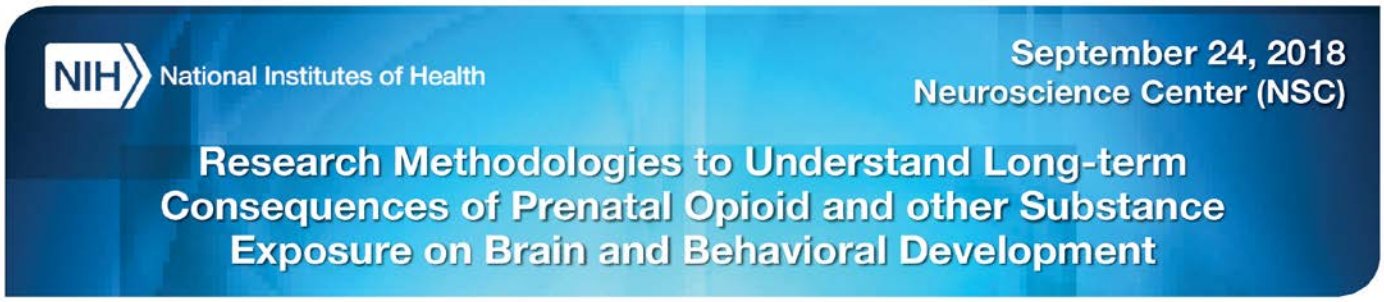

Meeting Summary

Background

On September 24, 2018, a panel of experts was convened to inform the development of a trans-National Institutes of Health (NIH) initiative that includes the National Institute on Drug Abuse (NIDA), the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), the National Institute of Mental Health (NIMH), the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute on Minority Health and Health Disparities (NIMHD), and the NIH Environmental Influences on Child Health Outcomes (ECHO) program.

The meeting's goal was to discuss research methodologies critical for conducting a large-scale longitudinal study on the effects of pre- and postnatal substance exposure and adverse environmental conditions on the developing brain and related cognitive performance and behavior, including risk for developing substance use and mental illness. This study will also establish normative brain development trajectories by recruiting nonexposed children, including those who are socioeconomically and environmentally matched.

The charge to the expert panel was given by the Directors of NIDA, NICHD, NIMH, the Deputy Directors of NIAAA and NINDS, and a Senior Program Official of ECHO.

The meeting focused on four major research areas:

- Developmental Effects of Prenatal Substance Exposure
- Measurement of Fetal/Neonatal Brain Development and Growth
- Measurement of Pediatric Brain Development
- Early Childhood Cognitive, Social, Emotional Development Assessments

In addition to these four sessions, panelists discussed other issues for consideration at a future meeting prior to summarizing the day's discussion.

Session 1: Developmental Effects of Prenatal Substance Exposure

Special measures will be required to capture the effects of prenatal substance exposure, due to the variability in teratogenic effects; types and numbers of substances used; and exposure frequency, timing, and duration. The following topics were brought up for discussion:

Teratogenic effects on neurodevelopment and other outcomes

- Physical features are good indicators of teratogenic effects—for example, there are distinct changes in facial features depending on the timing of drug exposure in the prenatal period that are associated with cognitive deficits later.
    - Three-dimensional (3D) facial imaging could be highly beneficial.
    - Two-dimensional (2D) imaging would also be useful and significantly more scalable and available in rural areas.
- There is a small increased risk for neural tube defects and oral cleft with opioid use in the first trimester.
- Alcohol and tobacco use affects growth and increases the probability of preterm delivery.
- Concerns were raised that mothers with substance use disorders (SUDs), including opioid use disorders (OUD), are much less likely to receive prenatal care and thus would be very difficult to recruit prenatally.

Considerations for the measurement of various exposures

- There will be a range of exposures in terms of substances used and the timing, frequency, and duration of use. There is a need to understand if deficits are specific to a substance or related to timing and severity of exposure.
- Adverse maternal and childhood experiences are associated with drug use and prenatal drug exposure and may strongly impact developmental outcomes. The contribution of these experiences to developmental outcomes must be understood before attributing specific outcomes to drug exposures.
- Opioid exposure is not homogeneous. There will be women who are using prescribed opioids during pregnancy, prescription opioids nonmedically, street opioids (some of which may be contaminated with other drugs), and/or opioid agonists/antagonists for treatment of OUD.
- There may also be differences among women who are using opioids to treat pain, women who are using opioids to treat pain and have an OUD, and women who do not have pain but have an OUD.
- Cannabis is becoming increasingly available and used, particularly in the first trimester to counteract morning sickness. All types of exposure need to be planned for, including accounting for differences in the potency and content of cannabis products and route of administration.
- There will be varying exposure based on ethnicity and geography, in terms of the prevalence of substances used and the doses administered.
- Polysubstance use will be prevalent, including co-exposure to other prescription drugs (e.g., selective serotonin reuptake inhibitors). Some exclusionary conditions may be necessary, but the consensus was to retain as much of the population as possible and consider using an informatics approach to disentangle effects.
- Pre-pregnancy drug and alcohol use should be queried as a less stigmatizing estimate of early pregnancy drug and alcohol. Recruitment pre-pregnancy would be very difficult and would impact cohort characteristics. It was suggested that participants be recruited during the second trimester, when possible, or shortly after delivery, if needed.
- Paternal or secondary caregiver use would also be a useful measurement and would remove some of the stigma against the mother.
- Repeated maternal interviews and biospecimen collections during pregnancy and at birth are important.

- The frequency of assessments is critical to understanding dose, timing, and duration, but if they are too dense it could impact retention.
- As neonatal abstinence syndrome (NAS) is starting to be treated differently, there may need to be new ways to measure it.
- There is a need to understand factors that may be confounded with exposure and ways to capture them.
    - Preterm birth and accurate age of neonate
    - Nutrition (e.g., during pregnancy; breastfeeding vs. formula)
    - Adverse childhood experiences (ACEs)

Biospecimens to collect

- From the mother:
    - Placenta
    - Hair
    - Blood
    - Saliva
    - Breast milk
    - Urine samples
    - Fecal samples
    - Skin swab for microbiome

- From the child:
    - Baby teeth
    - Hair
    - Cord blood
    - Blood
    - Saliva
    - Urine samples
    - Fecal samples, including meconium
    - Skin swab for microbiome

- From the father (if possible):
    - Blood for genetics

Biological and physiological measures of substance exposure or the outcomes of exposure

The Developmental Origins of Health and Disease concept should be considered. Fetal physiological measures can reveal prenatal exposures that may predispose the adult to metabolic syndromes or chronic diseases later in life.

Several measures were proposed, though this list would need to be refined to account for participants' tolerance and the utility of the measure. The measures proposed include:

- Growth trajectories in utero and afterwards, preterm delivery
- Micro-RNA profiles
- Cytokine profiles and markers of inflammation
- Oxygen and blood content in placenta
- Uterine blood flow, maternal and fetal heart rates

- 2D and 3D facial imaging
- Vocalization technology to capture caregiver tone of voice and speech patterns
- Cardiac orienting responses
- Visual processing issues with infants exposed to alcohol
- Auditory processing issues with infants exposed to nicotine
- Telomere erosion, though it is unclear if it can be measured prenatally
- Maternal methylation data, and possibly infant/child methylation data
- Fetal exosome, which could be a good predictor of later development
- Patterns of movement or motor development
- Eye movement in ultrasound

Session 2: Measurement of Fetal/Neonatal Brain Development and Growth

Assessment tools

- Imaging—Structural diffusion tensor imaging (DTI), resting state functional magnetic resonance imaging (rsfMRI), functional near-infrared spectroscopy (fNIRS), ultrasound
    - Registration of the whole uterus, contractions, oxygenation, using magnetic resonance imaging (MRI) to track all types of fetal and uterus motion
    - Placenta parameters, including how much fibrous tissue there is in the placenta, placenta consistency
    - High-resolution structural imaging from 18 to 36 weeks to get morphometric growth
    - Structural images to look at gyrus development
    - rsfMRI fluctuations in signal could indicate emerging neural activity, though this could also be due to neurovascular changes.
    - Due to the massive increase in myelination in infancy and early childhood, DTI would be highly desirable, though possibly challenging. Researchers are working on structural DTI and incorporating modeling approaches, like Gaussian mixed models, and other methods of network theory, like heat kernels, to get at more than just static information.
    - fNIRS is functionally portable, can be matched to atlases, and can be done in infants, and simultaneous fNIRS and electroencephalography (EEG) may be possible.
    - Cerebral blood flow is difficult to get in neonates; consider neurovascular coupling.
    - For the neonatal stage, consider techniques focused on network analysis, tissue microstructure, graph theory, optical techniques, and a screen for brain imaging in the neonatal intensive care unit (NICU), with the understanding that the methodologies are not fully developed for reliable measures to be compared across sites.
    - Postnatal measures are more mature, so quantitative imaging can be conducted and numbers compared at each age group without changing parameters.
    - Fingerprinting for brain imaging may be preferable to a fixed set of parameters, but this technique is still being developed.
- Central nervous system development using EEG markers of structural and functional connectivity
    - Look at phase coherence.
    - Can be done while infant is sleeping, starting from newborn to 1 month old.
- Autonomic nervous system maturation can be indexed by heart rate (HR) variability patterns.
    - Prenatal with maternal abdominal electrocardiogram (ECG), sensors in belly band to assess rhythms in the parasympathetic nervous system

    - Postnatal with HR and respiration inside and outside scanner
    - The position, sleep state, and fed state of the baby would need to be considered.
    - The utility of using respiration and HR variability to predict the sleep stage in infants was discussed.
- Ultrasound—Can get facial features better than MRI; hard to get gyrus folding pattern; used for head circumference, behavior; more dependent on the user, so hard to get reliable data.
- Wearable devices—Biological measures, sleep quality, and external environment. May also consider remote sensors to pick up environmental data.
    - Smart-shirts exist for infants that analyze child-parent interactions, parental/caregiver tone of voice, child sleep patterns, and ECG; sensors can be swapped out depending on the desired measurements.
- Machine learning and other techniques may be useful for integrating data for analysis to account for variability in sampling and measurements.
- Need to capture information from medical and developmental history (e.g., gestational age, physical measures, delivery issues).

Imaging considerations

- Frequency—From the first week to the third to fourth month, the brain changes significantly week by week, tripling in size and undergoing dramatic increases in myelination. Dense sampling would be desirable in the early stages, but must be offset by how much participant burden will be tolerated.
- Duration—30 to 40 minutes for toddlers; up to an hour for infants
- Optimization of measures for age/stage of pregnancy
- Signal-to-noise problems
- Unique infant sensitivities: e.g., noise
- Scanner issues
    - Large bore for women in the third trimester, but it is possible to use a smaller bore up to 36 weeks
    - 64-channel coil
- Standardization and harmonization; traveling phantom
- Fetal movements, working with machine learning measures to get patterns of motion, capture and quantify fetal movement. Leveraging high-quality databases may also help.

Session 3: Measurement of Pediatric Brain Development

The expert panel discussed considerations as researchers follow the subjects longitudinally. Specifically, members raised the idea that the measures will need to match the questions. For example, if the intent is to compare differences between substance-exposed and substance-non-exposed children, to predict later outcomes, or to understand mechanisms, the measures may differ. The goal is to understand the variability in developmental trajectories and how those trajectories may be altered due to substance exposure.

Measures

- T1, T2, DTI, rsfMRI, magnetic resonance spectroscopy (MRS), quantitative susceptibility mapping (QSM), others?

    - It is important to understand myelination, but there are different ways to get that data. Echo T2, structural DTI, QSM, a magnetization transfer (MT) pulse, and T1/T2 ratios were all discussed.
    - MRS takes a bit longer than a regular sequence and requires training to make sure it is acquired properly. A single brain region can be scanned in 10 minutes, while the whole brain would take 30 minutes to an hour. The amount of time required depends on the chemical being measured and the brain volume. Concern was expressed that the timing required for MRS would jeopardize collection of the other MRI measures.
- EEG has the advantage of being the same measure throughout development, scalable, and implementable in all kinds of settings.
    - EEG and resting state can be combined to pick up atypical patterns of development.
    - EEG with mother and infant when interacting
- There are some concerns about near-infrared spectroscopy (NIRS) because we do not have an extensive normative template, but a few sites could do it. The question was raised about whether a normative template was needed or whether this study could establish one.
- Visual evoked potential
- Event-related potential

Imaging issues

- Consistency of imaging modalities vs. changing modalities at different ages
    - Determining trajectory requires measures that have strong age dependence in that window.
- Frequency of imaging at different ages; would this vary by imaging modality?
    - Use dense imaging frequency early if it will be tolerated.
- Duration of imaging
    - It is unrealistic to expect imaging longer than 40-60 minutes in young children.
    - What modalities are most critical to get for every subject vs. a subset?
    - The high-risk subjects might not be able to stay as long in the scanner.
- Different brain growth trajectories depending on parental education. Need normative data across socioeconomic status as well as matching exposed and nonexposed.
- Different states of consciousness during fMRI acquisition across different ages (e.g., sleeping, watching videos)
    - Infants need to be imaged while asleep.
    - Could be easier to image young children while sleeping, but this removes the capacity for tasks and could require a significant time commitment per subject (3-4 hours).
    - Resting state is difficult to attain in young awake children, though there are some options of nonstimulating videos (Enscapes, Disney background scenes, etc).
    - Consider sleep state and phase of sleep; a concern is that not all sites have access to MRI-compatible electrodes that would be needed to perform such studies while in the magnet.
- Questionnaires should be included to assess when the subjects last ate and how they slept the night before, among other items.
- Standardization/harmonization of imaging platforms across sites for different modalities
- Motion detection and motion correction
    - It is possible to use neuroimaging of the orbits to assess gaze, as an indirect measure of eye tracking.
- Relating brain and behavior

    - Concurrent assessments, as can be done with EEG, event-related potential, and NIRS or
    - Separate assessments, as would be done with MRI
    - Relating one imaging modality to another (e.g., fNIRS to fMRI)
    - Structure and functional connectivity measures that are complemented with measures outside of the scanner

Session 4: Early Childhood Cognitive, Social, Emotional Development Assessments

The expert panel introduced work on other longitudinal studies in infants and young children to give examples of the types, time points, and frequency of measurements and assessments. This was followed by a discussion on the following topics:

Domains to be covered

- Executive function (working memory, cognitive flexibility, inhibitory control, delay of gratification, processing speed)
    - Some measures can be studied in 2.5-3-year-olds (frustration task, stress reactivity) that can be continuously used as the cohort ages.
    - Contingency learning, first few months
    - Eye tracking, can get very early behavior
    - Attention
- Social-emotional (temperament, behavioral inhibition, impulsivity, frustration)
    - The response to stress is relatively stable over time.
    - Measures of affective state
    - Vagal tone—a flatter fetal tone has been associated with worse NAS. Alternatively, one could measure the pre-ejection period via impedance cardiography, which allows for the separating of sympathetic and parasympathetic activity.
    - Emotion regulation can be assessed early on and can be examined for levels of reactivity.
    - Measures of early temperament, behavioral inhibition, and temperamental type rather than cognitive inhibition and impulsivity
    - Perception of social stimuli
    - Emotional N-back test could start as early as 5-7 years old.
- Child psychiatric symptoms (from early indicators of risk to symptoms and clinical outcomes)
    - Psychiatric outcomes
    - Social-emotional development and exposures
    - Infant-Toddler Social and Emotional Assessment (ITSEA)
    - Structured interviews with parents, and computer interviews could be integrated at later ages.
- Global Developmental Measures (NICU Network Neurobehavioral Scale [NNNS]), cognitive, language, motor development)
    - Neurobehavioral assessment at birth, NNNS. This assessment requires training, which could introduce issues surrounding feasibility.
    - Preschool psychopathology
        - The Schedule for Affective Disorders and Schizophrenia for School Age Children (K-SADS) can be used as early as preschool.
        - The Preschool Age Psychiatric Assessment (PAPA) can be used at 2-5 years old.
        - The Child and Adolescent Psychiatric Assessment (CAPA) can be used at 8 years old.

        - We will need to address the feasibility of interviewer based vs. computerized for this cohort.
    - LENA—language environment analysis
- Resilience
- Bayley vs. Mullen—The Bayley scales of infant development can be done at age 3 and is the standard, but there was some hesitation about its use and value. Mullen may be preferred, since it is broader. After the child is 3 years old, researchers may switch to the Wechsler Preschool and Primary Scale of Intelligence (WPPSI); however, it is unclear how effective these scales are for predicting future outcomes.

Caregiver measures

- Psychosocial and psychiatric assessments
- Parenting (observation)
    - It will be critical to monitor parent-child interactions with observational free play and environmental monitors, if possible; structured interviews for parent-child interaction can result in underreporting of problematic behavior.
    - Synchrony between mother/caregiver and child should be assessed.
    - Videos of parent-child interactions can be recorded—Can use machine learning techniques to support complex analyses of the data that yield interesting dynamic indices of the characteristics of the interaction, but this may require getting specific consents to use identifiable data.
    - There was discussion about enriching the sample for low caregiver quality in the nonexposed group to dissociate the effects of caregiver quality from substance exposure.

- Maternal/Paternal/Parental measures:
    - IQ
    - Possibly executive function, psychopathology, and cognitive flexibility
    - Definitely acquire from mother, and consider or attempt to collect from father or secondary caregiver, if possible.
    - Question was raised about scanning the mothers, but this is unlikely to fit within the scope of the project.
- Home and daycare measures
- Multi-informant
- Measures sensitive to developmental change

Adversity exposures

- Social class, education, stressful life events; exposures to adversity across time
    - Quality of neighborhoods, racism, and bullying
    - Family history
- Objective quantifiable measure(s) of stress/stress response to be used prospectively
    - Isoprostane levels
    - Cortisol—Collect saliva samples prior to emotion challenge; collect post-peak-arousal saliva sample
- Geocoding

Types and timing of measures

- A combination of maternal report, lab work, field work, and mobile/passive monitoring was proposed, depending on the measurement. There was strong interest in mobile and passive monitoring and a suggestion to support parental reports with biological or other objective measures.
- The possibility of staggering assessments was introduced and debated. However, it may be harder to understand individual differences if researchers stagger time points.

Cross-Cutting Topics

Design considerations

- There are practical issues concerning sample size, representation of the U.S. population, the number of sites and subjects at each site.
- Timing considerations:
    - Timing of the assessments to gestational age, real age, or developmental stages
    - Timing of the assessments to well-baby visits or other clinical visits
    - Assessment of all the subjects at all time points, or use of a subset
    - Time of day considerations
    - Staggering assessments
- A flexible assessment design was suggested to address the most critical measures for the given participant or condition.
- Consider assessing domains instead of a prescribed battery of tests. Start to bundle the battery, then scale it back to the things that have to be done.
- Consider participant burden for both the child and the family; be realistic as to the age of the subject and developmental stage.
- If technology/capabilities are not available everywhere, consider assessing on a subset of participants or using portable imaging technologies to reach high-risk populations.

Recruitment and retention

- Recruiting women while pregnant or immediately after their delivery and doing initial scanning while they are in the hospital may increase the probability of them coming back for later imaging.
- Recruit during the second trimester with few exclusionary criteria (e.g., those that will affect brain trajectories such as congenital heart defects, brain tumors), but allow for inclusion of high-risk women presenting at birth.
- Oversample for high-risk populations and for matching socioeconomic status, caregiver quality, and other conditions in the control group.
- Recruitment window should be large so changes can be made based on all of the unknowns.
- Access/transportation to adequate imaging facilities given participant locations
    - Subjects may be more likely to travel to a major research facility if transportation and accommodations are provided, and may prefer this to going to more local facilities.
    - Providing car service that has tracking may reduce no-shows.
    - It is possible to reach rural populations at hospitals in major medical centers because parents have to go there for services.
- It is very costly to do mobile scanning centers, but it can be done.
- Facilitate and utilize community partnerships; employ data collectors who live in the

communities that researchers are sampling.
- Consider how to provide feedback/service to families to increase the value of participating and improve retention.

Technology

- Wearable devices for adults and infants
- Environmental monitors to capture parent-child interactions and environmental exposures
- Mobile apps for collecting the reporting of substance use, nutrition, and other measures
- Informatics and machine learning strategies where possible to account for the variability in the sample

Other issues

- Harmonization with other large-scale studies (e.g., the Adolescent Brain Cognitive Development Study, Connectomes, ECHO) could be valuable.
- Plan for demises.

What are we missing that should be discussed at next meeting?

Legal and bioethical issues

- How does the study handle the need for treatment among participants—Embed an RCT? Build in an early intervention? Give referrals for services?
- What kinds of consent are needed to be able to use videos and machine learning paradigms to detect patterns in naturalistic behavior?
- What do we need to be aware of with respect to prosecution and loss of custody?
- What kinds of Institutional Review Board approvals are needed to follow these families when they move or are involved in foster care or child protective services?
- Are there unique concerns about data sharing among this population?
- What are the ethical implications for providing some results to participants or not?

Design issues

- How large of a sample size will we need?
- What exclusion criteria should we have, if any?
- How do we oversample to ensure adequate representation of high-risk groups, including those without prenatal care?
- How can we ensure that we obtain a geographically representative sample?
- What are the statistical implications of timing/staggering of assessments?
- Should we consider planned missingness in the design?

Conclusion

This meeting's goal was to initiate the discussion of the research methodologies to employ in a study of the long-term impact of pre-/postnatal substance exposure on brain and behavioral development. The focus of this meeting was on neuroimaging and other fetal/neonatal/early childhood assessments. Another meeting will be held on October 22, 2018, to discuss additional study design issues (sampling strategy, size, geography); recruitment and retention of high-risk populations; and legal, ethical, and social service issues. A formal Request for Information will be published in the NIH Guide to elicit further comments from the scientific and other interested communities on the optimal study design.