UNITED STATES FOOD AND DRUG ADMINISTRATION

- - -

ADVANCING THE DEVELOPMENT OF PEDIATRIC THERAPEUTICS WORKSHOP

- - -

Friday, April 17, 2015

White Oak Complex

10903 New Hampshire Avenue

Silver Spring, Maryland, 20993

The meeting was convened at 8:00 a.m.

MEMBERS PRESENT:

LYNNE YAO, M.D.
PETER COMO, M.D.
MARILEE ALLEN, M.D.
GAHAN PANDINA, PH.D.
RICHARD GARSHON, PH.D.
NATACHA AKSHOOMOFF, PH.D.
PHILIP SHERIDAN, M.D.
R. DANIEL MELLON, PH.D.
HEATHER ADAMS, PH.D.
ELSA SHAPIRO, PH.D., L.P.
ANN MCMAHON, M.D., M.S.

P R O C E E D I N G S

[8:00 a.m.]

WELCOME AND INTRODUCTORY REMARKS

DR. YAO: Good morning, folks. We're running a little bit late. We heard that there was bad traffic on the way to FDA this morning. And so what I would appreciate especially those of you -- those of you on the panel. We're going to take this opportunity to take five minutes and if you would like to go to the kiosk to preorder your lunch, then you won't have to wait in line, and you'll have it available. So if you'd like to do that now, we'll take five minutes. Those of you in the audience, you're welcome also to do the same, but we wanted to make we had the panelists able to do that first. Thanks.

[break]

DR. YAO: Good morning folks. I think we're going to go ahead and get started. I know we're running a little bit late and the -- we know that there was police action in one of the major arteries to get to White Oak. And we've already heard that one of our speakers is delayed. It sounds like he's stuck in that mess right now. So we wanted to give folks a chance to assemble. But I think we have the majority of folks we need to begin, so that's what we'll plan to do. So good morning, everybody. My name is Lynne Yao. And this is the wrong set of slides.

[laughter]

And we have the introduction. And we're really pleased to have -- I'm speaking on behalf of the organizing committee-- to have worked with our colleagues in the Division of Gastroenterology and Inborn Error Products to come up with this two-day set of workshops to try and advance some of the issues that are related to product development in children. And one of the things that rose to the top in our area of concern was really, how do we address evaluation of pediatric patients who have diseases that affect their neurocognition or behavior? And indeed, if they're taking a product that might affect their nerve cognition and behavior, how do we go about understanding those affects? So in thinking about that, that's how we sort of intended to set up the day.

I think underlying -- and I take every opportunity possible to present this -- that in general, when we think about product development in children, we believe strongly at FDA that pediatric patients should have access to products that have been appropriately evaluated. The pediatric patients deserve no less than what we would require and expect for any adult patient. And therefore, a pediatric product development program should include pediatric studies when use is anticipated. And again, this is from a guidance that is a global guidance document. We have two pediatric development, product development laws that we

use to guide all clinical studies that are related to product development in this country, The Best Pharmaceuticals for Children Act and the Pediatric Research Equity Act. And we've developed, I think, a very strong -- strong programs to get pediatric studies done because of these two laws. And in fact, since the laws were enacted in -- well, the first forms of it in 1997, but BPC in 2002 and PREA in 2003 -- we've had over 500 labeling changes related to pediatric information, use information. And that average is again about 37 labeling changes a year. Now this is -- this is compared to literally less than a handful of FDA-approved labeling that had information prior to 1994. So we really have come a long way. And this is just a graphic representation of that.

So as we've evolved pediatric product development, I would say in the 21st century based on these two programs, we understood that there was a need to really have conversations that were more related to pediatric-specific problems. And so we came up with this idea at FDA a couple of years ago -- just about two years ago -- to develop a series of workshops called ADEPT, or Advancing the Development of Pediatric Therapeutics. So one of the things that I wanted to highlight here is -- in terms of the objectives of these ADEPT workshops -- is that we really want to focus on development issues specific to pediatric patients. We want to make sure that we've included all the

important stakeholders in these discussions, academic, industry, patient, and professional groups. This is not intended to be an advisory committee meeting. So we are not seeking official advice; we are not seeking official advice, nor are we discussing specific product development programs except to provide illustrations. And I hope that folks understand that this is really to spur the discussion and hopefully come up with some reasonable next steps.

So what are the objectives for today? They're outlined here, and I just want to point out that in your agenda, you should see all of this information highlighted, so I'm not going to go through with it. What we've outlined today -- and our session chairs -- to set the stage for conversations in the afternoon. Session one is really intended to understand what are the current -- what's the current state of measurement tools to assess neurocognitive and behavioral development in children? Session two is to provide some examples where we've seen safety signals in animals or humans that alert us to potential needs for further study in children. And then Session three we feel like is going to be the meat. How do we study if we've identified an issue? How do we go about studying children?

Housekeeping. Order your lunch now if you can. The kiosk is right outside these doors, and it's $11. They'll take payment now. And then at lunchtime, all you have to do is show

them your little coupon, and you can get your lunch. Cell phones and noisy devices please, if we could turn them off now. Facilities and amenities, bathrooms, et cetera are all outside here, and folks out in the -- in the foyer can help direct you. And please see our staff outside. Denise Pica-Branco is over in the corner if you have a question in the room. And we have staff outside at the registration table for any questions that you might have.

Finally, a thanks to all of the distinguished panelists and speakers we have assembled today. I think it's no small feat that in your busy schedules, we've been able to get all of you together. And a special thanks to the planning committee, especially our outside planning committee members Heather Adams and Elsa Shapiro. Our project management staff Denise Pica-Branco. Two of the Division of Neurology Product staff Len Kapacala, Phil Sheridan, Peter Como for the Center on Devices. And then our OPT colleagues Diane Murphy and Ann McMahon. So with that, I would -- as we're getting the slides up -- as you can tell, I'm delaying just a little bit because we still have one speaker who's stuck in traffic. But in the meantime, if we could just briefly, if folks could -- we'll start with Phil. If you could just go around, I'd like to introduce our panelists. Phil, if you could just introduce yourself and where you're from, and we could go around the room.

Thank you.

DR. SHERIDAN: Hi, I'm Phil Sheridan. I'm a pediatric neurologist who worked at the NIH for a number of years with the Antiepileptic Development Program and the Development on Neurology branch. The last 15 years, I've been here with the Division on Neurology Products.

DR. YAO: I'm sorry. So panelists, if you push the red button --

DR. ALLEN: All right.

DR. YAO: The microphone will turn on.

DR. ALLEN: I'm Marilee Allen. And I'm a neonatologist and development pediatrician at Johns Hopkins and the Kennedy Krieger Institute in Baltimore. And I'm just here -- I'm here to give talks about early neuromaturation.

DR. WABER: I'm Deborah Waber. I'm a developmental neuropsychologist at Boston Children's Hospital and Harvard Medical School. And I'm here to talk about some of my long-term work and late effects of leukemia treatment in children.

DR. TOWBIN: I'm Kenneth Towbin. I'm a child and adolescent psychiatrist in the intramural program at the National Institute of Mental Health. My name tag is spelled T-O-B-I-N, but in order to protect whoever Kenneth Tobin without the W is, it's actually T-O-W-B-I-N. So cast your aspersions to the correct person. Thank you.

DR. COMO: [laughs] Good morning, Peter Como. My name tag got lost from yesterday's sessions. I'm a clinical reviewer in the Division of Neurological and Physical Medicine Devices in the Center for Devices and Radiologic Health. And I'll be chairing this morning's session.

DR. GERSHON: Richard Gershon, Feinberg School of Medicine at Northwestern University, and I'm here representing National Institutes of Health Toolbox for assessment of neurological and behavioral function.

DR. AKSHOOMOFF: I'm Dr. Natacha Akshoomoff. I'm a pediatric clinical neuropsychologist from the Department of Psychiatry and Center for Human Development, University of California, San Diego.

DR. SHAPIRO: Elsa Shapiro. I'm a pediatric neuropsychologist from the University of Minnesota. And I've been working in the area of inborn errors and metabolism for many years.

DR. DELANEY: Hi, I'm Kate Delaney. I worked at the University of Minnesota in pediatric neuropsychology for 17 years doing assessments of children, mainly with rare diseases.

DR. ADAMS: Hi, I'm Heather Adams. I'm a pediatric neuropsychologist at the University of Rochester Medical Center. That's Rochester, New York. And I work in Batten disease. That is a rare neurodegenerative disease of childhood, but using that

as a platform to engage in clinical research and other rare diseases of childhood.

DR. BARON: Good morning, I'm Ida Sue Baron. I'm a clinical neuropsychologist. I'm in private practice in this region, but I'm also the director of an outcomes research program on preterm birth at Inova Fairfax -- Inova Children's Hospital, Inova Fairfax Hospital.

DR. MELLON: Good morning, my name is Dan Mellon. I'm the pharmacology toxicology supervisor in the Division of Anesthesia, Analgesia, and Addiction Drug Products here in CDER. I'll be providing some insights into some of the non-clinical studies that we can do to get a better understanding of how compounds can impact the developing brain.

DR. BULL: I'm Marilyn Bull. I'm a neurodevelopmental pediatrician from Indiana University School of Medicine, Riley Hospital for Children. And I'm here to talk about the neurocognitive outcomes of the following environmental exposures.

DR. ELAYAN: I'm Ikram Elayan. I'm PHARMTAX reviewer in the Division of Psychiatry. I work with a lot of pediatric indications like ADHD, autism; review non-clinical data. So I'm here just to be on the panel.

SESSION I: NORMAL NEUROCOGNITION AND HOW IT IS MEASURED IN CHILDREN OF VARYING AGES

NORMAL NEONATAL AND INFANT NEUROCOGNITIVE DEVELOPMENT AND MEASUREMENTS IN THIS AGE GROUP

DR. YAO: Thank you, presenters. So our first presentation will be from Dr. Allen, welcome.

DR. ALLEN: Thank you for the opportunity to speak here today. I want to say straight off that I'm a little intimidated by all the psychologists in the office -- in the audience and in the speaker -- of the speakers because I know they know a lot more about evaluating neurocognition in older children. And I thought I would just start off by talking a little bit about where I'm coming from and what I'm going to try and do this morning. So I trained first actually as a neurodevelopmental pediatrician at the Kennedy Krieger Institute in Baltimore. And as I finished my fellowship, I was offered a job to do the neonatal follow-up over at Hopkins, which is across the street. And they encouraged me -- since I did take that job -- they encouraged me to also do a fellowship in neonatology so that, you know, I could understand the issues better. And for that I'm very grateful.

I've been doing neonatology full-time since then for

many, many years. And in addition, I've had the privilege of following a lot of the children that we've taking care of in our NICUs and their families over the years. So my insights are really from examining hundreds and hundreds and hundreds of babies and as -- especially during their first -- seeing them serially during their first years of life. So my objective today is to describe neuromaturation and early neurocognitive development and describe its measurement in neonates and infants and to discuss the many limitations of measuring early neurocognitive development. And I guess I can start right off the bat by saying that there are no good measures of neurocognitive development in the newborn. And it really takes watching them grow up and following their development over time.

But what I want to do is talk a little bit about the early development of neurocognition and some of the ways that we've tried to measure it, or at least tried to follow it. So this -- I always put this slide in my talks to remind myself that the babies that I'm seeing are on this continuum of development. And we have the ability to monitor the fetus a bit in utero because of the use of ultrasound. The preterm infant is actually an opportunity to follow neurodevelopment during what would -- should be the third trimester of pregnancy. And then the full-term newborn infant, I've started out by being taught that, well, you can't really tell anything about a

newborn and found, actually, there's a lot of early literature about how we can exam, evaluate newborns. And I've certainly learned a lot more since then. And then I'll talk about early infancy, but will not bring us up beyond that because the next speakers will be doing that. So as I said, I've been doing this a long time. That was before I had a little bit of gray.

So early brain development, so the brain's really complex organizational structure occurs during embryonic and fetal life. The -- then it grows very rapidly, and the human brain is actually 70 percent of its adult size by one year and 80 percent by two years. It's the neuronal circuits that are the mediators of the brain's diverse functional capacities. And the -- it's important to recognize that the structure and function of neuronal circuits continually changes and evolves from the time of the first contact between nerve cells throughout our lifetime. So there's tremendous number of developmental -- brain developmental processes that are going on during that third trimester and in early infancy.

The basics of cell proliferation, migration, aggregation, differentiation, and then glial cell development -- they support the neurons; the process formations, so they grow out their exons and dendrites; and then the one that I think has been receiving a lot of attention in basic research, synaptogenesis, so the connection of neurons to form these

neural circuits. And as part of that process, there are many, many connections that are formed, and then the ones that aren't used are -- die out. And programmed cell death is part of the developmental process, and that's called apoptosis. So there's a refinement of the synaptic networks that form according to how they're used. And then myelination, which continues well beyond infancy, is the -- development of the protective sheath around the neurons that make them faster and more efficient.

So my interest has been in neuromaturation, the functional development of the central nervous system, which is a very dynamic process by definition. And neuromaturation results from the continuous interactions between genes and, first, the intrauterine environment, and then the extrauterine environment. And of course, it's driven by genetic processes that are encoded in the DNA. So there's a really intimate relationship between central nervous system structure and function. And there's a very critical role that's played by movement -- fetal movements, sensory input, responses, in shaping the synaptic networks and also in organ and limb formation in the fetus, and then refining and pruning those neural networks to form. And the richer the environmental input, the greater, the faster, and more flexible the circuits.

So I'm going to start out talking about the very immature infant, the preterm infant, who wasn't really ready yet

to be born, and the kind of environment that the -- these babies are cared for. So they require a lot of support for their very immature organs and it's -- I can tell you we're not as good as the mother's uterus is in terms of supporting these babies, even their basic organ functions, but certainly their neuromaturation. So by the time that these preterms come to term or the full-term baby who's born on time -- so we have a newborn infant, and the brain circuitry is shaped by the complex interplay between -- again, the same thing, but during -- now on through infancy, the complex interplay between genetic processes, experiential inputs, and behavioral responses. So the newborn is not a passive recipient of input, and the newborn baby's behaviors influence her environment and illicit our input to them. These inputs and the baby's responses then shape the developing neuronal circuits in that baby.

So a newborn baby, you can't just walk by a baby like this. You're going to have some input into this screaming baby if, you know, if you see them, and that's the kind of input the baby is trying to illicit. Unfortunately, it's not all positive and supportive input, but that's another whole story. So there are multiple events that are taking place in brain development, and I just really like this slide because it gives you a sense of how very much is happening before birth is to occur, the term birth occurs. But even after birth, there's ongoing brain

development. The myelination, which is that protective sheath and the synaptogenesis, so creating and recreating and refining these neural networks and as part of that, again, there's this process, a programmed cell death, and there's a lot of basic research now looking at that whole apoptosis and its role in normal development.

So I'm going to talk about what we know about assessing neuromaturation and the insights that I've gleaned from my many years of examining babies. And it's very clear that normal or typical development proceeds in an orderly, sequential manner according to age. And for definitions, we use I think is important to share. So gestational age is the duration usually of the pregnancy, so a mother in an obstetrician's office, they'll talk about gestational age. And usually in a baby who's been born, when we talk about gestational age, we're talking about their gestational age at birth. Chronologic age is their age from birth that we all know. Postmenstrual age, which in the past has also been called postconceptional age, is that sum of gestational age plus chronologic age, and it's a measure of how long that infant has been developing.

The -- when we see these children in follow-up after birth, it's really too cumbersome to use postmenstrual age, which is usually discussed in terms of weeks, so we use their --

well of course, we always calculate their chronologic age, but in clinic, we also always calculate their age corrected or adjusted for their degree of prematurity, and that has been called the corrected age or their adjusted age. So as if you're calculating from when they were due to be born, not when they actually were born. And then of course, it's very important in the first few months of life, but the older the child gets, the less important it is to do that correction. So one of the things that's very clear is that there's individual variation and timing. So a baby who -- well, just like a 10-year old can be tall or short for their age, they can also be more mature or less mature for their age.

The same thing is true at every week of gestation. So there are some babies who are born at 26 weeks gestation, who are a little more mature than you'd think they should be, and those of course are the ones who are more likely to survive. But there are babies who are born at 26 weeks gestation whose skin is still pretty gelatinous, and their lungs are -- it's very difficult to ventilate them, and those are the babies who are more immature than expected for gestational age. And I've seen enough cases in which we know absolutely when implantation took place from artificial reproductive technologies that -- or assisted reproductive technologies -- that -- to know that yes, those babies truly were 26 weeks gestation, but they were

immature for their age. And that's part of the individual variation and timing that we see at every age.

And as in all of development during infancy, what is normal at one stage is very abnormal at another stage. But because development moves so quickly, it's really important to know their developmental stages, what's typical development in order to evaluate them. So this is the slide that I just put together. You're really not meant to read any of this. But it's -- I put it together to just kind of outline neuromaturation and how it develops in the preterm infant. Here if I can figure out the pointer. Here we go. Okay. So the -- how it develops by age first using their gestational age for the preterm infants and then -- or -- oops I meant to be -- I want to go back. I didn't mean to go there.

The -- and then the -- by term -- so the typically developing, preterm infant whose been pretty stable since birth looks a lot more like the full-term infant than you'd think. So there are things that are little bit different about their development, but for the most part, they would fit into these boxes. And just if you -- if we just take this one column, movements, the movement of a fetus or very, very preterm infant is very jerky and usually is primarily the large-limbs that are moving, or whole-body movements. And the movements become more refined and smoother with neuromaturation. So that by term, the

term baby has individual finger movements and really fairly elegant smooth movements. And then if we follow that baby out during the first few months of infancy, by around four to five months of age, babies have -- are able to reach and grab for objects.

But they also have, what's been described by Heinz Prechtl with his general movements score looking at the development of movement in babies, fidgety finger movements. And that's when they have kind of constantly moving finger movements, and these are normal. So they might be signs of a motor abnormality in an older person, but in a baby, those are normal. It's part about -- it's part of that whole process of realizing they have fingers and developing the ability to control them. So it's actually a marker of normal development to see these fidgety finger movements at four or five months from term. And then as they get older they are -- become more sophisticated at manipulating toys. And that's just an example of one aspect of their development. And as you can see, some of them -- some tone develops. Premature babies are very hypotonic until they start getting closer to term. So different aspects of development occur at different ages.

So what about cognition? Well, I wish it was this easy, but it's not. The -- one way to get at cognition is to look at their sensory responses. So babies as early as 24 weeks

gestation have auditory brain stem responses that can be recorded. So we know that the sound goes to the brainstem by that age, and they usually have some response to sound. They -- babies will blink to light as early as 23 to 26 weeks plus menstrual age, and that's even if they still have their eyes fused. They start responding to visual patterns; they like black/white contrasts as early as 32 weeks postmenstrual age. And their visual acuity and their visual attention rapidly improve towards term, so that by term, a newborn should be able to fixate, follow -- and follow an object, at least briefly. And they have a fixed -- because of the way their eye develops, they have a fixed focal length of about eight inches, which is right where a mother's face would be if she's holding a baby in her arms, and anything closer or further away will be more blurry to them.

But nonetheless, things are blurry to babies because they only have about 20 -- 200 or at best 20-over-150 visual acuity, and that rapidly improves, so that their vision is more like 20 over 40 by six months. They don't usually blink to light, which is a way that we're -- we use to look at whether or not a child sees, until about two or three months. So babies in the NICU, if you do a visual threatening -- well, they blink to light very early -- I'm sorry -- blink to a visual threatening gesture doesn't come in until two to three months after term.

So the sensory system really develops pretty quickly. The -- other aspects of neuromaturation take a bit longer. So we're used to looking at neuromaturation in babies in terms of their -- let's see -- in terms of their motor abilities. And that really progresses very rapidly especially over the first year.

Their fine motor abilities come in a little bit more slowly. And -- but continue -- actually, I didn't bring this out, this graph out beyond four, but it continues into adolescence. Fine motor and function can improve until adolescence; it's further refined. So this is a graph I made up myself, so it's kind of my concept of development. But a language takes a lot longer to develop as do the visual problem-solving abilities. You have to be able to manipulate objects for really -- to look at -- for us at least to look at how babies solve problems. And that experience of manipulation helps mature that part of the brain when they're ready.

So cognition is hard to assess partly because it hasn't developed yet. I think in the very young babies, looking at their visual ability -- their visual responses, their use of their hands, and their -- and looking at their acquisition of language is probably the best way we have at looking at their cognitive development. So early cognitive development really requires -- for it to be typical, requires a lot of different input. It requires input by all the senses. It requires some

mobility. You've got to be able to move around because the babies learn by experiencing their environment and also by manipulating objects. They develop fine motor function, but they also start understanding what it means to put something, you know, a block into a cup. And that sensory-motor integration is that visual -- is that processing of all these sensory stimuli and making sense of it in the brain and coordinating it with their motor movements.

And of course one of the most important things is communication, and that is how we generally assess cognition is by using communication with people. So initially for babies, it's non-verbal, but they do start with the acquisition of language as early as cooing by two to three months and then babbling, which is repetitive consonant sounds by six months, and we can follow that. They start responding to gesture as early as nine months, and by a year, they should be following gestured commands, like, "Come here," using your hand to demonstrate what you want. And within a couple months, they should be able to follow commands without your using the gesture. So those are all examples about the developmental acquisition of language and what we can look at.

The infant also learns by -- so experiencing the environment is really important for learning and then adapting to it. And there probably are critical periods of vulnerability

and -- which I could also say are critical periods of opportunity for working with babies during these critical times of their development. It doesn't mean they can't learn it later, but they're most receptive to learning at certain stages of their life. And I think an essential feature of cognition is flexibility. And I certainly haven't figured out, you know, mental flexibility. And I certainly haven't figured out how you measure that in a newborn.

So much of the audience probably already know most of these. This is just my attempt to put together a number of set -- cognition is the set of mental abilities and processes that are related to knowledge. And these are the things that we try and measure in older people, older children and people. And I like this graph; I just pulled it in to show this. And that's -- so executive function, which is how we organize ourselves, how we figure out what to do and monitor our progress, really doesn't develop much until the child at least is over a year. So there aren't ways really to look at executive function, which may be one of the most important aspects of neuromaturation of the brain. So there are a number of methods of examining babies, and I have been using the method described by Claudine Amiel-Tison, a French neonatologist and neurologist.

There are many others that have been used, and what they do is help select babies who are at the higher risk --

highest risk for having developmental problems. And this is just -- this slide just kind of lists some of the major people in the field who have developed methods of evaluating babies, starting with their classic assessment of deep tendon reflexes and then some of the primitive reflexes like the Moro, which is a startle reflex. And then there's a group of French neurologists who described muscle tone, the evolution of muscle tone in a newborn infant and then in preterm infants. And then at the same time, they also described responses that babies make to a stimulus, and Geselle's started that work in the 1930s. And some of that work has been further refined by Brazelton and others, and I'll just talk about some of the behavioral measurements in this minute. And then there've been a number of people who've looked at preterm neuromaturation over time and have tried to describe that.

So this is a slide from an early study that I did looking at -- looking for abnormalities in the preterm infant at term and looking at how that influenced their motor outcome. And there are many studies using other exams; there's nothing unique about this. But the really interesting thing I learned -- couple things that I learned: one is that if the exam is normal, in other words, we have some evidence that there's neurologic integrity, that child is unlikely to develop major disability either cerebral palsy or intellectual disability.

Here I have in terms of motor outcome, but I found the same thing in terms of cognitive outcomes. The more -- and that's -- I'm sorry, I have problems with my pointer here, so -- that's the far left column, but what's really interesting to me is the far right column.

So those are babies that were very, very abnormal on exam; it was not a subtle pickup. But only 50 percent of those children had cerebral palsy, and only about 30 percent had minor neuromotor dysfunction, which is when you have abnormalities on exam; you may be a little clumsy, but you're able to walk and function. And about 20 percent were normal when they were evaluated between one and two, so it doesn't mean they'll do fine in school, but in the first year, their development looked pretty good, and to me, that's just remarkable. And it's, I guess, one of things that's so rewarding about working with newborns is that they can have significant abnormalities, but there's that hope that they'll improve.

So I actually developed using -- really drawing on all of those people on that previous slide and the work they've done, but I've put together a maturity score that measures a degree of maturation. So on the slide where I showed all those things that change over time, putting all those items together and coming up with a maturity score. And this is the study where I looked at it and low preterm infants over time and found

that -- and these for kids who turned out to be normal, and these -- what I found was that there was, as expected, an increase in score with postmenstrual age. So we did another study looking at a larger group, and we found a correlation -- and I'm going to focus on neurocognitive development in the middle -- in the middle section of the graph -- where if they had -- the babies who had low MDIs, mental developmental indices, at 1.8 to 2 years, those babies all had lower -- mostly had lower neuromaturation scores or maturity scores, lower rates of neuromaturation, and then the babies who had MDIs over 70. This is -- I've looked at it in terms of maturation, but similar work is shown with, again, with abnormalities on exam.

So we can pick out the babies who are at risk for developmental disability, both cognitive impairment and motor impairment, but we can't in any way diagnose even cerebral palsy, let alone cognitive abilities in a newborn infant. So I put this slide in because I think if there's anything that we can use to look at cognitive development, this ability to look at visual fixation following visual attention is probably the most promising. And this work has been -- there are many people who are carrying out this work in an experimental manner. I use it as part of the Amiel-Tison exam when I examine babies. And I'm -- of course, it's very important to get them into the state of alertness, so that you can assess this, but it has been very

promising I think.

So there have been a number of assessments of neonatal neurobehavioral of a newborn and of also of preterm infants. They're -- pretty much most of them focus on description of that behavior and using that information for counseling people -- parents, sorry. The NICU neonatal network, or NNNS, scale was developed as part of the Maternal Lifestyle Study and looked at -- so babies who had been exposed in utero to narcotics and other illicit drugs and used it as a way of comparing babies after they were born. So some of these are being used to look forward as to how predictive they are, but I think that work, too, is still in progress.

So when we get to infancy and early neurocognitive development in an infant, remember it's a dynamic process, so we're just looking at snapshots in time. So much of it is nonverbal, but it's still communication. It's very state-dependent. You need that baby to be awake and alert and cooperative. And you really -- for a baby, you really need to realize the -- that their motor abilities and their sensory abilities really have an impact on an accurate assessment of their neurocognitive abilities. And most importantly, I think it requires a skilled examiner with extensive experience with neonates and infants. So I don't -- I don't want to spend too much time because I know my time is running out. There's --

there are developmental screening tests.

This is a list of tests that have been recommended by the American Academy of Pediatrics. They're cognitive testing. Most of the tests are based on the early work of Geselle where he developed a number of items that measure motor, language, adaptive, personal social behaviors, and for many of those things, he's really just sitting down and playing with the baby. And then from that, that work has been refined to the Bayley Scale of Infant and Toddler Intelligence, most recently restandardized in 2005 to the Bayley-III. There are concerns that this Bayley-III overestimates a baby's intelligence and that babies who are tested on the Bayley-III are not eligible for early intervention services.

But I don't have time to go into that, but that's kind of a controversy about the current use of the Bayley-III. There are other tests, the Mullen and the Griffiths, that are developed along the same line. And then I just put in here the Fagan Test of Infant Intelligence, which looks at visual processing and preference for novelty. And this is used at -- everywhere from six months to two years and is filled to not have the same cultural bias that many of -- that the other tests have. I haven't seen it very widely used. So in conclusion, there's no way to determine the intelligence of a newborn or a young infant, partly because it's inaccessible, but I think part

of is it's the -- it's the potential for cognitive development that we want to know. It is possible to assess their neurologic integrity and that we have to be very aware that the -- their -- the sensory abilities, sensory motor integration, motor function, communication, these abilities really influence their neurocognitive development and an accurate assessment. So thank you for your time.

[applause]

DR. COMO: Thank you -- thank you Dr. Allen for that excellent, comprehensive overview of development. Our next speaker did manage to get through the traffic mess [laughs] this morning. Dr. Gahan Pandina from Janssen Research and Development is going to speak to us on normal toddler, school-age, and adolescent neurocognitive development and measurement in these age groups.

NORMAL TODDLER, SCHOOL AGE AND ADOLESCENT NEUROCOGNITIVE DEVELOPMENT AND MEASUREMENTS IN THESE AGE GROUPS

DR. PANDINA: Thank you. Good morning. My name is Gahan Pandina, and I work at Janssen Research and Development, currently working on a project focused on autism spectrum disorders. I want to thank the organizers for putting these two days together. I'm sorry that I couldn't have joined yesterday, but I was actually at a meeting in Europe focused on autism spectrum disorders and a large project there that is developing new assessments and biomarkers to evaluate autism. So I'm here to talk about normal toddler, school-age, and adolescent neurocognitive development and measurements in these age groups in 20 minutes. Well, 19 minutes and 30 seconds. So I'm going to do my best, but I will also be on the panel and will be able to discuss this. Dr. Allen has set the table very nicely for me, so there's some things that I don't have to cover, which is wonderful. I also want to thank Dr. Yao particularly for inviting me today.

So I am a full-time employee of Janssen Research and Development and a J&J stockholder; those are my disclosures. The opinions I am going to be stating are mine and not those of the company I work for. So what will I talk about today? I'm going to give a bit of an introduction to cognitive assessment

in children. I'm going to talk about some assessment instruments, and then I'm going to talk about assessment issues, in particular measuring change over time and scores and normative data, and then testing in special populations. So I -- also I put this cartoon here because testing is often misunderstood. Everyone wants to know, "What's my number?" or "What's my child's number?" And the number, as we've learned over the past -- particularly last 25 years, is a much more complicated thing than one number.

First, childhood is a dynamic period of physical, psychosocial, and cognitive growth. When we talk about cognition, we talk about all of these things that are occurring in early development, this explosion of physical growth, which is so dynamic, the evolution, and really increase in speech. In particular, when we have speech and language, we're then able to access much better the inner world of the child. But I think there are many other aspects of their development that are important to talk about. Of course, there's a-- just a huge physical growth and physical achievement. That physical achievement is also highly variable within the normal range.

If you are looking closely at the very nice slide that Dr. Allen prepared on what is considered normal and the timeframe of normality, of when you would develop those things, single words, pulling oneself upright, walking. All of those

things have a very broad window. So as a psychologist, one of the things that we talk about commonly is confidence intervals. How confident are we that this behavior falls within a normal range? You can see here from toddler to, you know, peri, and then adolescent, many things are happening. Piaget of course talked about cognitive development as sensory motor preoperational, concrete operations, formal operations, the way that we interact with our environment, the way that thought develops, mental operations developing, and thinking about thinking, or metacognition, and then abstract thought, thinking about logical principles.

But you also have to reflect on the psychosocial aspects of development. This -- although Erikson talked about this early, Vygotsky also brought this through in more modern thinking and others have extended that work, to think about the interaction with the social world and how much is learned during that interaction. The main -- and maybe I'll just also mention that moral reasoning or, you know, thinking about how one's position in the world in relation to others is also an important aspect of development. All of these theories, however, have been challenged in part because they seem to march forward in very logical, conclusive stages when we know that that's not the case.

We know that children have fits and starts; they grow,

and then they recess; they learn a skill maybe very early in one domain and very late in another. So although we'd like it to be this structured and rigid and regimented, we know that it's not. So all of these theories fall short. So what I'm going to try to focus on a bit more today is cognition. There are a couple of major theories in cognition. The CHCR Cattell-Horn-Carroll model is a primary model that's used for most of the intellectual assessment literature that's developed over the past 30 years. There are other theories that have developed, including theories based on Luria, which are based on brain function.

So I'll talk about some of those too as we go through some of the testing processes that are used. I did want to pay homage though to other theories like the theory of multiple intelligences being developed by Gardner, past theory. Many of these are developed based on the shoulders of these other developmental approaches. And I think, as we talk later in the panel about how do we assess abnormality, it's still important to know that we're still learning how to asses normality. What's going on the brain over time? Well Dr. Allen took you up through the neonate into perhaps the toddler, and also talked about the NIH Toolbox.

There's some remarkable work that's being done in imaging right now to tell you about what's going on in the

brain. So what we know is that through very good investment through a large multi-center studies to atlas the brain in different brain regions -- I'm sorry you can't read what all of these regions are. But if you look at just the colorful panel on the left, this is cortical gray matter changes over time. And you can see decreasing from age 5 to age 20. This was an early study done by Jay Giedd and others in 2004. Since then, the Brain Development Cooperative Group has looked at development and research with the improvement in MRI research and our ability to detect and segment different parts of the brain and grain weight matter.

We've learned much more about that, so you can see while total brain volume doesn't change too dramatically -- that's the top, left panel on the right -- gray matter and weight matter are shifting dramatically. Gray matter is going down; weight matter is going up. There's a whole host of processes going on to suggest major changes in the brain that are underlying these physical changes that are underlying these great changes in physical growth and in cognitive adaptation. I also want to just note that the blue line is boys, and the red line is girls, and all of those panels on the right hand side, so you see that there are sex differences in the development of the brain and also in the timing of when these things occur. Some of those are due to hormonal changes; some are those are

due potentially to epigenetic effects, but all of these things are quite important in thinking about, what would happen if something were to go wrong, and what would happen -- at what point in development if there was an insult, or if a disorder develops?

So what are common uses of cognitive assessment in these groups? Well, the most common is school-based evaluations for academic problems. When children hit school age is where most of their cognitive deficits or psychiatric deficits become apparent. That can be in preschool, at around age 3, when early intervention evaluations are often occurring towards the end of 2 and in 3. But it often occurs when children hit kindergarten, and through the kindergarten year, they're often informally assessed through their physical and basic cognitive abilities. By the time they hit first grade, if major deficits are noticed, then they are often evaluated by a school psychologist or a child study team. These also track school progress.

The other major use of cognitive assessment is in clinical -- or psychiatric, rather, medical settings to diagnose or refer for specific conditions. Often, these might be first identified at school. You have the issue here of, who knows the child best? And who knows children best? So often a parent, especially if a parent has their first child, they know their first child well, but if they don't have expertise in children,

they don’t know what normal development is across the board. So other experts, particularly teachers, see lots of children all the time and can recognize these gross, emergent problems. They also are used to track treatment outcome, although that can be problematic for a number of reasons that I’ll talk about. They’re also used for research and cognitive development, disease biology, outcome, and biomarkers. These are major categories, of course.

What are the factors that affect cognitive development? I’ve tried to talk about some of these in the slide. I really don’t have the time to go into them dramatically. I think Dr. Allen was talking about developmental problems and medical problems that occur like intellectual deficit or other physical problems, neurologic problems. But even medical problems such as asthma or diabetes can have dramatic impacts on cognition. We’re talking effects in the range of a half to a full standard deviation of impact of what would be typically expected from their parents’ IQ. Genetics are a very good predictor of where a child’s intellectual capacity might end up. But they’re not the only predictor. And in fact, the environment has a big impact on whether children reach their full potential in terms of their cognitive capacity. There are of course psychiatric conditions and learning disabilities, language impairment, and language impairment

becomes quite important when we get to testing because if someone's language is impaired, it's very difficult to assess their knowledge. This is true particularly in the area that I work in, autism, autism spectrum disorder.

In addition, basic physical nurturing, nutrition, culture, and education all have impacts on cognitive development. So the first category of tests I'll talk about are IQ tests. These assess a range of a range of cognitive abilities, a broad range. They have a strong basis in cognitive theory. I mentioned the CHC theory, which is based on factor analysis, which is looking at multiple different, individual domains of cognitive abilities and analyzing to see how they hang together. What are the -- what are the major structures that hang together there? In addition, other theories like Luria and PASS, et cetera, that I talked about before, these tests have a large normative database and are available in many languages. They have age-grade and sex-based norms associated with them. You can see here, I've listed some of the major ones that are used. The Wexler scales, in particular, are used; the Stamford Benet, arguably the most common; the Woodcock-Johnson differential abilities scale; and also the Kaufman assessment battery for children. Many of these are available in short and long form, and these tests are usually employed by a school or a clinical psychologist or someone who's a psychometrician whose

been trained to administer them. They're interpreted by looking at the pattern of scores and detecting whether there are differences between different cognitive domains.

And I think, as we start to talk about what can go awry with cognitive development, it's important to recognize that these tests also, which I'll get to later, may have state-dependent and trait-dependent effects that we'll talk about. Achievement tests are a measure of academic performance and leaning. In contrast to what is sort of innate in the overall Gestalt of abilities of the child, these reflect what they've learned in school and in life. They yield scores also in domains and cognitive profile. They have large normative databases available, often age and sex and grade. When they're combined with IQ testing, they can identify discrepancies. Those discrepancies can be categorized as learning disabilities or learning deficits. I've listed some of the major achievement tests that are used here. It should be noted that there are long and short tests here. Short tests are usually brief screener tests, like the wide range achievement test takes maybe 20 to 25 minutes or up to 30 minutes, to a longer test like the Wyatt, which can take 60 to 90 minutes, depending on the ability of the child.

These also yield a pattern of scores based on their theory and can be compared and contrasted with IQ scores.

Neuropsychological tests were developed initially to assess brain trauma out of very classic neuropsychology. Halstead-Reitan battery is a classic one for adults. It's mixed been extended to children. The Luria-Nebraska tests have also been extended down to children. I'm mentioning the NIH Toolbox here only briefly because I know we have a speaker who's going to focus on that, which has a wonderful array of tools in their cognitive space. And also the NEPSY, which has emerged as a new instrument in maybe the last decade. These assess test theories based on brain function. For example, the Luria theory is like a triarchic theory where you have cortical arousal and attention, sequential and symmetrical processing, and then strategy as your three triarchic domains of function that it measures.

The reason why I mention these tests is that they're being more commonly used and more commonly used in pattern analysis to assess cognitive deficits, particularly in neurologic disorders. These assessment instruments have a wide age range, and I think it's important to note that one test cannot measure all, and it's important also to know that what tests are often measuring is compared to your same age peers and your same sex peers. How many of these items are you getting correct? And how is your performance differing from theirs? Their -- I won't go into this in too much more detail, but you

can see many of these tests cover a very wide age range. So, 2 and a half to 3, all the way up to 18 or adulthood. For some of the achievement tests, you'll ask, what is the normative data set? And they'll say it goes from 4 to 90. So imagine using the same tests for individuals from age 4 to 90. How accurate do we think that that would be, particularly at the tails to asses 4, 5, and 6 year olds, and 80 to 90 year olds? Important to think about that.

Other types of tests are specific functional tests, like for learning and memory or attention. I put some examples here like the California Verbal Learning Test. This is a list leaning test that assesses short and longer term verbal memory, and also recognition memory, a continuous performance test that tests attention to both impulsivity and sustained attention, and other things like the Ray Complex Figure Test, which assesses Gestalt visual-spatial organization and visual-spatial memory.

There are also computer batteries, and this is really an interesting area that's emerging over the last particularly 10 years: assessment of cognition through a standardized format using a computer screen. There are new paradigms that are being developed actually to assess individuals through computer games. We know that kids, in particular, love to play computer games, and we know that we can learn a lot about how people think when they interact with a more dynamic environment as opposed to

interacting just with a bunch of cognitive tests. There are also a number of experimental tests for emerging cognitive skills. I just -- I also want to pay mention to social cognition and emotional intelligence, just two of those that become very important as we start to get into the social domain. You can have someone who's very highly functional, but who has not good emotional intelligence, and that can have a big impact on their long-term ability to achieve their potential.

I wanted to put this up just as an example, although we're talking about cognition. As a psychologist, we use lots of information, so behavioral observation -- this particular picture was drawn by an 8-year-old boy who's typically developing. This picture is the same exact picture taught in the same exact class being drawn by an 8-year-old boy within a month of his age who has ADHD. So I wouldn't need to do extensive cognitive testing on this individual on the right to know that there's something drastically different from him, not just his drawing ability, than the child on the left. So we use lots of sources of information to tell us about what's normal and what's abnormal in cognitive development.

Scoring and profiling: I realize that this may be difficult to read, especially in the back, but what we're testing here is a normative distribution of scores. We want to look at how different from what's typically expected from the

child's expected development may be. Now you can look at what their average might be compared to their same aged peers, but what if you have someone who's being raised by scientists who have an IQ of 135? Would we expect that individual to have an IQ of 100? No, we know from all the research, we'd expect them to have an IQ of about 135, along with their parents, assuming that they haven't had any major insults? Even if they have some minor neurologic or medical issues, we expect that level of development. So it's all relative. And that's relative both in the truest genetic seen of the world, but also to the testing population they're being compared to. It also reflects the fact that we need to standardize the tests that we give, and those standard scores are very important when we interpret the data from cognitive studies. We'll talk about that more later in the panel, I'm sure.

We also use categorical descriptions, these are from the WAIS and the WISC that are used to describe different levels, and those are often very misleading because of the large range of variability. You can see the average range ranges from 85 to 115 in terms of IQ. The average IQ score on a major IQ test is 100, with plus or minus seven standard deviation. But 30 to 40 points in IQ is a huge, huge difference and has big implications in what the individual may achieve, so I think it's just important to note that what we're trying to assess here is

not necessarily average performance, but best performance. Do we think that this is a true reflection of the child's ability? And do we think, based on all the information and all the things we know about them -- the achievement of their parents, their parents' socioeconomic status, their education level, all of those things -- is this what we would expect from that child? Not just, how does this child compare to their same age peers at this particular time point?

Test selection and interpretation is also quite important. Assessment of change is a critical issue that's come to bear in psychology because the tests that are commonly used that I've been describing here are often used over a longer period of time. Most of the tests I've described today are not recommended within even a two-month interval. For IQ tests, it's generally thought that you can't give them any closer than six months, even 12 months. So can we do these to evaluate outcomes? Certainly not in a short-term clinical trial, or over the short term. The other issue that I'll mention is many tests are not appropriate due to their -- or adequate due to their coverage of language and cultural norms. We know that there are broad differences in education country to country, so as we're doing global programs and developing new treatments for conditions, there may be wide and vast cultural norms that are only basically covered in the testing norms that are developed

by the test creators.

We also have different difficulties with accurate estimation at the tails. For individuals who are very low functioning or very high functioning, the tests just simply do not have enough items, or enough -- we don't have enough to measure differences there. And at the low end, one or two items can make a difference of two to three of four IQ points, which then becomes problematic when you're assessing change. Do we expect someone's IQ is going to change by 15 points because they got four more questions rights on six different subtests? No, that's highly unlikely. We know that IQ is very stable. Last, the demand characteristics of the testing environment. People often wonder what they're doing when you being them in for testing. Kids are very used to being tested, as opposed to adults, who are less used to it. But even so, why am I being tested? Often, it's an area of crisis or challenge or as a result of a problem, not as a result of something good like going into a gifted and talented program. So that can create issues with testing.

Maybe I'll talk just for a moment about specialized populations, and then I'll end. Specialized equipment for very young children, I think as Dr. Allen was mentioning, very young children are a special breed all and of themselves. They are, as I said, dynamic, fits and starts. They'll say, "No more."

They'll just do what they want to do in the testing environment. So the tests are designed to be more interactive, more dynamic, but they also have more of a need for familiarity. The demand characteristics of the testing environment are important, so you must select the appropriate tests and stimuli, but you also need to have an evaluator who understands what normal is and what abnormal is. Often when we're trained as psychologists to do assessment, we're trained on what's abnormal, not normal. And you need to develop a good basis of normalcy.

And I think, particularly people who do physical evaluations of children over and over and over again like pediatricians or neurologists who see children with a wide range, are in a particularly good place to do that. And you need to learn what's developmentally normal. For clinically referred populations, you also have to have experience with them. The setting and referral source are important. For me, I mentioned -- I showed you that picture before of that child with ADHD, but behavioral observations during testing are often a big guide for us, and they really can tell a lot about the person's ability to do something. Often, people will tell you right off the bat when you start to test them, "I'm not good at math." And what you know is that they don't like it because maybe it's a deficit area for them, but it also means that you need to pay attention to when their doing it to make sure that they're

putting in the effort that they need, that they're not giving up, that they're accurate, that their reflections on their own abilities are accurate.

And then of course, in clinically referred populations, there is challenging behavior. I've has many people throw test equipment. I've had many people refuse to do testing, and I've had to do what's called a non-standard test administration. I had one individual with autism who had an IQ of 150 who refused to answer any of my questions unless I sang all of the items in a Puccini opera. So what do you do? Do you get a score by singing the [laughs] opera or not? So I think you have to rely on all of these factors, and then you have to be very careful with how you interpret the results. I realize that I'm just about out of time, so I want to say thank you, and I look forward to a great discussion today.

DR. COMO: I'm not sure if this is projecting. Thank you, Dr. Pandina, for that nice overview. And I'm surprised that the ADHD child kept it within the lines on his drawing. Our next speaker is Dr. Richard Gershon from Northwestern University Feinberg School of Medicine, who's going to speak to us about an exciting area of assessment: the NIH Toolbox for the assessment of neurologic and behavioral function.

NIH TOOLBOX FOR THE ASSESSMENT OF NEUROLOGICAL AND BEHAVIORAL FUNCTION

DR. GERSHON: Thank you so much. Okay I'm looking at the time there. Great. Let's just dive right in. Disclosures: I'm primarily funded by the NIH, the Department of Defense, et cetera. I do some minor consulting work for surgery boards and educational service providers. My background's in assessment. That's really my expertise, and as you'll see in a few minutes, that definition covers quite a wide range of types of assessments, but I partner with lots of people to do that. I'm here today to present work that is sponsored by the NIH blueprint neural science research. This represents 13 NIH institutes, which tax themselves 10 percent of their neuropsych dollars to do common projects, and the goals of this is to coordinate research efforts across the institutes, reduce the duplication efficiencies of scale, and then to do cross-cutting efforts.

Really, they've done three projects so far. The first one was to figure out, what are those common areas? And they found they really need a common set of measures. We'll get in that in a minute. The next one was we really need a reference set of imaging data. And so they sponsor what's called the Human Connectome. The Human Connectome is just about finishing

off a study of normal functioning of ages 20 to 30 year olds, where they used what as of today is cutting edge MRIs. My understanding is there's really only a couple of academic centers in the country that have them. They put people through a 3 hour, 3 to 6 hour MRI, so if you've ever had an MRI, 20 minutes is more than enough, and they give them various instruments as well, such as NIH Toolbox, to record that. And they actually have an RFA right now to take that work down to age 4 and up to age 95.

So the goals of NIH Toolbox were to develop a unified set of measures of a really wide range of things: cognition, emotion, motor sensory, you heard a nice overview of areas we're all concerned about. I'm going to get into areas people haven't discussed so much in terms of particularly motor sensory areas for use in large cohort studies in clinical trials. Now the interesting thing here is that people who develop instruments tend to be clinicians, and so they snuck in the fact that all these instruments could actually be used clinically. It wasn't the original goal, but ends up being the group that will help support these instruments further on. And the real though there should be common currency because you saw the previous presentation; it looks like it's not too bad. There are only four of five instruments in each of these areas, and the reality is there are hundreds.

And actually, if you were here yesterday, the subtext was we could develop additional instruments in many, many areas. My first meeting, I came out of the personality/clinical area, and then I was at an International Society Quality of Life meeting, and I counted. The NIH had sponsored 40, the development of 40 instruments to assess fatigue across various disease, many in the same disease. You'll notice the content underlying the hood is not all that different. And if for no better reason than that's not a good idea necessarily is the studies aren't comparable, so in the course of a trial, if five pharmaceutical companies are using five different measures of fatigue, no one can compare, how did those drugs work, one versus the other? So the thought of common currency -- also how do we do this in large studies?

Many of the tools had been presented and are quite expensive. And even if you might get a waiver for the first five years of the study as the Alzheimer's Consortium, which is actually meeting, and there's a group of us in town today because their meeting today, and they had a great idea; they have common data elements; they got them together; they got assessment companies to say, "we'll give you these really cheaply." That worked until year six, when it went up to a hundred some odd dollars per subject per year. And basically they're starting over from scratch.

So we created this effort. This is under our federal contract. We spent about $25 million over six years, but we took a lot of work from others. We had instrument development teams; we had 13 NIH science officers, and those who are researchers in the room can try to extrapolate what that means relative to trying to work with one. We had epi-biostats technology; we developed immediately in both English and Spanish. But probably more importantly is we had a multi-cultural team, and every instrument team had to produce an instrument that was viable across U.S.-based cultures. I learned things about cultures I never knew. We got rid of instruments that involved uncovering the foot because unless you're going to wear a white coat, maybe some cultures will let you look at their feet.

I got a call on a Sunday afternoon from a reviewer saying, "How dare you let that hairstyle be in one of your pictures. I will never talk to you again." I had absolutely no idea what they were talking about, but they found it very offensive. And also pediatric to geriatric, and so the goal of the Toolbox was to develop an instrument battery that was good for 3 to 85. I'll be honest; it probably does a better job for about 40 months, but it does work to about 95. That doesn't mean necessarily there's the same content under the hood. And so there're various instruments, and on some instruments,

however, we are able, using computer technology, to dynamically adjust the content, depending upon the level that someone's working.

But these early multi-cultural reviews have allowed numerous countries around the world to start translation efforts without great concerns for the multi-culture components. This was discussed yesterday a little bit that stimuli just don't necessarily make sense in different countries. So by the numbers, 80 institutions were involved in this effort; 256 scientists in staff. We used 20,000 subjects to get this work done, and we created 40 -- four 30-minute domain batteries and norm for ages 3 to 85. They said it couldn't be done. Most of our investigators said we couldn't get most of these areas down to 30 minutes, and generally we pulled it off. Okay, again, political across the age, but in their known intellectual property concerns, everything is in the public domain from that perspective.

Everything is psychometrically sound. We spent -- we spent the first year creating rules for what would be acceptable. We rejected many, many instruments that are in common use because it turns out that they weren't psychometrically sound when they were created, and many of them never had research evidence, and unfortunately, that applies to many instruments. Actually, probably nothing you saw in the

screen earlier today falls in that category, but many, many instruments used in health research simply were never validated in the population that they're used at the primary instrument today. These things cover the full range of a trait, and you know, today the group is talking about some pretty significant disease, but a major problem in the -- of any assessment is that an instrument has a floor and a ceiling.

So a floor says it simply doesn't go low enough, right? You just -- the instrument is designed to only go so low, and so you actually don't capture how poorly a person is performing, and the other issue is ceiling effects, and that is you don't capture how well a person is doing. So all common instruments for -- or most common instruments of physical functioning actually ceiling at something like, "Can walk around the block." Okay, so that means that if you're an athlete or if you're generally a healthy person, and you break your leg, and you are considered cured from an insurance perspective when you indicate you can walk around the block.

Well, just most of us in this room would not feel that that is the maximum physical functioning we want to have, and yet to date, there are very few instruments that cover that, so we set out to remove those ceiling effects. We were 90 percent successful. And actually, there are ongoing efforts right now to fix that. The Department of Defense got a hold of our

physical functioning measures and said, “Wait a minute. When a person comes back injured, they have to be able to walk -- hike 50 miles with a pack on their back, and unless you can capture that discrimination” -- but think about it, for them if they can’t do that, they don’t go back to work, so unless we have instruments that can assess that. So we actually have a group right now working with a set of health clubs where, you know, we’re thinking here: how do we deal with very minor levels of functioning? But we are approaching not patients, because patients can’t tell you what their active functioning is; we are approaching people who work out every day and military members and athletes to be able to expand upwards.

So an issue on this -- this is an anger -- looking at anger, aggression. The standard self-report instrument in this area is 12 items. The PROMIS anger CAT, which actually is imbedded in the Toolbox as well, takes and average of five items to take. So you’d think with five items, you’d get less utility, but they’re computer driven. So if you see a baseline, basically no effect at one month, both instruments demonstrate that whatever this intervention was has one a nice job; that’s great, but after three months, the study concluded because it was based on the aggression CAT that there was no improvement, and therefore, there is no reason to extend treatment to three months. Well, the problem with that instrument is that

instrument has a ceiling effect. It only can show progress to a certain level. So the CAT, the PROMIS instrument, which does not have a ceiling, actually demonstrates that that person continues to improve at that same rate, but a classic study in this area would have said there is nothing more to be gained; stop treatment; you don't need it. So it's something to watch out for.

Okay, what about these instruments? They are free of access, use control; they are royalty free. They've actually been completely free to date. The NIH has said to us, "You have to figure out a mechanism to maintain them." So there will -- there is -- there are neutral fees for technology, which are in the probably hundreds of dollars per year kind of range for individual use, and a little bit more for large studies. Many of them are open; many are just posted online. They can be used. Things such as the cognition instruments or to maintain what's called their class-III, you have to demonstrate that you -- as many, many of the instruments that were shown earlier today -- that you have the capability of interpreting them or using them wisely. So a person can't go and just take them and give them to someone and tell you what their number is.

A lot of this is based on item response theory and computer adaptive testing, and I could spend an hour on this area. This is my favorite area to talk about, but what it is

it’s a way of using the computer to greatly shorten assessments. It was mentioned a little bit yesterday; in the Toolbox, we use this to shorten vocabulary in reading. We get a clinically -- we get a very, very reliable measure of vocabulary in about 22 items in reading or reading fluency in about 20 to 24 items. All of the health-related, emotion items, both the NIH Toolbox, and I also do a lot of work with patient-reported outcomes measurement, an information system, AmeriQual. We’ve actually got algorithms down pat now that we can get a clinically meaningful score in an average of four items. And this is composed of 20 items and sometimes 90 items.

There was a nice presentation on that yesterday. What is CAT? It’s shorter. It’s targeted. It use the computer. It’s used by lots of people. We didn’t invent this. Really popularized by use in the military. Something like the graduate requisite exams, nursing boards, security dealers all take using CAT. Most of these places use CAT because it derives a unique instrument for every single person taking it, so you can’t cheat. So if I’m taking -- the army really needed this a lot. They test over a million students a year for what their aptitude would be for various roles in the army, and you can imagine that if you had the secret numbers and the answers the questions, you could get whatever role you wanted to, so they needed to generate a custom test.

So just a really brief example of CAT. If I'm doing a pass/fail test, and by the way, nothing in the Toolbox is a pass/fail. Our goal always is to find a level of functioning, but if you're doing a pass/fail test, typically I want to give a question right at the middle of that continuum. If I get that correct, I give a harder item; if I get that -- my animation, of course, going crazy -- if I get that wrong, give an easier item. Get that right, I pass the test. Why do I pass? Because I've demonstrated that I can generally pass items above a pass point. If all I care about is a pass point, there is no value in my giving lots and lots of items below that pass point. This has changed the face of things, such as the nursing board exam, which is given twice a year, which made it quite difficult to license nurses in this country because if you're busy in your hospital on the day of the exam, you then had to wait another six months. This exam, the nursing boards are now given every day of the year because we zero in on a person's ability level. In the case of the nursing boards, they have 13,000 questions under the hood. You can zero on in.

By the way, I'll show you a sad story here for a moment. If I take an item in the middle, I fail it. I get an easier item, I fail that. A little bit easier, I finally get one right. A little bit harder, fail, and the person fails. Why give this person another 90 items above the pass point?

They will not pass. Now I will tell you I am a consultant to many, many surgery boards, and we never pass or fail people in four items, but we do in 60, and this is as opposed to traditional exams, which used to be between 1,100 and 1,800 items. We don't get any benefit of that. It's unfortunate that people who've been in school 20 years very much are not really happy. Actually, they're not happy they've passed in 60 items either, but certainly fail. But we can apply the same -- whoops, I'll go back here a second. We apply the same concept -- do I have -- I'll never do this right. Well, what am I looking at?

That doesn't appear on my screen, so maybe look away. We can use the same concept to zero in on a person's level, so for instance, NIH Toolbox vocabulary test has hundreds and hundreds of items behind it. We give an item -- we start off an item based on that child's functioning grade level, and we can give items that are harder and easier and zero in, but if a child is gifted, we can quickly go higher, and if a child has depressed functioning, we can go lower within the, quote, "the same test," but very much getting different stimuli depending who that child is. Okay, what am I looking at up here? Change lamp?

MALE SPEAKER: We'll get an AV person.

DR. GERSHON: Okay, no problem. My lamp is going out.

Okay, the Toolbox has four domains in it, cognition domain framework. Now one thing to remember about the Toolbox, it was developed for epi-studies or ages 3 to 85, so you're going to notice here there's no visual memory, no visual motor task. There are groups that have now been sponsored to add additional tasks, so it is not a be-all and end-all for any one particular age, for instance. So quickly, executive function: now the Toolbox was released in October of 2012. It's been used -- I don't know. I want to say in six or 700 studies since that time. And most of them are ongoing because, frankly, the way sponsored research works, new assessments -- I always argue an assessment takes a generation to become -- if it's good, it takes a generation for it to be really ubiquitous. If it's bad, hopefully it disappears over time.

What I'm chatting -- I'm sending emails over here because the National Children's Study adopted the Toolbox for many of its assessment over time, and so they began the redevelopment of the -- I didn't even touch anything. I'm being controlled remotely. Okay well, if you'd give my slides, it'd be better. The entire Toolbox is coming out for use on the iPad. It goes -- it's in 10 institutions starting on Monday, and over the summer, they will all be available. My lamp is still bad. Somebody have the remote for the projector? Maybe not. Okay, great. No, that's still there, but it's not me.

You need the projector remote. There you go. No, no, no, it's not the computer. Okay, so whatever, we'll just pretend there's nothing in the middle of the screen.

We have -- we have -- I'll just run through a bunch of the tasks really quickly. Executive functioning, attention task. These have been derived for the iPad. Again, we find that the tasks grab attention down to normal 40-month olds. There's a group out of Phil; there's also a group out of Minnesota that will be released this fall. We really have pushed most of these tasks down to normal functioning for 2-and-a-half-year olds. And ironically, the errors I was worried about people being able to use the iPad were for 3 and 4 year olds and 85 year olds on up, and those are the groups that adjust to the iPad the fastest, so it's just a changing world. NIH Toolbox picture vocabulary test, again, about four minutes. Also available in Spanish, which is a completely different test because vocabulary in English and Spanish just don't even mean the same thing, but their good proxies for general intelligence.

Executive functioning, again, has some validation measures here. Working memory tasks, seven minutes. There's no long-term memory tasks just due to the time constraints. This is an extremely long test for the purpose of the Toolbox because we wanted things to be short. Episodic memory tests -- again, we don't have a lot of time here. Okay, language, we have an

oral reading recognition test. It's three minutes. It has three or 400 items under the hood, but it's just zeroing in on exactly the person's level; it does it very quickly. Processing speed, series of tasks takes 90 seconds. Again, very quick. By the way, it takes more time -- the majority of time on these tests it actually the instructions and orienting the subject to the stimuli. Once the test starts, we can zoom.

I was going to bring up some other frameworks, which really don't seem to be the focus of this meeting, but in the neuropsych community, and again, these were generally developed by neural psychologists, I was looking at areas that might be of interest to people developing issues. These are all cognition. You know, it's kind of your definition. You need cognitive functioning to be able to do any of the tests, so there's a motor domain, including dexterity, which I'll show you in a minute. Strength, balance, locomotion, endurance, again, we've created a battery which ordinarily would take two or three hours in the lab, this one about $100,000 worth of equipment. We've got it down to 20-some-odd minutes with -- under a new release, it'll be -- you'll need your iPad; you'll need an iPod. I'll get into that in a second, and about $50 worth of materials.

Now one is just nine-hole pegboard. This is off the shelf. You can -- it just does its job. By the way, we also use grip strength using a JAMAR grip strength dynamometer. It's

interesting that when we finished the test, JAMAR came back to us and said, “Can we switch to your norms,” that we actually had better normative data than they did. NIH Toolbox balance measure. If I had an hour or two, I’d show you videos. This is a test which typically takes $10,000 worth of equipment to do, but we’ve actually got it. We just put an iPod on someone’s belt. We have to do different standing positions. It’s a semi-adaptive test. We collect 15,000 data points in 6 minutes, and we can compare them normatively to existing things. It’s just, when we developed this test four years ago as custom hardware, it was manufactured in all sorts of strange countries around the world; now it’s build in to the standard capability of off-the-shelf, little, cheap hardware. It’s just amazing where things are going, which makes these things practical. Part of it, things aren’t practical. You’re not going to take a $100,000 piece of hardware to 20 sites and test people.

Sensory domains, vision. I’ve become an increasing believer that if we don’t test vision in audition, then we shouldn’t be using vision -- any visual stimuli or stimuli that require hearing. I’m a trained psychometrist from way back. I ran the testing center at Northwestern University in my early PhD. Days, and honestly, vision in audition was never part of our battery, and actually, I don’t think and standardized batteries require you to test it. But here’s the thing, if you

can’t tell when you give a test, are they getting a poor score because they don’t know it or because they can’t see it? By the way, we test vision -- we test corrected vision. We test vision with glasses on, with contact lenses on because that’s a person’s functioning visual level, and I have really bad news for everybody who -- I don’t know this because I don’t wear glasses. Well, okay, I cheat; when I read, I need glasses. When the 30-some-odd percent of kids have the wrong prescription, so they may look like they’re sitting there with their correct vision, but they cannot see. So is their IQ 80, or is it the fact they simply can’t see what’s going on? And ditto for audition. While it’s a much smaller impact in the population, it’s still there.

Olfaction, sense of smell, not as big an issue in the pediatric area. Neural gustation, sense of taste, but vestibular balance and semi-auto sensation, that’s touching capabilities. All things that are available in the Toolbox. Hearing, we have word and noise test. We’ve been now -- I’ve been hard at work with various groups on a royalty-free, pure tone audiometry that’s like kind of the tones everyone’s has in school. It does not exist yet in a good way. If somebody knows of one, fine, but I keep talking to people, and everybody want’s their dollar after it’s delivered, so we are still working on that with the National Children’s’ Study.

Static visual acuity, this looks like what the test would look like. It's not true. We actually put an iPad nine feet away on the wall, and we do different sized characters and a wireless keyboard to that iPad. The examiner then does that. We can get this down in like two minutes instead of a multi-length -- and by the way, these are somewhere between screeners. Most of our tests actually correlate with high resolution tests in the mid to high 90s, which in most cases, people would say is a replacement. I have to say that it's a screener or the International Neuropsych people get mad at me that, "oh my gosh, how are you able to do this in four minutes?" Well, we did it because we had a lot of money and a lot of smart people put their heads together and decided, "We just have to do this already."

Emotion Domain Framework. Very few people talked about this. There were a couple of presentations yesterday about self-report outcomes. Many of these domains come out of the patient-reported outcomes measurement information system. That's, gosh, 12 to 15 years in the making, probably $100 million over hundreds of sites, again, the goal of having common instruments. These are the umbrella areas. Psychological well-being. But when you've got neuropsychologists together saying, "What do we want to know about somebody?" Now I should tell you that we can test the average person in 15 minutes on all of

these domains and at a level congruent with an hour-and-a-half to two hours of assessment. It's because we're pushing the computer. So, if you are positive -- if you indicate right off the hand that your first couple items that you have a general positive outlook, we just skip all the negative -- you know, we skip the lower end of that scale, and we just zero in where you're at.

Positive affect, life satisfaction, meaning a purpose. Things such as social support, probably not as important at the very young ages you're testing, but we know, for instance with senior citizens, if they don't have social support, you can do a great job on their surgery, and if they don't have social support, we also know they'll be back in the emergency room the next day because they have nobody at home to help them. So we can do all the great things and fix them up, and we can do whatever, but if they don't have support systems outside of the hospital, you'd better believe they're going to be back. Medicare needs to know this because it just doesn't matter the quality of the intervention you've done. I'll let you go through the rest of these on your own. Sam, this is really AKA depression, but in the NIH appropriate way of stating it. Well, I wasn't supposed to say that, okay.

The norming -- we norm this on close to 5,000 subjects. We have one -- we re-test data. We did a stratified random

sampling at 10 locations around the country. We were able to get single-year age bands for 3 to 17, adult age bands for 85, census, balance, blah, blah, blah, and both English and Spanish samples. Since October, again over 500 studies. This is actually a little old; 73 primary articles have been published, which is amazing because that means that people actually used it and published it in under two years. That's a big problem with everything all of us are doing in the room. You could be doing something groundbreaking; you'll read about it in seven years, so in the meantime, we're all busy doing the same thing, very, very often recreating the same bit. I have no idea how the Toolbox decided on 579 articles. Again, it just -- it's just past its second birthday.

A couple of efforts I think are important. One problem with the development of new measures is that, how do they compare to all the work that has been done forever? There's an effort called Prosciutto Stone. Dave Sella's a PI on this effort, but its goal was to take tons of existing instruments and say, "Look, there's 20 years of research on the SF36, probably about 15 years; how does it relate to any new instruments because I've got 5,000 people here with this score? What do I do here?" So this literally takes data and has look-up scores from old instruments and new instruments and vice versa, so that that can be utilized. By the way, not all

instruments are that parallel, and then it says you can't do it. So it ends up with conversion tables like this.

Here's your CESD score; here's your PROMIS depression score, standard error; what's the bidirectional look-up table for that? And I will stop there. Oh no, thank you. One thing, I brought about 50 of these, this is www.nihtoolbox.org. You're welcome to take one back. It's a catalogue of instruments. While we prefer that people use these are batteries, almost no one does that. They cherry pick off the instruments that are important. When you're doing a large disease-based study, even as short as these are, you don't have time to give all the instruments you want to use, and I have an iPad with me. I'd be happy to show people at lunch what things look like. Thank you.

[applause]

DR. COMO: Our final speaker before our panel discussion is Dr. Natacha Akshoomoff. I hope I pronounced your name correctly. And from medical -- I'm sorry, from the Center for Human Development at UC San Diego, Department of Psychiatry, who's going to talk to us about the PING study.

THE PING STUDY: PEDIATRIC IMAGING,

NEUROCOGNITION, AND GENETICS

DR. AKSHOOMOFF: Thank you. It's a pleasure to have been invited here. I do spend much of my research time working with clinical populations, but I've been invited to talk to you today about our PING study and particularly to focus on how development of neurocognitive skills relates to the other measures that we've collected in this study. And I have no financial interest to disclose, other than a proud recipient of NIH funding, which I hope will continue. So the PING study was funded by the National Institute on Drug Abuse and has also received additional funding from other institutes. And this was part of the Recovery Act: Grand Opportunity, and there was a call put out to create a pediatric imaging genomics data resource, with a large group of typically developing children in the course of two years.

So we -- whoops -- we took on this challenge and were able to complete the study with a little bit of additional time. In about three years, our group was able to collect demographic, developmental data, family history information, cognitive data, and neuroimaging phenotypes as well genome-wide genotyping. And because this particular grant had a short timeframe, required that create -- we collect data on a large number of typically

developing children in a set budget. It meant that we had to really try to find the most parsimonious data collection so that we could complete the school with the money that was -- that was given to us, as well as to make it practical for collecting that much data across a wide range of sites.

We were successful in this effort. We recruited about 1,400 children, and we have now complete data on about 1,200 typically developing children who are aged between 3 and 20 years. This grant was to create a database that would be available to individuals, so we have accomplished that goal that researchers can go to our website and simply submit a request to collect -- to analyze the data that they choose to basically pull down the data as well as to use online tools to examine different relationships between the data that are part of this data resource. So we have a large group of investigators, and these are the primary investigators as well as staff across nine different institutions that have been successful in collaborating in this effort. And we continue to work together on analyzing the data and new projects that have followed this.

So what I'm going to talk -- what I'm going to focus on today is basically just giving you a general overview of how the neurocognitive measures that we collected as part of the PING study, what the outcome has been in this particular study as well as how that -- our investigations have looked at

neuroimaging results and differences in terms of genetic associations as collected as part of this study. So as Dr. Gershon talked about, we focused on the NIH Toolbox cognition measures in this test -- in this study. Those were the data that we collected across all the participants in the study. We started this study in 2009, so we utilized the first version of the Toolbox cognition measures. They have since been modified slightly, and the version that's available through the NIH Toolbox group is slightly different than the version that we used, but generally analogous. We have also continued to collect longitudinal data on the children that participated in our study at UCSD, and we have been using the newer version of the Toolbox on those individuals, so we will be able to look at some of the longitudinal data across typically developing children.

I should also add that all the children that participated in our study were considered typically developing. We definitely had some exclusionary criteria, but we did not screen for things that were more commonly seen in children, such as learning disabilities and attention deficit hyperactivity disorder. We got information about that from the families, and that is available as part of the data resource as well. But we excluded children that were significantly premature, who had other types of medical, chromosomal abnormalities et cetera. So

as you can see here, these are the major domains that are part of the NIH Toolbox cognitive ensemble, and I'm going to just briefly talk about these as they've related to our results in PING.

So we published the study, that was last year, just looking at the results from the Toolbox in our sample of children from PING, and I just wanted to show you mostly how the data look in terms of basically the raw score, so that you can see the focus of this paper was really on age-related changes in scores, and therefore, the underlying cognitive abilities reflected in these different measures. Here's the sample characteristics of the children that were in this analysis. We have data from 1,020 children between the ages of 3 and 20. We -- across our different data-collection sites, we were able to stratify age and sex. You can see that there are fewer children in the younger ages compared to the teenage years, and that's primarily because one of the major requirements of being in the study was that you had to be able to endure an MRI scan.

And so even though these are typically developing children, it's obviously a bit more challenging to collect good, high-quality MRI data from younger children, and so we have fewer children in those age ranges. Some of those collection sites did not recruit children in that young age because that was not -- they did not have staff or other reasons to be

recruiting children in the younger ages. But over the last few years, we've been successful, and we were proud of our efforts in this study because of our -- the expertise across different sites and being able to get good MRI data on very young children. And I can talk to people more about how we were able to do that, but obviously, these were typically developing children, and they were able to watch a movie while they were being scanned -- during the MRI scans. And the imaging data included high quality, high resolution T1, T2 images as well as diffusion tensor imaging and resting state MRI, if available, during the one-hour long scan protocol.

Our statistical team was particularly interested in looking across all of the studies that I'm going to be talking about today, not only how to analyze the effects of age across these different measures of cognitive abilities, differences in brain development, et cetera, but also, how are these influenced by other sorts of factors? So in the studies that I'm going to talk about, we looked at the influences of age as well as sex on performance and brain development. We also felt that it was important to look at socioeconomic status, and in our abbreviated information that we had from the children that participated in the study, we used household income and highest parental education level as a proxy for socioeconomic status. And then we were also interested in variations due to race,

ethnicity, but more specifically, in terms of genetic ancestry factors.

And other studies had looked at this as how those factors may be particularly relevant for differences that we see across individuals when it comes to measures of the brain, and controlling for this can help to elucidate other relationships that may -- this may be considered potentially a confound or a complicating factor, and so in these data that I'm going to show you from the Toolbox, we took these factors into consideration, and we also looked to see how much did they add to the variation in scores, as part of these statistical models. And I'm not going to go into the details here; this just shows you that for genetic estimates of ancestry, what my colleagues have done is they've compared the results from -- what they do is they compare the results from individuals to other estimates that have been derived from very large samples, to basically determine, for any individual, what is your -- what is your genetic ancestry? So, if an individual says that they -- if they self-identify themselves as white or Caucasian, to what degree does that adequately sort of match with what we know about the variations that we see in ancestry across different groups around the world? And so through reasons -- through methods that I'm not entirely familiar with, basically able to determine what is the best sort of estimate of your ancestry and

take those difference into account.

And what you can just see here is that across these different Toolbox measures, if we just look at cognitive ability, performance, or performance across these different measures, then certainly age and sex are important contributors to the variations, but we also found significant variance was also explained by SCS, as well as the combination of SCS and these genetic ancestry factors. So, all of these additional components of the variants are significant. Because we have a very large sample, they tend to add just a small portion of the variants, but these are important things to keep in mind when we compare individual results. And, as the other speakers have talked about, there's so much that goes into performance across cognitive abilities, particularly in children, and more studies, as they look at the influences of socioeconomic status and other factors, have found that these are important elements to take into consideration for children's performance, and this data from these typically developing children also demonstrate that, as well.

So, these complicated sorts of figures just show you the raw data here from our paper, across the children that we studied in PING, and these are -- so these are raw scores. They're not standard scores, and they've been adjusted for those different factors, and what you can see very nicely illustrated

is that there's quite a bit of change with age across these Toolbox measures. In the version of the Toolbox that's available to researchers now, these -- you can get standard scores for individuals across different ages, but when we look at our sample, what you see are these dramatic differences with age, and these are the flank or executive function scores here, and this is the measure of inhibitory control. What you could also see in these data is that there's -- that there's quite a bit of variation, even when performance has stabilized across the older ages, and that's of great interest to us in terms of, how do these scores vary across individuals, and what may account for that variation across individuals?

This is another measure of executive function in the Toolbox, the Dimensional Change Card Sort. You can also see these big changes with performance across age and the sequence memory test, with a lot of variation with some evidence of -- in that version of the -- of the Toolbox, some children performing extremely well. Here's the picture vocabulary test, which is a really nice measure for reasons that Dr. Gershon has explained; the reading test, with lots of variation in performance across the older ages; the processing speed test and working memory test. So, what we found in this study of just the cognitive performance in the Toolbox in PING was that all measures show strong sensitivity to age, but the age functions, as you can

see, varied across the different measures in terms of, how does performance change, and when did children show maximal performance at different ages?

The executive function measures, which we are particularly interested in the flanker and the dimensional change card sort test, exhibit this kind of bimodality in the youngest children, and so there's been more investigations in terms of, how well can those -- how well do those tests work with very young children? And, as I said, the socioeconomic status and the genetic ancestry factors explained modest but significant proportions of variability in most of these measures, and in some cases, it was much more substantial. So, those were important factors to take into consideration for looking at variations and performance and how they may be related to other types of markers.

This is just an example of the range of change that we see in anatomy across development and, as we've seen in other studies, we know that there's a dramatic difference or change in typical development in terms of the surface area of the brain, different measures of cortical thickness, and volume of different structures. And these are things, in terms of these variations in typical development, we are continuing to publish papers on what we've seen in our particular sample. In the current biology paper that we published in 2012, what Tim Brown,

the lead author, determined was that you could take these different measures of -- from the MRI data, and in a combination -- using them as a combination of factors, you can derive the child's age. You could also look at their birth certificate. It's a lot less expensive [laughs], but this was a really nice demonstration of how elegant these different measures kind of go together in ways that we don't fully understand.

And this is an illustration of how -- if you look across the different measures that are available in terms of cortical area, thickness, subcortical volumes, and the DTI measures of diffusivity across different tracks and regions of interest, et cetera, this is an illustration of how those different measures contribute to understanding the level of, basically, the child's age or their level of maturity. So, these are not the patterns of how these different measures look, if you were to look at the age of any of these individuals -- these individual measures at particular ages, but rather how they contribute to this sort of combination of factors to predicting a child's age. And so, you can see that with all of those differences in how these brain measures change as a function of child's age that they contribute to understanding the relative maturation of the brain at different ages.

This was a paper that was published just a month or so ago, received quite a bit of press; some of it we thought was

good; some of it was a little bit misleading, I have to say. But what Kim Nobel and Elizabeth Sowell did was looked at, how does family income and parental education relate to the brain structure measures that we have in the PING sample? And what I want to just highlight a little bit is how that is also related to the Toolbox measures, the data that we have, for the children in our study. And so, what they found was that, indeed, after you adjust for age and age squared and sex and genetic ancestry factors, there were -- there was a significant relationship between the surface -- measures of different -- the surface area of the brain in children and to what degree income significantly contributed to variations in the measures of surface area.

The press part of it was a bit misleading in terms of poorer children have smaller brains and, you know -- and so there is that finding here in this study, but what exactly that means and how that affects long-term outcome, et cetera, is the -- is the big question here. But we -- but we definitely found that there's this significant relationship between these measures, and other people, particularly Kim Nobel, in previous, smaller-scale studies had found that there's a relationship between income and different brain measures. And obviously, there are lots of mediating factors that need to be considered. What's interesting -- I think, at least -- from these results, is that, as you might expect and as I've talked about already,

performance on many of the Toolbox measures was also related to income and parental education or measures of SCS.

And after -- in this particular -- in this particular graph, what you can see is that when you take into consideration age and sex and genetic ancestry factor, that also income -- that there's a mediating factor of the surface area measures and a relationship to the Toolbox measures, so that -- so that helps to sort of look at some of these different, complicating models in terms of, what does it mean that the brain measures are related to income, and how might these things be interrelated? We had a paper that we published a couple of years ago from the PING data, and our colleagues were particularly interested in measures of executive function and focus, in this paper, on the Flanker measure from the Toolbox. And we knew from their studies and other studies in the literature that the surface area of the anterior cingulate region of the brain was particularly important for cognitive control, and so that was the focus of this study where they focused on children who were performing quite well on the Flanker measure, and how did that relate to different measures of the surface area of the brain and the fiber measures that changed quite a bit as children develop in this age range?

And by taking the performance on the Toolbox measure -- the Flanker test, in particular -- what you can see here is

that there's dramatic changes in terms of speed of processing with sort of a stabilization in the later ages of children, and that there's this kind of classic cognitive conflict effect in terms of cognitive control as measured by this -- by this particular test. What this figure shows here is that the surface area of the anterior cingulate, which was the reason of interest for this particular study, was significantly related to the degree of cognitive control shown by the children in this study, controlling for age independent of their speed of processing.

When we look at the underlying fiber tracks in the brain, the performance in terms of cognitive control was also related to the properties of how well those white matter tracks are developing, independent of processing speed. So, a nice example of how we can look at the relationships between performance on these cognitive measures and different aspects of brain development, not only in terms of the absolute structural size, but also some of the fundamental underlying properties of white matter tracks. And I'm going to just briefly talk about one of our other studies. One of the goals of collecting the GWAS data from PING was to be able to look at how gene associations may be related to different aspects of developmental phenotypes. So, one of our colleagues on the PING study, Jeff Gruen, has been particularly interested in reading

disability and language impairment and, using the data from a large study in England, what they were able to do was to -- was to identify the particular types of phenotypes that are related to when a child has a reading disability and a language impairment; what are the variations that have been associated with that? And then taking that information, applying it to the sample in PING of typically developing children.

And what they found was that -- I don't think I have a pointer, and you don't really need to know all these details, but in looking at a subset of the children from PING -- a sample of 440 individuals -- there were two genes that had been associated with having a reading disability and a language impairment: those first two genes that are up there. And what you can see in this particular table is that if you look at those particular allele variations, they were significantly associated with performance on the picture vocabulary test from the Toolbox, and -- but not on the oral reading test from the Toolbox. And these are kids who, by and large, do not have any type of a reading disability or a language impairment, and the idea is that this may help us to better understand, perhaps, how these particular gene variations may be related to aspects of language development.

What they also found was that if we looked at the fiber track volumes and -- here, what you see is the fiber track

volumes across the brain and how it's related to vocabulary performance from the Toolbox, as well as reading, that there were -- that there were significant relationships there once you've corrected for age, and handedness, and gender, and socioeconomic status. And I'm going to just briefly tell you about another example of a study that my colleagues conducted, again, using a particular variant. Here, they were looking at a particular variant that's been associated with schizophrenia, and the idea behind this type of approach is similar to the Eicher, et al. paper, which is if we look at a particular variant that we know is associated with a particular disease process, the questions is, why are those things related? Because we know that it's not a simple type of one-to-one correlation, but how might that variation be associated with differences that we see in cognitive test performance, as well as different aspects of brain development.

So here, there's a particular variant, and this is actually a much more complicated story, not that the other one wasn't also complicated. But by looking at this particular variation, what they found was that one particular -- one allele frequency that's associated with a higher risk was also associated with variations in terms of the specific brain regions that they were interested in looking at that were -- that we thought might be associated with development of

particular functions that might be more likely to be compromised in schizophrenia and other types of related disorders. And then they also found that this variation was associated with variations on the Toolbox performance as well.

So, I've just given you sort of a very brief overview from the results of our study, but I think that it's of great interest in order to better understand how these variations in brain development and genetic variations are associated with certain elements of important cognitive test performance in children. And the idea is that, by better understanding these or examining these properties and then applying them to different patient populations, we'll have a better understanding and the advantage of having a very large sample size for better understanding smaller clinical populations. Thank you.

[applause]

DR. YAO: Before we get into the discussion session, I'm going to modify the agenda a little bit to take into account our little bit of a late start. So, we're going to, I think, abbreviate, just slightly, the conversation and discussion section -- session for this session to about 10:30, and then we'll take a five to 10 minute break, and we'll reconvene for session two at 10:40. Now what that will do, everybody, is that I'm cutting down on your lunch, but that's always the first thing that goes. So, we'll have to cut down on the time for

lunch, which means that, at the break, it is very important for those of you who haven’t ordered a lunch to go over there and preorder a lunch. But with that, I don’t want to cut into any more, so we’ll start into the discussion period.

DISCUSSION

DR. COMO: Okay. That's okay. Lunch here isn't that great.

[laughter]

First of all, thanks to our first set of speakers this morning for those very interesting talks. The PING study is obviously very fascinating, and there's probably a lot more data coming from there. We've assembled a distinguished panel to discuss some questions that we've put together, and we'll also invite questions or discussion from the audience, too, but I first want to give our panel an opportunity to sort of discuss some of these questions. We really only came up with three major questions. The first one, as listed in your agenda, is, you know, which clinical areas and which measurement of neurocognitive or neuropsychological -- those two terms seem to be interchangeable -- have been the most useful?

And I think we've heard about which scales may be qualified for that. Perhaps more important of the discussion point is, what are the limitations of the available scales that we have now? And then, finally, we would ask the panel to discuss the gaps, if any, in the current standards of age-specific measurement of neurocognitive and behavioral function in children. So, I'll let anybody on the panel jump in as they

-- as they see fit. Any takers? I know Elsa's chomping at the bit to -- (laughs).

DR. GERSHON: Hang on a second. Taking off on Marilee had earlier, there's really a dearth of measures in the -- in younger age ranges. I mean, the -- I have been responsible, also, for the measurement strategy for the National Children's Study, and it was just an area where people were very excited to be doing things. I am happy to report National Children's Study was working on its serious measures in this area, and some might, come hell or high water, by the end of this early fall will release what they had done. But there's nothing miraculous in that area and unfortunately, particularly with some of the diseases we're discussing today, those would be the most valuable instruments to have.

I think a huge part you could see on any set of slides today, the variability at any age is always there. You know, I was looking at Natacha's slides, even just showing -- we always get hooked up with numbers, but as scientists, we get hooked up on a really little difference being significant. But sometimes, it's really a little difference, and people are all over the way and they have very productive lives and are very happy, so -- and then you get early childhood development where that variability is out the roof. So, it's not going to be something easily resolved, but it's certainly an area that we need to keep

working on.

DR. ALLEN: [inaudible]

DR. COMO: Could you turn your mic on, please?

DR. ALLEN: The National Children's Study -- have they found measures for neurocognition in early infancy or neonates?

DR. GERSHON: I know this much. I mean, there literally were dozens of different researchers working on things, and one of my takeaway tasks is to make certain they don't get lost. There is everything from -- a good goal was to take sometimes existing measures, which take a high degree of training to be reliable whatsoever. So, I forget what the one is that does puppet facial recognition task, but really takes a well-trained person. It takes weeks of training, and then you have to have that trained person do it, and migrating that to a task that could be delivered on a screen. And that group, I'm pretty certain, is actually going to finish their work by the end of the fall. We begged, borrowed, and pleaded, "Please let some of this work finish." But it -- because part of the thing is sometimes there are good tests, but good is relative, again, in this age range, but they're not there. I will -- I'm hopeful the National Children's Study will broadly advertise what's there. There are political reasons why that may not happen, but I am -- but at Northwestern, we'll release them regardless, a nice thing about being an academic.

DR. ALLEN: I look forward to seeing some of that. It'd be great.

DR. COMO: Go, Dr. Adams.

DR. ADAMS: I think that, you know, another area that I think about a lot and is pertinent to yesterday's workshop is the dearth of assessments that are available for children with different kinds of handicaps. So, for example, I noticed in the NIH cognitive domain of the Toolbox that only one task doesn't require visual skills to complete, and it's a supplemental task. It's the auditory verbal learning. So, for children who I study who are blind, the Toolbox, unfortunately, isn't something that we can tap into as much as we might like to. But I think, even for children with other kinds of physical disabilities and, particularly, as we go through today's workshop and look at assessing outcomes in children who are undergoing therapy for various diseases, that's going to become relevant.

DR. COMO: Dr. Baron?

DR. BARON: And on another practical matter, you know, looking at -- certainly, there's a lot of good things about the development of these tests, but looking at it from the preschooler age level, it's really not just a matter that the test may not be tapping. It's really about how you give the test. And if anyone thinks that a 6-year-old is going to respond the same way in the testing situation as a 3, 4-year-

old, it's not going to be the same. So I think, when you're giving a test like where you have to come in and maybe do it very quickly, and you're taking an anxious little child away from his mom and expecting it to happen in seconds, minutes, it's not the same kind of an evaluation. And the other part of that is, if you have only one measure for language and one measure for episodic memory, you're not going to get the breadth of understanding from multi-determined tests what really is an underlying neurological -- neuropsychological function. So, that's been the primary concern that I've had all along about that, is that the development of it must take into consideration that dynamic maturational level.

DR. COMO: Dr. Bull?

DR. SHAPIRO: I have -- Oh, sorry.

DR. COMO: Oh --

DR. BULL: If she wants to answer, that's fine.

DR. COMO: Well, go ahead. I'll let you fight it out [laughs].

DR. SHAPIRO: I just -- I just have a question. So, in clinical trials, one of the things that you need to do to see the effect of a medication or drug on a child is to repeat testing over and over again. I would like to know, from a Toolbox side, what kinds of practice effects are there? What -- has anybody looked at alternate forms of the test, so that

practice doesn’t interfere with the scores that the children get? Because you may have, as a result of alternate -- not having alternative forms, an increase in performance, rather than the effects of the medication.

DR. GERSHON: Sure. I’d like to, first, point out that the Toolbox is not a be-all and end-all. My life’s an assessment. I’m well, well, well aware it’s a type of tool that we can add to our -- the bigger Toolbox that all psychologists carry and people ado, and so, it has a lot fewer limitations than most other batteries, but it’s still limited in terms of length. And any test that’s not given with a person who’s making certain that the participant is acclimated to the testing situation is problematic. It was developed so you didn’t need as high level of training, but just the recommended trained and, even for the iPad version where she’s spending several hundred thousand dollars of producing all new training materials that are relevant to that, but even that’s not enough.

Okay, so the Toolbox -- there’s a fair amount of published literature. There are too many tests to run through. Memory tests are the ones that are most subject to practice effects. There are a minimum of three forms of the memory test out there, so you can do that. You can’t test weekly. Frankly, there is no such -- I mean, I’m -- if somebody has one, I’d love to see it, that, you know, has more than a form or two that’s

useful or that's been calibrated and then normed on the same scale. But the Toolbox has three forms on the primary memory test. You can see reach practice effects, and the literature is pretty significant. You give the memory task, the person does better on it a year later. The brain is a wonderful thing, and somehow this task, which seems irrelevant and you, and I looking at it offhand say, "Well, nobody's going to remember that." They do. Their scores are better.

So, the Toolbox has that. Many of the subtests from the Toolbox are dynamic, and so you simply don't get the same content. Twice, their randomization components in there -- so, the vocabulary test, the exact same person, exact same ability, will see a completely different set of items if they take it five minutes later. Is that perfect? No. It just happens to be better than a lot of things that we have available. So, again, I've -- last person (laughs) -- been testing kids my whole career. Kids are dynamic. The reality is there -- I would argue there's no such thing as a completely standardized administration of a neuropsych exam. Maybe the same administrator, but most of the diseases we saw yesterday, you're training people all over the world.

They're just not the same. The use of any test -- you know, the nice thing when we're doing trials is you're looking at averaged effects over people. When using it for pure

clinical purposes with a single individual, life changes very differently. You must give a deep battery. You're never going to know what's going on without doing that. There's wonderful power of statistics, however. If your real goal is to see how two or three hundred people move, the nice thing is a lot of all this smooshiness [sic], for lack of a better term, in the testing actually goes away, and we can prove what we need to prove.

DR. ADAMS: So, practice effects is one of the bugbears of repeat assessment in neuropsychology, but I wonder if one thing to think about with the Toolbox or any other approach, is to have a multiple baseline run-in so that you are giving multiple versions of the task, one after the other, and then, once you've maxed out on that baseline or that practice effect within a session, you go ahead and you administer it again, and that's where your data is coming from. And I just happened to see on the name tag list for today that Brian Harrow might be here from CogState. I know they have that build into their battery, which is a computerized battery. I don't know if he's in the audience or not, so I just bring that up as something to think about.

DR. GERSHON: Well, and most executive functioning tests in the Toolbox, that is built in. That's built into the time. It's not -- it doesn't take the initial score. Or, it

takes both the initial or an average or a change score to see what's there. Yeah, you've -- some of these cognitive constructs, that's just -- if you're not doing that, you're not doing anything.

DR. SHAPIRO: So, you really have a novelty effect also in when you're giving these kids the tests for the first time. They are not familiar with the whole process, and then you repeat it again and, often, multiple baselines for any kind of clinical trial is really important in order to get the sort of true performance that a child has. And I think that that's one of the things that needs to be built in to a Toolbox like this where you can do the test twice, for example.

DR. COMO: Dr. Bull?

DR. BULL: Just very quickly adding on to what Dr. Adams -- I think my question was very similar, and I was wondering if there's any discussion in the Toolbox arena of validating for specific populations, such as Down's syndrome, which has a burgeoning amount of research that's underway.

DR. GERSHON: Sure. There are somewhere between two and four hundred validation studies going on right now with the Toolbox. The reality of science is I don't know all of them, and we don't require people to report on exactly what they're studying. There is absolutely -- out of University of Washington, there's an intellectual disability study. They have

-- they have a cohort of Fragile X; they have a cohort of Down's syndrome; there are two or three autism studies in progress right now. The original norming sample's quite large. It actually included sample sizes of 30 to 50 of a lot of different diseases, so we have some pre-information how that'll work.

Yeah, as in any measurement battery -- I hate introducing a new measure because the reality is it should be an augment to what people are doing early on, until the validation of insists there that is, you know, valuable for the purpose you're using it for in a particular disease. The good news is that there are a lot of generalized measures out there that people use all the time, most of them not validated in the disease if you actually go back and check, and the Toolbox, again, correlates in the mid-90s with most of those. We have every good reason to believe that that will work. Is that proof? No. That's not proof. But yeah, sure. It's out there. It's being used all over. The nice thing is it's being able -- it's being used across the age span for that so that, hopefully, we can compare data over time, hoping Natacha goes back to her people in 10 years and, you know, gets the funding to re-look at them, and we'll see there -- yeah, that works on going and being rapidly published.

DR. COMO: I think it's important to -- probably don't need to remind this audience, but I don't think we ever

eliminate the practice effect, and what we try and do is go to great lengths to try and minimize the practice effect. And of course, in clinical -- randomized controlled clinical trials, hopefully some of that is accounted for by the randomization. Dr. Adams?

DR. ADAMS: Oh, I was just going to say that I don't mean to beat up on the NIH Toolbox. I'm very excited about it, and I was at some of the initial meetings that were held publicly about it. So, I think just because you were presenting it today it becomes a focus, but you know, the issue I raised in terms of tasks -- the limitations of the dearth of tasks for disability -- children with disabilities, I think, is applicable to all of the assessments that we deal with, regardless of who has developed them, whether it's NIH or Pearson or anyone else. I think another area that we need to have further development and thought about is tasks that do have cross-national, cross-cultural applications. We talked about that a bit yesterday in the extent to which we need to worry or not worry about having tasks be exactly, perfectly validated in multiple populations.

DR. GERSHON: And I -- I'm in mourning for what -- the National Children's Study because we had approval for revising all these tasks in people with compromised hearing, revising all these tests to be available for people who are blind, and we certainly can't -- you know, most current assessments of kind of

capability assume you have vision. It's not just the Toolbox and, obviously, people who are blind -- that's not -- that's not the issue. It's just the task. All assessment tasks we do are artificial. There are very few that aren't trying to project something here, and it's okay. I'm used to the Toolbox being a -- what is it?

FEMALE SPEAKER: Punching bag?

DR. GERSHON: Whatever [laughs]. A punching bag. But the issue is this, is actually, we're very forthcoming in that and looking at other tasks and trying to build off of what's there because, again, most of the assessments we're using just aren't that great. They're the best we have. And so, the goal is to keep building on them and, so -- but, I'm very -- I believe -- again, I said earlier, you need to give vision and hearing just to know if any test you're using is good, but we don't have alternatives. It's really horrible, and nobody seems to be sponsoring that work, which is -- just is -- we have to work on that.

DR. PANDINA: So, two things I wanted to bring up. One, you know, clinical population norms in other disability populations -- I know it sort of was mentioned, but I think one thing that we've discussed, especially when we're talking about psychiatric and neurologic conditions is, is there just a brain disorder present, and what would be the effect of a brain

disorder generally? And then what would be the specific effect of a specific disorder? And do we want to have specific testing around each of those? And would we like to have any specific normative data to just say, at least in one population, we know exactly what the effects -- the cognitive effects and the measurement that we're taking are. So, that's one point.

And another is state dependent effects. I remember being at a training one time for a clinical trial, and I had a coordinator who'd never given a clinical test, and then there was me who had given like 400 WISCs, and we were both being trained to give this cognitive test. And I thought, gee, I don't know what they're going to get from this data. It's going to be a real challenge because, you know, part of what experienced assessors are able to do is to -- is to both make the participant comfortable -- you can design tests that make the participant more comfortable or less comfortable, but it may be that that state-dependent effect is important, and we don't typically rate that in the testing session. We don't say anywhere, especially in clinical trials, how valid do we think this is of their -- an estimate of their ability because of X, Y, and Z factor? It's something we always do in a clinical report, so I think some estimation of their state and how valid an estimate -- we think it would be an interesting addition to add, even if it's just an experimental measure.

DR. COMO: Do -- oh, go ahead. And then I'll invite someone -- if anyone from the audience has a question, too, I'd like to invite them.

DR. WABER: Okay. So, I'll be brief. I just -- to change the topic a bit, one of the things that I haven't heard discussed at all in any of the presentations this morning that's really important with children and adolescents is questionnaires and parent and teacher input. And in our research, we always, always, always include parent teacher -- teacher, if we can get it, which is harder -- but parent questionnaires because a lot of the tests that we give, in order to get the reliability and validity and so forth, become so narrow in their scope that we miss -- I mean, this is true in spades for any achievement test that you give. You can -- I've seen kids who do totally fine on an achievement test and simply cannot do school. So, I think that -- and it's also -- it's expensive to bring kids into the lab. It's -- you get recruitment bias and so forth, and so I think there's also a place for these structured questionnaire measures, for parents and teachers especially, in the pediatric area.

DR. GERSHON: So, I -- just quickly, so –

DR. COMO: Okay.

DR. GERSHON: I represent a lot of those batteries. So, within the Toolbox, most of the instruments you'll see are

available both in pediatric self-report and in parent report, and then I refer to larger efforts in that, such as patient-reported outcomes measurement information system, Neuro-QoL, which concentrates on neurological disorders, and all those are also adding an observer report; not just teacher. These are well-validated, very short instruments that get there, and I would urge people, actually, to not get rid of child self-report, which is actually a growing field because there's so -- I mean, the bulk research has been on demonstrating that parents know best, and then you find out the kids care a lot about some areas, and they don't care about other areas. And we're -- you know, we talk about, particularly in the FDA, looking at the voice of the patient. Well, sometimes the patient's concerned with this, and yet, everyone's telling them they should be concerned with that, so it's -- I think it's a great point.

DR. COMO: So, with apologies, I think we have time for one question, but I certainly invite anyone in the audience to speak with any of our speakers or panelists during the break or over lunch.

DR. ALLEN: So, -- A.J. Allen from Eli Lilly, and I'm a child psychologist by training. I guess one of my questions kind of follows up on your comment, which is when we're dealing with the question of evaluating neurocognition in trials, we're dealing with a number of different situations. There are some

cases where there may be a disease state, like autism or whatever, where you're looking at some sort of impairment. Maybe you're trying to find out whether or not there's any additional impact of the drug on that from a safety standpoint or from a benefit standpoint. But that's one group that you're looking at. Another group maybe, when you're looking -- and I think one that hasn't really been covered here is -- so if I'm doing any pediatric study, ICH says I'm supposed to evaluate long-term safety in terms of neurocognition; doesn't matter what the drug is or the state, that's one of the concerns that out there.

We often get asked to do something with that. I can't necessarily try and find the 50 or 100 psychologists that are, you know, trained to do all this and pair them up with all the different sites that I may have this study going on. So, I guess one of the questions I have is, is there something -- and this may be something to maybe just carry on thinking about for the rest of the day -- is there something that we can give people as a screening test, you know? A way -- if we're concerned about cardiovascular safety, we can do pulse and blood pressure periodically and track how that's going; if we're concerned about growth, we can get height and weight and see how that goes, versus the growth curves. Is there something we can do to assess this long-term with the typical pediatric study?

We're not necessarily expecting that there's an impact on neurocognition, but we're trying to be due diligent and have some sort of trip wire or screen to alert us if there is an issue. Thank you.

DR. YAO: So, I'm just going to answer that real quickly, not because I have an answer, but for the -- as A.J. said -- for us to think about that moving into the conversations this afternoon. I'm going to go ahead and tell folks to run if you need to use the bathroom because we will reconvene in exactly seven minutes, so we can keep on trying to get back on schedule. Thanks.

[break]

DR. SHERIDAN: Help. How do I get these --

MALE SPEAKER: Oh, boy.

DR. SHERIDAN: How do I get the first speaker up here?

MALE SPEAKER: Well, let me see.

DR. SHERIDAN: It says welcome to White Oak. That's a big help, you know?

DR. YAO: I don't know. I can try and –

DR. SHERIDAN: I mean, I don't even see anything to click on there, you know? I was looking for a little icon or something. I figured I'd better not -- maybe I shouldn't even mess with it. That's the laser pointer.

DR. YAO: That's the pointer. How about this?

DR. SHERIDAN: This makes it go forwards and backwards, but you have to start with a slide set, so I don't want to do something real drastic. End of slide show. Lynne, help.

DR. YAO: There.

DR. SHERIDAN: How do we -- how do we get to the first speakers' set?

DR. YAO: I'll take care of it. I'll take of all the AV.

DR. SHERIDAN: Wonderful. Well, if you do all the AV, I'd be forever grateful.

DR. YAO: Yeah.

DR. SHERIDAN: Should we be asking people to take their seats or --

SESSION II: NEUROCOGNITIVE SIGNALS FROM ANIMAL AND HUMAN STUDIES: EXAMPLES OF THE EFFECT OF MEDICAL INTERVENTIONS ON LONG-TERM NEUROCOGNITION

DR. YAO: Yeah, if we could have folks take their seats, please. We're going to get ready with session two. So they're all loaded. So they're all loaded, and they're all in order. So this is the first one.

DR. SHERIDAN: Right. Deborah Waber. Okay.

DR. YAO: And I'll take care of it, so if you want to just go ahead. All right, we're going to –

DR. SHERIDAN: Well, good morning. I'm Dr. Phil Sheridan. I'm going to be sharing the section here, and our first speaker is Dr. Deborah Waber, who's going to be talking about novel approaches for assessing the survivors of acute lymphatic leukemia. Dr. Waber?

LONG-TERM NEUROCOGNITIVE EFFECTS OF THERAPY IN SURVIVORS OF ACUTE LYMPHOCYTIC LEUKEMIA: NOVEL APPROACHES

DR. WABER: So, thank you for inviting me. What I'm going to do is take you through a quick spin of about 30 years of research and, just to show you sort of where we came from and recent shifts that we've made in our approaches to this problem. Just for those of you who don't -- oh, I just also wanted to add that I added my colleague, Peter Cole, because as you'll see as I sort of work through this that we're getting to work that he led, and I wanted him to get credit for that. So, for those of you who don't know much about childhood leukemia, it was through the 1960s, a child diagnosed with leukemia really had months to live. It's one of the great success stories of cancer treatment and started with the breakthrough that -- what would happen would be that they would treat the blood, and then the cells would kind of lurk in the brain. They would -- and the drugs didn't get through the blood-brain barrier.

So, in the 1970s, they began doing what they called CNS prophylaxis, which was to treat not only the blood, but also the brain. And what they used to do that was cranial irradiation. Cranial irradiation opened -- had this -- had the effect of opening the blood-brain barrier so that the drugs could get through to catch those cells. And while it was a

great and very successful -- tremendously successful, actually; now 90 percent of children are long-term survivors -- it was -- one of the big problems that became evident early was the late effects in the CNS. And it was, clinically, very apparent to people who were working with these children, to the people who were treating these children, that when they got to school, many of them had very significant disabilities. So, that's where I came in. And this was -- I started working in this area in the mid-80s and I'm going to just show you some of the work that we did, just spinning through some of our publications.

So, this was the first publication that we had in the area in 1990. It was based on children who had been treated in the late-70s -- late-70s into the early-80s, and we looked at late effects, so by the time we got to them, we always sort of waited five years to start seeing the children. And it was quite apparent that they had deficits in their IQs, but also that -- I'm just going to try to find -- is this the -- I'm trying to find where the pointer is. Were you guys just using a -- yeah.

DR. YAO: You can use the pointer here.

DR. WABER: Use the pointer here, okay. Perfect. Okay. Okay, so one of the big things that we first discovered, and it was really -- you really didn't need statistics to see it -- was that the girls that we saw were much more impaired. Now,

the agents that are used, the main toxic agents, were radiation, methotrexate, which is an antifolate, and steroids -- prednisone. So, we saw these kids were much more impaired, and this just shows you the same distribution. So, what's happened through the years is, I mean, in cancer treatment, what people are looking for is to maximize the therapeutic index.

So, you treat as aggressively as you can, and then begin -- until you get a success in terms of the treatment, and then you begin to pull back to see how much you can pull back, thereby having fewer side effects, but maintaining the goals -- the gains that you made with treatment. So, what we've been working on through the years is, what accounts for the outcomes? And mainly, what we've been looking at is treatment variables. So, one of the first studies that we did -- and this is a Dana Farmer -- Dana Farmer -- Dana-Farber ALL consortium group, so it's a multi-site group. One of the first things that we found was that the combination of high-dose methotrexate and cranial radiation therapy was yielding a more -- greater impairment, and these are -- you can see that where I've put the red arrow. The other thing I want to sort of just briefly mention is that these are -- because we're doing a multi-site study, you have to kind of fly lean. You can't do the same kind of thing as you would in a clinical assessment. You have to sort of -- you know, somebody out there at all these different sites has to be

recruiting patients. You have to get comparability across sites. So the bad end -- what we want -- also wanted to look at would be comparability as the protocols change. We wanted to be able to compare cross-protocols.

So we use a very, very, kind of slim neuropsychological battery through the years. So here's a seven year follow-up to our 87 protocol, which would have been -- which meant they started doing the treatment in 87. Can't remember when we published the article. It was, you know, probably around 2000. But what we began to find was here we had children who were treated with 18 grades of cranial radiation. They'd come down on the methotrexate dose, no longer using those super high doses of methotrexate. But what we found is if you look at the IQs, even with radiation, we're beginning to sort of flatten out. So we get an IQ of about 100. We're not getting -- you know, once we come down to 18 gray, we're not getting that much more sparing, that much more kind of mitigation of the side effects. And we're going to continue to see this. Here, this is one of our favorite -- let's see here. Go back. Can't go back.

MALE SPEAKER: Hit that arrow?

DR. WABER: Yeah, yeah, yeah. So the right column is something called the Rey Osterrieth Complex Figure, which is something that we particularly love in our work. It's a complex

figure that the kids have to draw. And everybody, you know, it's -- no matter who you are, the kid's do very poorly on that. So you can see that there at the bottom. So, you know, in terms of our -- but we're getting very few areas where we're getting an increased level of low performance relative to what we would expect in the population.

Here we look at other kind of possible risks. Age at diagnosis, which we used to think was a big risk factor if they're diagnosed under 3 years of age. Not a big deal here. Males, females, standard dose, high dose methotrexate; I mean, we're not really seeing, for IQ at least -- the IQs are really pretty comparable. We're not able to really identify anything that's accounting for any bad effects and, in fact, you know, the mean IQ of our population, even with radiation, is 100. This was a subsequent -- we then started doing -- and we work in, kind of in tandem with the medical group in constructing these protocols. So they would build in randomized trials of different things that were supposed to mitigate these late effects. So this was a trial of hyperfractionated cranial radiation therapy because everybody was worried about the radiation.

So the thought was -- that the radiation oncologist has, "Well, what if we did 90 centigrade twice a day, rather than 180 once a day? Would that mitigate these late effects?"

And what we found was that we found a little bit of an effect for this visual learning. Here's a -- this is also a visual memory thing. But in terms of the IQs, they're almost exactly the same. There's no differences. And unfortunately, there was compromise of the efficacy of the therapy. So that's definitely a no-no. We do not want to go there. Next we compared -- we did a randomized trial comparing triple intrathecal chemotherapy with 18 gray cranial radiation. So the trend in leukemia treatment has been, as much as possible, to get rid of the radiation. And in order to do that they had to increase the intensity of the drugs that were given intrathecally in the spinal column.

So again, here we do this randomized trial. We're finding really not much. This is -- we're finding a little bit of an effect here for the vocabulary subtest of the WISC. The overall IQ is 100 here, 97 here, really not a significant difference. The only difference that I think was probably meaningful was rapid naming, which is just a cognitive efficiency measure. And I think we -- unfortunately, our numbers were not as great for that one. But I think, and that's a typical finding that you find after radiation is that processing is slowed down. But you know, big time, you know, we couldn't find anything that was really hugely interesting there.

This is another one that we did where we just compared

standard risk versus high risk patients. And here we can't really segregate what the particular agents are because the intensity of different agents is increased in the high risk group. And here we did get a number of significant differences. So there was about a four-and-a-half point difference in advantage and IQ for the standard risk patients. We found differences here on reading comprehension, calculations. So there were a number of differences. So we could see that, you know, if the overall intensity of the therapy was greater that, you know, we could see some differences. But it sort of, you know, it doesn't meet the standard of the randomized trial where you can really look agent by agent.

I was very interested in the steroids because we know that steroids cross the brain -- blood-brain barrier and have impacts on the brain. My friend Heather over here comes onto the scene. You can see her name there. She was -- Rochester was one of our sites. So we were very interested in that. We did a trial, as did the Children's Oncology Group, could not find much of anything. One of the subtests that we give of the IQ test where I put the red arrow, the matrix reasoning, which is a measure of fluid reasoning, we did find a significant difference. But overall, you know, the -- you know, there's just not much to write home about here. We thought that might be, you know, a place where we were seeing differences. We did

see -- if we look at, you know, sort of real world measures, we're finding that the dexamethasone group -- we thought the dexamethasone was going to be more toxic because it's a more potent steroid. And we are seeing a higher rate of special education in the dexamethasone group, but it was only marginally significant. So if we'd had 200 patients, we probably would have gotten a significant result because it is a -- you know, 12 percent increase is a meaningful increase, probably, therapeutically.

So over time, and this is, you know, we're going from 1985 to until about 2005. So that's a long time. What we've found is that late effects of ALL therapy have persisted, but they've become much more moderate as the therapies have been moderated as well. Many fewer children are being treated with radiation, which was thought to be one of the big culprits. Surprisingly, we couldn't really find that any of the treatment variables were accounting for much of the variability in the outcomes. We also found that sort of global patient demographics, the sex of the child, we found in the early years, probably if you increase at the very high intensities of therapy, females seem to be more at risk. And we never really figured out why. But once they came down on some of the toxicity of the treatments, we no longer saw the sex differences.

Age at diagnosis, which we, you know, again in the early years was a higher risk factor, no longer seemed, you know -- in our data, no longer seemed to be relevant. Yet we were able to see that there was -- there is variability in the outcomes among these children. And so we kind of pivoted to a different approach with the advent of, you know, molecular biology and genomics that began to say, “Well, are there children who are at a higher risk because of the way that they process these drugs that there may be some genetic risk that would increase the risk for some children but not others?”

So this is where Peter Cole came in. Peter’s an oncologist at Einstein. And the question that we asked was, “Why do some patients experience greater neurocognitive effects than others?” And there was, you know, kind of no one size fits all. We’d already sort of spent two decades trying to figure that out and really hadn’t come up with much. So what we did -- this is the publication that is in press in “Journal of Clinical Oncology.” But we looked at -- you know, we’ve examined different sort of families of factors. So treatment factors, we’ve looked at that. We didn’t really come up with much. There’re disease factors, environmental factors, such as socioeconomic status. We know that that is going to affect outcome, but that affects outcome for pretty much everybody, and then patient factors.

So the study that I'm going to show you, we focused on the genetic factors that might predict outcomes. We -- what we did was we basically repurposed a lot of data that we had collected. I mean, these data are excruciatingly difficult to collect because you're dealing with multiple sites. You've got to get a psychologist at each site. We were dealing -- we had -- two of our sites were in French speaking Canada. So we had kids who spoke French. We had another site in Puerto Rico. We had kids who spoke Spanish. So it was -- we were really threading the needle, and it was very, very difficult. And these are difficult, difficult data to collect. Just the recruitment thing for anybody who's ever done recruitment, it is the hardest part of any study. Collecting the data once you've recruited is a piece of cake.

So what we had was 983 patients who'd been enrolled on two successive Dana-Farber protocols. We had managed to collect outcome data after five years of treatment on 380 of them. And then there had been bloods that were collected at the time of remission for minimal residual disease, and they had been banked in a freezer. And so, literally, getting them out of the freezer [laughs] took about 18 months because the guy who had the freezer had left the institution, and somebody else inherited the freezer and what not. So this was almost a two-year process just to get the bloods out of the freezer, but we

did that. And then what we did was, obviously, the Venn diagram of putting together the neurocognitive outcome data with the genomic DNA. So we selected genes that were related -- that we thought -- we didn't do a GWAS because we didn't have large enough sample. But what we did was we selected genes that would be related to the agents that are used in therapy; so antifolate metabolism, glucocorticoids, oxidative stress, and neurotransmission. So -- and then looked at the functional polymorphisms, targeted genotypes that had a prevalence of more than 10 percent in our population. So this is -- you know, you don't need to read all these genes, but this was the set of genes that were investigated.

And then what we did was to estimate the effects of genotype on neurocognitive outcomes using multivariate regression. So, you know, going back to each of these four things, treatment factors, we controlled for cranial radiation. We controlled for what kind of steroid treatment. We also controlled for risk group of the disease. In terms of disease factors, we controlled for measures of socioeconomic status, which I think we had parent education. And then we also controlled for other patient demographic factors: age at diagnosis, sex, and race. And so what we found was -- we identified there was one particular gene, the NOS3, which if the child had the TT polymorphism, there was a fivefold increase in

low IQ, meaning IQ below 85.

These other two genes were associated with roughly a two-and-a-half-fold increase. Similar genes appeared for some of the other measures that we used. The Wechsler Digit Span, which is a member of working memory -- a measure of working memory. We had these two genes, not the NOS3. NOS3 came out for vocabulary. But for those of you who know the WISC, we estimated IQ based on vocabulary, and either block design or matrix reasoning, depending on when the -- which protocol, which epoch the testing was done in. So it's not surprising that NOS3 would come out because it is a component of IQ. And then we got from questionnaires -- we had a couple of attention variables and SO-C02a1 comes out as well as -- and then the COMT gene, which other people have related to working memory.

So these are just some analyses. As continuous variable, you can see here is the NOS3. We only had 30 patients with the TT polymorphism, but there's a very significant difference there. SO-Col2a1, that's for IQ. The middle panel shows vocabulary, and on the right shows the digit span. So that just gives you a sense of what the, you know, kind of the range of the differences is. So NOS3 -- and here I'm going to kind of plead that I'm only a psychologist, so if you want to ask a question about genetics, you can't [laughs] because I don't think I can answer, and this is information that I got

from Peter. But -- so we were interested in, what is this NOS3 gene which seemed to be so relevant? It regulates vascular tone but also protects against oxidative damage, which kind of makes sense given the treatments.

So homozygosity for the t-allele is associated with decreased activity of the endothelial nitric oxide synthase. And also -- which means that it doesn't have that protection and, therefore, we're seeing lower IQ. This is an article that was published a few years before we did our study from one of our collaborating institutions from the St. Justine Hospital. That's -- my timers done. I'm done. Okay. So what they found was -- I'm sorry. What they found was, and where he's highlighted it is, that the NOS3 -- they also found the NOS3 gene. Now I have to qualify that by saying this was a subset of our sample. So it's not like this is independent. But what they did that was more -- was quite interesting was they looked at IQ over time and found that children who had this NOS3 polymorphism actually showed a decline in IQ over time. So it wasn't that this was -- one of our big questions is, "Would anybody with this polymorphism have lower IQ than anybody who doesn't?" We went to the PING database, and you guys did not look at the NOS3 gene. So we couldn't use it. But I mean, that's a -- that's a big question that needs to be answered, obviously, and a qualifier.

But that's why this is so interesting because they actually showed that the IQ declined in these people. So to summarize, neurocognitive outcome after ALL therapy is variable. Susceptibility to toxicity may be influenced by common genetic variance, and the five variants that we did find that were associated with poor function are linked to neuroinflammation and/or oxidative stress. So we kind of feel like this is the next generation of late effects kind of work that can be done. Not just to look at, what is the toxic exposure, but then what is it about the host? What is it about the child that may make some children more vulnerable? Because those children -- one of the things that you never want to do is compromise efficacy. So you don't want to be doing any extra treatment or protective treatment on somebody who could benefit from a therapy and wouldn't be at risk anyway. So being able to ultimately identify who these children are who are at greater risk could lead, eventually, to some kind of therapeutic intervention. So I'll stop there.

[applause]

STRATEGIES FOR STUDYING COGNITIVE AND BEHAVIORAL EFFECTS OF ANTIEPILEPTIC DRUGS

DR. SHERIDAN: I will be the next speaker and talk about our experiencing difficulties with neurocognitive assessments in clinical development of antiepileptic drugs. I'm afraid I'm going to be raising more problems and questions than I'll be giving answers, but that's what the afternoon is for hopefully. Nothing to disclose, and these opinions are mine rather than FDA opinions. Epilepsy isn't a specific disease. It's a heterogeneous group of disorders more properly called "the epilepsies." And each of them produce a tendency for recurrent seizures. A seizure is an occasional, sudden, excessive discharge of cortical neurons that disrupts ongoing cortical function.

They may be convulsive or non-convulsive. There're many different seizure types. In general, we divide them into two major categories: those that start focally in the cortex and those that arise from deeper in the central cephalon, producing primarily generalized seizures. And each of these can be divided into many, many subtypes and sub-subtypes. In general, the partial seizures are caused by hypoxic ischemic injury, brain malformation, tumor, infection, trauma. In contrast to the primarily generalized seizures, which in general, are

genetically determined, although there are exceptions to both of these. But the point is that when we're assessing epileptic patients, we're dealing with a very heterogeneous population. Most of the patients have normal intelligence. Some have above normal intelligence.

But as a total population, they will have impaired cognitive abilities compared to a population of age matched, education matched, healthy subjects, even if the epileptic patients are not treated with antiepileptic drugs. So we have built in confounders when we're trying to determine if the antiepileptic drug is contributing to neurocognitive problems. Why is this so? Well, the underlying etiology of the seizure may also cause neurocognitive impairments, or the seizures themselves may cause neurocognitive impairments. There is psychosocial affects from having epilepsy due to restrictions or due to societal stigmas that still persist.

And then, finally, we have what we're really interested in today, the effect of the drugs. The seizures cause neurocognitive impairment. The age of onset of the seizures is important both in terms of total lifetime duration of exposure to seizures, and also because of the critical, age-dependent stages that Marilee Allen was talking to us about earlier today. The frequency of the seizure, duration of the seizure, severity of the seizure, and the mechanism of the

seizure is important. With regard to state-dependent performance, it's important to know whether the patient is in between seizures, interictal, having a seizure, or having just had a seizure, which will cause different degrees of neurophysiologic dysfunction. And finally, we worry about the accumulative cortical damage, which some types of seizures could cause. The goal of the antiepileptic drug therapy, of course, is to eliminate the seizures with the fewest adverse effects from antiepileptic drugs.

Now these drugs work by a variety of mechanisms, but what they tend to have in common is that they reduce neuronal irritability. Now unfortunately, this also means that they are likely to interfere with physiologic neuronal excitability involved in the developmental synaptic growth and connectivity during infancy and childhood, and in ongoing cognitive function at any age. Usually, the neurocognitive adverse effects, as detectible by routine neuropsychological testing, seems slight when we're looking at monotherapy; that is a single antiepileptic drug used in the standard therapeutic range. But of course, for a particular patient, it may be necessary to have multiple drugs, and perhaps, at higher than standard therapeutic range.

So for each individual patient, there's a balance of risk versus benefit. How do we currently study the

neurocognitive adverse effects of antiepileptic drugs as we are developing them? Well, we begin with nonclinical studies. Clinical trials then come, which gives us an index on short-term adverse effects. And finally, we do clinical open-label studies looking for more long-term adverse effects. Nonclinical studies are going to be addressed by one of the later -- I guess our last speaker this morning -- this -- Dan Mellon, and also by our panelist Ikram Elayan. They, in general, are involved exposure of an animal, usually a rat during the pre- and post-natal developmental stages, and then the rat is assessed as an adult.

A different category of nonclinical studies is the juvenile animal studies in which we consider what age child is going to be developed. Or, excuse me, what age child is going to be exposed to the drug and pick the developmental period in the animal model that would correspond to that, and see what the effect of the drug is at that specific developmental stage. When juvenile animal studies are done, besides the sensory motor cognitive function testing, typically a detailed neurohistopathology is done, looking for a developmental insult such as neuronal apoptotic changes. Hopefully, these studies will alert us before we start testing the drugs in the patient population, and also give us information that we can’t get directly from working with patients. Given the nonclinical background, we then advance to clinical trials. Some phase-two

trials are done for safety and tolerability.

Then we get into the pivotal trials for efficacy and safety, phase-three trials. Typically, these trials will have some measure of cognitive function and behavior as secondary endpoints. The advantage of these studies are that they're randomized, double blind, placebo controlled. Unfortunately, there's no agreement as to what specific assessment tools to use in these trials. We usually leave that to the sponsor, and if the sponsor picks a reasonable set of testing, then that's what is done. The third type of study, or the long-term neurocognitive assessments, over typically six to 12 months of observations. Now these are not randomized. They're not blinded, so that we have a different type of information here, but still helpful. We also have to realize that when drugs are initially introduced, both in the efficacy studies pivotally -- pivotal efficacy studies, and in these long-term, open-label studies, the patients are on antiepileptic drugs concomitantly in the background. And we're typically testing the new drug as adjunctive therapy. And that, of course, confounds any observations of cognitive deficits that are picked up.

Now there are, in addition to these, many studies in the literature in which researchers have tried to look at the cognitive effects of already marketed antiepileptic drugs. The shortcomings of many of these studies are very well summarized

by Dr. Kimford Meador in a chapter in the Wiley textbook of epilepsy. I'll very briefly list some of the points that Dr. Meador makes. There are experimental design problems, subject selection bias. Typically, these studies don't randomly assign to treatment groups or adequately match from either one drug to another, or placebo to the drug. There's nonequivalence of various clinical variables. For example, controlling for exactly how high the antiepileptic drug blood level is in a particular patient, or what seizure frequency they're experiencing around the time of testing. The nonequivalence of certain dependent measures, for example, did the treatment groups perform similarly on the dependent measures prior to treatment? Other design issues, sample size often is not driven by any statistical power considerations.

Selecting the meaningful test assessment, of course, is open to debate and will strongly determine whether you find any significant findings or not. The concern over test/retest effects we were discussing a few minutes ago. Finding the proper statistical analysis is crucial. And finally, if you find a statistically significant finding, does it really mean anything clinically? We are concerned that neurocognitive adverse effects are probably most important in childhood because even a slight affect could have a severe cumulative effect over many years and also because, as was discussed earlier this

morning, the developing nervous system is probably more vulnerable to adverse neurodevelopmental effects of the modification of neuronal excitability, which of course is how antiepileptic drugs work.

So what is the future then, in order to give us a better appreciation than we have now? Well, we hope that in the future, there may be some agreement on certain basic assessment tools that would be used in all studies to allow some degree of cross-study comparison as hazardous as such cross-study comparisons may be under the best of circumstances. In addition to these, it would be prudent to have studies that are specifically tailored to the particular, specific antiepileptic drug, or epileptic syndrome, that's being studied in a particular trial. We hope that these approaches will be discussed further this afternoon. Thank you very much.

[applause]

Now our next speaker is going to be an encore performance by Dr. Marilee Allen who will be discussing assessments in the Neonatal Intensive Care Unit.

NEONATOLOGY FOLLOW UP:

LONG-TERM EFFECTS OF NICU TREATMENT

DR. ALLEN: Somehow it went over here. Okay. There we go. Well, thank you for giving me to -- giving me an opportunity to discuss something that I think is very problematic. And that is neurocognitive outcomes in acute infants and what influences these outcomes. And more importantly, what are we doing in the NICU that we could change to improve outcomes? So this is a very difficult question based on how you define my title as to NICU treatments. So we can describe the outcomes of many, but not all, reasons for babies who are in the NICU and the results of that care. But if we're trying to talk about the effect of different specific aspects of neonatal care on neurocognitive development it becomes much more difficult. And we -- the only thing that I know to do is to look at long-term follow up and randomized control trials, which we know is a very expensive and difficult proposition when you're planning a trial.

So the NICU, lots of reasons why babies go to a NICU. Prematurity is a big one. More than half of our babies that we admit to the NICU are premature. But they range from the extremely preterm infant born at the lower limit of viability, who has often multiple organ system involvement as we try and

keep them alive, to the late preterm infants who are there for immaturity of some -- one of their -- at least one of their systems, often respiratory distress. That's one of the most common reasons for admitting a baby to the NICU. And they might need support to maintain their temperature and gavage feeding sometimes to begin with. But they're there for a relatively short period of time. But we still use drugs on some of these babies that have not really been tested systematically. So the only reason why I have a copy of this slide of this baby, who's really a very ho-hum baby in our NICU is because he's my son.

So there are also full-term babies in the NICU. And some of those full-term babies can be even sicker than the smallest preterm infants. This is a baby who had -- has neonatal encephalopathy, resulting from some asphyxial event and -- around the time of birth. And he is being treated with a hypothermia protocol. So we use whole body hypothermia, which is just a little less expensive than selective head cooling. And we're able to use relatively inexpensive cooling mattresses that they often use in operating rooms. And it's actually something that needs to be kept -- these babies need to be kept at a precise temperature for 72 hours and then very, very slowly warmed up a degree at a time following that 72 hours.

And of course, these babies can have multi-organ system failure because it isn't just the brain that's injured

with an asphyxial episode. But it can affect all the other organs. So -- and there are other reasons why full-term babies need to go to a NICU. Babies with congenital anomalies -- many of the babies that we have in the NICU need surgery for an anomaly or even have multiple congenital anomalies that affect many system -- many organ systems. And babies who have infection come to the NICU for IV antibiotics. And those babies can be very, very ill or relatively not so sick, but need the IV antibiotics. Respiratory failure is a big -- another big cause, and cardiac problems are another. So lots of reasons why babies go to the NICU.

I've just chosen a few of these groups to put in this slide, looking at what the risk factor was that they were in the NICU for, and then some of the reported rates of cerebral palsy and intellectual disability. The top half of the table are different gestational ages at the time of birth. And the rates of both motor and cognitive impairment grow up -- go up with every decrease in gestational age category. And that's really seen in virtually all outcomes, including survival, and organ injury from prematurity, and complications of prematurity, and then outcomes.

There's a line there for babies with respiratory failure. And you know this is a -- it's difficult to come up with these figures because everybody has different ways of

defining these populations and -- but actually that's kind of true for the preterm studies, too. And of course, they all have different ways of following them and determining the outcome variables like cerebral palsy or intellectual disability. Different ages that they see them at. And then, just to give you an idea of -- HIE is hypoxic ischemic encephalopathy. We actually -- I've actually been taught now to use the term neonatal encephalopathy because they're all neonates and because we can't really precisely determine what that insult was.

We're really looking at its effect on the baby. And the severity of it, mild, moderate, or severe, is really staged based on how the baby is presenting. How sick is that baby? How -- are they comatose? Are they having seizures? Are other organ systems affected? Are all part of the staging of it? And then we've only been using the hypothermia for less than a decade. So the bottom line is looking at those babies who had either moderate or severe neonatal encephalopathy and were treated with the hypothermia. And I actually don't -- haven't put together any of those results in any detail. But there were several randomized controlled trials that showed an improvement in survival with the hypothermia, and they looked at short-term outcomes.

So why are preterm and very sick neonates so vulnerable? Well, certainly the time of birth can be very

traumatic for some babies. And the more premature they are, the more -- the more vulnerable they are to traumatic injury. There's lots of exposures to toxins in medications, and I would really emphasize medications. Many of the medications we use, like diuretics, have not really been properly studied in different gestational age groups or even in neonates. But we've been using them so long, we're loath to stop using them. And they have some affect, we think, on the infant's illness. We talked a little bit about the hypoxic ischemia. They may not be presenting with it, but some of these babies -- premature babies for example, have very -- a lot of respiratory problems. And so during the course of their hospitalization may have events in which they have some hypoxia or ischemia. Infection and inflammation again, many of these babies are very vulnerable to infection. And even just an inflammatory response without infection can cause injury to organs, but especially to the brain.

For these kids who are critically ill, it's really hard to provide them with good nutrition. So we're probably not so good at that. And there've been some very smart people who've been working for a very long time trying to come up with ideal ways to feed babies either parenterally IV or enterally, often using feeding tubes. And you know, we work on improving that all the time. But there's also some basic defenses that we

all have that aren't quite so well developed in babies. And especially when they get sick, they can lose those defenses. So the blood-brain barrier is one thing that in very sick babies, their blood-brain barrier may be disrupted. And it doesn't protect the brain the way it does for the rest of us most of the time. We all have automatic mechanisms that control cerebral blood flow. So when our blood pressure goes up, there's a control there so that our brain doesn't see that really high pressure. And when it goes down, there's a control there that keeps the blood flowing through our brain. But for babies, that can be very inadequate. So, and especially as they get sicker and sicker, that -- they become very vulnerable to injury from blood flow.

And then there's a number of neuroprotectors that we all have. And many of -- many of them are common things we know about like cortisol and thyroid hormone. But for many of these babies, they're not making adequate amounts. And for the very premature babies, we don't even know what "adequate amounts" is. We don't even know what they should be. And then the last one that I'm very aware of is just the disturbance of normal development, normal neural maturation. We don't provide the kind of neuromaturation support that some of these babies need, especially if they're very, very preterm or if they're very sick. So our isolettes, for example, are very imperfect. They

can certainly keep the babies warm and maintain a thermoneutral environment for them. But they're not providing the kind of tactile and other stimuli that they would get either in utero or being held.

So, brain injury. Brain injury, those babies are more likely to have cerebral palsy and/or intellectual disability or both. And different types of injury -- the rates are different for different types of injury. So the -- at the very -- I tried to do it in order of severity, but that's very difficult because of the different kinds of injury. So the top line is germinal matrix hemorrhage when there's a very vulnerable, richly vascular area of the preterm brain that is vulnerable to hemorrhage. And that's the smallest amount of hemorrhage you might see. And then intraventricular hemorrhage is when there's hemorrhage into the ventricles that we all have, the fluid-filled sacs on our brain. And if they're really dilated, they can develop posthemorrhagic hydrocephalus, which is this one right here. Oops, I'm sorry [laughs]. And those babies need shunts and can have many complications from their shunts, and that creates further injury.

And then this next line here is looking at infarction, which is hemorrhage or infarctions, so poor blood flow or hemorrhage into the brain, the brain substance. And then there's white matter injury or periventricular leukomalacia,

which are little cyst formation where there's resorption of brain tissue that can be seen on neuroimaging. And then some of the -- there's some -- there's been some recent studies that have shown there're cerebellar hemorrhage in some of these babies. We didn't know that because our head ultrasounds, the neuroimaging we used, didn't really look at that part of the brain until very recently. But there are also babies who are premature who had completely normal neuroimaging studies. And those babies still have a higher risk of cerebral palsy, and especially of intellectual disability, than do full-term babies raised in similar socioeconomic status groups.

So, two things: you can have very severe brain injury, and like this infarction here, or a lot of extensive injury that causes resorption of brain tissue, and you don't always develop cognitive impairment. And then you can have no signs of brain injury and still have cognitive impairment. So let's talk a little bit more about the kind of cognitive impairment we see. So, and I'm going to be talking here about prematurity because that's where the most information is. There's a lot of studies about this now. And for those other subgroups, there's smaller groups of kids, and there's not many studies that go out to school age.

So if you look at their IQ scores, these population of preterm children, no matter what age group -- what gestational

age group you select have a normal range of IQ, but the mean is lower for the most part. And this is even shown in young adults, l9 to 35 years old. The -- if you go to the smallest, most immature babies born before 26 weeks gestation, their mean IQ for that population is only 84 compared to controls of 104. So, and especially if you look at it -- so there're more babies with -- who go on to be children with very low IQ scores. Or if you look at the other way around, which kids have IQ scores above 85? It's about, depending on the study you're looking at, 49 to 83 percent of preterm, school-age kids, versus 98 to 100 percent of full-term controls. So there is a big difference.

And very, very large studies have shown that for every week of gestational age and IQ, every week down, lower gestational age, the IQ decreases by somewhere between 1.3 to 1.7 points, which doesn't sound like so much except for when you get down to the extremely low gestational age at birth babies. So of course, babies who are born preterm who have signs of neonatal brain injury have higher rates of intellectual disability. And there have been some really nice neuroimaging studies that show -- that correlates with degree of white matter injury, reduced brain volume -- this is in, you know, school-age kids and adolescents -- and changes in the gray and white matter density. And it isn't just limited to IQ.

So if you look at preterm adolescents, adolescents who

are born preterm who have no cerebral palsy and have no intellectual disability. If you look at them as a group, they still have overall lower full-scale verbal and performance IQs. They can have lower receptive language scores. Many measures, looking at language, looking at visual, motor, perceptual skills, spatial memory tests, and then of course the academics, reading, spelling, written language. And a persistent finding in many studies is arithmetic. And then there've been some really nice studies now looking at different measures of executive function. And have found that these adolescents who are born preterm, even if they don't seem to have a cognitive impairment or motor impairment, still have lower scores on these tests of executive function.

I just like this to illustrate for some of the range of visual perceptual difficulty some of these kids have. So the top line is what they're asked to draw. The middle line is a representative sample of full-term controls, and the bottom line are babies -- are school-age kids who were born with birth weight below 1,500 grams. And as you can see, I mean that kind of representative sample shows a lot of problems with being able to just copy fairly simple figures. And I always think that if they have so much trouble with these simple figures, what does -- what's reading like for them? So long-term outcomes. There've been now a couple studies that have looked at longer-

term outcomes in adolescents and young adults. There's a higher rate of attention deficit disorder, somewhere between two to five times greater. It's not -- it's equally high in males and females in many studies. And it seems to be more of an inattention than a hyperactivity.

So I have personal experience with this because this is my son. And he had -- definitely has ADD, but so does his father --

[laughter]

-- and his sister who was full-term. So I'm sure there are interactions there. I don't know of any studies that have specifically looked at, it but I think this whole field is rife for that kind of study that looks at genetic contributions. He also had dyslexia. And we could recognize he was struggling with reading, and we got him tutors. We got him tested and then got him tutors that at least helped him feel like he was working on it. He made it through high school by working harder than everybody else. He had about four years of having blue hair, which he told me was because he worked so hard, he had to have something that made him look a little more cool. And of course, blue hair was cool, at least in those days.

But he was able to get through high school, I think, fairly unscathed. But then he went to a very competitive college. And relative to the other kid -- the other college-age

kids, he was at a real disadvantage because they were all smart, and they were all efficient. And he was smart, but not nearly as efficient, and he had a lot more problems in college. So, he did manage to graduate. He's now in graduate school for math, which is better for someone who's dyslexic than philosophy, which was his undergraduate major. And I'm sure -- I mean, he's a wonderful person. He's going to have a great life. We're so lucky. But I think being aware of what his problems were during his childhood and trying to get him through the school system where everybody's supposed to do equally well on everything is very important.

Because you know the most -- the worst thing that can happen is to fail tests over and over and over again because it really affects self-esteem and even your willingness to take that next test. So I think it's really important, just clinically, to follow these high-risk babies as they go through childhood. Just kind of a summary about neuromaturation in the NICU infants, so it certainly proceeds despite brain injury. The effects of injury are often not apparent right away. But of course, the worse injury, the higher the risks of motor and cognitive disability. But there is a remarkable plasticity in young infants, in -- that allow areas of the brain to take over the function of the part of the brain that was injured. And there are certainly -- it's been well demonstrated now in MRI

studies of adolescents that this often leads to alternative neuronal circuits that can lead to fairly -- to adequate function, but probably not the most efficient.

And that's what can get some of these kids in problem when they're competing with their peers. There's a very nice series of studies by Laura Ment and her group at Yale who've looked at language outcomes from a -- of children from a -- that were involved in a randomized controlled trial of a neuroprotection agent, indomethacin. And they've shown that many of these adolescents, the ones who have -- especially the ones who have problems with language processing, are still using both sides of their brain to process language instead of the usual shift over to the left side of the brain. And I'm guessing that's probably because there's injury or dysfunction of some of the areas on the left. But it does make them less efficient. And of course, preterm infants who have no sign of brain injury are still at risk for neurodevelopmental problems.

So how do we sort out the treatments that we use and what their effects are? It's extremely complex, and frankly, an overwhelming task. And it requires long term follow-up to actually see what the cognitive outcomes are. And the other thing that's complicating it is that what we do in the NICU that's part of neonatal intensive care isn't one thing. It's constantly evolving, and it's different in different NICUs.

It's even different with different neonatology attendings. So there's always a change in what neonatal intensive care is. So the only way that I know to look at this is a large, randomized control trials with long term follow-up. And I'm going to just quickly run through examples where this has been helpful for us to have some neurocognitive follow-up after randomized control trials.

So my first example is the use of high-dose dexamethasone, which is a corticosteroid for bronchopulmonary dysplasia, which I prefer to call chronic lung disease in preterm infants. Very high doses for 42-day period of time, which is fairly long in these very preterm babies, was found to be fairly effective in getting babies off of ventilators. And they sometimes had a relapse. So there was some controversy about how effective it was in the long run. But we all like getting babies off the ventilators, so we all started using steroids. And then about 10 years or so later, there were reports -- I actually have it better outlined in the next slide -- there was a report -- in 1998, there was a report from a randomized control trial that was done in Taiwan that showed that at two years, there was an increased rate of cerebral palsy in the kids who were treated with dexamethasone during the first week of life.

So that led to a lot of concern about what we're

doing. And after reviewing, doing a systematic review of many, many trials, Barrington came up with -- well, Barrington published a study that was a systematic review, and this was his conclusion: that for every neonate who was treated with dexamethasone, it probably prevents 10 cases of chronic lung disease. That's not to be overlooked because chronic lung disease is a very bad thing to have, and it affects a lot of your function in life, including your cognitive abilities. But there were 12 more cases of cerebral palsy. So there was definitely a downside to it. So that led -- that led to a long period, almost a decade of not using steroids.

But now there's some reevaluation of that because now we're -- then we saw a whole lot more lung disease, and there're lots of problems with the chronic lung disease. And there are questions now about, "Well, can we use a different steroid? Can we use lower doses? Can we use it for a shorter time and still see the lung affect without affecting the cognitive abilities?" And there was actually a recent meta-regression study that was done that demonstrated an effect modification by the likelihood of chronic lung disease in your population of premature babies. So first of all, we kind of agree that the studies are showing that you shouldn't use it too early, so not before a week of life. But if you have high level of chronic lung disease in your population of preterm infants, meaning if you select the

sickest babies who are most likely to get lung disease, it can be beneficial. And if they -- if it's a fairly low risk, you probably shouldn't be using it. So exactly what the parameters are really varies a lot, but the outcome series have certainly helped us in how we use it.

And I skipped over this; this is just -- this is a table from the study from Taiwan where they also follow kids out to age 8. And they did find long-term cognitive effects in not just IQ scores, but in -- three of the four subtests had lower scores in the kids who were treated with the high doses, long course of dexamethasone, and they also found significant differences in some of their academic performance. Okay. So, the -- let me go back one more. I guess the -- I'll quickly go through the evidence about inhaled nitric oxide because that has taken a good deal of my time a few years ago.

So inhaled nitric oxide is used for kids who have respiratory failure after birth, and there are plenty of full-term kids who have that. Our kind of ultimate treatment if nothing else works, is to send them to ECMO, extracorporeal membrane oxygenation, which is the heart, lung bypass machine until their lungs can get better, and when their lungs get better, they can come off. But of course, we prefer not to do that because there are lots of complications from ECMO, and so inhaled nitric oxide is something that's been used for

persistent pulmonary hypertension, which is part of the respiratory failure package in most premature babies -- full-term babies who present. And it can help treat and lessen the disease. So there were a number of randomized control studies that were done that showed that -- two of them looked at outcomes -- developmental outcomes. And they found -- one looked at one-year outcomes and the other looked at two-year outcomes and -- at least, that's what's in the literature -- and they really found no difference between the groups who were treated and those who weren't.

Well, the question comes up about, what about preterm infants? We see a lot of respiratory failures in preterm infants; should we be using nitric oxide? So, I was part of a group who looked at that systematically. And we found that there were actually many, many, many studies that looked -- randomize control trials, more than you'd think there would be -- looking at nitric oxide. Now this is a very important question because inhaled nitric oxide is a gas that's delivered into the respiratory system, and it's very, very, very expensive. So it has financial implications, and what they found is -- so there are a number of studies. And I just wanted to point out that, you know, most of them have follow-up to about two years. But some of them have very, very small numbers of kids, and generally, they're looking at the cognitive scores

that are below 70.

And only one of these studies really found an advantage in the group of children who were treated -- so fewer cognitive impairments in children who were treated with the nitric oxide. And when we did the meta-analysis looking at MDI, mental developmental index on the Bayley below 70, we really didn't find a significant effect. It doesn't mean that in some circumstances nitric oxide -- inhaled nitric oxide shouldn't -- you know, might not be helpful in preterm infants, but as a use for all of us in these preterm infants with respiratory failure, I think it's not -- it's probably not going to be beneficial. We also found -- those studies did find no difference in death either -- in death or disability.

And then the last one, I just wanted to briefly mention is a very well conducted study, looking at caffeine for apnea of prematurity and preterm infants. So, preterm infants forget to breathe. Caffeine helps them remember to breathe. They did a very large multi-site, randomized control trial. They did a follow-up to age 2 -- 18 months to 24 months, and what they found is that MDI scores -- low MDI scores of various categories, but especially MDI's below 85 were higher in the group that got caffeine. Wait a minute. I'm sorry, the percent of kids who had MDI scores below 85, that group -- that was higher in the group who did not get caffeine. In other words,

there was a benefit to caffeine in terms of their cognitive scores. So for me, that helps me as a neonatologist because caffeine is a relatively benign drug; it's pretty easy to use; it's pretty safe; it's inexpensive. And actually, the one other thing they found was that it made incidents of lung disease less. That's not a finding I expected them to find. So, if it has a beneficial effect, I'm going to be a lot more comfortable using that medication.

So in conclusion, and those are just some examples. There are more, but not a whole lot more because there are a lot of randomized control trials done where they don't do neurodevelopmental outcomes and especially not neurocognitive outcomes. So to some extent, we can describe the long-term effects of neonatal intensive care in general populations of NICU infants, but it's really difficult to determine how a given intervention influences neurocognitive development. And the only way I really know is a randomized control trial, and if we're talking about cognitive scores, it needs to be long enough to determine neurocognitive outcomes, which makes it very expensive.

So first of all, they have to look at neurocognitive developmental outcomes, and until we have more accurate measures of early neurocognitive development, we really need to be following up for a long time. And you could discuss; school

age, adolescents, young adulthood. You know, it depends on, well, how much money you have probably. And then the other thing I would say is that we should be determining sample sizes to provide enough power to detect significant differences in neurologic and cognitive impairments because most studies do not do that. Thank you.

[applause]

DR. SHERIDAN: Thank you. The last speaker in this section will be Dr. Dan Mellon who will be discussing the importance of non-clinical studies in identifying possible neuro problems that would arise as adverse effects from drugs with a case study.

CNS SAFETY EVALUATION IN DRUG DEVELOPMENT: SIGNALS FROM ANIMAL STUDIES THAT SUGGEST THE NEED FOR FURTHER INVESTIGATIONS. CASE STUDY: ANESTHETIC-INDUCED NEURODEGENERATION IN THE DEVELOPING BRAIN

DR. MELLON: Thank you very much. What I'd like to do today is give you a bit of a whirlwind tour on what we do from a non-clinical perspective to try to evaluate the safety of drugs and understand what some of these drugs might be actually doing to the developing brain, which is, as we've heard this morning, an extremely complicating and challenging prospect. And I'd like to discuss a little bit of a case study about how we've been working and trying to better understand what anesthetic agents do to a developing brain. And as Dr. Sheridan indicated, this drug class as well as the antiepileptic drugs is specifically designed to actually reduce to neuro-excitation state, and as a result, they may have a potential impact on brain development. Of course, I also, since I work for the agency, do need to remind you that these opinions are really mine. They do not express the views of the agency, and I do not have any financial interest to disclose.

So from a non-clinical perspective, when we first have a drug product that we're looking at to better understand

whether or not we're even going to go into humans in the first place, prior to that, we have a bit of information we can use and leverage to get a better understanding of what the risks might be to the central nervous system. We have a very extensive pharmacological binding profile, which gives us an understanding of whether or not it's going to be binding to receptors that are commonly associated with the central nervous system. Frequently, we'll get some functional assessments to know not only does it bind, does it actually activate or inactivate and antagonize those receptors?

And ideally, we'll be getting a pretty good amount of information associated with tissue distribution. Does the drug actually penetrate the central nervous system and, therefore, present a potential concern? Prior to even going into humans, we also do a pretty extensive central nervous system safety pharmacology study, which is frequently referred to as a functional observation battery. Sometimes an Irwin test, named after Dr. Irwin who helped to characterize these, and that is really a very extensive evaluation looking at reflexes, looking at responses, and at all sorts of behavioral assessments to give us an idea as to whether or not a drug is going to show something. It would look a lot like an impact on the brain and functioning. Those are actually done with a bit of timing so that you can understand whether or not they're reversible, and

we can get a better understanding of how a drug might be behaving.

The meat and potatoes, if you would, is actually coming in our general tox studies where we do have the luxury of looking histopathologically at the brain, looking at both macroscopic and brain weight, but mostly the histopath to give us a better understanding of whether or not there's any evidence of either inflammation or any evidence of neurodegeneration within that tissue. I will say that we generally do not have any cognitive evaluations that are conducted until generally after phase three, which is pre- and post-natal development study. That study, which is sometimes referred to as a segment three study, does include at least one if not more learning and memory functional assessments. And of course, if we are going into a pediatric patient population from the get-go or earlier during drug development, many circumstances will dictate the need for a juvenile animal study. That is definitely a case-by-case assessment.

So let's look at that pre- and post-natal development study. In this scenario, we can actually expose pregnant animals to a drug from the time of implantation throughout weaning. So the pup is actually exposed in utero, and possibly, depending upon how the drug partitions via the breast milk. In general, for most drug products these studies are conducted in

the rat. That gives us a lot of benefits because we have a lot of historical control information. But certainly, if the product does not work we have a lot of work in that model, which is the case for a lot of biologic products that may actually require the use of primates. In terms of the pup evaluations, we certainly look at pre- and post-weaning survival and growth. We do utilize the characteristic physical landmarks. That gives us a better understanding of whether or not there's some developmental delays, including unfolding of the ears, an incisor eruption. They do include a sensory function and reflexes, including like an acoustic startle or a surface-riding reflex. And we do include behavioral assessments.

There -- well, just as in humans, it's been very much debated as to what's the best test, or battery of tests, that one could use to get a better understanding of learning and memory, particularly in these animal models. So as a result, the international community has not come up with any particular recommendations. However, there are several tests that are more commonly used than others. When we do feel the need to try to evaluate particularly brain development, it's important to realize that that standard battery of studies really provides us with exposure evaluations, at least in the rat, throughout the end of their third trimester and the beginning of weaning. And our general tox studies generally don't start up later until the

animal is a little bit older.

That's actually done intentionally because, in many circumstances, we want to make sure that we understand the impact of the drug on the reproductive tissues as well as part of that general process. So you don't necessarily want to do those studies in animals that are immature; you want to make sure that you can get those endpoints. The downside of that is that there is a gap in that evaluation, and the juvenile animal study is designed to really bridge that gap and identify some of the concerns that we may have for a particular drug product.

So, what kind of non-clinical cognitive testing can we do? Well, we can't get rats to take the SATs, or we can't, you know, utilize the Bayley, but we can try to challenge them and see how well they respond to some of these processes. And in a large part, we can utilize the various types of maze apparatuses to get them to learn a behavior and see how well they perform and recall the task that's put in front of them. And to some extent, this really is very much like some of the games that one would play in kindergarten if you would want to put it into that perspective because all of these mazes really include aspects that are essential to learning in memory process. It involves the ability to retrieve previous experiences and utilize them for the task at hand.

One of them that we use a lot is called the Morris

water maze, and the way this particular test is done is there's a fairly large pool of water in the middle of a room that has cameras over top. And we can keep an eye on everything and control the environment. And the animal -- rats are tremendously good swimmers. They may not necessarily like to swim if they don't have to, but they're really good at it. So, one of the things that this particular test can do is within this pool of water, there is a submerged platform. Well, if you were in a pool of water, and you didn't really want to be there, you would try to get out of it. And that's exactly what the rat does.

And in this case, they're normally going to swim around for a little while until they can figure out one way to get out, and they will eventually find this submerged platform. This particular room is set up so that there are images, either shapes and colors in various positions around the room, so that the rat can actually orientate themselves within the room. When you place the animal in there, the first thing they typically do is go around the edges. But eventually, they find the hidden platform. And you can actually measure the amount of time it takes the animal to reach that platform. Or you can measure the distance that the animal actually swims to get to that platform.

Well, as you can imagine, if you take this animal and you put them in this environment repeatedly, you know, every

couple of days, they get better and better at figuring out where that platform is. They realize that if I swim in the direction of where the -- you know, let’s just say, in this case -- the red triangle is, I'm getting to that platform much faster. So they learn where that is; they learn how to get there, and they become much more efficient at it. And this process, as we know, actually involves specific structures within the central nervous system. So this assessment gives us an idea of how well the animals behave.

There are other, more complicated, different types of assessments that can be done. One of them is the Cincinnati water maze. This also is a situation where you have a starting point in a maze; there's always dead-ends. But there are ways that you can navigate the process to actually get to the end of the particular maze. And by manipulating the environment and manipulating the circumstances, we can get a pretty good understanding of how these animals respond, how well they respond to it over time. And I'll show you some data that will help illustrate what this type of study will actually show us. In that process, we're going to look at what the anesthetics do to the developing brain and the challenges that we have been struggling with for years now trying to better understand this.

And this story probably starts-- well, it probably started earlier than this, but one of the bigger papers that

came out was in 1999 in Science when John Olney's group actually was looking at an NMDA receptor antagonists, mostly in the context of trying to better understand their potential utility for ischemia and damage that can grow there. And they were looking at a post-natal, day seven, rat pup. They used a compound that's not approved, although it was investigated for use in humans called MK801, which is a very potent NMDA receptor antagonist, and they found that when they administered this drug to animals during their peak brain development, they had widespread neuron apoptosis. And as we heard this morning, there were actually these little black dots, our tunnel method of staining. Even in a normal brain development, there is some amount of apoptosis that takes place as a normal process. And you can see that there are some segments in here in a few cells. Well, there's obviously a very big difference between the saline treated arm and the MK801 treated arm in terms of the magnitude and the number of cells that have undergone apoptosis during this process. This paper looked at several other compounds, one of which we do know is approved, although not approved for use in pediatric patients, but is commonly used, and that's ketamine.

So what does this age and window of vulnerability look like? The next year, Dr. Olney has -- and his group were very much involved in trying to understand the impact of alcohol in

the developing brain. And I'm sure many -- everybody in this room knows that alcohol is actually an NMDA receptor antagonist and a GABAergic potentiator. And what they were able to show is that if you gave alcohol to these rat pups in terms of their brain development that they found very profound changes that really focused from the early on in the embryonic stage to about the post-natal day 14, which correlates with the timeframe where the rat's brain actually is developing more often. If you actually dosed them later on, you didn't see any of these changes, and as you can imagine, different structures within the brain showed different levels of vulnerability depending upon the timing. So when you compare this time period in the rat, this correlates actually with what is believed to be the comparative brain growth spurt. And this is simply just simply based on the size of the brain, and we heard about that this morning. So the rat's brain growth spurt actually occurs post birth at around post-natal day seven. In humans, in the scale -- and this is a challenging scale to look at because the scales are very different.

In rat, the bottom here is in days, and humans it's months, so you can see that in humans, we actually have a brain growth spurt that's growing during the third trimester and during the first few years of life, but peaking at around birth for the most part. These particular -- the challenge of the

non-clinical arenas that we need to look and be aware of when these changes are taking place when we're trying to understand the impact of the drug and the translation. So does this change? Is this just simply a lot of apoptosis that's going to happen anyway, and we're just getting them all out of the way now because they're redundant, or does this actually leads to functional consequences? And the first study that actually reported functional deficits was in 2003.

Vesna Jevtovic-Todorovic, who did some work with Dr. Olney early on, was looking also at a post-natal day rat pup. Six hours' worth of an anesthetic regimen that would be a cocktail that includes nitrous oxide, oxygenized fluorine, and midazolam, and they were looking at a variety of endpoints, including some behavioral assessments. So again, we'll look at data of the Morris water maze, and I'll walk you through this. In this case, they actually measure the distance that the animals swam before they reached that hidden platform that we discussed. The dark -- there it is -- the dark circles here show the control animals, and these are the trials. At first, it took a little while for them to figure out where it was, but once the control animals figured it out, they were pretty adept at getting to where they needed to be to get up on that platform.

What this illustrates, however, is that the animals

that were exposed at post-natal day seven -- and this is being evaluated at post-natal day 32, so the drug is not on board anymore. They actually had -- took a lot longer to figure out where that platform was. So they fell behind their littermates, what appears to be simply because they were exposed to this anesthetic regimen. They actually went and looked at these animals later on, and the animals who were exposed to the anesthetic regimen still fell behind their littermates, suggesting that this is a prolonged, if not permanent, change in their ability to actually manipulate and maneuver within this particular task.

Does this happen in primates? Well, our colleagues in the National Center for Toxicological Research were able to evaluate this using a rhesus monkey model to try to better understand whether or not this same phenomenon was taking place here. They evaluated -- they exposed pregnant monkeys to -- during gestational day 122, which is about in the third trimester, and also at post-natal day five and 35. We were -- you know, these are not the cheapest studies to do, so you can imagine that we wanted to make sure that we would see something or we would not, and so we utilized ketamine by intravenous. This is an anesthetic dose of ketamine; it's a 24-hour exposure, and they also utilized the exposure for three hours in post-natal day five animals to try to understand whether or not this

was going to occur. And the dose of ketamine that was used is a little bit higher than what would be used in humans; however, this is the dose that's required to actually anesthetize them.

So this is a clinically relevant physiological state that is being induced with ketamine. And what they were able to show is that they also saw evidence of neuronal apoptosis as measured by fluoro-jade. So, if the animals were exposed during the third trimester, there was a significant increase in the number of neurons that were undergoing apoptosis. The same was true at post-natal day five. By post-natal day 35, you could anesthetize the animals for 24 hours, and they were -- there was no evidence at all of apoptosis, so again, suggesting there's a window of vulnerability here that it correlates in the primate with that period of brain growth. The good news as well, post-natal day five, even during that period of brain growth, a three-hour exposure to ketamine did not show evidence at least in terms of apoptosis.

So does this have functional consequences? Well, we have been very fortunate that the NCTR has available and has designed an operant test battery, which again is a series of challenges and puzzles basically that primates can undergo as part of a learning exercise and be test varying aspects of their cognition. And this is a complicated panel, and I'm not going to have time to go through all the details, but suffice it to

say that they go through four different processes and at an incremental repeated acquisition assessment, which gives an assessment of learning. In that case, it's basically trying to figure out a sequence of levers that you press and get a better understanding of when you get those, and you press the levers in the right order, you actually get rewarded with a banana-flavored food pellet.

There's also a condition of position responding process, which helps understand color and position discrimination. There's motivational assessments that can be utilized with this panel and utilized in a progressive ratio assessments, and a delayed match to sample assessment, which is a short-term memory. And what that one is, is an image of a shape that is generally placed in this metal plate. The image is removed, and then different images pop up, and if the correct correlating image does pop up, the animal can press a lever and they get rewarded with a food pellet, and they also get a light signaling that they got it right. So, the animals are put into these learning assessments, and they're done over time, so you're testing them repeatedly as the animal develops. And the way this does is it gives us a pretty good idea, basically, of how well the animals are learning.

The beauty of this assessment is that you can actually use the same panel with kids. In fact, Dr. Merle Paule, who's

the director of NCTR -- this is a picture of his son who is much older now -- but they do the same thing. They don't quite respond as well to banana-flavored food pellets, but they do a great job with nickels. And as a result, when they've been able to test this and show that the accuracy of the response in terms of an overall test score using this apparatus appears to correlate quite well with the overall IQ scores. So how did our primates do? Well, this is an assessment of the primates, and what this is, is again a training score to see how well they do.

And they get certain points if they accomplish the task and do each one of these tasks in a progression. So they -- each challenge is put forth with them, and what this particular graph shows is that in red, the animals who are exposed to ketamine -- this is a 24 hour -- were lagging behind their classmates, if you would, in terms of how well they were performing on this particular apparatus. The lag actually continued for quite some time. These animals are still with us, and I think they're about six or seven years old now, but they're continuing to be -- finding ways of trying to challenge them and see how well they respond. But these particular deficits are pretty significant, and if you speak to Merle, he'll tell you of all the drugs that they've looked at from a neuro-tox perspective. This is a pretty significant magnitude difference that does have implications regarding how these

compounds may ultimately have an impact on human brain development.

So as a result, as you can imagine, when a lot of these data have been coming out, we've been trying to get an understanding of, is this actually happening in humans? And I'm going to show you one example of a retrospective clinical evaluation that was done in Mayo Clinic by Randy Flick and Dr. Wilder and Dr. Warner and his group. What this does is, they went back -- they have the unique aspect of being able to go back through all the health records and scholastic records in Rochester county outside of Mayo and look and see what was -- you know, whether or not the individuals who were exposed to anesthetics were having any challenges or were being able to be diagnosed with any type of learning disability. In this paper, which came out in 2009, the dotted line is an individual child who is exposed before the age of four, who's not exposed to an anesthetic, versus children who were exposed to a single anesthetic regimen, meaning they had to have one surgical procedure.

This doesn't take into consideration duration or anything of that nature, versus children who had to have multiple procedures that required anesthesia. And this is an overall learning disability diagnosis, which does suggest that perhaps one exposure may not be that much different than none,

but certainly if you were exposed a couple of times, there was some evidence that you would have a greater likelihood of being diagnosed with a learning disability later in life. Now obviously, there's a huge amount of challenge in interpreting these data because this does not disassociate the fact of an underlying disorder in the first place, but nonetheless, the challenge still exists: how do we know whether the anesthetic is contributing to it or not? This is just one study, and there's multiple studies who have tried to evaluate this retrospectively.

And more often than not, the authors do conclude that there appears to be a correlation with the anesthetic regimen that had the exposure. As a result, the agency has actually been -- established a public, private partnership with our colleagues at the International Anesthesia Research Society that we refer to as SmartTots. And that organization is designed or intended to try to promote further investigation and research into this to find out whether or not the anesthetic or the sedatives that have to be used in that NICU setting -- as we discussed, you know, just a few minutes ago -- whether or not these compounds actually do have an impact; and perhaps ideally find out perhaps some of these might have less of an impact; some of them -- there may be things we can do to intervene and prevent that if this is indeed happening in humans.

There are ongoing prospective studies that I would like to mention very quickly. Three of them are ongoing as we speak. One is called the GAS study. This is being conducted by our colleagues up in Boston as well as at an international site that is trying to compare the effects of a spinal anesthetic versus a general anesthetic for inguinal hernia repair. That procedure is a relatively short procedure, so it doesn't necessarily take some of the longer exposures that we would be, but we are trying to get a better understanding of whether there's a greater risk using either one of these particular methods. There's a study referred to as the PANDA study being conducted by Dr. Lena Sun and colleagues, which is actually trying to minimize the impact of the patient population by looking at twin cohorts.

So, twins, where one of the twins were actually required to surgically procedure and the other didn't. So, you take into consideration the socio-economics; you take into consideration the family environment and the home environment. You try to eliminate all of those and try to determine whether or not you can disassociate an impact there. One of the studies I think is really interesting, and there's a lot of work coming out of Mayo, but there's a study that's just getting under way called the MASK study, and why I think that's so interesting is that because Dr. Randy Flick and his colleagues are going to

correlate and work with our colleagues at NCTR and utilize that operant test battery in children who have been exposed to an anesthetic regimen. And so we have the unique opportunity of being able to compare what the primate studies were showing to what the non-human primate studies were showing, the humans. And we can compare both of them and see how well they respond and really utilize the information that we have to try to strengthen our translational research.

I would be remiss not to acknowledge the fact that we've heard a lot today to note that the brain is not just developing when it's growing in size. It is changing, and the synaptic connections that we've heard and that was described this morning are changing throughout life. And as a result, we cannot forget the fact that the adolescent brain is also going through a process, and it is unclear yet what some of these compounds might be doing under those circumstances. There are some very elegant studies looking at architecture that is a little bit more of a subtle change than perhaps neuronal apoptosis. So it's possible that exposure to various compounds, whether they be anesthetics, antiepileptics, or anything else have the propensity or possibility that they might alter the way those neuronal circuitry are actually utilized as we evolve in our world.

So with some concluding thoughts, I'd like to say that

honestly the non-clinical assessments that might actually trigger concern and suggest maybe we might want to think about doing some cognitive assessments in people include, in a large part, a basic pharmacology distribution of a drug. We can -- we can, you know, leverage that knowledge tremendously. We have the luxury and the clinical role of looking histopathologically and doing some preliminary cognitive assessments. And I think, based upon that information, in many circumstances, I strongly suggest the need for a juvenile animal study, possibly two, depending upon which patient population a drug program is actually going to target. The adolescent risk I think is probably very different from the neonatal risk, and if a drug is going to be used all the way down to that age range, which in the anesthetic world they are, we probably need a couple of different assessments because the brain is undergoing different changes during those time periods. Clearly, from a non-clinical perspective, we need to do both histopath as well as cognitive functional assessments.

I will say we do not, at this point, do social behavioral interaction studies on animals, but we can. And I think given some of the data that we're seeing, that might be a pretty good thing to start considering. I do think that the anesthetic story -- and I think you can -- the same can be said for many of the examples that we heard today that the non-

clinical data in the anesthetic world is quite compelling. I showed you two studies, but there's a plethora of studies out there that are reinforcing the exact same concepts that those studies illustrated. And it does provide us a really unique opportunity to look at how the same phenomena, whether it be a different test -- because I have actually had anesthesiologists tell me, "Well, we don't put our babies in water mazes." And it's like, that's true, but the exact same process seems to be -- and the exact same brain structures seem to be utilized to some extent in these different parameters. And the rat did certainly predict what was happening in the monkey, and we're still doing studies. But the concern is that the primate is actually -- may very well be predictive of what's taking place in the human.

I would like to acknowledge my colleagues here in Cedar who have assisted in putting together this whirlwind tour, Dr. Elayan and Fisher and Woo; and Karen Davis Bruno, our fearless leader; Sharon Hertz, my division director; Riga Roca [spelled phonetically] and Allison Lynne, who’s involved in the SmartTots arena as well; the folks down at NCTR who've done a tremendous amount of work in primates, including the director Bill Slikker, Merle Paule, Cheng Wang, Sherry Ferguson. And I would acknowledge, if you are interested in this particular subject matter, SmartTots has a website that does provide a lot

of updates on what anesthetics are doing and some of the research that's taking place and are trying to promote a better understanding of that, and their website is www.smarttots.org, so I encourage you to take a look at that. You can even sign up for email alerts if you're very much interested. So with that, thank you very much. I appreciate your time.

[applause]

DR. Yao: So I just did a quick time check and our best efforts to get back on schedule by the lunch break have failed, so we're going to [laughs] -- we’re going to -- and we don't want to shortchange the discussion here, so we're going to take the prerogative of moving lunch back a little bit. So, I'd like to be able to break for lunch at about 12:40 or so, but I think that we ought to reconvene at about 1:15. So, we'll do 12:40 to 1:15, should give us 40 minutes or so to get lunch. So with that, Phil.

DISCUSSION

DR. SHERIDAN: Great. Well, I would particularly like to hear about the animal signals since the afternoon discussion will probably focus more on the clinical. So we have a panelist -- Ikram, do you have any comments that you have, given what you've heard this morning?

DR. ELAYAN: Well, our thinking here is, what do you think as clinicians when you see our data? So, for pediatric studies and for the sect three studies, we do all of these, and we have signals sometimes that does show there is deficits in learning in these animals. And what we could do is just put it in the label. And that's what we think that as a clinician, we could probably use, or parent, for clinical trials. But is that -- is that what you want from these studies, just what we see in these animals, or what other things do you expect? And so we look, you know, sensory motor functions and learning and all of that, but how can you do that for on your own purposes? And, you know, all of the time we, you know, communicate with our clinicians, and we tell them our concerns, especially if we have neurohistopathology. That is toxicity that we are concerned with, and we try to limit the dosing in children based on these findings. But what other things as a clinician that you can use from animals that, you know, you want to use?

DR. SHAPIRO: I have a question. I deal with children with very severe problems and inborn errors in metabolism. They have multiple, multiple surgeries. I mean, the average number of surgeries by the time they're 9 or 10 years old is about nine. It's about one a year. Something like that, and so the anesthetic risk is something that we know about in these children, but we can't do anything about it. So one of the questions that I have, and maybe the animal data can help with this is, are there age -- these windows of age where we should avoid anesthesia? Are there times when anesthetic makes a difference -- a bigger difference for these children? Because we can't not do the anesthesia, but the question is whether there would be a time period, a window, when that would be better? And so the animal data would probably shed some light on that, but that needs to be translated also into human data. And so, you know, the corresponding windows of development.

DR. ELAYAN: Right. I'll let Dan discuss this because that's his area, but I will talk about other things that we do from a psychiatry and offer ADHD and autism and all of that. So we would try to, developmentally between an animal and a human, compare the ages and windows of vulnerability in children and animals. That's the best that we can do from whatever the literature has, whatever we know about the brain development in the human and the animal. That's our best guess, but as far as

anesthesia, I think Dan has better experience with that.

DR. MELLON: So, that's the important question is, what is the most vulnerable period? And, you know, in this -- at this point, the best we can extrapolate from the animal data is that the period of the most rapid brain growth, that synaptogenic period on the graph that I showed, appears to be the most vulnerable time period. And there are interesting data that are being generated that look at endpoints that are not just neuro apoptosis, but some of the synaptic architectural changes, such that the neurons are still there, but they're having dendritic spine differences. And we don't know whether or not that means that you're rewiring or establishing neuronal connections that were not there before that might translate into much more subtle changes.

If that's true, and we're not certain yet, that window might be a little broader. But our best estimate is that certainly the youngest of our patient population are probably the most vulnerable and that at least, you know -- with the primate studies from post-natal day five to post-natal day 35 probably is about under six months, but we've been pretty cautious about our interpretation because that's only one endpoint of brain development. So we generally get an idea that maybe under 3 years of age are probably the most vulnerable. The under six month are probably even more vulnerable if this

does indeed translate, but we recognize the fact that this is all a risk benefit perspective.

These kids who are that young are not going to have, you know, optional surgery. This is things -- these are procedures that need to be done. So the best advice we could ever provide, I think at this point, is if at all possible, minimize the duration if you can and minimize the dose of the drug that you can because this is dose and duration dependent. And the younger you are, the more vulnerable you may be. There are some wonderful data out there though that I think is a little bit reassuring as well. And that is, there's a neat series of studies in the rodent that were done that if you could produce these lesions in a rodent, but if you put them in a highly enriched environment, they compensated.

And they were able to overcome these particular deficits, such that they did not manifest into to more prolong. We're hoping -- it's very limited data at this point, but we're -- I think, you know -- one of the good things is that we're going to be interacting with these children as much as possible because they are more vulnerable anyway. The more enrichment they can get, the more hopefully adaptable the brain will be and the better off they'll be to minimize any impact if there is. We don't -- we don't know for certain how well this translates. It looks as though -- we just had a science advisory board last

November on that particular subject. The science board was pretty convinced that the animal data is quite compelling and that it's likely that the same phenomena will be taking place. That's part of the risk benefit that has to occur.

Again, as you point out, these procedures have to happen, and they have to happen with an anesthetic. What I think the animal data will hopefully help with is perhaps identify regimens that might be less problematic than others. And that's one of the things that's being discussed now. There are some drugs that look like they may not be as problematic. The opioids, for example, certainly have impacts on the developing brain, but right now, they don't seem to be showing the exact same type of thing that something like isoflurane and sevoflurane or even propofol would be showing. Dexmedetomidine is an interesting drug; it's an alpha two agonist. It doesn't seem to be showing the exact same profile, but I will say neither one of those drugs has been fully characterized like ketamine and sevoflurane have.

So we don't have the data yet, but we're trying to generate it as quickly as we can to find out, are there some procedures that you can use either a caudle block instead of doing a general? Are there some things that you could use, maybe a combination of an opioid plus an alpha two, that would avoid the need for a general anesthetic? And with a little bit

of luck -- we're all keeping our fingers crossed -- we might be able to find that. And the other area of investigation, I think is hopeful for the future -- it doesn't helps us immediately -- is that, are the things that we can do to intervene that would allow us to block the neurotoxicity if it translates into humans, but still accomplish the task at hand?

And I think that's where there's a lot of area of research as well. And hopefully, at the end of the day, and that's really what the SmartTots initiative is all about, we'll be able to find a way to do what we need to do better. And that is our objective; unfortunately, we don't have answers right now. But if you could say anything, the shorter duration at all is possible, the, you know, the least amount of anesthesia that you need to accomplish the task and help the kids that need help.

DR. SHERIDAN: Great, and we have some questions at the table here.

DR. SHAPIRO: If I could just make one response to that. And so there are other things that limit what we can do with these kids because they all have airway problems, so we can't use sedatives. And so the -- you know, you have a lot of confounding factors in this sort of thing. And separating out what the effects might be of the anesthesia is almost impossible, although we can see it in our long-term outcome

data.

DR. DELANEY: The other thing is people -- parents will postpone and then pile up their procedures. So if a four-hour anesthesia -- period of anesthesia is more detrimental then -- because of the airway issues, they're worried about the risk of surgery and doing general anesthesia. So when we see parents -- they're not emergency surgeries, but they know it has to be done, so they stack them. When they visit one of the medical centers that can handle the airway and sophisticated procedures, then they do it all at once, so the child is under sedation, general anesthesia for four hours, versus if you spread this out, can you wait until they're 4 years old for an orthopedic knee repair or whatever it might be.

DR. MELLON: I think that's the advice that one could give is that if it's a procedure that could be delayed a little bit, then delaying it a little bit is probably better for the brain development. But there's a lot of ramifications of that. So, a good example of that is if you have a cleft palette repair. I mean, you could delay that, but there are ramifications of not repairing it as well. There's social stigma ramifications; there's, you know, nutritional ramifications, so it's a -- it’s a -- it’s a very challenging scenario. And I think that the conversations have to take place with the physicians, with the parents, and very difficult

decisions have to be made in that regard and --

DR. YAO: So I'm going to take the prerogative to shift a little bit because I hope that what Dan and his talk and Ikram and her comments have pointed out is that his is an area of active and intensive research right now. But I don't want to spend a lot of time talking about, you know, what would be a recommendation because really, the point here is that, as Dan is stating, we don't know -- we need a lot more information. And really, I think what that calls for is that if we have patients of a certain population that are being exposed to these anesthetic agents right now for reasons that we know justify that risk, how would we potentially follow them up? That's really the question, rather than say let’s make a recommendation today about whether to use them or not.

DR. ADAMS: So, I think my comment and question follows from what you just said and what Dr. Shapiro said, which is that these data are fascinating, and I really enjoyed the question about, you know, clinician to the water maze researchers, how we talk in the same room about this. I think that a question I have is when we have children with inborn errors of metabolism or other diseases that disrupt CNS development, central nervous system development, I wonder if we also have to think about in these children who already have vulnerable brains as a result of their disease, whether the

window of vulnerability is different, whether the type of vulnerability is different, and whether the extent of vulnerability is different. And so healthy rat models will give us a really good understanding of the basic process that's occurring because of these exposures, but then we also have to look at models of animals that more closely resemble the diseases that we're working in where there's going to be, perhaps, excessive vulnerability.

DR. REAMAN: I think one potential limitation, and correct me if I'm wrong because this is really a question, is that there really is tremendous variability, so I think Dr. Waber's presentation clearly demonstrated that. So are there any pharmacogenomics or genetic predisposition correlates that can be done in that might better inform human studies, so that we can really rationally and with a sound rationale develop clinical, longitudinal evaluations?

DR. YAO: So before -- Dan, before you -- or Ikram, before you attempt that, Greg, could you introduce yourself because you're --

DR. REAMAN: Sorry. I'm Greg Reaman. I'm a pediatric oncologist and associate director of the Office of Hematology Oncology Products.

DR. ELAYAN: I will just address both the animal model and the genetics. So a lot of times animal models that we have,

you know, we don't really know if the etiology in that animal model is similar to the human etiology. And that's another problem, especially in psychiatric disorders. Maybe other disorders, maybe there is some genetic correlation between the human and the animal, and that's probably -- so far, what we see when we see these studies, we don't have any genetics that go with it, but I'm sure people are looking into that. But for other things that there's no correlation between the human -- it's just like, let's say, an ADHD model in an animal. It's a spontaneously hypertensive animal that, you know, is hyperactive. So what is that related to ADHD? We don't know. So how valid is the animal model to the human disease is very important. And then go from there, how is that related genetically to the animal disease and the human? You know, what is the correlation?

DR. MELLON: And I'll be honest; I am unaware of any genomic process that has been explored in this arena. The phenomena mechanistically that seems to be explaining what's happening is really an overall neuro-excitation state of that given neuron and whether or not the neuron is receiving the signals that it needs to stick around. And if it doesn't, then it undergoes an apoptotic process that's fairly well characterized, so the interesting thing from a neurodevelopment perspective that we're seeing is that this is happening in rats;

it's happening in monkeys; it’s happening in rabbits and guinea pigs, and mini-pigs, and we've actually even seen it in flatworms. So the -- it seems like it's a very basic phenomena that mechanistically is very translatable. And -- but -- so I'm not aware of any evaluations to date. That's an excellent question. I wish I had an answer for you.

DR. REAMAN: Because I think -- but what you're doing with the classive agents, specifically anesthetics and things that have a primary mode of action on the central nervous system, would be appropriate. But I think we have to be concerned about the broader classes of drugs for which we don't know what potential central nervous system toxicity will be.

DR. MELLON: Well, I think that's an excellent point, and when you really put it into perspective, as anybody who treats little kids around here knows, most of the drugs that you have available to you are not actually approved for that indication. We have very few drugs that are, and that's one of the efforts that's being put forth by the agency is to get better information, so that we actually have useful information for you. With the anesthetic agents, these are very old drugs. And many of those development programs probably didn't have a segment three learning and memory component to them in the first place. So, that's part of the reason why we're able to leverage the expertise of our colleagues down at NCTR to try to do that

and evaluate some of these aspects. And I think what this illustrates very nicely is that we need be trying to -- again, I have a bias, but we need to be trying to leverage the animal data as much as we can, so we know what kind of endpoints we should be looking for in the clinical studies and the type of behaviors to try to correlate the phenomenally fascinating assessments that you have available to you back to the animal models. So the animal models can get better and be more useful. And we have this beautiful communication back and forth from the clinical to the non-clinical to allow both sides to learn from each other and improve both processes. And that translation is critical, so that we can better predict this before we go into a clinic, and we know what we're getting into, so you can make those appropriate risk benefit decisions, which are extremely challenging.

DR. SHERIDAN: Are there any questions from the floor before we conclude? And please identify yourself.

DR. SHAMUS: Adam Shamus from BioMarin. Just one comment related to genetics and animal models. As people up here know, even when we're looking at some rare sorts of diseases, the animal models, both small and large animal models, the brain structures affected can be quite different than what's affected in the human. We see in some model, there’s cerebellar defects specifically, and then in humans, we know there's very

little cerebellar defect, for example, so even when we know the genetics, it's very challenging, you know, to have a lot of faith in the translatability. Just a comment.

DR. SHERIDAN: Good. Well, certainly one lesson I'm taking away from this is that there's a possibility of a continued interaction back and forth between the non-clinical and the clinical. It's not a matter of the non-clinical doing their work and then going home, and then the clinical goes forward; that there will be questions back and forth and that each can illuminate the other, so that's important to keep in mind this afternoon. Time to break for lunch, Lynne?

DR. YAO: Just a couple of housekeeping announcements. So we're at 12:37 right now, and as I said, I'd like to reconvene if we could at about 1:15. 1:15, 1:20 at the latest. Any of you who need transportation to or from this facility to the airport, if you could just meet out in the lobby real quick, we'll see if we can try and help you with those types of arrangements if you don't already have them. And again, you should be able to pick up your box lunch if you preordered outside. Okay, thanks.

[lunch]

SESSION III: TOOLS AND STRATEGIES IN THE EVALUATION OF NEUROCOGNITIVE AND BEHAVIORAL OUTCOMES IN PRODUCTS USED TO TREAT CHILDREN

DR. YAO: Okay, folks. We're going to try and get started again. I'll give folks just a few seconds to get -- come in and get settled. I do think we've schedule the day such that we can have, I think, a really, really interesting and robust conversation in the afternoon. We really intended to sort of squeeze things in, pack things tightly in the morning, which would give us more time, actually, to talk and discuss things in the afternoon. So we are running a little bit late, but I think that we'll be able to have that conversation and still get out on time. As I know, many of you have to get back to your home bases after this meeting. So I don't want to delay anymore. We're starting session three, and we have guest session chairs for this last session of the day. Heather Adams and Elsa Shapiro have agreed to be the session chairs. I'll be running the podium again. I wanted to introduce my colleague, Ann McMahon, who'll be giving the first talk. To set the stage for the afternoon, just some considerations on long-term safety and pediatric trials that have been submitted to FDA. We'll be followed by Dr. Bull's talk on the environmental exposures and neurocognitive effects. And then finally, sort of the wrapping

it all up to move forward with the discussion is Dr. Shapiro’s talk. So with that, I’ll let Ann start. Thank you.

REGULATORY CONSIDERATIONS RELATED TO EVALUATION OF LONG-TERM SAFETY IN PEDIATRIC PATIENTS

DR. MCMAHON: Thank you, Lynne. It's been a terrific day so far; I've really learned a lot from everybody. Great talks. I'm going to -- I have a few words to say about considerations of long-term safety in pediatric trial submitted to the FDA. My name is Ann McMahon, and I work in the Office of Pediatric Therapeutics in the Office of the Commissioner at the FDA. Let's see. I have no financial interests to disclose. I do want to say that the opinions that I'll express are my opinions alone and not necessarily those of the FDA. So I'll touch on definitions of long-term studies in children and purposes of long-term studies in children. I'll talk a little bit about preliminary data on duration of pediatric studies submitted to the FDA and categories of long-term designs of post-marketing studies submitted to the FDA, and I'll end by talking a various few slides on regulatory mechanisms in post-marketing submissions.

And there's a koala, waiting for what I have to say. So thoughts on definition of long-term studies. Basically, there is no definition of long-term studies of pediatric trials. There are -- in animal studies, the length of the study depends, in part, on length of the generation and on other factors, and

in children, studies depend on -- the length of studies depend on the length of the disease perhaps, the length of the intervention that you're -- that is of interest, perhaps on the age of the child, and perhaps on the outcome of interest. So there is no one definition of long-term, and it's something that I'm not sure has gotten a lot of thought and maybe would benefit in this area from thinking more about what long-term means.

So purposes of long-term studies. They can be -- long-term studies can be valuable in assessing both efficacy and safety, of course. And here, we're talking about safety. And some examples of safety endpoints that might be of interest after exposure to medical intervention are malignancies, or here, we're talking about neurocognitive outcomes. So we asked what the duration of the pediatric studies submitted to the FDA might be. One question of interest in this preliminary look at the data was how the different disciplines, or divisions, differed in study length. So far, we have looked at 117 pediatric studies. So this is a preliminary look for us. There were -- the two divisions that had the greatest median duration of pediatric trials were transplant and ophthalmology and oncology. And you can see that the number of trials in each is very small. So we'll be -- stay tuned for further analysis of this.

Now I'll turn to the designs of longer term pediatric

safety studies that are submitted to the FDA. So registries is one type of study that is longer term in general submitted to the FDA. In contrast to clinical trials, registries are not randomized, and registries employ an observational study design that may specify protocol for data collection. There are many minimal inclusion -- there are, generally, minimal inclusion and exclusion criteria, and the data are gathered when a patient presents for care at least generally. And registries measure outcomes under conditions of "the real world" and may be more generalizable to the general population than, say, clinical trials.

Observational studies draw on inferences about the possible effect of treatment on subjects, and assignment of subjects to a treated group versus a controlled group is outside the control of the investigator. Randomized control trials in the post-market setting are generally larger than in the pre-market setting. There's less confounding than other post-marketing safety studies. The interventions and outcomes are well documented, generally, and expensive -- and these studies, or rather trials, tend to be more expensive and difficult to carry out in the long-term.

To be complete about post-marketing submissions to the FDA, I'll talk about post-marketing requirements and commitments. Post-marketing requirements and commitment studies

in clinical trials occur after a drug or biologic product has been approved by FDA. Under various statutory and regulatory authorities, FDA can require manufacturers of certain drug products to conduct post-market studies in clinical trials. The 2007 Food and Drug Administration Amendments Act, or FDAAA, specifically provides FDA with authority to require drug manufacturers to conduct post-marketing studies and clinical trials to asses possible serious risks associated with the drugs.

So post-marketing requirements include studies and clinical trials that sponsors are required to submit under one or more statues or regulations. Post-marketing commitments are studies or clinical trials that a sponsor has agreed to conduct, but that are not required by a statute or regulation. So summary -- a summary of the studies that I mentioned above is that registry studies can theoretically be conducted for long timeframes, but do not have a comparator arm. Observational studies will allow for a comparator group and long timeframe, but have uncontrolled confounding. And randomized clinical trials control for much of confounding, but are likely to be smaller in size and shorter than is optimal due to their expense.

So some next steps, from my point of view, in developing long-term pediatric safety studies would include

defining long-term, as I mentioned before, as that relates to pediatric safety studies, and developing expertise in pediatric study designs that facilitate long-term and pertinent studies. So in conclusion, we are at the beginning of determining the definition and purpose of performing long-term safety trials in children -- trials or studies in children. We work -- we need to go into innovative study designs that will accommodate long-term considerations for children. And I want to acknowledge Dr. John Murphy and Lynne Yao for their nominating me to present here, and thank you for your attention.

[applause]

DR. YAO: So our next presenter is Dr. Bull, and here is her talk.

DR. BULL: Can I do this one? I can do this?

DR. YAO: Yeah.

DR. BULL: Awesome. Well, good afternoon, and I too want to thank everyone for inviting me to participate in this really wonderful meeting. I've learned so much, and I feel that much of what I'm going to share you may have already heard as part of your back experience, but my challenge was to share with you the environmental exposures as it relates to neurodevelopmental outcome, and to start, I unfortunately, have no financial interest to disclose.

[laughter]

And I thought I would start by giving you the -- sharing with you several examples of environmental exposures that we know or believe to have had neurodevelopmental outcomes, and I was going to start with the ACEs study. And I hope you have heard of the ACEs study. But if you haven't, it was a part of -- involved over 13,000 adult subjects who were part of a comprehensive, extensive healthcare survey, and it was published way back in 1998. But what they determined -- and this has become really a gold standard study as it links childhood experiences to adult outcomes. And what they determined was that over 50 percent of the adults evaluated had at least one of seven specific risk

factors that they surveyed, specifically child abuse and neglect, household members with substance abuse, mentally ill, suicidal, or imprisoned household member, or history of violence against the mother.

And looking at them, the adult health outcomes -- and there's a lot of detail that's provided in the full study, but just to share with you that one and a -- 1.4 to 12-fold increase in alcoholism, drug abuse, depression, ischemic heart disease, cancer, chronic lung disease. So there was a strong guided relationship between the exposures to a childhood adverse effect in multiple health risk factors as adults. So this was the first study that actually linked childhood experiences and adult outcomes and confirmed what Frederick Douglas said over a century earlier that it's easier to build strong children than to repair broken men.

That led to the American Academy Pediatrics publishing the technical report and policy statement in January of 2012, and these were published by Dr. Shonkoff, who many of you well know from his history as Neurons to Neighborhoods publications some time ago. But this is -- if you're not familiar with toxic stress components and the relationship to childhood outcomes, I would refer you to this, at the AAP.org, and you can go to Policy and look at toxic stress. And what they're relating, and review, in a very extensive way, is the biomarkers and

physiologic indentations and disruptions. This is a very busy slide, but look at the health and development and outcomes and the potential relationships. And what that, Dr. Shonkoff says is, “The time has come to expand the public’s understanding of brain development and shine a bright light on its relation to the early childhood routes of adult disease.” And I think that’s in part certainly what we’re here to talk about today.

So I want to take a few historical looks at environmental effects, and the first being malnutrition, and many of you are familiar with the Dutch Hunger Winter study that really is the first natural study that was almost perfectly designed in the sense that the German-imposed embargo on food to Western -- six Western Dutch cities between the Fall of 1944 through May 1945. And the healthcare and recording of what -- even the food rations given to people in that time, were very carefully recorded. And the study was abrupt at onset and abrupt in its conclusion.

So it showed that there were an increase in, not surprisingly, congenital neural tube defects, spina bifida, and anencephaly for infants that were conceived during the famine, and at least the two-fold greater cumulative risk of schizophrenia as adults in persons’ conceived at the time of this famine. And so, there were also two-fold increases in schizoid personality disorders, and the other types of emotional

disorders did not appear to be affected. So in terms of fetal under nutrition and health, there was also an increased incidence in central obesity in women, double the rates of coronary heart disease, increased risk of metabolic disease, five times the risk of breast cancer, and there was also some early now signs of possible accelerated cognitive aging in the subjects that were born at the time of that famine.

So this is our first evidence of epigenetic programming through prenatal famine. It also has shown that the -- there were -- have now been recently studies in specific methylation of effected genes. So the epigenetics coming out of this are also extremely interesting. Another malnutrition and development study is the Barbados study, and there were 129 children hospitalized within the health system of Barbados in the first year of life; they had [foreign language]. And they were evaluated at ages 5 and 11 and compared with another 129 well-nourished comparison children. And there was strong correlation between infantile malnutrition and acute deficits, specifically about acute 12 points lower. And the contribution of malnutrition appeared to be independent of socioeconomic factors. Well, the same authors, they then looked at these 129 children later and were able to follow-up with 85 percent, or 109 of them and evaluated them between the ages of 9 and 15.

And for this program study, they looked at -- they

talked to the teachers in the school system. And by the way, I should say Barbados has a 97 percent literacy rate, so FYI. The several healthcare-related findings were their delayed sexual maturation, only in girls, and reduced physical growth in both boys and girls that had experienced that first year of life malnutrition compared to the normal controls. But their teachers also reported -- and they were blind to the nutritional status in the first year of life -- behavioral deficits, including attention deficits, reduced social skills, and emotional instability that were independent of the IQ. And there was slight correlation with socioeconomic factors and conditions in the home because those were also looked at, but it was not a strong correlation.

Another known, well-recognized environmental contaminant is lead. And the children's -- comparing children's intellectual function with lead exposure has been increasingly of concern, and this comes from an international pooled analysis looking at a longitudinal cohort of birth to 5 to 10 years, and their outcome measures were full-scale IQ. The covariates assessed were home inventory, evaluation, sex, birth weight, birth order, maternal education, maternal IQ, and the lead-associated incremental decrements were greater when the maximal lead was less than 7.5, compared to maximal lead levels of over 7.5. And I'll give you the specific ones in a summary slide at

the end of the presentation. So there was -- what's important here, there was no evidence of a threshold for adverse consequences for lead. So it's a continuum.

Well water and arsenic and child IQ has also been a hot topic lately. And well water, and maybe I should just share with you that that's really how I got to this meeting in the first place was because there's been so much in the consideration about arsenic and rice, and that's been in the discussion of the Consumers Union. The FDA has it under consideration at this present time, and some of the data that we have comes from well water. But why I got here is that I think it's -- Consumers Union has recommended no more than one serving of rice a week is fine, and I raised my hand and said yes, but what about these very small, premature infants that Dr. Allen has talked about this morning? We're giving eight servings a day of rice to these babies as a thickener for children with dysphasia. So that made me very wondering: do we know? What do we know? What should we be recommending to our parents who come to us asking, "I just heard rice isn't so great. What are we supposed to use?"

And if you want to know where that stands right now, it's kind of a long story, but it's still under evaluation. We really don't know what that means to babies. But the issues related to IQ come from well water arsenic, and in this

particular study, this was done is Maine. There have been other studies done in Mexico and in Bangladesh, which add other confounding factors for sure, but the 270 children studied in the state of Maine, Augusta area specifically, where the well water arsenic is known to be above the current EPA guideline of 10 micrograms per liter. In fact, it goes up to at least 115 in some of those settings. And they looked at families who'd resided in these homes an average of seven years.

Water samples were taken at the point of a kitchen faucet. And the WISC four was administered. They adjusted for home scores, maternal education, maternal IQ, and the number of children in the home, and they found that water arsenic was significantly, negatively associated with full scale IQ and working memory, perceptual reading, verbal comprehension, with a range of IQ point reductions between four and a half and six and a half. So some of the problems with this study were they didn't -- they didn't quantify for or control for the amount of water that children drink, and there are some other issues as well. But this is what we have currently in our literature.

Another major environmental exposure is alcohol. It is probably -- it is believed to be the single most common cause of intellectual disability in this country and probably elsewhere in the world as well, and it results in abnormal cognitive functioning, difficulties with abstraction,

mathematical problems, poor attention and concentration, memory deficits, as well as behavior problems. And the effects of relative alcohol exposure have been attempted to be valuated, and at least one drink a day throughout pregnancy is associated with an increased risk of IUGR in some studies. Maternal age, poverty, health, fetal susceptibility probably play a role, and there're probably genetics roles because we know how that alcohol effects is extraordinarily frequent in Native American populations, for example.

So we do know that the potential harm to the fetus is greater with more alcohol and that there's no safe, protective level to the fetus. Why? Because alcohol freely crosses the placenta, and it's a known teratogen. It causes chronic, nonreversible sequelae with multiple birth defects; no dose response relationship has been established, but binge drinking clearly puts you at a much higher risk. There are also, as I mentioned before, other risk factors like older maternal age, ethnicity -- African-American, Native-Americans -- genetic factors, and some polymorphisms about alcohol dehydrogenase are protective.

So the most common cause, as I said, of intellectual disability with a prevalence at .5 to seven cases of thousands live births. Well, I took this from the FDA website. I think it's been there a while because it uses the terms "mental

retardation" instead of "intellectual disability," which we [laughs] don't do, but I show it because it reflects everything that we talked about this morning. You can see the embryonic period for potential teratogenicity, but note that the brain neural tube defects, intellectual disability, and central nervous system goes right up all the way through this entire period of gestation.

So we also know that it effects prenatal growth retardation at or below the 10th percentile for age. This is the definition of fetal alcohol syndrome. There are also fetal alcohol related disorders and defects, and there's a whole category of things, but to be diagnosed with the full syndrome, you need prenatal growth retardation, central nervous system, abnormalities, and specific dismorphologic features that are detected on examination. And if you look at these children, they have specific similarities across a whole spectrum of ethnicity. So these are -- the things to point out include the -- what we call an undifferentiated philtrum, sort of a flatness to the period -- to the position between the nose and upper lip, a thin upper lip. And when we measure eyes, there's a shortness of the palpebral fissures. So there are known standards, and we use those as part of the definition.

There are also behavioral features of fetal alcohol syndrome, and I mentioned them before in terms of irritability,

regulatory problems with sleep, hyperactivity as children become more mature; and in adults, inappropriate sexual behavior, legal problems, substance abuse, these problems have had their initiation in fetal life but extend through adulthood. Another well-recognized, now, teratogen or adverse effect is the organophosphate pesticides, and this is a study that looked at IQ in 7-year-old children. It was a longitudinal birth cohort study of mothers of immigrant Hispanic background, living in an agricultural -- either living in or working on agricultural environments in Salinas County, California. They had evaluated their cord blood or maternal blood prenatal metabolites, and these correlated with lower cognitive scores at age 7. There were no apparent confounding effects of other neurotoxicants because we studied for those, and the associations were linear, and there was no threshold present.

This -- the psychological assessments were done by native Spanish speakers who were specifically trained for this evaluation. And what's also of interest, this was only a prenatal effect. Postnatally, they watched the mother's urinary diarchal phosphatase levels related to the organic phosphate pesticides, and they were not consistently associated with cognitive scores at age 7. So a similar study was done in the -- looking at air samples in the city of New York, where the organophosphate chlorpyrifos was used as a residue for

infanticide until it was banned in the year 2000. So these were children enrolled in 1997.

They had 265 children, pregnant mothers. There was no maternal diabetes, hyper tension, HIV, tobacco, or documented drug use. The tobacco was self-reported and documented, confirmed by maternal cotinine levels. Maternal blood and cord blood levels measured the prenatal exposure to this insecticide, and they evaluated the children at age 7 and found for every standard deviation increase in the chlorpyrifos levels, a 1.4 percent reduction in full-scale IQ, and working memory reduction of 2.8 percent. So there was no apparent, significant interactions with covariates, including maternal education, maternal IQ, quality of home environment, and other chemical exposures, specifically including tobacco. Well, we've learned this this morning, but I think a picture's worth 1,000 words. This is what happens to the brain between 30 weeks and 40 weeks; just that 10-week period has a huge change, and we know that it -- brain myelination continues post-conceptually at a very rapid rate, and it continues well into young adulthood.

So the additional studies have looked at that and structural maturation of brain regions, which are absolutely essential for successful development of our cognitive, motor, and sensory functions, and that MRI studies looking at white matter density in fiber tracts increases their age related. And

there's also evidence of gradual maturation in late adolescence and a childhood in adolescence of these fiber pathways.

So let me raise, then, a really hot topic, which with increasing legalization of marijuana, may or may not increase the use in children, so -- and I don't have strong evidence for this, but I think it's worth us thinking about because it's going to become extremely important for all of our adolescents. Knowing that those brains are still developing, there are -- have been increasing studies looking at the long-term effects of marijuana on the brain, showing that there's less bilateral orbitofrontal gyri volume, changes in functional connectivity, and this appears to be associated with earlier age of onset abuse and probably differential effects of initial and chronic marijuana use. So looking at -- this is a study that looked at altered brain morphology in recreational users.

Now what your definition of recreation is [laughs] -- their definition was use four days a week of, at least, about 11 joints. And comparison, they're non-using comparison population was 20 young adults with less than five lifetime exposures. Now you could argue whether five lifetime exposures is ideal either, but that was what they were able to use. There was tendency for greater alcohol use in the marijuana users, but they showed increase the whole brain gray matter, density of the left nucleus accumbens and the left amygdala. Specific areas related

to pleasure areas for brain. So another marijuana outcome was a study in South London where they looked where there is a high availability of a high potency cannabis substance.

And the proportion of their patients with first episode psychosis attributed to the high potency cannabis case contextual study. And they showed there was three times greater risk of a psychotic disorder in those using high potency cannabis versus no cannabis. So the thought is that the high potency cannabis was more available, and therefore, it increased the psychosis, but the use of cannabis in population at risk of psychosis -- are they self-medicating is another considerable variable. And that's not no. There are also studies looking at related episodic memory deficits and hippocampal morphology differences in adult and individuals in schizophrenia subjects and shape differences.

And it's considered that the observed differences could be biomarkers of neurobiological susceptibility to certainly the schizophrenia, but also to poor memory. So we know that social and behavior problems, lower academic achievement, working memory, visual scanning, cognitive flexibility and learning appears to be related to the number of episodes of lifetime marijuana and correlates with overall lower functioning. And the American Academy of Pediatrics in March of 2015 published their policy statement relevant to the

legalization of marijuana, and we anticipate that will be followed up very soon with a policy statement for healthcare guidance.

So I told you I would share with you the relative magnitudes of associations from the literature. And arsenic and water, for exposures equal to or greater than five micrograms per liter, resulted in the decrement of IQ of 4.5 to 6.5. Blood lead, in that multinational-international study, was if you looked at maximum blood lead ranging from 2.4, which is the lowest level pretty much that we can quantify, to 30 was 6.9 points with 3.9 of that being attributable to the 2.4 to 10. And 1.9 to 10 to 20, and then 20 to 30, 1.1. So this has resulted in the CDC saying that the top level we should be concerned about is five, and that's sort of at the very -- at a low end; organophosphate pesticides, prenatal exposure, seven.

Just quickly, there are many other environmental exposure that one could talk about. Specifically, folate supplementation can reduce the likelihood of neural tube defects for about 85 percent. That would be in China. In this country, it's resulted in about a 40 percent reduction in neural tube defects. There are many other factors, not just folate, that play a role. Essential fatty acids have been discussed in brain function, and I know many of you probably know a lot more about this than I, but its early in its evaluation stages, but

increased mental processing composite in the Helen study looked at 4.1 IQ points by age 4.

Vitamin A is a nutrient essential to early embryogenesis, but it's also a teratogen to the fetus when it's in excessive doses. Iron, anemia compromises oxygen deliberate to the developing fetus, especially the hippocampus and can alter myelination, so these are other areas of great interest and research. This is just my simplistic -- over simplistic, I'm sure -- representation of attribution, relevant to age, looking at sensitivity because one of the things we're asking is, how long do we have to study? And so, if you have certainty for the confounders, and since they go down with time, but they -- sensitivity increases, and at some point, they intersect and at the optimal place for evaluation.

So what are the considerations? Well, there're covariant, and we've heard them all this morning. Maternal IQ and education, gestation, birth weight, birth order, gender, socioeconomic status, and I want to just stop there, just for a second, because that changes. And the evaluation parameters that we use for that often were, "You're on Medicaid." Well, since the Affordable Care Act, we have lawyers in early practice, single parent, who's on Medicaid through the Affordable Care Act because her environment, her income meets the new standards, and yet she's providing us a home environment

for the child. So just have to be aware that the criteria that we've always used in the past might not necessarily be the case.

Cultural validation, language is an important issue, but there are also cultural things, and we talked yesterday a lot about English in UK versus English in the United States and how that can be affected. But I just want to give you an example of, back in the day a few years ago, we were doing Stanford–Binet's on our very low birth weight, preemie outcomes, and I had a little girl. She was five. We were doing her Stanford–Binet, and the psychologist came to me afterwards after she administered the test. And she said to me, "You won't believe this, but," she said, "One of the questions was, 'What are the four seasons of the year?' And the little girl just brightened up. 'I know. There's pheasant season; there's fishing season; there's deer season; and there's bow and arrow season.'" Now you want to give her credit for that, but she wouldn't get the credit. But in rural Indiana, that was her experience.

And there's strength of evidence about which there's no universally accepted standard. Many people, that's a highly arguable topic. So my conclusions are that cross-sectional evidence really can't clarify our questions of causality. Longitudinal or interventional studies are needed. Brain myelination is such an important point; we've heard it all

today. It occurs on a continuum, prenatal through at least early adulthood, that the stability of the intelligent quotient -- and you all know this better than I -- but usually not until they're just 5 or 6. And that longitudinal study is useful to examine these changes over time and relationship to biologic, environmental factors are important. So I'm going to stop there, and with our goal being happy, healthy children.

[applause]

DR. YAO: Okay, our last formal talk of the day is from Dr. Shapiro, and we're hoping with this talk to set the stage for a conversation that will take place for the rest of the afternoon.

WHAT ARE THE MAJOR CHALLENGES IN THE EVALUATION OF LONG-TERM NEUROCOGNITIVE AND BEHAVIORAL OUTCOME?

DR. SHAPIRO: Thank you very much for inviting me, and I am going to be, I think, repeating a lot of what people have said this morning and this afternoon, but I just want to put it all into a sort of a context of, what are the major challenges and evaluation of -- and I'm going to add something here -- both short-term and long-term neurocognitive and behavioral outcomes? I'm going to spend most of my time on neurocognitive and just a little bit on behavioral outcomes because that's a very complicated area. Let's see. This is a different one than -- oh, here it is. Okay. I got it. This side, right? Oh, okay. Okay. Thanks. Thanks. Okay. So there are my financial interests. And I'm going to talk a little bit now about what I'm going to be talking about. So I'm going to sort of outline the challenges and talk about deciding the concept of interest to measure studies of drug effects. I'm going to talk a little bit about how to evaluate outcomes, age considerations, timing and repeated measures, quality control, and disease-specific considerations. And then I will review a little bit about the tools for measuring outcomes and the role of computerized assessments.

So I just want to sort of challenge everybody to think

about how we can best measure these outcomes in the most efficient and precise way. So the challenges in clinical trials with children are, one, finding appropriate and sensitive cognitive outcomes. And I want to emphasize the “appropriate” because sometimes our standard measures are not what we want to use. The challenges of lack of appropriate measures for measuring certain kinds of concepts and certain kinds of functions: lack of normative data for certain kinds of testing, and in multi -- in some centers, there just are small numbers of any appropriate children to measure these things, and there’s -- either will necessitate a long period of recruitment in trials or multicenter studies. Studies also may require very long-term follow up, even many, many years; for example, in children with cancer and children with neurogenetic diseases to detect treatment effects. And I want to also add to the mix the lack of trained pediatric psychologists to work in clinical trials. That’s something that I think we need to pay more attention to.

So let’s start about discussing the concepts of interest to measure. And I’m going to use a psychological term, the set of terms, state versus trades. Now usually, that kind of distinction is a descriptor for personality measures, talking about how the person is at the moment versus their long-term traits. But I also want to think of it as a descriptor for cognitive abilities because there are abilities that are more

easily altered by the situation that are temporary, caused by external factors such as fatigue, illness, or medication. These abilities are more likely to be in the domain of processing efficiency. So examples are things like attention span, memory encoding, reaction time, motor speed. They're much more easily altered by the child's state at the moment, rather than their long-term traits. And traits are abilities that are stable, long lasting, and that represents static aspects of brain function; their predictable and their enduring basic abilities. And examples of that are visual-spatial ability, vocabulary, problem solving, and of course, IQ as a general measurement.

So that's one of the things that we have to decide. When we're looking at a -- the effects of a medication, are we looking at the things that are going to be effected right away, or the things that are going to be effected in the long run? And then of course, there are short-term effects. Will the treatment have an immediate effect? How enduring will that effect be? And that will require the state-type measures such as motor speed or attention span. Long-term effects: will the treatment have an effect on the child's future performance with repeated treatment exposure? Will it have a late effect on enduring traits? You know, we can't tell whether something's going to have an effect on reading ability unless the child is old enough to be able to read, whether the brain has matured to

this state -- to the point where there might be an effect on reading. And those effects might be silent initially.

So in young children, we need to really think about, should we be looking at the rate of development of cognition? And does a treatment slow the growth of cognitive abilities, and which ones does it slow? In young children, do we use growth parameters, such as raw scores or mental ages, rather than standardized scores, such as the mental development index on the Bayley, or any other score that might be determined by normative data? In older children, how do we decide which neurocognitive abilities to measure? By previous research? By hypothesized mechanism? Something that we know about, maybe from animal studies? Do we use adult data to choose what we should be measuring in children? I think that gets us into a lot of hot water. Or if it's a pediatric disease only, do we have natural history data that will allow us to choose the right concept of interest to measure?

So what is the criteria for good measures? Does it measure what you want it to measure? Does it have good reliability and validity? Are there good normative data? Is it appropriate for age and level of functioning of the sample? What are the timing considerations? Are there practice effects? I brought that up earlier. Are there alternate forms so that you can repeat it every month, for example, if you have a very

active medication that you're looking at? Most children's tests don't have alternate forms. Can it be used across the whole time period of the -- of the trial? Will the child outgrow the test and have to be moved to another test? Is the battery short enough to avoid fatigue effects? And most of the time, our batteries are much too long, and they are much too elaborate for a child to complete at one setting. So we need really short, focused batteries. So, let's think about the challenges and solutions for young children under 6.

Young children require skill testers with familiarity with the test and ability to manage behavior to get valid testing. It's extraordinarily important. And you have to validate those results because children are so effected by the environment, by their fatigue state, by all kinds of things, and often one of the things that we have found is that we do parent report measures to validate a child performance measure, so that might be one thing or the rating of the motivation by the tester.

Another thing in young children under 6 is there are very, very few good measures of attention and memory under age 4. Mostly, we're limited to tests like the Bayley, which measure cognitive and motor development only, and I've had a little experience myself in developing tests for children under the age of four, both for memory and attention. It's very

difficult and very time consuming, and it’s a very expensive and lengthy process to norm and validate a new test. So we need to -- this is a tremendous need in the field right now, to develop good, normative data tests for both attention and memory. And again, I’m going to emphasize the testing, especially in young children under 6, needs to be short to avoid fatigue, and I think that it’s also extraordinarily important to remember that when you're doing something like this, it is not a clinical evaluation; is a -- we’re looking only to look at the effects of a treatment.

So, I think that we have this morning a very good graph of how tests transition from one age to another. I think you presented that, and I think that’s very important. Moving from a Bayley to a WPPSI to a WISC is a very, very difficult thing, and we don't have good tests. And I think that the NIH Toolkit may offer us that opportunity to be able to take some of these measures and move right through the various age ranges. So, one of the things that we always do in our trials with children with neurodegenerative diseases is select tests with the largest age ranges. Motor tests are a great example. The Purdue Pegboard has a very large age range, and we use that when we’re looking at motor function because it has the largest age range.

And there are other tests that have more narrow age

ranges, and we -- even if the normative data is less good, we would choose the one that has the large age range. And you need to select tests which measure the same thing for all age ranges. For example, the Wechsler tests all have verbal and performance components; they all have vocabulary block design, digit span similarities, with age norms, so those are -- those are good measures in that respect. But you go to, for example, the Stanford-Binet or the DAS, and you find a test that has different tests at different age groups. You may come up with an IQ score or some score, but the tests differ from one age level to another. So you need tests that are consistent across all ages.

How often should children be tested? Well, that differs by age and by test, so younger children can be tested more often because the rapidity of their cognitive growth means that tests items can be different at different -- are different at each developmental level. So you have a child coming in at three months, or let’s say a year, and then you’d re-test them three months later, and the whole item set is going to be different than the item set was before because they're doing so much more three months later. So there isn’t, you know, a practice effect, let’s say, in young children because the items are different. IQs test, after age 6, we don’t give them very often. We don't want to give them more often than yearly. So,

why is that? Because problem solving tasks have remembered solutions.

Once you give a block design test, the kids can remember how they did it the first time. So if you do it too closely in time, you're going to have some practice effects. Now, verbal learning or encoding need alternate forms always because kids can remember, even a year later, what the items were on a verbal learning test like the California Verbal Learning Test, which has only one form. So there are many adult tests that have alternate forms, for example, the Hopkins Verbal Learning Test. We don't have too many of those in childhood. There are tests that don't have practice effects. Attention tests, working memory tests, and motor speed tests generally don't have much in the way of practice effects. So how often should children be tested? Well, it really depends on what you're trying to measure and what the effects that might be seen in that particular medication or that particular treatment are. We need more tests specifically designed for clinical trials with alternate forms.

Now, I want to mention one other thing, and I mentioned that earlier today, and that has to do with the novelty effect, and a good practice is to give tests once before a base line in order to see whether there’s a novelty effect. The initial testing may be much better than the second testing

because -- or it might be the other way around. There might be a novelty effect, or there might be familiarity with the materials that alter the performance on the test. We -- I think I mentioned this yesterday that we were giving some of the measures of memory on attention to children in Uganda, and we were watching -- they were children with Cerebral Malaria.

And we were watching them on video tape and comparing to the children that we had in Minnesota, normal group of children, and they were much more attentive, but they had never seen this kind of computerized test before. They had never seen the materials before. It was absolutely fascinating to them. So you know, their eyes were fixated right where we were measuring their attention by their eye fixation, and there was absolutely no inattention in those children. And so we realized that we needed to get them acclimated to the test before we could actually see whether their attention span was being measured or not.

In multicenter trials, it is very important to minimize noise by having really good quality control, so we need testers that are blinded to the medication condition. I think that’s extremely important. Testers must have training on the test. We need scoring reliability. We always re-score all tests independently. Random observation of the tester or videotaping is a good way to get standardization, and perhaps in

the discussion later, Kate Delaney, who has done a lot of this in different centers, can relay some of her experience in reducing the noise in these multicenter trials. And in any single center, minimizing the number of testers is always best.

There are disease-specific and age-specific considerations. Some treatments may require follow up studies that track children for years if there are developmental effects. For example, late effects of chemotherapy. And so then you have the problem of the challenge of test comparability over time if you're looking at a longitudinal kind of outcome. Some treatments are for diseases that may have age-specific effects. And then an advantage of disease-specific clinical trials, cooperatives, or networks is their greater ability to collect age and disease specific normative data across multiple trials.

So, let me run through some of the specific measurement tools. So everybody uses IQ tests. They're used a lot; they're well normed, but they're probably insensitive in the short run unless subtests are examined. Measures are available across the life span, and I've mentioned this before. One of the things that we have found is that abbreviated forms of these tests, for example the Wechsler Abbreviated Scale of Intelligence, is a good choice for a clinical trial. The challenges with IQ tests is that they're lengthy. They average

about two hours to administer a whole thing. They're not continuous across age periods. And the problem is that the test manufacturers, the test developers, are constantly updating them.

And one of the things we know in any clinical trial is do not change the test midstream in a trial, from one version of a test to another. We've seen that happen, and we get completely different results when we're moving from a Wechsler three to a Wechsler four, for example, and the WISC. We had that and a trial in adrenoleukodystrophy, which is an inborn error of metabolism, where the performance IQ in the WISC-III was a great predictor of outcome, and then you move to the WISC-IV, and your performance IQ was completely different, made up of different tests, and we couldn't use it anymore. It no longer predicted the long-term outcomes. So you need to stick with the test that you have used at the beginning. IQ tests are best in measuring late effects.

Memory testing; really important; working memory, digits forward, digits backwards, easy to use in any child over the age of 3. You can't do digits backwards, but you can do digit span, and it's an easy way of assessing working memory. Encoding is a different question. So transforming information to be remembered into a form that can be stored is what we mean by encoding, and mainly it's list learning or paired-associate

learning, where you're remembering, say, a picture with a word. There's lots and lots of tests. So the only test that actually does have some alternate forms is the Rey Verbal Learning Test in young children. The California Verbal Learning Test, the WRAML, and so forth, they don't -- there are no alternative forms.

Now, in adults, the Hopkins Verbal Learning Test has alternate forms. It's not officially normed under 16, but we used it, and we recently collected norms on that HVLT for children from 8 to 16, and we'll be publishing that soon, but we really -- we don't have anything for under 8, and so that's really important. We use a paired-associate learning test from the KABC, the Atlantis test, which is -- or the paired-associate learning from the KABC, and the Atlantis subtest on the NEPSY, but again, no alternate forms. So you do need alternate forms for encoding tests, and maybe then NIH Toolbox is going to be able to provide that for us. We have not explored that, and I'm not familiar enough with it to really say that this might be a good substitute or a good way of looking at memory as a medication effect. It's very difficult and expensive to collect normative data, so it's not something that you do casually.

Executive functions are really important, and attention, there are lots of computerized continuous performance test: the Connors, the TOVA, the IVA, the Quotient. They're

sensitive; they give a lot of information, including vigilance, impulsivity, response time. Some of them have not been used for non-ADHD research and generally not very good for children under 6, so we really need tests for children under 6. Inhibition, there are Stroop-type tests and CPT's that can be used. There are Tower tests in the planning and problem solving. There's a whole list here of different kinds of measures of executive function, and I think that the -- and again, NIH Toolbox offers some good measures of executive function. They're good, easy-to-administer measures that are sensitive probably to short-term effects.

There are parent behavioral reports, such as The BRIEF. We haven't found them to be so useful in clinical trials. We like it for general kinds of clinical usage, but it is something to consider. I mentioned motor speed inaccuracy and peg boards, and there's some evidence for practice effects. It needs that pre-baseline testing that you need to do in any clinical trial. Language and visual-spatial ability generally are covered in IQ tests. Language measures are so dependent on socioeconomic status and culture that we tend to avoid them in all clinical trials that we're doing, and using as much visual material as you can, unless you have a visually impaired population. Visual-spatial ability using measures like block design from IQ tests is pretty good, but these tests, both

language and visual-spatial ability may be sensitive to late effects only and not immediate effects.

Let’s get to behavior. So we have in the psychologists armamentarium, their Toolbox, there’s the BASC and the CBCL. That’s sort of what is used. They all have parent, teacher, and self-report measures. Then there’s the BRIEF inventory, the Behavior Rating Inventory of Executive Functions. They’re marginally sensitive to short-term changes due to medication or illness. The BRIEF may be more sensitive, but it’s limited in its scope. What we really need are the quick behavioral measures related to the short-term changes that might be behavioral changes that might be related to treatment or medication effects. Finally, can you use computerized assessments? They have great advantages. They're fast and efficient; they’re standardized, and they're reliable in their reliable procedure.

You don't have as much as an examiner effect. There’s automatic scoring; it minimizes training, and touch screens are very easy to learn. But they're not appropriate for younger children. You need the examiner interaction to motivate the child. You can’t judge task engagement; you can’t judge motivation. There are errors, also, due to variable computer configurations and operating systems. I know that we use the Cantab a lot, and sometimes it works, and sometimes it doesn't

work, and so that’s another problem. And then there are also practice effects relating to familiarity with the computer response. So basically, we need more research that needs to be done regarding reliability and validity in children.

So, my final challenge is the lack of pediatric neuropsychologists or developmental psychologists trained to work in clinical trials or even in the field of longitudinal research. They just don't learn about clinical trials in graduate school. We need training in clinical trials and outcomes in medical settings at the postdoctoral level. I would say that this has been one of the things that I have found to be one of the biggest challenges in doing multicenter studies is to find neuropsychologists in various centers who really know what a clinical trial is all about, who need to -- they don't understand the need for rigger and for short, precise, focused evaluations, and this is a great need in our training programs.

So, in summary, this is -- these are my sort of final recommendations here. That task batteries need to be short, focused. Short and long term outcomes need to be treated differently. Test selection needs to be age specific and possibly drug and disease specific. And keep in mind that the effects of drug and disease can be confounded. We need to minimize transitions from one test to another due to age. We need more tests that cover a wide age range. Normative research

on memory and other practice-sensitive measures is needed. Development of computerized assessments for older children are needed. Use of pre-baseline testing to decrease novelty effects should be utilized. More emphasis on quality control in multi-center studies in neurocognitive assessments and the need for more psychologists trained in clinical trials. Thank you.

[applause]

DISCUSSION

DR. YAO: So, I was hoping to maximize the amount of time we have for discussion to sort of postpone the break or take the break, if you will, ad hoc, or as needed, PRN, so if you really need a break, please feel free to get up. We're not going to shame you in any way. Just get up and take a break, and that goes for people on the panel and people in the audience. But I really do want to make sure we have a chance to discuss the important discussion points we've outlined, so I would propose that we would move forward with that if that's okay, and I'll turn it over to Elsa and Heather for the discussion.

DR. SHAPIRO: Okay. You. Go ahead. You.

DR. ADAMS: Oh, well I mean, I wanted to thank Elsa and Dr. McMahon, and Dr. Bull. Thank you all for your talks and your contributions. My question is provoked by your talk, Dr. Bull, but I'd like to hear everyone's perspective on it. I'm curious about how this can be done.

MALE SPEAKER: Oh, that's not good.

[talking simultaneously]

DR. ADAMS: Well, I'll just speak loudly. How about that? So, I'm curious about the concept of interactions between and among difference exposures. So, for example, I was thinking

about the study that was done, the Seychelles Child Development Study, and this is a study in which mothers and children who lived in the Seychelles who have high exposure to methylmercury because of the high seafood diet were studied for many years. And they looked at the prenatal exposure to omega-3 in utero and then followed the babies developing -- you know, longitudinally activated or born based on the mother's exposure. And what they actually found was that there really wasn't a direct, primary effect with methylmercury on neurodevelopment of these children, and in fact, the thinking is that the high seafood diet where they have the exposure to the polyunsaturated fatty acids, the PUFAs offset the potential negative effects of the methylmercury. And so I think about that, you know, you can have a combination of many different types of exposures, some good, some bad, some benign that either amicably or applicatively [sic] or in some other way add up to what you observed in your study.

DR. BULL: Well, I'm aware of that study as well, and I just ran out of studies that I could talk about, and I think you're absolutely right. I think we have to look at the multiple variables, lead being another kind of baseline event in the lives of the world, some places more than others. So knowing those things, they have to be looked at, anemia similarly.

DR. ADAMS: Yeah, I mean, I think not all exposures are bad, but --

DR. BULL: No, exactly [laughs]. Yes.

DR. ADAMS: -- different exposures are going to play differently in the sandbox with one another.

DR. BULL: [affirmative] No question.

DR. SHAPIRO: I would like to bring up another issue in the work that you're doing. And that is, I've done some work in lead overburden in young children, and one of the things that we found out was that when mothers found out that the children were lead burdened, okay, that one of the things that happened was that the higher SES, slightly higher SES, smarter mothers did something about it and confounded the results of the outcomes of the tests. It became a circular kind of a situation where high socioeconomic mothers, once they found out about the exposure of their children, were more likely to intervene at that time and then mitigate the effects of their children's cognitive performance.

And so, you know, this is -- it's sort of a dilemma because you want to inform the parents immediately; you have to. I mean, it's required. And then you have the problem of differential intervention by the parents according -- and it really, really worked out that the higher SES moms and the higher -- the more educated moms were more likely to change

their children's diet, to have their children start washing their hands, to make sure there wasn't lead in the -- in the -- in the child's environment than the lower SES groups, so then you had that confounding effect. So have you seen that in other studies, and is that something that people are aware of?

DR. BULL: I think people certainly are aware of that. And part of the problem with lead is that many of the mothers in the low-so have no choice. There's nothing they can do. Limited, extraordinarily limited in what they can do. The higher socioeconomic mothers will move. I live in the historic district where we have a lot of lead, and it's an upper socioeconomic area where there are children with upper socioeconomic families with leads of 19. And unbeknownst to their families, who then they just stopped, you know, got out. But the other study that looks at that to some extent or has a similar confounder is the arsenic well water study in Maine because the -- and that was one of the things they looked at was mothers, you know -- do they use an alternate water supply? Are they using filters? There were a lot of other things these mothers automatically were doing, thinking that it was better for their health. So those are confounders that have to be considered.

DR. SHAPIRO: Right. Right. But it needs to be part of your trial that you --

DR. BULL: Absolutely. That's correct.

DR. SHAPIRO: -- have to build that in to a trial. Right.

DR. YAO: I'd like to sort of go back a little bit. I think I heard it, but I heard it in ways that I'm not sure I -- I think I heard, but I just want to clarify. And I would appreciate if folks, maybe we might go around the room and -- or at least the folks who have been doing the studies in children. If we had a product, if we had a drug that was going to be used to treat a child in the NICU, and we were concerned from an animal study or from other information that there might be an effect on this patient's memory or executive function, some measure of cognition in the future, that there was something that we were worried about, when would you test? How would you follow? What would you do?

I mean, I think that gets to the heart of it. Do we have enough, even right now, that can -- that we could even expect, and if that were the case, because I've heard this too, if we were to say, "Well, you know, these tests are really good," how would we operationalize it? What are the problems with operationalizing that? In, again, in the -- with the constraints that we've heard about, and Kate, I'm especially interested in hearing some of your thoughts on that. So I would be happy, you know, to sort of start with Gahan and go around.

DR. PANDINA: Sure. So I think one of the things that we always face in long-term clinic trials is discontinuation, and it also depends distinctly on when we start the trial and for how long that is. I think when we're talking about neonatal treatments though, that's a completely different thing, right? It's an upfront treatment, and then how long would you follow them for, and what would be the paradigm that you'd use? So, but you still have the problem of attrition, and how do you capture the most lives and follow up for the longest period or for the period, at least, in which you should be able to see a change in trajectory? And I think that, I mean in part, that depends on what you think you're going to see over time. Obviously, the parents that I see clinically are different than the people that I'm looking at in clinical trials, and my parents want to know, "If I take this now, when he's 25, is he going to be able to have children and, you know, will he go to college?" I mean, questions that we can't answer, right?

So I guess it's -- I'd be interested in hearing what the panel has to say, other people on the panel, but it's weighing against what we can assess over time, how many patients we can capture, and what's going to be the sample size that we'll need to make some inference or conclusion about that. No matter what the measures are that we have, that's usually the thing that we're facing, the challenge that we're facing, is how

many people will we be able to get back and endure and how much, not just investment of resources, time, and money, and effort, and energy, but just how willing are people to be followed over that period of time? And I think that just requires a different system all together when you're talking about neonatal exposure.

Just one more caveat, I was speaking yesterday to some folks in our baby center group, and one of the people there, who is a researcher, is a mother of three children, two of whom were in the NICU for several months because they were born very prematurely, and she was talking about her follow up. So she was going back at three-and-a-half years and still following along with parents who had had similar experiences at the time. So there may be opportunities through other kind of venues, through social media, through other places, but that would require a much more creative approach than we have now. It doesn't lend itself easily to that sort of clinical trial follow up.

DR. SHAPIRO: It needs a structured kind of approach like, you know, some of the cancer follow ups, you know, the long term outcomes in cancer, you know. These kids get seen on a regular basis, and you know, they get a lot of personal attention, and so there is a lot of ways of getting them back into the clinic. And so you need a structure for that and I, you know, I think that there are structures within some of the

NICU follow up centers, and I think that that needs to be done. I know that, for example, in the work that I do on inborn errors of metabolism, you know, we structure these kids to come back on a yearly basis. They come back once a year, and we have a, you know, three-hour visit with them, and that's a yearly thing. That's part of the NIH Rare Disease Clinical Research Network, so we have funds to do that. But there aren't always funds to have these structured, long-term follow ups of whatever it is that you're looking at, right?

DR. YAO: Natacha, I'm very curious about, as you're listening on the conversations, about how you think the imaging might help, or I mean, is that going to be any easier? Is there any correlate that we have established enough that -- I know that the same concerns that we have about ease of administration of the test, which hamper the evaluation in the youngest patients, is of course the same problem in the data that you've collected because you've got to sedate, or you've got to give a small infant anesthesia to get an MRI, but you know, I'm trying to figure out where -- when you say three hours of tests, you know, once a year, I'm already hearing some of the things that I've heard from yesterday and today, you know. Fatigue, they flew in; it gets very hard; the more that you have to do, the less people will come back. How can we build in and what kinds of things could we do to make that period maybe more flexible

and shorter, as Elsa was saying, so that we could get these in?

DR. AKSHOOMOFF: I just have a couple of comments. One is that for children that are high risk infants and who've been in the NICU, particularly the preemies, there is great effort to try to get them into follow-up clinics across the country, so that like in California, that's part of the healthcare system. Even there, it's stunning to me that it's hard to get families to come in. So either you've got families that are really, you know -- they're usually more educated, or they're really scared, and they're very good about coming back to those visits, and then other families that never return, or they come in only when they start to get worried as their child is maturing in that short timeframe.

So one thing that you said that struck me was if we had evidence from an animal model about something like learning or memory, if there was some reason to believe that we might be able to capture some element of that when the children are younger, even if it may not be entirely analogous to sort of a higher level learning or memory problem that would affect children, you know, that we could see when kids are a little bit older, it would probably be a fairly quick little assessment that you might be able to piggyback onto those sorts of follow up visits through a clinic, you know, often times with a psychologist or a nurse practitioner, or something of that sort.

In terms of the imaging issues that's a big question now, too, because studies have been done particularly with high risk infants, you know, while -- before they leave the NICU, they -- you can have them sleep and take an MRI scan and that sort of thing. But you're inevitably probably only going to capture the cases where there's a pretty -- a bit more of a prominent kind of, you know, sign, high white matter, you know, intensities, or things like that. And so it's a good question. You know, other avenues that people haven't talked about are EEG and other kinds of brain imaging elements that are usually a whole lot easier to do with kids and can now be done pretty --

DR. SHAPIRO: ERP's, right?

DR. AKSHOOMOFF: Yeah. And it can be done a little bit -- much more cheaply now and easily, if we develop those techniques.

DR. ADAMS: Dr. Baron?

DR. BARON: Yeah. Well, I want to make one quick comment before I make a longer statement [laughs], and that is I'm not the only person up here who hasn't given a talk, so I'm able to thank everybody who did [laughs] because it was really an excellent day, and so I'm very happy to have been able to listen to all of you. You know, I'm doing research on NICU patients, and the focus when I developed the battery was executive function and memory attention. We begin at age 1, and

we begin with -- we begin at age 1, and we are doing comprehensive testing at ages 3, 6, and 9 so far. We're bringing the 9 year olds back this year.

There was a real focus on 1 and 2 years as a way to bring people in because what we do is we are giving the parents questionnaires that we find very helpful. Not that -- there are ones that haven't been mentioned yet today. And that lets the parents know that we're not the follow along clinic, but we are a group that wants to follow them. And then when they come in at 3, 6, and 9, we have a very comprehensive battery that assesses every domain very quickly.

And the point I wanted to make is you've heard about a lot of standard tests today and a lot of good tests, but actually in my battery, I don't think I have any of those tests. [laughs] Okay. And my population's different than Elsa's so -- and may be different than yours. So there are many other tests that are out there that have been used for a long time that have very good -- well, I wouldn't say they have very good normative data, but they are normative data. And if you have a large control group coming along at the same time like we did, we felt we could use the standardized test and bring in some experimental tests, which have worked beautifully, but which you all don't know about because they are experimental. But we have our data because we have 700 normative controls.

And so what I wanted to bring up is that in deciding what test to use and how to follow them, you know, I did a lot of research before I created the battery in terms of, what did my population look like in the literature? And what I found was executive function, attention, and memory were going to be the high hits, so I included more of those measures, but then balanced it out by adding the ones that may not be as relevant. They've turned out to be relevant, but it covers every domain. The tests themselves don't take more than five, 10 minutes each.

So you can pick and choose and create a battery that's highly focused for your endpoint without having to go to these batteries that take two hours, and that's my personal perspective. I've said this for years. I've written about it. I am not a believer in giving a two-hour test when I know only one or two subtests are going to matter to me. So I think that there's a far wider range of measurement instruments than you've heard about and that if you're consulting with a neuropsychologist or, you know, someone who's really familiar with the armamentarium that's out there, you can shape your battery much more effectively.

DR. DELANEY: Dr. Baron, I just want to add something. I think you brought up a good point about the parent questionnaires. I've been involved with some study design around NICU follow up and clinical trials, and that's one of the

questions about attrition, and how will we bring the families back? Will they be -- am I mic'd? Yes. Will they be motivated? How will we keep them in the study? And that was one of the things we talked about, and I suggested and actually doing -- and I know there's a debate around phone interviewing and phone questionnaires, but to keep the parent engaged. So if you're seeing them at one year, two years, and five years, for example, keep them engaged, keep in touch with them so that -- and the chances are you'll have them coming back for that. And also for these trials, there's always a question about, do we provide the information to the parent and the results? And I think with these kids you have to.

DR. BARON: I was just going to say, I forgot to mention we give every parent a report, and it isn't of all the data, of course, but it's of the pieces that they can take to the school, and our feedback is that that has made a difference in their child receiving services. So they come back because they know they needed it to get the services, and to continue it, they want to come back and go through it again.

DR. SHAPIRO: I agree with that 100 percent because we find that one of the reasons that we have patients staying in our longitudinal studies, our natural history studies, is that they get a report at the end of our evaluation --

DR. ADAMS: It needs to be a two way street.

Absolutely.

DR. SHAPIRO: -- and I think that's quite important that parents feel that they get something out of it. And if it's just information, that's good enough.

DR. MCMAHON: Yeah, I just wanted to throw out there this idea that I think was, if I got it correctly, that people were alluding to this morning, that it seems like if you have hundreds and hundreds of tests, many of which a lot of people agree are very good, do you get as much data to draw on as you might be able to get from a few tests that everyone agrees are very good and have lots of data? And I just wanted to sort of see what you thought about all that.

DR. BARON: Well, I use -- there's another battery Elsa did mention called the Differential Ability Scale, and the reason I like it so much, aside from the fact that the child no longer has to fail multiple times before you move on to the next subtest, they can be succeeding at the end, and you still move on because it's based on a different system of how, you know, the scores are calculated. But within that battery, there are those core subtests that we are concerned about that have, you know, semantic memory and are contributing to the IQ. But there's another whole series of tests that come in that battery that are neuropsychologically meaningful, and once you start to use it -- I think most people I know who use it, use it for very

young children because it begins at 2 years and 6.

But it goes to 17 years, 11 months, and when I started using it clinically with the older children and adolescents, I said it’s just as good for them because it has some of those -- it has the Stroop test; it has short term memory, auditory, and visual. And so compressed in that battery -- and you can pick and choose whatever diagnostic test you want for your own particular research interests, but because it’s all in that same test, as well standardized as a Wechsler.

With really colorful, beautiful materials for the child to get engaged with in a much better way than a Wechsler, you can, within 60 to 90 minutes, run through all kinds of important, serious domains of neuropsychological function and also have an IQ that correlate .84 with the Wechsler. So that is a core piece for us, and we count on that, but then we add to it because it doesn't do everything, and that’s where all the other five and 10 minute tests come in. And so if a child or a project is developed that can’t take that timeframe, you at least have a core battery that’s neuropsychologically and psychometrically sound, and that’s how -- that’s how we’ve dealt with that.

DR. ADAMS: I want to pick up on something that -- oh actually, Debbie, you've been waiting a while to speak. Go ahead.

DR. WABER: That's all right.

DR. ADAMS: Please.

DR. WABER: [laughs] I just had a -- I was trying to get back to you initial question, and I think part of it gets back -- I think we -- as a neuropsychologist, I am just overwhelmed by how many tests we have and that we over test and do way more testing then we should. And that gets back to insurance companies because that's all the paperwork you need when you bill. And then you get into experiment-wise error because if you have too many tests, some of them are going to come out, and then you don't know what to make of the ones that do or don't come out. So I think it's really important to be kind of -- I think it's important to go back to developmental psychology and to think about, you know, if there are markers.

You talked about EEG, you know. Are there markers of risk in, say, in a pre-term that would then -- that you know are going to be highly correlated with some kind of outcome at five years or something like that? And really try to identify what might those be to identify your higher risks, so that you can get a sooner answer to your question. You can't wait five years. I mean, I know, for example, that many years ago Dennis Mulphy [spelled phonetically] -- he's in Illinois now or in Nebraska or something -- did some very nice work on auditory processing in -- I think they were like newborns. Very, very --

DR. ADAMS: They were NICU preemies. I did some of those --

DR. WABER: Oh, you did? [laughs] Okay.

DR. ADAMS: I did some of those ERP --

DR. WABER: Yeah. And then was able to predict language processing.

DR. ADAMS: Yeah, he looked at auditory processing, speech sound perceptions during -- from infancy through the early years and then had kids come back successfully, and then by the time they were in school, he was doing tests of reading and finding that speech sound perception in the preemie was predictive of reading development --

DR. WABER: Thank you.

DR. ADAMS: -- reading skill.

DR. WABER: Thank you.

DR. ADAMS: Yeah.

DR. WABER: So, you know, so I think that if -- going back to the basic research and getting, you know -- because I think most of us as neuropsychologists come out of a tradition of clinical psychology, which is probably not the right tradition for doing this, so thinking about like, what are the developmental precursors of later elaborated behaviors that are meaningful? The second point is that I think that we -- what we do in the lab or in the examining room and stuff, we have to be

sure that it has real world relevance. And so I think, you know, I think I would take issue a little bit with what you said also about the questionnaires because I find that in a very impaired person, no they're not discriminating, but as you -- that you didn't find like The BRIEF all that useful is what you said.

DR. SHAPIRO: Yeah.

DR. WABER: Okay.

DR. SHAPIRO: Of the tests, I like it the best.

DR. WABER: Yeah. Okay. But I think that, you know, some of these things, the questionnaires can give you a very good insight into what the real world kind of functioning of a person is, and in my mind, that’s what you really care about. So for example, we did one study on steroids where we used the CBCL and actually, you could -- we administered it over the phone every week, if you could put up with that --

[laughter]

-- and found a very nice little, you know -- what happened during leukemia therapy, they have one week on -- a five day bolus of steroids, and then they’re off for two and a half weeks. So, you know, during the steroid week, we saw a big bump in behavioral problems, and then it went right back down again after the steroid week. So you could use some of the

questionnaires; you can do things over the phone. And the final thing is I think, you know, even with the cancer population, if you don't have a population that has ongoing need -- like for example, with leukemia, kids are very sick for a few years, and then they get better.

And if things go well, they go about their lives. I mean, they have this, you know, the family of course has a traumatic history, but they go on. And what we found is that in the early days, when the population as a whole had bad problems, it was very easy to recruit, but as the kids got better and better over the years, it got harder and harder to recruit, even with the reports and even with the phone calls and all that kind of stuff.

And so I think, you know, people have busy lives now, and so I think we have to think about ways of doing this kind of work that don't require that somebody take a day off work, come bring their kid in, take them out of school, and so forth, and so I think that we have to be very creative about some of these more questionnaire kind of things that you could do in a half hour. Phone conversations might get you as much -- you know, you have to validate them against all the regular stuff we're supposed to do, but I think in terms of feasibility, you know, you just don't get a representative population if you insist that they have to come into your lab --

DR. SHAPIRO: I agree.

DR. WABER: -- because no matter what you do it takes them a whole day.

DR. SHAPIRO: I agree with that. I wanted to ask you if you had ever considered taking the CBCL items that were sensitive to the steroid use and creating a scale that might be even shorter and easier to administer because I think that might be the way to go. I think you --

DR. WABER: Well --

DR. SHAPIRO: -- people need to be creative about taking --

DR. WABER: Yeah.

DR. SHAPIRO: -- you know, not giving that whole long thing --

DR. WABER: Right.

DR. SHAPIRO: -- and just finding a way of measuring change in a very short and easy way. I want to say two more brief things.

DR. WABER: Yeah.

DR. SHAPIRO: I didn't mean to say that those tests were not good. I think that they're fine for long term outcomes. I just don't think they're really good unless you're doing what you're doing, measuring every week, and maybe getting a subset of items. For getting sort of an overall look at drug

sensitivity, for example, I think you can use it as a long term outcome measure, and maybe if you had a subset of items that were good, that would be great.

DR. WABER: Well, we tried the subset of items, and our statistician like, you know, would say, “No, you can't do that, so we have to start the study” --

DR. SHAPIRO: [laughs] Yes, you can.

DR. WABER: -- “all over again.” And we couldn’t, you know -- you’re absolutely right, but --

DR. SHAPIRO: Right. Right.

DR. WABER: No, they wouldn’t, you know -- the statistician people wouldn't let us do that.

FEMALE SPEAKER: Yeah, go ahead.

DR. YAO: Can I ask a follow up question? And, Ken, I’d be interested to hear your thoughts on this to but, so this is the question I think Elsa has already asked and many others have sort of thrown out there, but you were talking about, and I’ll use the same example of memory, so how long is long, and when would you measure? I mean, I’ve already heard that, you know, as Ida Sue mentioned, well, we’d start at one. Like is there any reason to start before one at all, and is there any reason -- when would you actually test, and how long or what -- maybe, it’s too broad a question, but maybe just some sort of parameters?

DR. SHAPIRO: I think it depends on the disease.

DR. ADAMS: It depends on the disease, and it depends on the treatment, the exposure. So if you want to think of the treatment and think about the half-life of the treatment, how long are you expecting the direct effects of the treatment to last? What are you expecting the treatment to do mechanistically to brain development, and how long do you expect that to unfold? And then also when do you expect this cognitive skill to come online in a robust way? Because each of those things may operate on a different timeline, and for probably every disease and every, you know, drug or biologic or device that you're considering, you're going to have to ask that in a slightly different way because that intervention is going to push and pull at those timelines in different ways.

DR. TOWBIN: Yeah, I think that's a really good answer. I think that that covers the water front really well. One of the things that I've been surprised about today is that we haven't said anything about adaptive functioning. Maybe that takes us out of this realm, but I wanted to follow up on Dr. Waber's comment. No one has said anything about that, and are we of one mind that we all now feel so confident about the way to measure that, that we don't have to worry about it? I don't know if it's in the Toolkit, but --

DR. ADAMS: We need another workshop for that. No, we

seriously do.

DR. TOWBIN: -- the population [laughs] -- It's, I think, a very serious problem in all of the populations that we're talking about. For the populations that I know most about, things like IQ are not so sensitive. Higher-functioning individuals with autism spectrum disorders, those with mood or anxiety disorders can look very spiffy on many cognitive measures, but their adaptive functioning is poor, and it doesn't begin to get at that. Some of the executive function measures do provide some window into what some of the problems are, but I think there is a way in which this is going to have to be addressed as much as some of the other things we've been talking about. And I guess just one more thing. I've been somewhat quiet [laughs]. And I think that I was looking from a more 10,000 foot sort of view about this.

I was very interested in what Dan had to say about the animal studies, and I was thinking about how those animal studies might actually inform us about regions and functions that we need to be concerned about and how that would dovetail with investigations that help. Those would inform us about the kinds of subtests and functions that we would then be looking at in neonates and older individuals. Animal models allow us to look at, well, what would long-term exposure be? Dr. Sheridan talked very nicely about the way in which we've thought about

this for anticonvulsants. Many of the patients he sees and I see are on medication for years. You know, we talk about things like atypical antipsychotics in individuals with autism, but now these individuals are on these drugs, not for months; they're on them for years, and so we want to be looking at, what the accruing effects of those kinds of agents? There are no studies looking at very long-term, late effects of these agents. So there's going to have to be a means to do that as well.

DR. ELAYAN: Sure. Well, I think I have to interject here a little bit, but that's fascinating, and that's very interesting because what I'm hearing you here, I relate exactly to what we do in those animal studies. So, you know, the timing, when you give the drug, for how long, the long-term would -- you know, during the treatment. We stopped the treatment for, you know, a good recovery period. So the drug is out, and then we look at the long -- that's what we call long-term effect for the drug. As far as the -- what part of the brain that is affected, if we know what a drug's doing, that's what we look for, and neuro-histopathologically, and we try to, with these mazes and all of that, try to use the ones that are specific for that brain region.

So, you know, I'm thinking of, you know, at least that's where we start. And that’s when my question came to you. You hear all of these things from us, but what do you use these

data from a clinical point of view? And it's fascinating to hear. It's a lot of things that you can, you know, think of what tests to do, and all of these, you know, measurements that you need to do, but definitely our part, at least, we think we're trying to reflect what the humans are going to be exposed to, for how long, and all of that. So hopefully, we're doing the right thing the first step before you do come into the action here.

DR. TOWBIN: So what's interesting about this is I don't that we'll be able to put much stock in brain morphometry, MRI studies. I really think that the kinds of insults and the kind of dysmorphology that's going to show up there is not going to be so informative for the kinds of things that we want. I think functional, magnetic resonance imaging is going to be much more important, and indeed, we have now, through virtual reality mechanisms, ways of doing water maze tests on individuals in fMRI, and it's very much the same water maze test that you can do in rats. This is really the kind of linking, I think, of what RDoC is trying to do by looking at biologically relevant kinds of things and then how they can be carried forward to human studies. And I think that's where we need to go as well when we're thinking about these kinds of neurocognitive outcomes.

FEMALE SPEAKER: Exactly.

DR. PANDINA: One other challenge maybe that I would put out there. I think -- first of all, I agree that there are other supplemental tests that you would want to do functionally for the kinds of questions that you would want to ask about: memory, or attention, or verbal learning. I completely agree a standard -- I was confined to the title of my talk, but I think that the types of tests I was mentioning at the end there, many different types of tests that you could build in that measure more specific functions, and again, different parts of the function that could be relevant. The other challenge though that we face is intervening treatment and intervening occurrences. So in autism, for example, I was just talking to a father the other day who has a child, 17, developed a seizure disorder at 17. So, if you're following him longitudinally, he suddenly developed a seizure disorder, and you're looking at him in a safety cohort for something earlier, you know; well, what do we say?

Could we introduce -- now we introduced an AED. He now has a new treatment on board. How does that affect your evaluation of the safety and the likelihood of the relationship to an earlier event? So I think the complexity of these longitudinal studies are lots of things happen in between. Kids also in these high-risk groups engage in higher-risk behaviors and have negative outcomes associated with the behavior

associated with the illness. So I think finding a way to -- you know, you're trying to split the atom [laughs], and you're trying to also answer a number of different questions, and that becomes a challenge when you look at these very, very long-term studies. So I think extrapolation, and I would wonder about this sort of cohort approach.

The one really nice thing about many of the longitudinal studies that the speakers here have been involved in are looking at these large cohorts where they're not standardizing a treatment over time to see what the naturalistic course looks like, and then be able to compare. Unfortunately, we don't have those cohort studies for all of the populations of interest. So you don't know what the naturally-occurring incidents -- that type of adverse event would be in the population, with or without your treatment. It makes it very difficult. So I don't know if the panel has any thoughts about that. The other was just the duration of untreated illness. So if you go a long time without getting your ADHD treated, we know that your likelihood of having a better outcome is lower. With autism, it's true as well. If you go undiagnosed until the age of 5 or 6 because you were somewhat verbal, the likelihood of you having a good outcome is less; the likelihood of you having fewer people get out of the diagnosis is less. So.

DR. YAO: All right, well, I'm going to ask --

[talking simultaneously]

DR. SHAPIRO: I do. I have a --

DR. YAO: Do you want to go?

DR. SHAPIRO: No. Go ahead.

[talking simultaneously]

DR. SHAPIRO: It just was another topic.

DR. YAO: Okay. Well, this maybe will close the first question I had. So maybe I didn't ask it the right way because I totally get that, depending on the DZs and the area that you think is most affected, or when you think the timing will occur of the deficit, that those all play. But you know, as I was -- as I have thought about this, it -- and I'm going to take this, again, crude example again, but if you had a situation in which you were worried about memory, okay, just memory, and I'll take that baby in the NICU again. I mean, how short is short, I guess, I'm going to say? I mean, how -- at what point could you ever expect to have, you know, any information that would be helpful in terms of estimating that product's or that drug's impact on memory? Like would you -- could you get away with something that was 12 months later, and then expect that I wouldn't need to worry if everything was clear at 12 months? Right. So how short is short, I guess?

DR. ADAMS: So what kind memory are you talking about? That would be my first question.

DR. YAO: And again, this is the pediatric morphologist asking the question. The answer is yes. Yes, memory.

[laughter]

And I would -- please, you know, fill me on, you know, if you had a specific kind of memory you were worried about, when, you know -- how long would that take to manifest?

DR. ADAMS: Well, to give you an example from the setting of pediatric morphology, one of the studies I'm involved in is a study of chronic hypertension in children, and my primary collaborator is a neuromorphologist. And so we study memory in a number of different ways. We also study executive function and processing speed and attention, and these are in older children, not preemies or NICU babies. But we get different signal from out different memory task. So depending on whether we're thinking of rote verbal learning, or visual spatial learning, or some other types of learning and recall. So I think it depends, and I know we keep hedging with that for you, but I think it really does depend on all of these different factors.

DR. BARON: The other problem is, even if you've got something that you thought you could do in a short period of time, that may be true only for that short period of time, and then as the child matures, that may not be persistent. And so -

-

DR. SHAPIRO: And you're measuring different things sometimes.

DR. BARON: Right. Right. Right.

DR. SHAPIRO: So, you know, we developed a preschool memory test, and it was great. We were able to get some good data, and then we sort of looked at, okay, these long-term outcomes. How did it relate to, for example, the Atlantis on the KBC? Well, the correlations were about .3. That's not adequate if you're looking at something that’s, you know, an encoding type of task. So you know, you got great results on the early tests, and you got good results on the later tests, but they were measuring something that was really different in the child. So, you know, that's a big problem.

DR. WABER: The other problem is -- and I think it's sort of insoluble, is the developmental problem because development is a hierarchical process. So children don't come into the world with a memory module, and a language module, and all these little independent modules. Cognition is constructed in the course of development. And so what happens is that, you know, you can have sort of a minor sort of aberration that then sends off, you know, into another trajectory, which then, you know, loops out into yet again another trajectory. And what we, as grown-ups from, you know, this level say, "Oh, that's

memory." But it may not have been memory at all. You know, and what you call memory is, you know, as Elsa was saying -- I mean memory is a crude word that represents a lot of different processes depending upon the context, and you can get -- It's not even clear that with one agent, you're going to get a signature, you know, even if you're 10, 15 -- you know, the way -- I mean, I think animals are a little bit different that way, although I am not intimately related to them.

But my understanding is that -- and also, you know, that you can get sort of a behavioral signature, although I know there are individual differences in rats. But you know, by the time somebody's 10 years old, you know -- though I think what you're left with is just, you know, getting back to your question about adaptation and so forth, is you can't say, "Okay." You know, it's rare that you're going to find an agent whose toxic effect is going to be, you know, when I'm 10 years old, I'm going to have a memory deficit. That happens to adults who, you know, whatever. And so, you know, I think you start -- you look at it from different levels and say, "You know, adaptively, is this person different in a meaningful way from somebody who, you know, was in the control group?" or whatever, and it can unfortunately take many years for that to play out.

And you know, my own personal sort of -- and I come at it -- I'm not an infant person, but I kind -- my kind of

personal view of it is if you look at an 8 and 10-year-old, and there is no difference, who cares what they were like when they were 3? Because ultimately, you know, it all sort of came out in a wash. On the other hand, you could get a minor difference at 3 years old that, you know, spins out so that, you know, it gets pretty major, and those are things we simply can't know. So you have to make, you know -- I think the answer is unless it's something -- I mean, my own clinical experience is, if it's something big and bad, you'll see it when they're 3. If it's subtle, you may not see it until they're 10 or 15, or you may never see it at all. And so I think it's really hard to answer your question.

DR. PANDINA: I do think though, you get -- I'm sorry, go ahead.

DR. BULL: No, I -- just this is a fascinating discussion. Thank you so much. But I would add, from the environmental exposures issues that they may be something totally different. So all of the things in terms of behavior, function are age-related, but from the arsenic group, bladder cancer at age 20. I didn't even mention it, but that's another issue related to arsenic. So we're not going to follow these seven-serving-a-day preemies with rice cereal until they're 20, but that's something that we need to be aware of.

DR. SHAPIRO: Well, why not?

[laughter]

DR. BULL: Money, money, and money.

DR. SHAPIRO: Right, but it would -- I mean, those kinds of things that, you know, if we know something about what's going to happen in the future, that it's so -- even if we do it every three years or something like that, it would be so interesting to be able to have a structured follow-up of, you know, many of these studies.

DR. BULL: And I think -- I just want to come back to what I think you asked the very first time, and I thought we were going to get around to it, but in terms of what do you tell the parent, I think you asked that at the outset of the study. If you're going to use something in the NICU, what should you tell the family about the long-term study that you're considering relevant to that drug? And Sue raises the whole IRB informational sharing information.

DR. ADAMS: To add to that is just a complexity that we don't have time for in this workshop or at the time of day, which is not just what you tell the parent, but what the parent hears.

DR. BULL: Right. Right.

DR. ADAMS: So when you have a child who's been diagnosed with high-risk ALL and needs to go on therapy immediately, and you're going to enroll them into a trial, which

is just what you do for kids with ALL. They all go on trial, and that's why we have such great success rates. And then they have to make decisions and read a 50-page consent form and all of that, it’s, you know, what do you tell them, and what do they hear? I think that's very important, and it is a big discussion for another day.

DR. BULL: And I teach my medical students that if when you've told them all those things, they've heard 50 percent of what you heard, you did an awesome job.

DR. ADAMS: Yeah. I mean, if they're heard three things -- if you can get them to understand the three most important things, and some of our consent forms now, for the more complicated studies, were moving to a questionnaire at the end of the consent form. So before consent can be completed, our IRB might suggest that we do just an information check on the last page to make sure people understand the core elements of the study before they sign on the dotted line.

DR. SHAPIRO: I want to follow up what --

DR. PANDINA: I was just going to say, I think trajectory assessment is important. So I do think that those period tests, whenever they stop, and you're doing a subtraction from what the normal trajectory might be from a typically developing individual, and then from what the population might be that's not receiving drug A or drug B, or intervention A or

B, is still important. And I think you ought to always -- I mean, as a clinician, you're always thinking about, how close have we gotten them back to a good developmental path that they're going to stay on forever? My endocrinology friends always talk about, you know, if I know what their height and weight is, and I see that that's normal for five years, I'm feeling pretty good about them. I can make some estimations about their cognitive development. I might not be perfect, but generally, I see it in those two things.

If I see an abnormality there, a change there, there's something going on that I want to investigate. So coming up with whatever that subtraction is, using those core measures that you have, even over a briefer period of time, is still not nothing. It's still important data. And I mean, we face that all the time in clinical trials when people ask us, "Will this drug still be affectatious?" We do a discontinuation study at six months. It's still affectatious, but we can't do a double-blind study for 12 months or at 18 months because that would be unethical. So you know, how long can you carry this out? I think the same is true with safety studies. You can start to think about where that deviation is important. What's a clinical meaningful deviation? Maybe that's a good way to start thinking about it, irrespective of how long we can follow them and how long it's feasible to follow them in longer-term

studies, especially in longer-term drug studies.

DR. SHAPIRO: But I wanted to follow up on something that you were talking about, the adaptive scales, and talk a little bit about what parents want, you know. I think we talked a little about this yesterday, and that, you know, the IQ is fine, but maybe in addition to that, they're also wondering, “How well is my child going to function, and how much disability is my child going to have?” I guess I'd like to hear what people think about, you know, adding the adaptive disability scales to whatever the relevant -- in relevant studies. Some of them don't require that, but others may.

DR. ADAMS: We do that in the Batten disease research. We have adaptive measures, both on our disease-specific rating scale, and also I give an abbreviated version of the scales of independent behavior revised, which is a standardized assessment of adaptive skills. And not only has it been really helpful to get an understanding of, you know, just children's day -- the day-to-day impact of the disease on children and their families, but we've also been able to do some hypothesis testing now, looking back at our natural history data. So for example, we tested a hypothesis of a different disease progression for females versus males and found that, amongst the things that were different, were differences in terms of independent ADLs, and that girls had a much more rapid progression of disease and

lost their independent ADLs more quickly than boys did. So I think that those sorts of measures are very practical and useful and give you a sense of disease impact, but then also allow you do some hypothesis testing about what the disease is doing to these kids.

DR. SHAPIRO: Anyone else have thoughts about adaptive functioning?

DR. DELANEY: Dr. Baron, have studies been done on early parent, you know, perception and hope for -- I'm thinking of someone I visited who was about to give birth, pre-eclampsic 24 weeks, and she said to me, "I just don't want her to have learning disabilities," and I thought, you know, and I didn't say anything, but thinking if she would survive. And she did. She survived and is thriving at age 10, amazingly. But I'm just curious, are there -- is there research around that asking parents in the NICU, obviously a very stressful time -- not in the NICU, pre-NICU. What do they hope for, and what do they --

DR. BARON: Well, let me tell you a little bit about our experience, which seems a little bit different than some of what is in the literature because we haven't talked about care center factors, and you know, I started this 10 years ago not knowing much about NICUs exactly, the way I should have, and thought all NICUs were the same, and they're all different. So what I've learned is that, in our NICU, our outcomes are really

quite good, and our 23-weekers have a mean IQ in the average range from our study. We have over a 90 percent survival rate for 23 weeks, 24 weeks, and 25 weeks, going up a little bit in the 90s for each. And we have children in each of those gestational stages that have IQs in the 120s, mid-120s. So there is certainly some children who have IQs much lower, but we are blown away by the competence of some of these children.

We even had a 22-weeker who had an IQ of 122, and I'm thinking about writing it up as a case study. So the fact is that what I've learned is that gestational age and birthweight are really not the answer. Sure, they contribute, but factors other than those matter hugely, and we have a mean higher education, higher income – Washington, D.C. area population, 20 to 30 percent of our children, English is not the first language at home, which has made for some really interesting data from when we saw them at 3 to when we saw them at 6, and I'm writing that up now. So that there are these factors that are forgotten sometimes, but there are so many of them, and they're really complicated, and if you have a -- I can tell you about if you have a 23-weeker or a 24-weeker who comes through our NICU and has not had surgery and just was in the NICU gaining weight and doing well and not having medical complications, not having neck, not having other -- they could do really, really well.

So then I read the literature, too, and I see from

other places that that's not the outcomes that are being reported. Some of those outcomes are being reported in current years, but if you actually look at what their birth cohort year is, it's back when there was a different standard of care. So their treatment is really different then what our treatments are. So you have to be very careful, I learned, about reading the literature because if it's published in 2013, and it's about adolescents or late young adults, they may have been born, you know, before surfactant treatment, or they might have been born at some point in time which has no relevance to a child being born today. So there isn't a really simple answer, and I will say that even the children that have average IQ, who are at the lower gestational age, that does not mean they're fine. They have deficits, too, but they might -- the bar might be raised, you know, so it's a little milder. So our perspective is that these children should be resuscitated actively. They can do well, and with good maternal education, good, you know, income, some of the beneficial socio-demographic factors that we count on, there could be a really optimal outcome. So that's been our experience, and I know it's really different than communities where it's a very different socio-demographic group.

DR. DELANEY: So that's even important in considering sites for trials. I mean, this is something that probably isn't even considered if they have the facilities, if they have the

physicians, if they have -- for a NICU study. Getting a facility like yours would be optimal, but if there's that much difference --

DR. BARON: Right. We think we eliminate some of the confounds. Now I will tell you, in the very beginning, in the very first year when I was looking at the literature, and I looked in one, without naming, places. I looked at one place, and I looked at what the mean IQ of their control group was. It was 84, okay?

DR. DELANEY: A control group?

DR. BARON: Yeah. So, you know, our control group is usually in 110, 11, or 12. So I just went [laughs], you know, spent a long time going through this and realized that, in fact, the decisions that the NICU makes are very different in our NICU maybe than in another NICU, and that's been explained to me in multiple ways. So the treatments they choose to do and the consistency, and the first hour protocol they have in the delivery room -- everything from the delivery room through for the first 60 minutes, it's so intense, and practiced, and rehearsed that everybody's doing the same thing for every baby. Well, you know, my understanding is you could have a neonatologist come to the crib of one baby, give a nurse an order in another NICU, not ours, and then another neonatologist comes to the next baby and gives the nurse a different order for

the same kind of reason.

That doesn't happen is my understanding, as in our NICU. It's very -- and it's also 24-hour, board certified, neonatologist care. There is never a time when residents or fellows are left alone on the floor. So there's a lot of things that have been done, and I do know that they are participant in the, you know, Vermont-Oxford network, and I think they have about the lowest infection rate of any place, and that's huge. And then some of the variables that other people have always published that this is a complication, we have a high rate, and we've never found it to be a complication. So it's been fascinating, but we do feel like it's a little bit different, and I'm completely -- in fact, next week, Pediatric Academic Society's V23 to 25 week poster's being presented. And it's just remarkable to me how these children did, 170 of them.

DR. ADAMS: I'm going to invite people to think about a change of topic. Oh, go ahead, Else. Did you want to --

DR. SHAPIRO: Yeah. I would like to ask you to talk a little bit about telemedicine and how that might be useful in some of the work that we're talking about because when you were talking about the telephone conversations, it occurred to me that in some of these particular follow-up, longitudinal studies of outcome that that might be a good solution for not bringing people back into the --

DR. ADAMS: Yeah. I would be happy to talk about telemedicine and remote assessment. Before I do, so that you can all help me remember, the topic I wanted to shift to was the topic of education and training that Elsa brought up at the end of her slides because I think, you know, we all who do this feel that there aren't enough of us out there, but we struggle to get people interested and trained up, so that we can really rely upon them to give us high-fidelity data in multiple sites, and that limits our ability to do our work. So I know that as a field, you know, we need to think about ways to bring training in clinical trial research to more psychologists who are doing this type of work. And so I just wanted to make sure we have a chance to talk about that as a group.

In terms of telemedicine, I gave a talk yesterday. I'm happy to send you the slides. You weren't -- I don't think you were here, so I can send those to you. And it really was just an overview. Remote assessment could be anything as simple as sending a survey by mail because you're remotely capturing information, and they mail it back, or using telephone, or using computers in various ways, shapes, and forms, whether it's to connect with someone live with video and audio, and you're live steaming, or having automated tasks in the cloud that get pushed to people. But you can also now -- I didn't talk about this yesterday. You can use the internet, and email, and those

technologies to push questionnaires to people in an easier-to-complete way. So -- in our institution, a number of studies that are done use -- not necessarily mine, but throughout the institution, use REDCap now, and REDCap was developed by Vanderbilt University.

A lot universities now have accounts for it. It's a way to develop forms that you can push to people online, and they can -- you know, they get sort of a secure link. They fill it out. It's compliant with health privacy laws, and it comes back, and it goes right into a database, so that you don't then have a CRF that you have to hand enter into a database itself. REDCap is okay for simple surveys, but it doesn't have the functionality to really do complicated things. But I could well imagine a scenario where you might have a questionnaire you might want to push to people every three months, or six months, or however often, and you use item response theory to perhaps streamline the completion of the assessment. And it's completed by a parent, or a teacher, or the participant themselves if they're able.

You could use it for behavioral assessments. You could use it for adaptive function assessments. And then, you could think about some of the live video assessments for cognitive testing or even using things like PDAs and tablets. So there was a study published last year by Doufeu [spelled

phonetically] et al. I can't remember the journal, but this was a study that looked at the correlation between performance on a lexical decision making task. And this is a task in which participants look at a word, and the only thing they have to do is just decide, is this a real word or not? And the dependent variable is time. And so the traditional way to give this task is you kind of show the cards to people; they make the decision, and you know, right or wrong, and how quickly they do it.

So they looked at the correlation between a direct, face-to-face administration of that task, versus that same task delivered on a smartphone that a person could just download to their own personal smartphone and complete it just by tapping yes or no on the touchscreen. And they found that the correlation was probably .9 or better. There's a really nice graph in the paper that shows that. And what they also discovered is that it only took them about four months to get data on several thousand participants, and it probably took them four or more years to do the equivalent data collection for the in-person norming of this test. So there are lots of ways you can think about using remote assessment, both in gathering your outcomes, but perhaps also doing your initial validation as well.

DR. PANDINA: Maybe I can mention something about this as well. Two things. One, about 10 years ago now, I conducted

study in depression. I won't talk about the drug, but I'll talk about the result. We did phone-based assessments of a depression ratings scale, and then we did in-clinic assessments, and we did the phone-based assessments in intervening weeks as well as the weeks that they came in the clinic. And what we found is we had a lower placebo response on the times when they weren't in the clinic. There was a better separation between drug and placebo in patients when they weren't in the clinic. Now you know, you can talk about what the benefit or the weakness is to not seeing the patient more frequently in the clinic is, but you know, patients don't want to come to the clinic. Subjects don't want to come in if they don't have to. So that's the first piece, and I'll tell you the second piece after you respond because you --

DR. ADAMS: I was going to just tag on to that that one of the -- one of the items I touched on yesterday was that there are -- there is the idea, especially if you're dealing with rare diseases or conditions, whether it's prematurity or something else, where you really want to have someone who has expertise in the type of assessment that needs to be done and the type of condition that needs to be assessed. The more raters you have, and the less experience they have at multiple sites, the more variability you're going to introduce into your measurement that you can't afford to deal with. And so if you

can streamline the number of people who are expert in both the measure and in assessment of these types of subjects to fewer people, you're just going reduce -- you're going to reduce the amount of unnecessary variance that you gather in your observations. And so that's going to lead to a better measurement of your affect.

DR. PANDINA: So another example I'll give. We're right now working on autism spectrum disorders, and we're creating a system to capture information. And one of the things that's quite interesting in some of the pilot work that we've done is that parents will capture a lot of information, even daily, when they can do it easily. So if you can get them to do it on this, and you only ask them a few questions, they'll do it daily, and they don't mind if it takes them two minutes. If you can feed a bit of that data back, the data that's important, to them, so that they can also see it, so you're collecting outcomes that are important to you and some outcomes that are important to them, they're even more likely to do it.

So I think as we go forward in thinking about informationally adaptive functioning, and often parents are quite good reporters on the child's adaptive functioning, especially if you can ask them the right questions, and then you can marry that with clinician measures. There are ways to do this that will also save, you know, time and effort in terms of

-- and I'm sorry to have missed your talk yesterday. I was coming back from Europe, so. But it's very -- what we're finding is that if you can make it relevant to the parent, rather than centering the evaluation process around the doctor, but around the parent, and having them be -- also get something for them and be able to use the information themselves without compromising your trial by using a measure that they can see that's going to somehow confound the outcome.

DR. ADAMS: I was at an NIH workshop a few years ago. Elsa, you were at this workshop, and I apologize. I can't remember the title of it. But the focus of the workshop was to do with clinical trials in rare diseases, and at some point during the workshop, a parent who was in attendance stood up and said, "You're not using us enough. You're -- you know, you come to us. You ask us questions about our child's disease, and we give you the answers. But let us be sort of coinvestigators with you. We can do it. If you train us in how to do the assessment and how to use the measures and how to, you know, conduct evaluations, you know, without bias to the extent that we can be without bias, you know, give us a shot. Let us help you because could be doing for you than we are."

DR. SHAPIRO: That's a mandated part of the Rare Disease Clinical Research Networks that parent advocacy groups and parents must be involved in every longitudinal study that is

funded by NIH. So that's been -- it's been really helpful because they've been helpful to us and to other studies.

DR. SHERIDAN: With regard to that, and also linking up to a topic we were talking about previously -- that is, not only testing cognition, but also testing behavior or adaptive abilities. We're sometimes confronted with question, “Well, look, on this cognitive endpoint we chose, this is statistically significant different. But what does that mean clinically?” Now does it -- does it -- we've traditionally told ourselves, well, if you not only have that difference in cognition, but you can also show the difference in some kind of behavioral outcome or in a quality of life outcome, then you could be more confident that it's meaningful. And this would be whether you're talking about something good like a treatment for Alzheimer's disease, or something bad like an adverse effect from some treatment. And is it help -- or would more involvement -- well, first of all, is that a correct principle? Does that ring true, or are we off the mark? And then secondly, is this perhaps a role for increased involvement of the patients and their families? Any thoughts?

DR. SHAPIRO: Yes [laughs]. Right.

DR. ADAMS: Absolutely. So, you know, the work that we've done in our team in Batten disease, some of the most interesting questions -- research questions that we've been able

to answer have come from families and their observations, and they’re sitting down over a cup of coffee with us and saying, "You know, I've wondered about" fill in the blank. And we go, and we have our data set, and we can evaluate it. But we never would have even thought to ask the question had a parent not come to us, you know, who has 24/7 expertise in their child's disease. So yeah, I think the more we can involve families and really defer to their expertise, the better.

DR. TOWBIN: I think you've landed on something quite important that was said earlier, which is -- and others have echoed throughout the morning, this difference between statistical significance and clinical significance because many of the things that we've seen in some of these studies show five or six IQ points difference. Well, you know, clinically it's a little hard to know what somebody with an IQ of 94 versus 100 really, how they're functionally different, even though statistically we can get there. So I think that at least what I care more about is how people are doing, and how then those other things correlate with abilities and other functions, this kind of unfolding of things. If a difference at one age then correlates with very much greater differences in other functions later on, then it's meaningful to me. But as you said, if it comes together later on, then it's not all that significantly different, and I think sometimes the numbers obscure what the

clinical significance may be.

DR. SHAPIRO: And you know, it's very interesting that in our research on mucopolysaccharidosis type I, Hurler syndrome, and the attenuated forms, the Hurler syndrome children have a poor prognosis without treatment, and so they're given haematopoietic cell transplantation, and most of them now are surviving, and they're doing okay. From our standpoint, they're not doing great. They're like mean IQ, about 84, something like that. When you ask the parents, they think their kids are doing terrific. I mean, they are really feeling great about them because their initial expectation was that these kids were going to die, that they're going to be handicapped throughout their life. In contrast to that, we have the attenuated group of patients who now are having enzyme replacement therapy, but the expectations for them were that they would have a relatively normal life.

But in fact, the enzyme replacement therapy is not so effective all the way down the line, and so some of these kids are really struggling. In adolescence, they have a lot of psychiatric problems. They have a lot of adjustment problems. And they're -- when you ask patient satisfaction, or parent satisfaction, or parent expectations, they're really disappointed. And so, you know, you've got the higher functioning kids with lower -- you know, the parents had higher

expectations, but now they're sort of not so high, and the others are feeling better. So it really has to do with how, you know, they come to perceive their children and what their expectations are, and then they have different levels of satisfaction depending on what their initial expectations were. So that's an example of how parent expectations may be very important.

DR. BARON: I think it's a really important point because we are always look at those kind of data in clinical work, in reviewing journal articles. Is this statistically significant, or is it clinically meaningful? And I think your question points up that if the IQ measure is the endpoint, then you didn't have backup from something else that was of concern for that patient population, and that other piece might have been able to answer that. So when I teach to students or talk about what we do with neuropsychologists, I make the point that there is no single test that's going to tell you really anything.

You really have to think about what is the backup support for any hypothesis you're going to raise about brain function. And an IQ number, you know, an IQ test, you know -- an IQ is what an IQ test measures. And then, there are the other functions that are important to us as neuropsychologists, and we can measure those, too. So it's important to have

something a little more specific, a little more tailored to the problem at hand. It could be adaptive. It could be coming from some other domain like memory. So I think, if it's only an IQ that you're relying on, you will run into this problem of what is meaningful when you have a small difference.

DR. ADAMS: So for the holdouts in the audience. Are there questions or comments that you want to add?

MALE SPEAKER: Sure, I'll ask one, and it's again coming back to this question of you've got a trial where it's sort of -- you don't know -- you're not necessarily expecting an impact, so a little different from what Lynne described, but you know, we're expected to assess neurocognition or behavior as part of the trial. Does it make sense to maybe, you know -- we don't know what we're looking for, but maybe that's a case where the adaptive behavior scales might be the best option, and I guess I'm curious what people think. Doing that during the course of a follow-up of a year or two.

DR. SHAPIRO: I think maybe if there was a way of getting a very short adaptive function, I think that would be terrific. But I do think that if you're looking at the effects of a medication, that you really need to have something short and brief, and here is where I think that the NIH Toolbox really would be a great way to go because you could select from that a very short kind of assessment of a lot of different areas, and I

think that that might be appropriate if you're not knowing at all what your medication might -- what kinds of affects you might have. So if you, you know -- you would have a combination of sort of a series of neuropsychological tests, and maybe a brief adaptive function.

MALE SPEAKER: Okay. That's what I wondered about because it seems like, given the question about small-effect changes on a neuropsychological test, if you don't have something to correlate with, you may be a little bit uncertain of what the value really is.

DR. SHAPIRO: That's right. I think that's right. Anybody else have comments on that?

DR. TOWBIN: You know, really just a couple of things that I think one would want fairly broad coverage of behavioral kinds of things. CBCL kinds of things, I think, were also pretty quick and pretty useful and now can be pretty easily automated, so that people can do them even through a secure portal. So having some sense about a range of well-validated behavioral kinds of measures like that can be important. And then academic stuff, depending on the age range, can also be useful as sort of a just general screening marker for individuals.

DR. SHAPIRO: But if you're looking at short-term effects of a medication, you're not going to see effects on

achievement very quickly, even though --

DR. TOWBIN: I fully agree with that. I was thinking more how long out the investigation is going to go for some of these.

DR. SHAPIRO: Right, and I don't know what medication we're talking about, but you know, it might be something that was acute, or it might be something that had long-term effects. Those would be different.

DR. TOWBIN: But I fully agree that if you're doing an eight or 12-week medication study, the educational things aren't going to be all that useful.

DR. SHAPIRO: Right. And I don't think IQ is going to be that useful. I think something like attention, or executive function, or motor speed, or something like that would be the most important thing, and maybe some very quick behavior change. I think that would be --

DR. TOWBIN: I think at least in those studies, the IQ measure is going to be part of the inclusion criteria before you even get in the door.

DR. SHAPIRO: Right, but I don't think it should be an outcome measure.

DR. WABER: And also, it depends on -- a year is different, depending on how old you are. So if you're a two-year-old, a year is a really long time. If you're an 11-year-

old, a year is not a big deal developmentally. So that's just another -- it's another complication of doing things with children.

DR. SHAPIRO: Right. Yeah.

MALE SPEAKER: May I? So this workshop's been amazing. I've really enjoyed the conversations, particularly for such as in inborn errors, the genetics where we've got childhood dementia, where you can imagine the primary endpoint has to be neurocognitive tools, batteries, perhaps secondary endpoints. They also need to be of this type of detail. But there's a lot of pediatric drug development that I don't necessarily consider that's in that domain. If it's a safety assessment, I'd like to open up the thinking of what's real-world, and not just what's -- you know, yesterday we had a parent describe that participated in the trial.

It was clinical trial trauma because there was so much to go through, including neurocognitive testing. And when I think about real world first, the patients I see in my clinic, you know -- and I know there's a lot of wiggle and fuzz in this, but you know, school feedback is report cards. I have parents who have kids with tremendous IQs but horrible grades. I have parents who have great students who have very mediocre IQs. What is the role for the school feedback to the parents? Shouldn't that be part of some of this data that we collect?

Does it have any value, or does it have no value? Of course, given what's happened in Atlanta, this is not good timing to ask about school-based standardized tests, but I'd love to hear people's thoughts.

DR. SHAPIRO: It's very difficult these days to get schools to even participate and to give you feedback on an individual child. I don't know what everybody else's experience is, but that's my experience.

DR. TOWBIN: I couldn't agree more. And also, a teacher who just finished their training and somebody who's been doing it for 20 years is not the same kind of rater for many of these kinds of items.

DR. SHAPIRO: Right.

DR. ADAMS: There's different types of data you can get from school. You can get standardized assessment information that you send to them, and they send back. Then, there's the stuff that just occurs as a part of the education, the report cards, the school assessments, and those are very difficult to use because -- not just from teacher to teacher, but from school to school, and district to district, and tax-base to tax-base, and state to state. The educational policies and standards are so vastly different from student to student and grade to grade that you really, you have apples, to oranges, to bananas, and everything in between. We, in our study of

hypertension, which is a multi-center study, we have four sites, and only two or three of the sites are able to get teacher ratings. We have the teachers fill out the BRIEF, and one of the barriers has actually been a barrier with the district schools themselves, which have policies about not permitting research activities to happen that engage their teachers because their teachers are very busy, and they need to protect their teachers' time.

DR. SHAPIRO: It's difficult.

MALE SPEAKER: I appreciate that. I mean, part of me just wonders, you know, like is it good enough just to know that child's working at grade level or not?

DR. ADAMS: Grade level is different. So I'm in New York State, and I have kids come to me from Florida, for example, and what we would call grade two -- sorry, what we would call, say, you know, grade two in Rochester, New York, Florida might consider grade three. So are they at grade level? Yes. For the state and the district, they're in, they're at grade level. But they could be doing completely different things.

MALE SPEAKER: Okay. Thank you.

DR. CRAENIK: I also think it's been a great workshop. My name’s Dan Craenik. I'm actually with the Center for Devices and Radiological Health and the Division of Radiological Health.

Thanks. I find that you have a lot of same discussion that is had in the imaging community and yet sometimes look to the imaging for answers, but I just wanted to comment that imaging also is looking this way for answers for validation, and it has the same questions about, what's the link between something you see on the radiological image to the patient's feel, fit, function, survival? How predictive of it is it? Is it predictive over time? There's a huge question of clinically meaningful changes.

You might see something on an image that has no impact on the patient whatsoever, and you might have huge differences in your patients and almost nothing that's showing up on the images, and then it's also faced with the same challenges of -- I've found it kind of, I guess, refreshing or disturbing that -- how you talked about the changes in clinical scales over time. Imaging technology's changing easily as quickly. It's the same thing. If you start a longitudinal study, you want to use state-of-the-art today. Five years from now, I guarantee, not anything close to state-of-the-art. How do you compare your baseline from today to your follow-up five years ago? And I mean, I'm not overly familiar with the kind of pediatric population we have in development, but adding development to that, in my mind, just complicated the issue unbelievably, meaning your patients are also changing dramatically over time,

and I just kind of throw that out as just a comment and kind of a refresher. So thank you.

DR. SHAPIRO: So there's another issue, too, about scans, and that is if you have multi-center study, having comparable images from different centers. We have five centers in our study, and we get them to send us, you know, images of control subjects, and we have to send them back, and send them back, and send them back, and sometimes our imaging person has to go there, and finally, we may bet some comparability, but there's still -- it's fine for volumetrics, but if you're going to do be doing DTI, it's very, very difficult to have some sort of --

MALE SPEAKER: Sometimes we look at multi-reader, multi-case studies because there's variability between how different physicians interpret the same images. I would say -- you say volumetrics are close. I would -- I've looked at lots of different software packages. The software package you choose will change the volume you get. Can you compare between studies --

DR. SHAPIRO: They have to be adjusted. They have to be adjusted, you know.

MALE SPEAKER: -- and the acquisition is an added complication. So you inquire at one site, and somewhere else, they use different parameters. One has GE, one Siemens, one

Philips. No, I mean, can you compare the -- they don't even use the same language for MRI when they talk about setting up the pulse sequence, which means that what you think is the same parameter on different systems might not be, and then some parameters have different names, even though it's basically the same thing on the different systems. And I know it's just a real challenge, and I just wanted to --

DR. SHAPIRO: Yeah, it is a --

MALE SPEAKER: I guess the big thing was I heard that a little -- kind of the neurocognitive using image to validate, but we'd like to use the other way also; and there's a huge amount of variability in both.

DR. SHAPIRO: Right, and you know, in order to be in our study, you have to have a Siemens, you know. When we started a Trio, and some of them now have SkyRIS or something, PRISMA or something, but you know -- and the same software package. So that's the only way we could get comparable data from one center to another; and I think that that's not a cheap or easy thing to do, and you need a central reader or one person to do all the volumetric analysis in one center, right?

DR. TOWBIN: And I suspect that you didn't have a number of controls that you flew to each site should all be scanned in the same scanner, using the same sequence --

DR. SHAPIRO: Our controls were all --

[talking simultaneously]

MALE SPEAKER: -- so that you could have --

DR. SHAPIRO: --done in our institution.

MALE SPEAKER: -- if you -- interrelated reliability, if you will.

DR. ADAMS: But I heard of that being done in one study, where it was an adult study where one of the investigators went to each site and popped into the machine --

[talking simultaneously]

DR. SHAPIRO: We do that. Yeah, we've done that.

DR. ADAMS: [inaudible]

[laughter]

DR. AKSHOOMOFF: Yeah, we did that in PING. We had research assistants go; they loved going to Hawaii --

[laughter]

One of our sites was in -- the physicists had to work very hard to make sure that --

[talking simultaneously]

DR. AKSHOOMOFF: So the data have been all centrally processed within our core at UCSD and site is one of the covariates to try to deal with that issue.

DR. TOWBIN: But I guess the point is that it's really expensive and complicated just to get to that, and so we're not quite at the point where we can, you know, easily have multi-

site studies for MRI or FMRI or DTI.

DR. SHERIDAN: I have a sort of an FDA reviewer type question. The way the FDA's structured for -- a certain disease or certain drug class is likely to be reviewed, whenever there's a clinical study protocol, by a limited number of people within a particular division within the agency, which leads to the dream that, perhaps, there could be a certain bare-bones, standardized set of measures that would be done in all such trials for a particular drug class or a particular indication such as, perhaps, onset epilepsy, say, that couldn't be pre-specified that all the studies will do these. It might be a Bayley in the young children and then some kind of Wechsler intelligence test in the older kids, combined with some measure of behavior; and then that way, you could do cross-study perspectives. But then I hear things like Elsa saying, “Well, you know, if you're interested in the short-term effects, you should be looking more” -- and I may have this wrong, but things like attention and working memories compared to IQ, that the IQ is something that you would see in the long term, and not just your six-to-twelve-month, long-term studies, but real long-term studies, like how are people doing five years from now? So is it sort of a pipe dream to think of trying to have a small handful of tests that you're going to say, “Everybody’s got to do these?” Is that sort of looking in the wrong direction? I'd

like to get some reaction to that.

DR. SHAPIRO: I think that would be okay if you were, depending on whether you were looking at short-term or yearly effects. If you're looking at yearly effects, you might want to do some tests that have more, like IQ tests or something like that. But short-term effects, if you're looking at the effects, you know, down the road in three weeks or six months, you need to do something that's going to get that. But you could have it. I don't see any problem in having a standard battery. What do other people think?

DR. TOWBIN: I think you're cursed with needing both.

DR. SHAPIRO: Yes

MALE SPEAKER: I think that, especially for anticonvulsants because the expectation is somebody might be on those for quite a long time, that you would need the short-term effects because if something is happening soon, you would want to know that. You would want to be able to track that signal. And then long-term, because it may be something that accumulates rather slowly before you see the effect, I'm thinking of, like, topiramate and things like that.

DR. SHERIDAN: Or something that took effect in the young child when the nervous system was developing, but you cannot appreciate that until they reach a certain level of maturity down the road, and you see, oh, the wrong connections

were made five years ago because of this drug.

DR. TOWBIN: And then the other suit is that it's not like there's just one kind of epilepsy that you're using that drug for; and so, you know, the effects of where that lesion is and what kind of epilepsy they have is certainly going to be important, too, as you pointed out.

DR. SHERIDAN: Plus our drugs end up being used for pain control, also used for mood disorder, and many other indications.

DR. TOWBIN: How well I know [laughs].

DR. YAO: So, I'm going to ask or maybe make a comment or ask a question that will lead into the last area that I think Heather wanted to kind of bring up, and then I would propose we go into the final comments. So Brian brought up the question about school, and could you use information from school, and I think it sort of feeds in nicely to the concern that has been raised throughout both days of the conversations that in order to have information -- okay, I've heard lots of problems already: timing, testing, disease, treatments, you know, all of these play in. But it sounds like we have a pretty big need for skilled professionals to actually be both considering what of all these things is the most important, and then to administer the tests that we believe would be important. So I would like to hear comments maybe because this is -- you know, we’re going

to -- we want to write something, so this has got to go out there, and then maybe we'll be the sounding call to get more funding or to get more training. So I would like to hear folks' comments about how you think that that -- how we're going to get there, how are you going to get the people you need to do this?

DR. SHAPIRO: Well, I mean, you know, I'm on the adjunct faculty of the psychology department, and I train a lot. I have in the past trained a lot of graduate students, interns, and then post-docs, and most of them come from the psychology department knowing zip about clinical trials for sure, but even things like longitudinal studies or natural history studies or, you know, follow-up of different conditions. And you know, they know something about psychometric theory; they know something about, you know, behavior genetics; they know something about a variety of things, but this they don't know. And the question is the psychology departments are very entrenched in the curricula, I think, in general. I don't know what other people feel, but I think it needs to be at the postdoctoral level. That's my feeling, that graduate students are not getting it, and that maybe there needs to be some kind of effort to train psychologists at the postdoctoral level to do this kind of work. I don't know; what do other people think?

DR. ADAMS: You know, I was telling Elsa earlier that I feel very fortunate that I just sort of on accident landed in

a neurology department that has a very strong emphasis on experimental therapeutics, and in fact, the Department of Neurology at the University of Rochester has, I think, the only NIH through NINDS funded experimental therapeutics fellowship, and so we've actually had a couple of neuropsychologists go through that. I have not, but I've interfaced, you know, with them and the faculty and now fellows in the fellowship now will come to me if they have questions about cognition. So I think that postdoctoral level training is good because the folks who are going out into private practice are not necessarily going to be the ones who are tapping into for this work, but the ones who are doing post-docs and are attaching themselves to academic missions are going to be the ones we really want to capture.

So I mentioned the NINDS fellowship, and they also have a clinical trials methods course, which is structured differently now. It's in its second sort of batch of funding now, and I think NINDS, just because my experience with that institute is -- I have experience with that institute -- they've done a lot to train neurologists in how to do clinical research. They had a course of small clinical trials, which was webcast, and I and a neurologist attended it through webcast. So if there are ways for -- if there are opportunities for training and funding for training for psychologists that's similar, I think that would be really valuable, and fellowship is a good

time to do it because you have some protected time to engage in that level of learning.

DR. DELANEY: Another resource is CTSIs and CTSAs. They usually have a nice online or webinar that are available for basic trials and --

DR. ADAMS: Yes, ours as well.

DR. TOWBIN: But I think the didactic exposure --

DR. BROWN: I took -- what? Go ahead. Sorry.

DR. TOWBIN: Oh, sorry, after you.

DR. BROWN: I was just going to say that I totally agree with that. I really can't even add to it because I think that's exactly right; that is the right level. What I'm seeing in graduate schools around here that even is that assessment has been minimized, and that's kind of shocking [laughs] after how much of an emphasis it used to be. So students are coming in at graduate level and have never really given a test and don't really know how to do it. So I think that that's part of it, The other part of it that I see that I've been concerned about is if you're not in a medical center with neurology and neurosurgery, no matter what population you're seeing, if you're only seeing learning disabilities and ADHD, there's something missing in your training, and we're seeing more and more people who have headed in that direction and are losing sight of the neuro-and neuropsychology, and so it's really important to bring

that to attention as well as the rigor of the protocols.

DR. TOWBIN: I was going to say that I think you are all a rare resource for this because it isn't going to be a didactic exposure that's going to give people the tools to move forward with this. You have to work with a group that's doing the work to know how you're going to do it yourself later on. There really has to be an immersion. So even in a postdoctoral program, unless it's really dedicated to learning about these kinds of techniques within a group that's doing experimental therapeutics of all kinds, I just don't see it as a likely outcome for a more generic training.

DR. BROWN: I'm not aware -- I mean, I wouldn't know this, so I'm not aware if there's anything like a one-year fellowship that comes after your post-doc to do clinical trials anywhere.

DR. SHAPIRO: Well, what we've done, the Rare Disease Clinical Research Networks have the post-docs, so that people can apply for them, and actually, the person that's sort of proceeding me at the University of Minnesota was a licensed, OMAL disease network fellow, and she has learned about clinical trials and has learned how to do this sort of thing. So that's one way, but that's one way, but that's like one person, you know? I mean, we need dozens of persons like this, and there just aren't people in the field. I'd like to ask Debbie about

that. Do you have other people that you rely on in the cancer world?

DR. WABER: No, fortunately I've got Heather, but --

[laughter]

DR. ADAMS: And the only reason I got -- and I think studying the kids on the data Dana-Farber protocols was my introduction to this, and that's what lit the spark for me, and it was only because I was lucky enough that I was being precepted by Donna Palumbo, who is a neuropsychologist on faculty in a neurology department who was also pursuing expertise in clinical trials. But she was the only one, and because I was, you know, her servant for the year, I was lucky enough to do this stuff. But that's how specific it ends up becoming.

DR. WABER: Yeah, I think so, and you know, I think the other problem is, especially with these kind of rare things, so you might, you know, you have like many, many sites. I mean, I've been on studies where there have been like 40 sites --

DR. SHAPIRO: Right.

DR. WABER: And you can't possibly train 40 people. So, I mean, I've seen studies where, you know, they'll have nurses doing it, and they'll have, you know, because there's just nobody, and I -- finding even a psychologist, you know, at a site sometimes can be really tough.

DR. SHAPIRO: Well, Kate's had a lot of experience going around to sites --

DR. DELANEY: Yeah, but 40 is a lot.

[laughter]

DR. WABER: I've never -- I wasn't -- I was not PI on it, but you know, in my institution, we have -- I have started what we call a behavior core for multisite studies so that -- because this was a really recurring problem, and we would be asked to participate in multisite studies, and then, you know, whoever the local person was, you know, the GI person or somebody was like, "Wow, okay, what do I do now?" You know? So mainly, what we serve are these studies that have a small number of patients at multiple sites, and I think institutions need to do more. The old GCRCs used to have some of that, and I guess some of the CTSAs now do, but I think a lot of the people who run these things don't necessarily see the merit in why. You know, they have their DXA machine or whatever, but they don't really get why they would need a psychologist as part of a CTSU.

DR. SHAPIRO: Right.

DR. WABER: And so it's been a real battle within the institution to keep the funding for it, and you know, it's difficult.

DR. SHAPIRO: The old GCRCs used to have somebody dedicated to --

DR. WABER: Some of them did; ours didn’t, but some of them did. Some of them did, but I think, you know, the more that, you know, there can be lobbying from other sources that these things, you know, that these big institutions, if you're going to be in, you know, multisite trials, this needs to be a lab that's available the same way that you have blood labs.

DR. SHAPIRO: In a clinical trial, you can't use rotating graduate students --

DR. WABER: Right, right.

DR. SHAPIRO: It's just not allowed. I mean, I just -- I would never allow that on any study that I'm working on. And people always want to do that.

DR. WABER: Yeah, or they'll say, “Can the nurse do it?” You know?

DR. DELANEY: One of the -- one of the things I have seen, in defense of some of the psychologists who are being recruited to participate in these trials around the world, is they may see one or two patients, and the demand on their time for that one or two patients involved all of the study training, not the neuropsych or neurocognitive specific training, but everything else that goes with it with being -- they're not the PI, but they're being asked to do the work for maybe one or two patients, and then they're also not -- yesterday during my talk I spoke about this, including them, doing a cross-site if they

have the time and if they can participate in calls about the patients they're seeing for the trial and keeping them involved. So they're not just being recruited by the PI, the geneticist, or neurologist, or whomever, to just, “We need you to see this patient” [laughs]. And it's a clinical trial and do all of this, you know, EBC training and everything else. So that's one of the observations that I've made, and anyway, I think considering there, everything else they have going on, and --

DR. ADAMS: Yeah, I think including people is really important, you know. CKiD, Chronic Kidney Disease study, has done that, where they have periodic check-ins where it'll either be a conference call or, you know, even one or two times, they had -- at the conferences, the psychologists were invited to attend and supported to attend, so I know that in past years, they've been able to do that. The other thing I was going to comment on was to your point, just flushing out what you were saying about psychologists may have just a few subjects, and they have to maybe spend more time on the training and certification for the study than they do actually seeing the subjects, and then you might -- and this has happened to me, you have a scenario where if it's a multi-center child, the sample size has been accrued before your site even comes on board. So now you've just spent a lot of your time doing this training, and you don't even see one subject, and many local contracts are

a widget model where the psychologist doesn't get paid unless and until they see the subject. It's not percent effort; it's, “We'll pay you X amount of dollars for the subject that you see.” So there's no compensation for the time that was spent and taken away from your other duties in order to complete the training for a study you never actually see a subject for.

DR. PANDINA: I have participated on both sides of the fence. I actually started during my doctoral training doing lots of clinical trials. I got pulled in to do schizophrenia-cognitive assessments in some early phase-two trials, and then went on to become a rater in some pediatric and some adult trials, and then did cognitive assessment and sort of ending up running defacto are -- for our psychiatry team. Most of these were in psychiatry and neurology trials, so it's a little bit different than when you're just being pulled in when it's really, you know, something like a kidney disease or some other cancer where really the specialist is the oncologist or the protocol team. But one thing that is really true is, first of all, I had a lot of really good experience, but it hadn't been modeled how I should use that experience.

So I had a lot of good training in design, and I'm a clinical psychologist by training, had a background in neuropsychology, did a minor rotation and doctoral fellowship there. But the challenge was how do you -- I think, in graduate

school courses, we have psychopharmacology courses embedded within psychology programs; we have clinical research design embedded within psychology courses. It would not take very much to extend that knowledge to clinical trials for pharmaceuticals, generally, so -- and where they're used. I don't think that is too much to ask. It also can't be anathema, which coming from a very strong behavioral, cognitive behavioral program and now ending up working for a pharmaceutical company for 14 years, people think that's a little weird, but it actually was pretty natural for me because I understand the diseases. I understand the conditions, and I appreciate we want to know who can be treated and who should not be treated with these medicines, and if we do that better, it'll work better.

DR. ADAMS: And I think those things are not at all mutually exclusive, but highly complementary. I -- the other hat I wear is as a clinician who provides cognitive behavioral therapy, and I'm always wanting to provide evidence-based intervention, so and unless we have rigorous clinical trials to establish an evidence-base, you know, we're not -- we’re just kind of left shooting in the dark, so we need that.

DR. PANDINA: I'll also say, on the sponsor side, when we ask sites, “Gee, you’ll need a psychologist to see patients,” they freak out. I mean, generally they're not happy with us, and this is in psychiatry where the department should be right

there, and we know they're working the department, but they, you know -- we have residents who we have do our rating evaluations. That's set up in our program. We already know how to bill for that, so that there's this sort of push-pull about how we get involved. And then it's seen as pushing down the credential. So can you give us something that you can just train the people, and we can just train our coordinator to do it? We don't want to have to bring another person in, and that makes it very, very difficult on us as a sponsor. We basically have to say, “Okay, we'll allow a master's level person to do it. Okay, a master's level with mental health experience. Okay, a psychiatric nurse who has one year of mental health experience who's never given a psychological test in their life is now the same as the psychologist who's been doing this for 15 years.” So we have to make the protocol simple enough for them both to do it equally effectively. And I know, as a psychologist, we're probably not going to easily achieve that, right. It's not going to work exactly the way we think. We're going to get better assessments from one group than another, and that's going to be some sort of confound in our data. So I mean, we've ended up defaulting sponsors to computerized cognitive batteries for certain measures because they get away -- get us away from so many problems, but then it limits our flexibility on the other side to say, can we use some of these more experimental kinds of

tests that may be more appropriate in individual situations?

DR. SHAPIRO: So one idea I had was, a lot of pharma companies are really wanting psychologists to do this kind of work, and I think that, you know, there just aren't funds out there. But why not get pharma companies to sponsor fellowships in clinical trials for psychologists? I think that might be a possibility. I know that in our institution that Genzyme sponsors a fellowship, a postdoctoral fellowship, in pharmacology for their training people to do infusions, for enzyme replacement, and all kinds of other things that have to do with that. It seems to me that maybe that's a source of funding for people to learn how to do clinical trials. The other obvious is NIH, another -- you know, and integrating it into a regular program, but I think there's more than can be integrated into a regular graduate program maybe. I don't know. I don't know; what's your thought?

DR. PANDINA: My experience is it has to come from the field, so if pharma's going to sponsor it or in any way participate, there has to be some give and take. It has to be a request from the field that it's something that's needed, and there has to be a mechanism to do it to make sense for everyone, and that mechanism is an important one.

DR. YAO: Right, which is, I think again, just to wrap up this portion, sort of why I asked the question and why we

wanted to have the workshop because we wanted to hear from folks in the field, what was really necessary to advance this science, and so, I think all of these we hope to bring forward as needs that we really do need to have addressed in order for us to appropriately move into the field of study of neurocognition behavior in pediatric patients who are going to be -- who have, you know, again, underlying diseases that need treatments that effect these areas, or that are taking drugs for other reasons that could have specific target or off-target effects that would be important to follow. So thank you very much for those insights. I think, you know, we want to move on to provide enough time. It's right at 4:30, and we want to build in a little bit of time here. Thank you for those handful that are in the room. Please recognize that there are many dozens and dozens of people online watching, so we still have a pretty significant, substantial audience that have stayed on. So, in the last half hour, I was going to hand it over to Ann who I think was going to provide some final remarks and then make one last ask of the panelists.

CONCLUSIONS, SUMMARY OF MORNING AND AFTERNOON SESSIONS AND NEXT STEPS

DR. MCMAHON: Thank you. Turn this on. I just mainly wanted start by thanking everybody that gave such great talks. I learned a huge amount yesterday and today, and all the panelists that are here, I want to thank you for being here and for coming; and the steering committee members, and I know that Elsa Shapiro and Heather Adams were participants on the steering committee and all the FDA participants. It's really been a terrific workshop. I've learned a huge amount, and I think all the talks were really fantastic, so thank you. And I just wanted to go through really briefly the topics that I would like to talk about in the last half an hour, and I'm going to ask you all to talk about them. And the topics are highlights of normal neurocognition in neonates children and adolescents; key points about animal models and identifying medical product-related, neurological signals in children; key points related to clinical examples of long-term studies in children; key points about neurodevelopmental outcomes of environmental exposures in children; and key challenges in the evaluation of long-term, neurocognitive and behavioral outcomes; and where to from here?

Really, where do you suggest that we take all this information? What's the next step? So I think what I would

like to do is go around the room and primarily focus on the non-FDA folks here, just to say here and on the panel, and just to say if you have, say, one or two or three take-home points about those topics that I have listed, what would those be? There will be a transcript, which will be public from this meeting, so we're going to be able to revisit all of what we've talked about, and I think it would be very helpful as a last point on the transcript, say what you all think are the most important things to take home from this work. So why don't we -- should we start over on this end?

DR. ELAYAN: Sure, if you want FDA people and non-clinical, so I'll just give you my perspective about this, about the non-clinical data that we collect. We feel that we are trying to model as much as a human condition and the timing and the brain development and the behavioral in animals before we go into humans, and the data that we collect. We, you know, participate with our clinical folks in the divisions and talk about these things and how they can use that information for their own purposes. And of course, we have it in the labeling, and all that needs to be conveyed to the patients and the parents, and if there are two drugs with different aspects of what effects then your behavioral, there's a choice to go for less toxic or less effects from animals. So that's our main focus and our main point of view. That's at least what I'm

thinking from non-clinical.

DR. MCMAHON: Thank you.

DR. BULL: Well, I come exclusively from the clinical standpoint, so that’s very difficult, but one of the things that was a highlight for me, I think, was the correlation -- or what I perceived to have heard was the close correlation between animal and primate, not human studies, and I'm wondering if a greater emphasis on that in terms of clinical applications would be, and not in drug trials, but in other areas perhaps would be of value. And I'd like to hear more about that, actually, in terms of what we can learn from the animal studies outcomes that we could, actually, share with our patients and utilize. I'm sure there are many other things, but that -- and also the whole issue we just discussed about training from clinical trials, we're just getting started in some of those areas, so I know I've come home with a whole list of things that I'm going to be looking for as we approach the staff that will participate. So those are my thoughts at the moment.

DR. MCMAHON: Thank you.

DR. MELLON: I found the whole conversation fascinating, and I have to be honest: this is not my area of expertise, so I’ve learned a heck of a lot. One of the things that I think is really interesting about this, at least from my perspective, is that we have a unique luxury in non-clinical

world where we can be a little bit more invasive and evaluate some of the structures in the brain that might be impacted by either a disease state or certainly in a therapeutic intervention perspective, and I do think it is extremely helpful to try to identify which structures may be impacted and how, and try to help understand from that translational perspective how that might manifest itself to help direct, perhaps, a more focused screening processes in a clinical room. We recognize the fact that the animals are not always going to behave the same way, so we're still probably going to have to throw the blank, you know, the net a little bit wide to try to evaluate it.

But I think the challenge I think for all of us is to learn from each other, going back to try to understand what the clinical data is telling us, that then can be taken back and try to better understand what the animal data is telling us as well. And so I do see this as a very significant two-way street to try to understand brain development and translation across species. The one thing I think would be my wish list that I would ask of the folks in this room and anybody online is we have a challenge in the drug world frequently where we see a histopathological finding, and we want to try to track it. The challenge frequently is, what is the right control group? And in the animal model, I have controls that are untreated, but in a

disease state, that doesn't always exist, and when you have to give an intervention like an anesthetic, there's no good control group.

What would be really helpful is if we had faith in sometimes using the individual animal as their own control. If you had a, you know, a baseline measurement and then can track it. The huge challenge that we've all discussed today is, how do you do that when you start out in a NICU, and you translate that throughout the lifetime of an individual? So one of the challenges I think of the field to me would be to try to evaluate the utility of the various interventions and tools that we have available to us for infants, for 2-to-3 year olds, to 6 year olds and 12 year olds, and see how that correlates over time. And honestly, if after a little while, we realize we're no longer seeing an impact, then we've either adapted and recovered, and it's not a problem. But trying to compare those tools across the age ranges might be really helpful to be able to leverage the baseline and values for any individual prior to an intervention to truly understand what the intervention might be doing. Not a simple task, but one that's a noble effort.

DR. BROWN: Said so well. I guess what I've heard is that there's really a lot of uniformity in opinion here, and there's no disagreements, and there's no complications in the sense of people arguing a position versus another, but what we

don't have is that fine tuning. And it's that fine tuning for each specific clinical population that's of concern that really needs greater attention, I think, in terms of what is really known, what has been done, and what can be arranged to answer these questions in a brief period of time, and which populations really could be looked at further down the road. So I was really impressed with the fact of how everyone's point of view actually came together, and that's the piece that, I think, is missing for me is this ability to -- you know, like you said, it's not just memory; it’s not -- you know, we have to break it apart into its subparts. That's what we do as neuropsychologists in terms of our domains. We need to look at that in each population to get to those answers.

DR. ADAMS: I think that --

DR. MCMAHON: Thank you.

DR. ADAMS: Yeah, than you. Thank you very much, and thank you especially because you set up my comment perfectly by observing that fine tuning is something we need to work towards and think about because I think, also, that in addition to fine tuning as much as we can with each disease, with each condition, with each exposure, with each drug or device or biologic that's being considered, there’s going to be -- there are going to be unique and thorny issues to sort out and addressing questions that the FDA needs to address or that investigators need to

address. So I think that, as much as investigators, whether it's industry, academia, and the FDA as well, from your perspective, can engage with us as early and as often as possible, just as you want investigators to engage with you as early and often as possible when it comes to thinking about determining what endpoint measure is going to be your primary endpoint for a trial. I think, you know, we may not be experts in the disease in question or the drug in question, but we might have a way in thinking about how to approach neurocognitive or neurobehavioral assessment that could complement what you're trying to accomplish. And so if you can engage with us early and often, we can be pretty helpful, I hope.

DR. DELANEY: Thanks. That was a good lead-in to what I was going to say, keeping the experts involved and carefully planning these protocols and study designs, assessing what is important to assess, and not trying to get everything about that particular child, and also thinking about ways of being flexible or including measures like you have in your research. And I guess doing all this while keeping these parents in mind and what their -- what their hopes are and their outcome and their goals for these unsuspecting [laughs] infants, toddlers, and children that we're seeing for these research clinical trials.

DR. MCMAHON: Thank you.

DR. SHAPIRO: Well, I think I'm just going to repeat

what other people have said, in a sense. I mean, things need to be disease specific; we've talked about that. We've talked about age specificity. Timing, that is a very, very important thing, and so I think we really need to pay close attention to the timing of when things occur to the child during development because that is going to really influence outcomes. And I think, you know, whether it's exposures or whether it's cancer treatment or whether it's the emergence of a disease, whatever it is, timing for the intervention is really an important factor that we need to consider. Another point is the parent involvement and the adaptive and disability kinds of approaches. We really need to incorporate that into the work that we do because I think that parents need to know, you know, that their expectations, you know, how they're being met.

And then finally, I would say that, you know, we should work toward standardized approach. I think that Dr. Sheridan asked about specific protocols, and I think that's all well and good. The only problem with having these very standard protocols is that sometimes, you know, we come up with an innovative idea, something that all of a sudden occurs to us that's different, that we've observed in a group of patients, and we want to add that to a protocol. So I think that standard protocols, but with some flexibility, is really going to be an important idea for the future. And then finally, I just want to

reiterate this training thing again, and I think we really need to focus on getting more people trained in this kind of thing because I think there are so many new treatments, innovative treatments, you know, things that are specific to genetic subgroups, a whole variety of things. We just need people who are trained to be able to handle that.

DR. MCMAHON: Thanks.

DR. AKSHOOMOFF: A couple things that I wanted to add is that it's true -- somebody mentioned earlier that for lots of treatments that are used with children, we have, unfortunately, very little data about neurocognitive effects, drugs that are directed towards changes in behavior, controlling behavior, et cetera. Unfortunately, we have limited data about the shorter term effects and certainly long-term effects, and oftentimes clinically, I have heard parents say, “Well, I don't want to start my child on that medication,” one that's widely used for ADHD, for example, or something, because we don't know the long-term effects, and you have to agree with them. Of course, you know, there are other reasons -- there are other things to consider in terms of quality of life and the long-term impact of not having the child involved in something that could potentially have a really positive effect.

The other thing that I learned today was just thinking about, even with anesthesia, about the importance of at least

having some measures of -- some aspect of neurocognitive abilities, particularly for children, you know, where maybe the medication or the treatment itself is not supposed to target, you know, improvement in terms of behavior or development, but could potentially have inadvertent negative consequences. And although on the one hand, everything that we've talked about has made it sound like that would be impossible to ever look at [laughs] quickly and easily in the course of a different trial, I don't think it necessarily has to be that way, particularly as a screening, if you take a screening approach, and have some very, perhaps, simple sorts of measures that could be followed up to determine how that effects safety and long-term outcomes for children. I talked about our work that was involved where we've been looking at normal development, and the theme has always been, in our work, the variability in normal development.

And as I think Elsa and other people have mentioned today that when you look at younger children, the general rule is that variability tends to be even a bigger concern in terms of what's normal, and what's a sign of potential difficulties? And so the other component of this in terms of helping us develop better clinical tools for -- in terms of interventions, both behavioral interventions and drugs and things of that sort, is to keep that in mind for where basic research that involves pediatric development might also be key towards -- with this

sort of information it might be useful for having the sorts of databases or information that clinical trial researchers can plug into, rather than having to try to get lots of information about normal development. And then I also would like to echo that I think having people that are pediatric psychologists, and particular neuropsychologists, with some expertise being included as part of the team for developing, or at least informing development of clinical trials. And if there really is a need for professionals like that in the field, and you feel like people could make good careers out of that, then that's important to tell directors of clinical training programs. I was just at a meeting a few weeks ago in our university where we have a clinical program with a neuropsych specialty, and, unfortunately, not too many students do pediatrics, and the director of our program said, “We need to figure out what people are going to do because not everyone's going to get an academic job; that's pretty clear.” So if you have a PHD in clinical psychology, what are some opportunities? How can we train people to be ready for the future?

DR. PANDINA: So there's not too much to say that's not been said already. First, I want to say thank you for inviting me as a sponsor from a pharmaceutical company to participate in this forum. I think it's so important to have forums like this where we can discuss, with FDA and academia

experts in the field, an area of such importance, pediatric cognitive assessment, and safety and also, in efficacy, just thinking about pediatric cognition and how we incorporate it into clinical trials. I think it also allows us to think about how that will let us understand cognitive development in pediatrics, and when it goes awry and we have to intervene, what happens, and how do we study that more generally? Because we don't always talk about that, even in the field absent of medication or intervention.

I do want to raise, and keep in mind, approaches that we have to gather the information, so new tools and technologies, how do we gather information? What's important? How do we make it easy for parents to give us the information that we need? How do we give the information back to them? Oftentimes, even if they leave with a report from a longitudinal study, say, where they come in from a clinical trial that's a sponsor-initiated clinical trial, what are the key things that they would want as follow-up? Some centers that we work with say once you're in a clinical trial with us, you stay in with us, and you get clinical care from us. Other sites say, you know, we have people that come in for clinical trials, and then we never see them again. So if there's something that we can do on the sponsor side that will facilitate continuing that relationship between you and the patient, since you're the ones

who are caring for them -- we're the ones sponsoring the trial and trying to get the data -- that would be helpful for us to know.

The other thing is, you know, thinking a bit about the trajectory of illness, someone who cares deeply about pediatric patients and wants to see more of this research done, having a pathway to understand, when can we intervene? When can we start developing medication trials? I know that there have been great dialogues about that recently in the area of autism where I have been focused the last couple of years: how can we start clinical trials in autism rather than relying on going to adults first and then extending down to the pediatric population where maybe the change we're going to make is less? You know, in a place like autism, if we're going to intervene with a drug or a drug behavior combination, then we might at a very early end of an age spectrum. And I also, I think, learned that many people are thinking about translational work: how do we bridge from animal to human? And how do we think about the tests that we're doing? I just came from a meeting on autism where that's what it's focused on, translational biomarkers, and so that that was very refreshing to me, so thank you very much for the opportunity to participate.

DR. MCMAHON: Thank you.

DR. TOWBIN: Well, you know, first, what a pleasure to

sit with very smart people and hear their thoughts about complex ideas, so I enjoyed my spectating and being a part of this. I wanted to thank particularly Lynne and Ann for making this happen because I don't think that we would all be together here if FDA hadn't kind of pulled the resources together to make it possible, and to think that FDA is worried about this and thinking about it in an active way, I think, is a huge compliment to the work that the FDA is trying to do. I am really biased when it comes to children. I think neurocognitive assessment in children is particularly important. I know the Toolbox would like us to be looking at geriatrics, but at least for right now, for all the things that we're talking about, kids are a developing organism, and the concerns about what happens during development is more crucial in pediatrics than it is throughout the lifespan, acknowledging that development doesn't stop when you're 18 or 50 or beyond.

I think that the approach of the Toolbox is the right thing. I really appreciated how it wants to take advantage of the technology, the way Dan was just saying, that the reduction in time, the integration of different measures, using well-validated, highly-reliable measures. I think that's right for these kinds of screening things, and so I really applaud that the NIH has brought that together and is trying to make that something that we can all use. I do think that we can't push

aside everything that we've learned about clinical trials in terms of the kinds of questions and the timing and why we're doing what we're doing. I think that, in some respects, it's hard to have a generic recommendation about neurocognitive assessment. I think it's really by the disorder, by the drug, by the issue, by the neuroenvironmental toxin; I just don't think we're going to say one size fits all, any more than we can say, you know, somebody comes in who's 10 years old; what's the neurocognitive test you want to give them?

It's going to depend a lot on the question and what you're evaluating. I would like to see that we could draw a closer link between what we know from animal studies, and there would be a way to look more deeply at the pre-clinical research for every agent that we use to begin to look at, what are the regions of the functions that we think could be compromised by agents? And so using that as sort of a guide for where we want to go, I think, is quite important. I think that we didn't say much about mechanisms today, but I do think that it wouldn't be hard to twist anybody's arm to say that what we're looking at when we're doing these neurocognitive batteries are, if you will, behavioral results of complex actions in the brain and that we don't yet know enough about the mechanisms.

And to the degree that I've been involved in this over my time across the way, I know that you can see the same

behavioral result with very different brain mechanisms; and so one of the things that would be so interesting is to look at people who have had an anesthesia pre, post, and see, what are the differences in doing the same tasks or something that's very similar? I also believe that, for things like autism and for a number of complex disorders, it's compensatory mechanisms that breakdown that are as important as the deficit in the single mechanism, and we don't know enough about those compensatory mechanisms. So that's useful, and then I'll just reiterate my hope that we will think more about adaptive, real-world kinds of functions. So thank you very much for letting me be here with you all.

DR. MCMAHON: I've just been told that cabs are now waiting, so I did want Dr. Waber to have the last few words if you -- thank you [laughs].

DR. WABER: Thank you. So I guess I have to be quick because my cab is waiting, but this has been really so stimulating listening to everybody kind of grappling and never being able to come up with an answer, and I think part of the problem is dealing -- is development itself, that we're dealing with a massive developmental process that's going on, and we're trying to kind of nip in at various places. But there's this huge mandate that's going on that we, you know, have to grapple with that's always kind of undercutting whatever we're trying to

do methodologically, so I think that's one place where the pre-clinical work is so important because animals have such short lifespans. And so you can ask all kinds -- do all kinds of manipulations and ask all kinds of questions that we simply can't that can inform, you know, how we look at the development -- you know, what if you give the drug to a neonate? You know, what is the relevant, biologic mechanism? I also heard a lot of measures being thrown around, and I think it's really important to be balanced about how much weight we put in the measure versus the strategy and the theory.

And you know, measures come and go based on what a publishing company decides they want to do that year. But you know, I think as scientists, we have to have our eye on the strategy and the study design and all that kind of stuff and not be too troubled by how we measure various things. I really loved the idea of having similar batteries across different things just because, even if it's not useful for the particular disorder you're studying, if you have this huge compendium of data on it, it begins to make more sense. So that's where I think the Toolbox can be incredibly useful, and I would vote for that being part of any battery and anybody who can do it. And finally, putting in a plug for functional significance, and you know, what are meaningful differences? What are meaningful, clinically meaningful kind of outcomes that we might be

measuring and not getting too hung up on thinking more about the effect size than the P-value? I guess is how I would put it.

DR. MCMAHON: Okay, thank you, and thank you to everyone. It was a wonderful two days, and safe travels. Thank you.

[applause]

MALE SPEAKER: It was a pleasure.

[applause]

(Whereupon, at 4:57 p.m., the meeting was adjourned.)

INDEX