**Title:** Sex Differences in ASD: An Examination of the Female Behavioral Phenotype.

**Core Contributors/Work Group**: Casey Burrows PhD, Rebecca Grzadzinski PhD, Tanya St. John PhD, Kevin Donovan, BS, Julia Parrish-Morris PhD, Young Kinh-Nhue, PhD Jed Elison PhD, Joseph Piven MD

**Background**: Autism spectrum disorder (ASD) is diagnosed in 4 times as many males as females (Fombonne et al., 2009). This discrepancy increases when considering individuals with average or above IQ (Nicholas et al., 2008). Over the years, the field of ASD research has found that females with ASD are diagnosed later (potentially missed in early development) particularly when language and cognitive abilities are intact (Dworzynski et al., 2011; Giarelli et al., 2010; Wiggins et al., 2003; Shattuck et al., 2009; Goin-Kochel et al., 2014). In addition, females are more likely to receive an alternative developmental or psychiatric diagnosis (e.g., language disorder, ADHD, anxiety) instead of ASD (Hiller et al., 2015).

*It is likely that the current diagnostic criteria for ASD are male-biased and that, in order to adequately represent females with ASD, modified diagnostic criteria is needed.*

**The overarching aim of this project(s) is to extensively characterize the ASD behavioral phenotype in females compared to males.** For the longitudinal aim, we propose to identify data-driven subgroups using longitudinal data from early development to 2 years in a sample of children at familial risk for ASD.Letting the data guide groupings may inform sex-specific diagnostic and prognostic features.The Infant Brain Imaging Study (IBIS; PI = Piven), a familial risk sample, provides a unique opportunity to examine the female ASD phenotype without the potential ascertainment bias associated with community samples.

**Methods:**

We propose restricting our analyses to high familial-risk participants (those who have an older sibling with ASD). We will use data from the 6, 12, 24, and 36 month visit. Below is the data coverage for the proposed sample.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Instrument | Males | Females |
| 6 months | AOSI | 176 | 131 |
| 12 months | AOSI | 198 | 146 |
| 24 months | ADOS (Mod 1 and 2) | 221 | 160 |
| 36 months | ADOS (Mod 1 and 2) | 116 | 72 |
| First 3 time points | - | 156 | 115 |
| All 4 time points | - | 94 | 53 |
| Total Timepoints  (sum of first 4 rows) | - | 711 | 509 |

*Measures*:

We propose to examine longitudinal trajectories of observed social communication (and restricted and repetitive behaviors), as observed on the Autism Observation Scale for Infants (AOSI, 6, 12 month visit) and Autism Diagnostic Observation Schedule (ADOS, 24 mo visit). Both the AOSI and ADOS involve social interactions with a trained examiner who rates the child’s behavior in different domains. They are separate measures, but assess similar, developmentally appropriate, aspects of social communication and RRBs.

**Analytic Approach, longitudinal analyses:**

* Data reduction:
  + Option 1: identify codes that are similar across the AOSI and ADOS, determine if they function similarly at each time point to be able to combine in longitudinal analysis
    - We have ~4 codes that index social affect and are similar across measures
  + Option 2: identify codes that index social communication (and RRB, though restricted range) across the AOSI and ADOS
    - Casey still TBD which codes, but should be closer to 10-15 codes per instrument, then standardize by time point
  + Goal: create social communication (and RRB) composites that can be compared over time across measures
    - For now, we’ll focus on social communication since there are more codes, and the behaviors are easier to observe within these contexts
  + Statistical questions/considerations: how do we ensure we can compare items?
    - Could do a z-score, could use typical average as baseline
      * Big question: should we use full non-ASD sample as “typical” comparison? Or low-risk sample?
      * Sex-specific? Separate standardizations by sex? Need to look into grand-mean vs. group-mean centering?
      * Certain amount of variability we need to see in typ?
* Longitudinal analyses: we propose conducting **growth mixture modeling** to identify subgroups that vary in their intercept and slope of social communication
  + Goal:
    - Model longitudinal data based on diagnostic categories (ASD vs. no-ASD [use V24\_dx\_dummy, aka put atypical group with no-ASD group]) as a baseline model – see whether that captures longitudinal variability well
    - Identify subgroups based on data rather than diagnostic criteria – growth mixture modeling to identify latent profiles across time
      * Question of whether we should run this separately in males vs. females
    - *Note*: we should control for IQ in these analyses as well as age at each visit (most visits have a pretty tight window, but 36 mo visit had a longer possible window for the visit)
      * Will need to discuss how to do this, since we have IQ measured at each time point. Or we could create a simpler index to start
  + Statistical questions/considerations
    - Using the 36-month visit
      * Several sites only brought certain kids back for this visit, but others brought all kids (SEA, CHOP)
      * There is also more variability in age window at this visit.
      * We could potentially look to see whether data violates “missing completely at random”
      * BUT we think this will add considerable power, especially having an extra data point for longitudinal analyses
      * *Note*: we will need to decide how to model “ASD membership” variable if we include 36 mo visit – some kids were not dx’d at 24, but did receive a diagnosis at 36
        + *I would recommend just classifying by “ever dx’d with ASD”* – Casey can create this variable in new dataset
    - Imputation?
      * Shown in simulation studies to be valid when MCAR, but we only actually need MAR assumption.
        + Once you use all v’s in imputation, fact that person has missing value is random
      * Get a dramatic increase in power
        + Do HR samples have more missing data? BUT does that actually matter if we are only looking at HR
      * Module 1 vs. Module 2
        + The ADOS has slightly different activities and codes based on child’s language level, which results in them receiving a different ADOS “module” – single words gets Mod 1, flexible phrase speech gets Mod 2.
        + More kids will get module 1 at V24, but some do receive Module 2. At 36 mo, most will receive Module 2
        + If we include V36 we need to make sure we find a way to appropriately combine Mod 1 and Mod
        + Include them for now and then can re-run if we don’t include across all time points?
    - Should we model separately by sex? Or combine across sex and then see if there are subgroups that emerge in the full sample
* Examine differences in subgroups by sex, IQ, diagnostic outcome

**Implications:** The results of this work have the potential to inform modifications of diagnostic criteria for females with ASD that is both sensitive and specific. This work may also provide novel information about the “female protective factor”, suggest unique treatment targets for females with ASD, and underscore the need for deeper examinations of the biological underpinnings of ASD in females.