medical/Physical Working Group has made slower progress because of their charge, which is to find biomarkers that would be useful in DS research. The group combed the literature but found that very little information is available. He suggested following the models of the spina bifida and cerebral palsy associations, which lobbied Congress and obtained funding for patient registries. There is now enough data in the spina bifida registry to do studies and begin to pin down the natural history. The spina bifida registry has received continuous funding for several years.

Dr. Hooper said it is important to pay attention not only to DS outcomes but also to the mediators and the moderators. Otherwise, it is impossible to correctly interpret clinical trial outcomes. A large natural history study is needed, and it should begin at birth and continue over the lifespan. Multisite studies are needed to investigate some of the issues participants have discussed, including using new technology to measure DS outcomes. Testing companies will need to be involved in this research. A multisite study on alcoholism sponsored by the National Institute on Drug Abuse provides a template. Investigators in that study use the same outcome measures. There is also a need to minimize bias in subject samples, which tend to attract high-functioning families. Finally, it is important to measure clinically useful outcomes that will help the DS individual with adaptive behavior and learning.

Vince Randazzo said it is necessary to educate families, and that is something the advocacy groups can help with. Parents may be reluctant to put children in studies that break up the DS individual's routine.

Dr. Handen said that it is important to identify four or five DS biomarkers to investigate. Each biomarker should be investigated by age group, including preschool, childhood, adolescent, young adult, and aging adult groups. Understanding the natural history is also a priority.

Dr. Conners said there are so many measures and so much work to be done that it is overwhelming. There is a need to narrow the number of outcome measures. It will also be important not to set the research methods bar so high that it stymies progress. For example, if a measure is appropriate only for some subpopulations, the work on those subpopulations should be allowed to go forward even though it cannot account for the entire DS population.

Sara Weir, M.S., said that education is key. The DS health care providers are the ones that families trust, so they need to be part of the conversation. Advocacy groups are committed to doing more to educate families.

Michael S. Raffi, M.D., Ph.D., welcomed Roche to the table. He also said the measures and biomarkers that group members presented would be valid outcomes to use in phase II studies.

Summation of Day 1

Dr. Tiina Urv, NICHD

There is consensus that the DS outcome measures need to be defined but that there is much work to do to reach that goal.

Day 2, April 24, 2015

Goals for the Day

Dr. Tiina Urv, NICHD

The task for the working groups on the second day is to develop three short-term and three long-term goals for future clinical trials. Short-term goals are those that could be completed within 18 months.

There is a clear need for a better natural history of DS. It would be important to first identify the information already available before planning new studies. For example, there is a New York study that has 20 years of data on people with DS that could be tapped. There are many articles and studies containing data that could be used.

Final Working Group Breakout Sessions

Cognitive Working Group

Dr. Ropacki said that the group should begin to identify a cognitive battery, stratified by chronological age or mental age; find ways to encourage investigators to pool long-term data; and develop a national registry. The registry should obtain consent and pre-consent for next stages so that there is no need to go back to the Institutional Review Board (IRB).

Dr. Harpold said that DS has a registry, DS-Connect™, but that it is patient-centric in that members of the DS community enter their own information. Investigators who register in DS-Connect™ can access the data and use it to design studies.

Dr. Edgin said that it is difficult to get families in the registry, because they think they are signing up for a clinical study.

A participant asked whether there were any pharmaceutical studies about to begin and suggested that the group could suggest cognitive measures for those trials. That raised the question of what outcome measures are being used in pharmaceutical studies going on now.

Dr. d'Ardhuy said that Roche selects tests based on their reliability, stability, and clinical relevance, among other factors. The tests selected depend on the disease or condition under investigation. Roche gets assessments from caregivers and clinicians and tries to obtain a global view of functioning. Among the tests they have used are Cantab and BRIEF.

There was some discussion of the pros and cons of BRIEF-P, with some participants saying that the norms are weak and that parents have said the questions are too difficult for their children. Dr. Lee said that her group has used BRIEF-P for school-age children with DS, comparing them with typical children of the same mental age.

Dr. Hooper said the Patient Reported Outcomes Measurement Information System (PROMIS) includes item banks that are designed for treatment trials. Investigators can choose items from different age levels to construct a test for various aspects of cognition, including inhibition and set-shifting.

Dr. Hooper said that he likes the idea of developing a test battery, which could be a short-term goal. The working group has already identified tests that are ready to go and tests that should not be used. The group could also identify cognitive domains in which there is no appropriate test available. Identifying tests that are ready to use or contraindicated, or areas where none are available, would be useful for upcoming clinical trials.

Dr. Lee agreed and said that when the group identifies tests for immediate use, they should realize that they are identifying tests that are good enough to use. There are no perfect tests available.

Dr. Abbeduto said the group should first define the criteria they use to decide that a test is ready to use and should articulate those criteria in a white paper. The group could also provide some metrics for the field, such as standardizing how tests are administered. He suggested they write a white paper laying out what they look for in the measures they recommend.

Dr. Edgin said that the group could also discuss their criteria for reporting results, such as how many children have been tested using the measure and how many of those children were at floor.

Dr. Abbeduto said that another issue is the window of time that is appropriate between the test and the retest. Dr. Hooper said that TRT should always be included.

Dr. Abbeduto said he would like to have a catalogue of data related to DS. Investigators could consult the catalogue and possibly combine their data with the data of others.

Dr. Ropacki said the Coalition Against Major Diseases has developed standard reporting forms that the group shares through open source. This standardizes data collection, allowing data to be pooled.

Dr. Hooper said that setting standards for testing and data collection would help but would require some data management expertise. Dr. Abbeduto said the data sharing requirements must be specified up front.

Dr. Ropacki said there is a common data elements standard that the group could adhere to. Standardizing the data is imperative, or the data will be impossible to share. Dr. d'Ardhuy said there would also have to be an agreement to share the data with other investigators.

Dr. Abbeduto suggested the group recommend that there be standards for reporting data. Dr. Harpold suggested the group put this down as a short-term goal. Dr. Hooper said there is no need to invent the data sharing standards, which are already available, but that the group should say how the standards would apply to the DS population. The group could create a matrix of test characteristics.

Dr. Abbeduto said that this task will require a review of the literature before standards could be set. For example, it would be important to find out what a reasonable TRT is on a test involving a child with DS. It might not be possible on a given test to achieve a 0.9 TRT in the DS population.

Dr. Edgin asked what cutoff levels should be set. For example, how much floor would the group tolerate?

Dr. d'Ardhuy said that his team at Roche had an 80 percent floor on the Stanford-Binet. They replaced that test with the Leiter, which still had a very high floor. Dr. Mervis said that there is a need to extend the existing norms downward six standard deviations so that the test can be used with DS children. Dr. Edgin said that a high floor would produce an artificially high TRT.

Dr. Krinsky-McHale said that it is not possible to have one test that would be appropriate for all levels. A test battery covering a wide range of abilities would be more useful.

Dr. Hooper said the group should set criteria for selecting tests. They should name the tests that fit the criteria and are close to being usable but that need modification. They could also identify the gaps where a domain requires measuring but there is no appropriate instrument. This is the short-term goal. Dr. Lee agreed, but said they should name the tests that are the best available right now, even if some of those fall short of the standards. The group agreed that when there is no test that fits that standards and the best available tests are named, the group should include suggestions for how the test could be brought up to standard.

Dr. Ropacki said there must be a standardized way to administer and score the tests. All groups have to do it the same way.

Dr. d'Ardhuy agreed and said that some of the disappointing results Roche received on one drug trial may have related to inter-rater reliability of the test used across sites. Also, it is necessary to have a test administrator sit with parents as they complete certain surveys.

Dr. Edgin said that she worked on a multisite project in which the investigators trained all of the test givers. The investigators videotaped the test sessions and still saw quite a bit of variability. To avoid this problem, the test has to be easy to administer. Many tests are not easy to administer and the population is difficult to work with.

Dr. Abbeduto said rater training is an important issue. But the problem is not limited to poorly trained raters; there is also a lot of inter-rater reliability variation among clinicians. One solution is to identify any tests that might require a high degree of rater training.

Dr. Abbeduto suggested that there be an analysis of subtests to see which are the most sensitive to change and which are not. Subtests that cannot detect a cognitive change should be scrapped.

Dr. Harpold said that there may be multiple levels to the standardization issue. If standardization is to be a goal, perhaps the task should be further broken down.

Dr. Abbeduto said that the group has identified three goals:

- Specify the principles used to develop and evaluate the outcome measures;
- Identify the outcome measures that are ready, or closest to ready; and
- Identify the measures from which investigators can obtain data to evaluate the measure.

Dr. Hooper said that he would like to add one more to the list: identifying where the gaps are.

Dr. Ropacki recommended that a standard agreement be made available to groups that plan to pool their data. The agreement will make it clear how and when the data can be used. A participant noted that NIH requires the sharing of all raw data. Dr. Abbeduto agreed and said that the agreement should not only be about pooling data, but also about how it will be used. The Fragile X investigators used their pooled data to re-norm the ABC for that population.

It would take time to pool data and use it. Drawing up a standard data-sharing agreement could be a short-term goal, with the implementation a long-term goal.

Dr. Abbeduto said that data pooling can be a multistep process, with the short-term goal being to pool data that are already available. The long-term goal would be to create a consortium to collect and share future data and to create a toolbox of measures that are stratified by age and level of functioning. It is also important to have a battery that is applicable across languages and cultures.

Dr. Urv said that a number of investigators have included individuals with DS as controls in intellectual and developmental disability studies. Perhaps those investigators could share their data.

Dr. Mervis said that it is difficult to get people with DS enrolled in studies. Her studies contain many more children who have Williams syndrome. The parents of children with Williams syndrome want to join studies, but DS families do not. Family doctors tell the DS families that there is nothing more to learn about DS and that enrollment in studies is futile.

Dr. Harpold said that the fact that family doctors are discouraging families from participating in studies shows that the three groups at this conference—medical, behavioral, and cognitive—need to work together. The doctors in the medical group could be helpful. One participant said that the families may believe that all studies are drug trials, which is not the case.

Dr. Mervis said it would be helpful to find a way to inform families that not much is known about how to treat DS children and that studies could help learn more.

Dr. Abbeduto said that he completed a study in which parents reported that medical providers had been insensitive to them, leading them to mistrust the medical community.

Dr. Harpold said that DS families do not really know what the research is about. Dr. Krinsky-McHale said that investigators need to reach out to the community and explain the research in plain language. Investigators should go to the meetings that parents and advocates attend.

Dr. Edgin said that investigators should target domains that are important to this population, including talking and sleeping. Dr. Hooper agreed and added adherence and transition as topics that would interest parents. Dr. Harpold said that quality-of-life issues are important to families.

At this point in the discussion, the group defined their short-term goals:

- Specify standards for data collection and testing. The standards should evaluate the test's adequacy and reporting. (The group would produce a paper on this topic.)
- Provide list of measures for current and imminent clinical trials. (The group would produce a second paper on this topic.)
 - Measures that are good enough for now
 - Measures that are not recommended
 - Identification of the gaps
- Conduct a survey to identify common data measures across sites.

The group also outlined its long-term goals:

- Create a toolbox stratified by age and level of functioning.
- Create a consortium to collect data.
- Create a battery that can be used across languages and cultures.

Dr. Harpold said that one way to reach parents and convince them that research can benefit children with DS is to work through groups such as the American Academy of Pediatrics (AAP), which communicates with pediatricians and publishes professional guidelines.

Dr. Ropacki said that investigators must present their efforts as something that is good for the community and families. Dr. Harpold said that investigators must address parents' underlying fears. A study will not fundamentally change a child's personality, but it may improve the family's quality of life. Dr. Ropacki said that parents might also be interested in having an in-depth evaluation of their child, which many studies could provide. CHARLOT, for example, provides information on factors like diet and depression.

Dr. d'Ardhuy suggested the group share their efforts with DS investigators in other countries; overseas investigators may be interested in participating. Dr. Harpold agreed and said the group should publish information online about their efforts and state that the group is open to collaboration in other countries. Other members mentioned upcoming meetings where it would be possible to spread the word and engage more people.

One participant cautioned that involving partners from other countries, including lower-resource countries, could make the standardization difficult to maintain.

Dr. Ropacki said he is willing to bring word of the DS effort at this meeting to the Clinical Trials on Alzheimer's Disease (CTAD) conference and asked whether the Alzheimer's community would want to work with DS researchers.

Dr. Harpold said that some would welcome the collaboration, but some would not. He suggested Dr. Ropacki make the suggestion at CTAD. That could capture a new group of investigators who could be great allies. Working on DS and Alzheimer's in tandem could be fruitful for both areas.

Dr. Reeves said that Alzheimer's disease has a lot of funding, and he urged the group to reach out to the Alzheimer's community. There is a high prevalence of Alzheimer's among people with DS, but understanding why some people with DS do not develop the disease could provide some useful information. In addition, people with DS are part of the Alzheimer's community, and there should be interest in improving their lives, too.

Dr. Lee said that the field of DS is very far behind in neuroscience studies compared with autism and Alzheimer's. There have been few imaging studies of the brains of DS children and adults. The DS field knows more about the aging DS brain because of investigations by Alzheimer's disease researchers. Part of this is the difficulty of successfully scanning DS children. Her group had a 60 percent success rate at scanning.

Dr. Krinsky-McHale said that the IRBs will not allow children to be sedated in order to undergo imaging. Dr. Lee agreed that they were not allowed to sedate the children.

The group identified the following questions in relation to possible cross-cutting (with other fields) long-term goals:

- Are there ways to engage parents in research?
- How can we advance human neuroscience?
- Is it possible to obtain a grant to address issues related to measurement in a multisite study?

Dr. Hooper asked whether they should explore the issue of the publications in greater detail and find a way to organize themselves to start that work. Dr. Lee said they may be able to publish a series of papers in a special issue of the *American Journal on Intellectual and Developmental Disabilities (AJIDD)*.

Breakout Group Final Presentations

Medical/Physical Working Group

Dr. Jeffrey M. Reznik, Urban Family Practice Associates

Dr. Reznik said that his group's short-term goal is to work out a model for diagnosis and treatment based on organ systems. Within each organ system, the group identified three or four problems, such as sleep apnea and immune dysfunction.

For example, the group identified five methods to diagnose sleep apnea, which is associated with the cardiovascular and respiratory systems. Four of the diagnostic methods need more work before they could be used in this population. One is ready to use now. The group found two biomarkers associated with sleep apnea, both of which could be tested for now. They also found three conditions associated with sleep apnea that, if treated, could help improve the condition.

Dr. Adorno said that the group would also evaluate the appropriateness of treatments for DS individuals. For example, they will determine whether DS children need a specially fitted continuous positive airway pressure mask. The group will continue to develop this model, and will do it for several age ranges because different problems characterize different stages of life.

Another goal is to partner with other groups that are working on related Medical/Physical issues.

The long-term goals are to evaluate comorbid conditions in the individual and to identify tests that are the gold standard for testing each condition. The group also hopes to link comorbidities to phenotypes. For example, whether a DS individual who has a congenital heart defect would be more likely to have sleep apnea or score lower on cognitive measures. Clarifying these relationships could suggest new treatments and new areas for basic research.

Another long-term goal is to identify problems in aging DS individuals and develop guidelines for treatment in consultation with other groups. Finally, the working group would like to apply its model across all organ systems.

Dr. Worley said that the group also identified the following pathophysiological traits that could be primary outcome measures for drug studies:

- Dysmorphogenesis
- Inflammation
- Immune abnormalities
- Free radical injury
- Decreased energy production
- Metabolic abnormalities
- Stem cell depletion

Dr. Urv asked what the group meant by “measures” when they said “partnering with other groups developing measures.” Dr. Reznik said they meant partnering with other groups that are doing research. Dr. Videlefsky said that they propose collaborating with other organizations to develop guidelines for taking care of adults with DS. There are pediatric guidelines but no adult guidelines. The second part of that is to look at current research studies that include biomarkers, diagnostic tests, and clinical outcomes to see how they can build on those studies. The guidelines are for clinical care, which will drive clinical trials to make them clinically relevant. Dr. Parisi further clarified that the group will identify the evidence base for clinical guidelines developed so far and, if there is no evidence base, determine the outcomes measures that should be developed to confirm whether those guidelines make sense.

Dr. Parisi also said that there are studies of biomarkers of Alzheimer’s and DS already underway. The working group hopes to partner with those groups to learn what they have found out and to tap into that knowledge so that nobody needs to reinvent the wheel.

Dr. Ropacki recommended internal collaboration among the three working groups. A participant asked that Dr. Ropacki share what his research group has found so far in using his biomarkers on the DS population.

Cognitive Working Group

Dr. Nancy Raitano Lee, Drexel University

Short-term objectives:

1. Specify principles for standards for data collection and evaluating measures for adequacy and reporting. The group would produce a paper on this.
2. Provide a list of measures for current and imminent clinical trials. The measures would be classified as “good enough for now” or “not recommended for use.” The list would also identify gaps where there are no appropriate measures. The group would produce a paper on this.
3. Identify what measures are being used across research sites. This would entail developing a survey to find out what other sites are using. It would be preferable to use outcome measures that have norms for DS, but the first step is to find out what investigators are using.

Long-term goals

1. Create a toolbox, perhaps using domains of functioning, and stratified by age and level of function.
2. Create a consortium to pool data across sites.
3. Create a battery that has applicability across languages and cultures.

Cross-cutting long-term goals:

1. Engage parents of DS children in research.
2. Advance the neuroscience of DS in humans, particularly in children.
3. Obtain a program project grant to address issues related to measurement across domains and multiple sites.

Dr. Abbeduto explained that they decided to rate measures as they did because the measures that investigators use have not been standardized. The working group hopes to set some standards for the measures that will be used, such as defining an acceptable level of TRT and the minimum window of time between test and retest. The group would also make recommendations for test manuals, training test administrators, and reporting details of testing, including how many subjects there were at floor and how they were recruited.

Dr. Conners explained that the group settled on the “good enough” category for tests because members felt it was important to present some measures that could be used, even if they have some weaknesses.

Dr. Abbeduto said that he would be happy to facilitate a group if others were interested in partnering on outcome measures.

Dr. Parisi asked about the first short-term objective and whether the paper the group planned to write could include the behavior group. Dr. Lee said that her group had talked about integrating their work more closely with the behavior group. Dr. Parisi said there might be ways to integrate work on the biomarkers and imaging from the medical group. Dr. Lee agreed.

Dr. Harpold said that it is important that participants agree about where things stand in DS research, what steps should be taken next, and what benefits can accrue from the next steps. Having everybody on the same page will help in a variety of ways, including making it possible to engage more with families.

Dr. Handen said that it would be helpful if the work of the groups had some common cores, such as defining which tests are appropriate to use within which age ranges. The three working groups should use the same age ranges.

Dr. Hooper said that the cognition working group knew that there is no “one size fits all” in testing. That is why their first step is to define what they want in a measurement tool. Then the group could identify measures that could be used for different age groups. Dr. Lee said that members of the group work across the lifespan and are considering age ranges as they proceed.

Behavior Working Group

Sigan Hartley, Ph.D., Waisman Center, University of Wisconsin

Long-term objectives:

- Identify current or developing technology to provide naturalistic measurement of target concepts, including tests such as LENA (Language ENvironmental Analysis)
- Expand psychometric properties, sensitivity to change, and normative data for key measures in DS.
- Apply principles of advanced quantitative analysis to best characterize change in clinical trials.

The field has many paper-and-pencil tests. The behavior group recommends tests, such as LENA, that involve using newer technology. There are also new multisensor wristbands that could be used in trials with DS individuals to track repetitive behaviors.

The group suggests pooling data to develop better norms. There is also a need for measures that give more than a global score, because a global score gives little information about the variability within or between individuals. In addition, more sensitive measures are needed to track changes in clinical trials.

There is a need to know the natural history of DS and identify developmental changes, as well as to embrace more quantitative strategies. There is an opportunity to combine longitudinal and cross-sectional data to answer some of these questions.

Dr. Capone presented the short-term goals, which were summarized in the following:

1. Short Term Goals

- Stratify concepts by age to evaluate measures accordingly

Age	Concept	Self-Regulation/EF	Maladaptive Behaviors	Psychiatric/Mood	Sleep
0–5 years	Development Early Maladaptive	EF- lab BRIEF-P	VABS		Actigraphy, CHSQ, PSG
6–12 years	ADHD, Ext. ASD, Int Puberty	BRIEF P BRIEF	ABC, Conners, NCBRF, VABS		Actigraphy, CHSQ, PSG
13–35 years	Mood, Anxiety, Psychoses, OCD, Movement	BRIEF	ABC, Conners, NCBRF, VABS	ADAMS BFRC/PNSS NPI	Actigraphy, PSG
35 years & older	Pre AD/AD		VABS	ADAMS BFRC/PNSS NPI	Actigraphy, PSG

- Take into account specific behaviors (RB) and conditions (AD & Dementia)
- Identify measures Ready to Roll and dro

The chart contains some major themes in DS, such as self-regulation, maladaptive behaviors, mood, and sleep.

The behavior group stratified the concepts by age ranges, choosing birth to 5 years old, 6 year to 12 years old, 13 years to 35 years old, and over 35 years old. The group identified some of the instruments that could be used in a clinical trial for those domains and within those age groups. The tests could be used to measure the effect of an intervention on a target behavior or to screen patients in or out of a study based on behavioral criteria.

There is a need to collaborate with the medical and cognitive groups. Sleep issues are common to both groups.

The scales listed in the slide provide an impression of behavior for the previous few months. The behavior group suggests gathering real-time data over a 48-hour period. Real-time monitoring would capture behaviors not captured by current rating scales and would help build a richer data set.

Other short-term goals included working on collaborative efforts.

- Identify additional colleagues to join the Behavioral Working Group, including parents, DS experts, and topic experts.
- Identify collaborations with cognitive and medical groups by concepts and age, such as the following:
 - Common data that can be combined for advancing measures;
 - Sleep apnea and behavioral outcomes; and
 - Biomarkers and behavioral outcomes.

Dr. Dykens said that it is essential to build collaborations with parents, other DS experts, and with experts in related fields, such as catatonia. A program project grant would be one way to do this.

DS-Connect might provide a way to disseminate information, involve parents, and identify collaborations. The three working groups at this meeting could also determine what data they have in common and identify a systematic way to share data. One way to do this might be to use a REDCap survey.

Dr. Rafii went through a variety of clinical biomarkers that could be used in clinical trials of DS, including heart rate, electrodermal response, inflammation, and oxidative stress. Research in Alzheimer's disease has already identified some biomarkers using imaging and metabolomics.

Dr. Hartley said other short-term goals include connecting clinical trials to the real world and developing ways to include parent ratings in clinical trials. Clinical trials should be accessible to a wide range of DS individuals, regardless of their reading or language levels, and should cross national boundaries. The outcomes of clinical trials should have relevance to clinical practice.

AJIDD is planning a special issue to focus on the DS natural history. Participants should think about what data they have available that can help answer questions about natural history.

Dr. Dykens said that all federally funded autism research data is shared in the National Database for Autism Research. The behavior group suggests creating a similar database for DS that would include phenotypic data. (There is already a database containing genetic information, dbGaP.)

The final suggestion is that existing epidemiological resources for people with DS be linked. For example, some states collect a wide variety of health information. Electronic medical records also have rich data sets. It is possible to get information on people with DS through these and other sources. There may also be federal repositories that can be mined for DS-related data.

Summation of Day 2 and Next Steps

During this portion of the meeting, the three working groups were asked to discuss the aims that could include all three groups.

Dr. Capone said that collaboration across the three groups was the first suggestion and will require ongoing discussion after the meeting ends.

The working groups could publish papers together. The cognitive group suggested papers that could involve other groups. The paper that came out of the Fragile X meeting could be a model. Keeping the working groups alive is very important because members have already started the effort to identify DS outcome measures.

Dr. Hooper said that laying out the standards by which to evaluate outcome measures would provide a sound foundation for the work that working group members will do next.

Dr. Urv said it took several years for the Fragile X paper to reach publication and asked whether it would be possible to publish a paper out of this meeting sooner, suggesting that an interim step could be taken, such as publishing a short paper on the NICHD website.

Dr. Abbeduto said he would not favor doing a quick paper for the web, because it might be necessary to backtrack from it later. The first paper on standards (TRT, reliability, etc.) is the most important and could be completed within 8 months if about six people work on it.

Dr. Hooper said that the *AJIDD* timing might drive how quickly the paper comes out. Dr. McKrinsky McHale said that the group will need funding to continue with this effort.

Dr. Urv said that publishing the papers will have to be a grassroots effort and that the funding will follow the papers. Study groups are more likely to look favorably on a grant application when there are papers to back it up.

Dr. Abbeduto said that the Fragile X paper did help the group develop outcome measures and to obtain funding after the paper was published. The paper legitimizes the group and gives it a structure. A group of about five people can accomplish this, although it would be best if Dr. Urv were to act as the organizer and continue to encourage forward motion.

Dr. Reeves said that the Trisomy Research Society has a conference coming up that could provide an opportunity for working group members to present their work and to network with DS researchers from other countries.

Dr. Harpold said that he likes the idea of posting a preliminary report online, because it might prompt more people to get involved in the effort.

Dr. Urv said that the groups could post a short summary on the NICHD website and invite others to get involved. Dr. Abbeduto suggested summarizing the short- and long-term goals the groups had developed, but others rejected the suggestion because it might tie the groups to the goals. Dr. Esbensen suggested that information could be posted to the DS-Connect™ site. The information also could be sent to email listservs that include people from the DS community.

If the aim is to publish the papers as quickly as possible, the *AJIDD* special issue is not a good choice, because it will take time to develop that issue.

Dr. Urv said it might be a good strategy to publish in journals that people doing clinical trials, including those in the pharmaceutical industry, often read. Dr. Esbensen said that the members of the cognitive

group would likely read *AJIDD*. If the first paper has to do with measurement, *AJIDD* would be a good place to publish. Dr. Urv agreed, but said it is also important to reach people who are running pharmaceutical trials and those who are outside IDD.

Dr. Harpold said that it would be good to work on issues of obtaining consent from research participants. Researchers could make good use of guidance that addresses how to get a wide consent that could cover a range of data uses and cover participation in future studies.

Dr. Urv agreed, saying this is one of the larger cross-cutting issues that the group did not have time to address. Other issues that could not be addressed in the limited time were common data elements, natural history, and registries.

Dr. Visootsak asked how the pharmaceutical industry can help to advance this field. Dr. Urv said that the industry can help by educating DS researchers about the industry's needs, Dr. Capone said that the industry could help fund the field testing of measures for clinical trials, and Dr. d'Ardhuy said it would be valuable to share data sets.

Dr. Parisi said there is an opportunity to collaborate with groups developing biomarkers for some subpopulations, such as the aging population. Dr. Videlefsky asked whether there any free software programs that would assist collaboration. Dr. Dykens said that REDCap is a secure, web-based, easy-to-use free tool that is can capture data, run surveys, and schedule patient flow. Dr. Esbensen said that Medrave, for medical information storage, is another option. Dr. Worley suggested that the electronic medical record Epic is a good way to share, and another participant said that an effort to streamline the extraction of information from medical records is underway.

Dr. Edgin asked that Dr. d'Ardhuy make a list of the types of measures that the pharmaceutical industry would like DS researchers to use. These include tests that have an alternate form and that exist in multiple languages. Dr. Visootsak said that she and Dr. d'Ardhuy may contact a group of meeting participants to share that information.

Dr. Parisi said that some participants appear to be suggesting a consortium effort to collect natural history data and other sorts of data.

Dr. Urv said that they have a long list of things to get done. She asked that participants contact her about working on the papers so that they can begin the next phase. Dr. Urv also thanked participants for their hard work leading up to the meeting and in the meeting itself.

The meeting adjourned at 1 p.m.