THE STUDY TO EXPLORE EARLY DEVELOPMENT (SEED): SEED PHASE 3

Table of Contents

FIGURES & TABLES

I. INTRODUCTION

A. Background

The Children's Health Care Act of 2000 (Appendix A) mandated CDC to establish autism surveillance and research programs to address the number, incidence, correlates, and causes of autism and related developmental disabilities. Under the provisions of this act, the National Center on Birth Defects and Developmental Disabilities (NCBDDD) at CDC funded five Centers for Autism and Developmental Disabilities Research and Epidemiology (CADDRE) in FY2001 and FY2002, including:

- Kaiser Foundation Research Institute
- Colorado Department of Public Health and Environment
- Johns Hopkins University
- University of Pennsylvania
- University of North Carolina at Chapel Hill.

CDC participated as the sixth CADDRE site.

The five original Centers were renewed in FY 2006 for another 5-year funding cycle (2006-2011). The core CADDRE activity was to implement a multi-site collaborative epidemiologic study focused on autism called the Study to Explore Early Development or SEED. The first phase of SEED (now known as SEED 1 - CDC IRB PROTOCOL #4169) was initiated in the 2006-2011 CADDRE funding cycle. A second phase of SEED (SEED 2 - CDC IRB PROTOCOL #6230) was funded under a subsequent third funding cycle (2011 – currently ongoing with completion in 2016). This phase of funding was announced as an open

competition; however, the original five CADDRE grantees received the awards. Again, CDC served as the sixth SEED 2 site.

Both SEED 1 and 2 had the same primary scientific aims: 1) to characterize the autism behavioral phenotype and associated developmental, medical and behavioral conditions of autism, with a special focus on identifying homogeneous autism subgroups for etiologic analysis, and 2) to understand the genetic and environmental risk factors for autism with a special focus on immunological, hormonal, obstetric, gastrointestinal and socio-demographic features.

A third phase of SEED, (SEED 3) will be funded and initiated in 2016. With this phase of funding, CDC funded five sites:

- University of Colorado Denver/Anschutz Medical Campus
- Johns Hopkins University
- University of North Carolina at Chapel Hill
- Washington University in St. Louis
- University of Wisconsin System, Board of Regents

Again, CDC served as a sixth site.

Project activities for SEED 3 will be ongoing from 2016 – 2021.

B. The Study to Explore Early Development (SEED): Statement of the Problem

Despite significant advances in our understanding of the clinical features of autism spectrum disorder (ASD) and substantial progress in establishing ASD prevalence studies across

multiple populations (CDC 2014), for the most part the causes of ASD remain unexplained. The most significant advance related to etiology has been recognition of the strong genetic influence on ASD occurrence, although no specific genes have been identified. The cause of ASD is probably multifactorial and one likely scenario is a multiple gene interaction possibly in response to certain environmental stimuli. The evidence from epidemiologic, neurologic imagining, and other studies indicate that the causal events leading to the disorder most likely arise prenatally.

In the face of these considerable gaps in our understanding of the causes of ASD, large population-based epidemiologic studies of ASD etiology are lacking. SEED was designed to address this critical need. The strengths of the study are:

- Case-control design with multiple-source ascertainment of case and comparison groups that include representation from diverse population subgroups;
- Confirmation of developmental status of all subjects based on standardized clinical evaluation procedures;
- Uniform subject inclusion criteria and data collection protocols applied across all study sites;
- Large sample size and study power;
    - Ability to stratify by autism symptom profile to investigate etiologic heterogeneity
- In-depth exploration of multiple research domains
    - Joint collection of genetic and environmental data

During the original SEED planning phase, selection of the research domains of interest was based on a lengthy process of review and consensus among CADDRE investigators. The process included extensive literature reviews from which written working papers were prepared for discussion and debate, and culminating with an in-person meeting to obtain final group consensus. From this process, specific research domains were designated as high priority for SEED based on the strength of their reported associations with ASD and recognition of the outstanding research gaps in each area, balanced by appropriateness of the SEED study design and feasibility of obtaining relevant data.

The original high priority research domains of interest for SEED include:

- Investigation of the ASD phenotype
- Infection and immune function, including autoimmunity
- Reproductive and hormonal features
- Gastrointestinal features
- Genetic features
- Sociodemographic features

Many of the domains are also linked by different hypothetical causal pathways leading to ASD. Of particular note, there are a number of potential cross-cutting hypotheses involving the infection, immune dysfunction/autoimmune, hormonal/reproduction, gastrointestinal, and genetic domains. Thus, one benefit of selecting multiple domains is the ability to examine not only the independent relationship between ASD and factors from each main domain of interest, but also the interaction between different, but possibly inter-related, domains.

To illustrate, the immune system, endocrine system, and brain communicate by chemical messengers (e.g., nuerendocrine stimulus by neurotransmitters, neuroendocrine and endocrine release of hormones) that transfer information to receptor end organs. Further, autonomic innervation of primary (thymus, bone marrow) and secondary (spleen, lymph nodes)

lymphoid tissue demonstrates structural links between the CNS and immune system. Immune cells possess functionally active receptors for a wide variety of neurotransmitters and hormones (e.g., catecholemines, serotonin) and immune-derived signals to the brain are largely mediated by cytokines such as interleukins and TNF alpha, e.g., immune cell derived Interleukin 1 can affect brain functions by activating noradrenergic neuronal pathways and hypothalamo-pituitary hormone release. The neurodevelopment and regulation of endocrine function has intimate links - in the CNS, as well as in peripheral tissues – with the immune system.

These interdependencies among the immune and endocrine systems and CNS may underlay important etiologic mechanisms in autism. For example, serotonin (5-hydroxytryptamine, or 5-HT) is a monoamine neurotransmitter synthesized in serotonergic neurons in the central nervous system (CNS) and enterochromaffin cells in the gastrointestinal tract. Elevated platelet serotonin is found in about one third of cases of autism. Increased production of 5-HT by enterochromaffin cells in the gut have been considered as a mechanism for elevated whole blood 5-HT levels in autism (Minderaa et al. 1987). There are very strong connections between the gastrointestinal (enteric) nervous system, the central nervous system, and the immune system. Gastrointestinal abnormalities may serve as a useful phenotypic subtype or endophenotype in genetic studies of autism. Alternatively, investigators have suggested that the reported hyperserotonemia in autism may be related to autoimmune mechanisms. For example, one hypothesis is that the serotonin binding site in myelin basic protein may function as an auto-antigen (as it has great structural similarity with the 5-HT receptor) which may lead to auto-antibody production against the 5-HT receptor.

The focus of SEED is on preconception, prenatal, perinatal and early postnatal risk factors. These factors will be considered within subgroups defined based on certain child traits, thought to be associated with possible etiologic variation in ASD. These traits include child sex and cognitive function. Biologic samples are collected from both children and biologic parents. These samples will provide DNA needed for genetic analyses as well as biomarkers of exposures and, potentially, disease.

C. The Study to Explore Early Development (SEED): SEED Phases 1, 2, and 3 Project Overview

SEED 1 was initiated by CADDRE in late 2007. Participant enrollment and data collection activities were completed for all sites by early 2012. A total of 3,899 children and their families were enrolled in this study phase; over 97% of these children (n=3769) were study "index" children (defined as children initially target for enrollment into case and control groups); the small remainder of enrolled children are siblings of index children who were enrolled for special studies Among the SEED 1 enrolled index children, ~2600 completed the full study protocol including an in person clinic assessment.

The original SEED target sample size (in part, guided by resource expectations) was deemed good for many principal analyses but was not ideal for many genetic analyses. It was also recognized to not be optimal for many autism subgroup or stratified analyses. Thus, in light of the potential sample size shortfall in the initial SEED implementation (now called Phase 1), it was decided to build on the established SEED data resource and implement subsequent rounds of SEED data collection. Therefore, SEED 2 was funded in 2011. Enrollment and

data collection began in 2012 and all data collection activities were finalized in early 2016. The final numbers of children who enrolled in SEED 2 and completed the study protocol are similar to those for SEED 1. SEED 3 will be funded in the 2016 – 2021 funding cycle to further increase the total SEED pooled sample size for investigation of high priority hypotheses and enhancing SEED's analytic potential. The goals in implementing both SEED 2 and SEED 3 were to maintain the same basic study design and general protocol integrity to ensure that data pooling can be achieved across all sites and across all SEED phases for this purpose. Modifications that were implemented in the second round of data collection included: 1) recruitment from a more recent cohort of children than SEED 1, and 2) more focused (streamlined) data collection. Modifications that will be implemented in the third round of data collection will include 1) recruitment from a more recent cohort of children than SEED 1 and SEED 2; 2) more focused (streamlined) data collection (further reductions in data collection to discontinue collection of data that are deemed sufficient in terms of sample size in the first 2 SEED phases); and 3) strategies to improve recruitment and completion rates.

The primary goal of this protocol is to describe SEED Phase 3 methods, although in most important respects the methods for SEED Phases 1, 2 and 3 are similar.

D. The Study to Explore Early Development (SEED): Overview of High Priority Research Domains in Phases 1, 2, and 3

In the original design for SEED, the CADDRE investigators collaborated on an extensive review of the epidemiology of autism, including the current state of knowledge for each of the

proposed high priority research domains (Newschaffer et al., 2007). Each of the original high priority research domains for SEED and the rationale for their selection are summarized below. Specific hypotheses in each domain to be considered for analyses are found in Appendix B.

For both SEED Phase 2 and Phase 3, it was decided to retain the original primary and secondary scientific interests of SEED, but with a somewhat narrower data collection focus. The rationale behind the narrower data collection was driven by three considerations: 1) the need to direct resources towards enhancing the recruitment process thereby boosting enrollment (and thus the need to lower data collection resources); 2) an effort to reduce participant burden thereby possibly also improving enrollment and data collection completeness (i.e., success in completing data collection on enrolled participants); and 3) certain data collected in SEED 1 only or both SEED 1 and SEED 2 are not needed in subsequent SEED phases of data collection because the primary research questions and analyses pertaining to those data items can be sufficiently addressed with sample sizes already achieved in SEED 1 and SEED 2. Ultimately, the SEED Phase 3 data collection is a balance between scientific and resource considerations and items that were retained in SEED Phase 3 are primarily either 1) necessary to characterize the ASD phenotype according to current research standards (e.g., standardized psychometric instruments administered during the clinical developmental evaluation) or 2) cross-cutting multiple SEED priority scientific areas (e.g., maternal interview, maternal, paternal, and child health history forms). Specifically, data collection for the following original SEED primary and secondary areas of interest will be reduced in SEED Phase 3:

- Some aspects of maternal health will no longer be collected for any of the children enrolled in SEED. We will discontinue collection and abstraction of prenatal and labor and delivery medical records. However, we will continue to collect fairly extensive maternal health data via the maternal interview and Maternal Medical History form that is completed by the study child's mother. Although the maternal medical record data are important for select in depth studies of maternal risk factors that cannot be fully captured via maternal interview, the medical record data previously collected during SEED 1 and 2 are deemed sufficient for these types of studies. Moreover, the medical record data collected in SEED 1+2 have been and will continue to be useful in helping us understand the validity of many data items from interviews and self-administered forms that are dependent on maternal recall.
- Some aspects of child health will no longer be collected for any of the children enrolled in SEED. We will discontinue collection and abstraction of neonatal and pediatric medical records. However, we will continue to collect fairly extensive child health data via the study child health history form that is completed by the child's mother. We will also continue to ascertain limited neonatal data such as birth weight and gestational age on study participants from the birth certificate files, and we will continue to collect data on the child's anthropometric measurements during the in person visit.
- Data collection will be greatly reduced for the majority of children recruited and enrolled in one of our three study groups -- children with other (non-autism) developmental disabilities (referred to as the other DD group or simply the DD group).

Children with an indication of a DD but without a past diagnosis or special education classification of autism or ASD were enrolled in SEED 1 and SEED 2 for several reasons: 1) these children were screened for ASD at study entry because it is known that many young children with an ASD will not yet be diagnosed, but will receive services for other more general developmental conditions such as developmental delay; thus the DD group is an important source of ASD case ascertainment for SEED (see below for full details); 2) children in the other DD group without an indication of ASD serve as an important second control group for etiologic analyses. Although our primary analyses typically assess case-population control comparisons, we also assess case-DD control group comparisons and DD group–population control comparisons to better understand whether our study exposure ascertainment might have been influenced by recall bias and to assess whether any associations we observe in our case–population control comparisons are unique to ASD or rather, are observed more broadly for many types of DDs. In SEED 3 we still need to include a DD group for case ascertainment. However, we will reduce the amount of data we collect on children in the DD group who do not screen "positive" on our initial autism screening assessment.

Given the more limited data collection in SEED 2 than SEED 1 and SEED 3 than both SEED 1 and SEED 2, a range of selected hypotheses within each research domain and the associated data collection elements are detailed in Appendix B to clarify the specific SEED Phases to which they pertain.

1. Investigation of the ASD Phenotype

One of the difficulties in ASD etiologic research arises from the substantial variability in ASD symptom severity and presentation, and in co-occurring behavioral, psychiatric, and medical conditions (Carlsson et al., 2013; Close et al., 2012; Levy et al., 2010). This phenotypic complexity is often apparent from the very early stages of development in children subsequently diagnosed with ASD. Most preschool children with ASD have more cognitive, adaptive, behavioral, and social delays than children with other DD and typical development; although constellations of symptoms may vary among individual children. Severity gradients within the two diagnostic domains of social communication and restricted interests and repetitive behaviors (RRB) also range from mild to severe and do not always follow the same pattern across domains (e.g., some children may have severe social communication deficits and few RRB while other children may have mild social communication deficits and many RRB (Georgiades et al., 2013). Reducing phenotypic diversity by classifying children with ASD into more meaningful sub-groups may thus support the search for potential causes.

It is widely accepted that both genetic and non-genetic factors are associated with the development of ASD (Bailey et al., 1995; Hallmayer et al., 2011; Miles, 2011), although distinct genetic mechanisms have been found for only 10-25% of all children with an ASD (Abrahams & Geschwind, 2008; Geshwind, 2011; Miles, 2011) and neither genetic nor non-genetic mechanisms are well understood. Genetic contributions may lead to common familial characteristics associated with autism (Piven & Palmer, 1997), such as certain personality traits and psychiatric disorders often found within the two diagnostic domains of ASD: social communication and RRB (Constantino & Todd, 2014). This leads some to suggest the

presence of a broader autistic phenotype (Le Couteur et al, 1996; Bailey et al, 1998, Sasson et al., 2013). The heterogeneity of the phenotype may also be the product of different etiologic pathways leading to ASD.

The study goals in the area of the ASD phenotype are to identify: 1) the distinctive features of children with ASD, compared to children in the other DD comparison group or general population, related to: physical traits (e.g., greater head circumference), including dysmorphic features (collected in SEED 1 only); medical conditions (such as, tuberous sclerosis, Fragile X, epilepsy, or sleep problems); developmental problems, including language delays and intellectual disability; and behavior difficulties (such as, self-injurious behaviors or attention/hyperactivity problems); 2) the distinctive features of parents or siblings of ASD children, compared to parents or siblings in the other DD comparison group and general population, related to: parental psychiatric/affective problems (such as, diagnoses of depression, anxiety disorders, or schizophrenia); medical conditions (such as, tuberous sclerosis, Fragile X, epilepsy, or sleep problems); developmental problems, including language delays and intellectual disability; and behavior difficulties (such as, attention/hyperactivity problems); and 3) discriminating features of children with ASD, with and without regression, related to: language skills, cognitive or adaptive delays, medical (e.g., history of seizures or gastrointestinal symptoms), physical (e.g., growth impairment) or genetic traits. With these data it will be possible to identify more uniform ASD symptom groups for:

- Elucidation of the complex ASD phenotype; and
- Investigation of specific risk or etiology factors.

2. Infection and Immune Function, Including Autoimmunity

Reports are common of infectious-related exposures as an etiologic mechanism (e.g., Chess, 1971; Deykin & MacMahon, 1979; Ivarsson et al, 1990; Piven et al, 1993, Atladottir et al, 2010), or immune dysfunction including autoimmunity as a feature in ASD (e.g., Singh et al, 1993; Gupta et al, 1998; Comi et al, 1999; Jyonouchi et al, 2001), but the evidence is often fragmentary or methodologically weak. Further, the infection and immune function/autoimmune hypotheses related to ASD have been inter-related. For example, it has been suggested that viral infection (as one possible mechanism) may induce immune dysfunction including autoimmunity, or autoimmune mechanisms may increase susceptibility to infection and subsequent damage to the CNS. Assuming the abnormal immunologic features are real it's unclear whether they lie on the causal pathway to ASD or are part of the ASD phenotype (Korvatska et al, 2002).

SEED will attempt to address gaps in the evidence linking an infectious exposures and the development of autism. These gaps lie primarily in the systematic documentation of both prenatal and early postnatal infectious disease history or infection-related exposures (i.e., type, timing, frequency, symptoms such as fever, and duration)

SEED will also study whether children who have autism are at higher risk for infectious diseases than children in the general population or children with other DDs.

The study goals in the area of infection are to identify whether, compared to the other DD comparison group and general population,

1) Mothers of children with ASD are more likely to experience during pregnancy or through the end of breastfeeding: a) clinical illness from infections (e.g., STDs, Group B strep), or b) different treatment histories for infectious illness during pregnancy (e.g., prescription medications such as antibiotics);
2) Children with ASD, from birth up to the 3rd birthday, are more likely to experience clinical illness from infections

SEED will also attempt to address gaps in the evidence for immune dysfunction, including autoimmunity and immunogenetics. These gaps include systematic documentation of clinical measures of immune system abnormalities in children with ASD, and their families (e.g. family histories of autoimmune disorders).

The goals in the immune function area are to identify whether, compared to the other DD comparison group and general population:

1) a) Children with ASD are more likely to have a nuclear family history of autoimmune disorders, allergies, and/or asthma – overall, by specific conditions, by specific nuclear family members, b) Occurrence of autoimmune disorder in mothers of children with ASD is time-related to pregnancy, c) Nuclear family history is present, if present, is associated with specific ASD subgroups;

2) Children with ASD are more likely to have an autoimmune disorder, allergies, and/or asthma.
3) Mothers of children with ASD are more likely to have been exposed to Rhogam or vaccines (such as influenza vaccine) during pregnancy.

Due to new discoveries and hypotheses likely to arise prior to laboratory analyses of SEED samples, no recommendations for specific immunogenetic markers have been made as yet. Possible fruitful areas that may be considered in the future include MHC genotyping, C4B (Warren et al 1991), and low resolution genotyping with respect to DR4 (Daniels et al 1995; Lee et al, 2006; Johnson et al, 2009).

3. Reproductive and Hormonal Features

One of the key epidemiologic features of ASD is the marked sex bias, suggesting that prenatal hormonal factors may play a role in ASD etiology. A variety of other prenatal characteristics such as reproductive and pregnancy complications, maternal age, and prenatal endogenous (e.g., testosterone) or exogenous steroid exposure (e.g. therapeutic medications, contraceptives) also suggest an association between ASD and prenatal hormonal features (Geschwind & Behan, 1984; Manning & Bundred, 2000; Funderburk et al, 1983; Eaton et al, 2001). The long-term developmental effect of infertility treatments, the incidence of which has risen markedly in the last decade, also warrants investigation.

Postnatal hormonal abnormalities have also been reported in children with ASD, notably hyperserotonaemia (Cook, 1990), but also abnormal levels of oxytocin and vasopressin

(Modahl et al, 1998; Insel et al, 1999), and stress (HPA axis) hormones such as cortisol and ACTH (Tordjman et al, 1997). Elevated serotonin levels have been linked hypothetically with autoimmune mechanisms in ASD and also cellular immune dysfunction (Krause et al, 2002). Further, cortisol, ACTH, and β-endorphins have been noted to lead to immune-suppression.

The goals in the area of reproductive and hormonal features are as follows:

1) Assess whether mothers of children with ASD have, compared to the other DD comparison group and general population: a) different menstrual and reproductive histories, including reproductive failure or treatment for infertility, b) different clinical course of index pregnancy, including complications, c) different patterns of exogenous hormone exposure, including infertility treatments involving hormones or contraceptive use, during the index pregnancy or through the end of breastfeeding, d) different endogenous hormone levels during the index pregnancy, indicated by clinical conditions, such as hypothyroidism, or morphologic features in the child, such as different ratios between the length of the second and fourth digits. (Manning & Bundred, 2000; Ronalds, et al, 2002; Manning et al, 2002; Hönekopp, 2012)

4. Gastrointestinal Features

It has been commonly reported that gastrointestinal (GI) symptoms are a common co-occurring feature of ASD (Barton & Volkmar, 1998Horvath et al, 1999; D'Eufemia et al, 1996; Buie et al. 2010; Cheldez et al 2014). However, there are few systematic data on the

actual type or frequency of GI disturbances in ASD children and none derived from a population-based study with comparable data for children without ASD. Also, specific gut-derived hormones (Nelson et al, 2001) that can influence the CNS (Ferezou et al, 2002), endocrine (Mazzocchi et al, 2002), and immune system (Ganea and Delgado, 2002), in addition to GI function, have been observed to be elevated in ASD children.

Many components of the GI area based on detailed parent-report of child GI symptoms and accompanying diet will be addressed by SEED Phase 1 data only, whereas analyses based on more general parent-report of GI symptoms or problems will be possible with all three phases of SEED data. The goals in the GI area are to determine whether:

1) Children with ASD are more likely to have GI symptoms than children in the other DD comparison group and general population;
2) Children with ASD and GI symptoms are more likely to have a history of regression, greater cognitive delay, and a family history of GI or autoimmune disorders than ASD children without GI symptoms, or children in the other DD comparison group and general population;
3) GI symptoms are associated with dietary patterns (SEED Phase 1 only),
    a. children with ASD are more likely to have restricted diets than children in the other DD comparison group and general population,
    b. restricted diets in ASD children are associated with specific measures of abnormal nutrient intake or behavior (e.g., temperament)

4) GI symptoms in ASD children are associated with candidate biologic markers or genes for ASD

5. Genetic Features

A unique SEED contribution to ASD genetic studies will be the attempt to assess the simultaneous contribution of maternal (parental) genotype, maternal exposures, child genotype, and child exposures. From these four sources different etiologic models including heritable factors (single gene, additive and epistatic), nonheritable factors, and their interactions can be developed. Further, a case-parent/control-parent design has been adopted to provide the greatest flexibility in investigating both main effects and interactions between the heritable and non-heritable factors of both parent and child.

The most inclusive approach to identify genetic variations that play a role in these etiologic models would be to examine all genomic variation for relationships to other etiologic factors and autism phenotypes. While this is technically not feasible, genome-wide association studies that query 1 million or more sites in the genome (per person) are possible. The SEED sample will provide an excellent opportunity to pursue such a scan. In this way, we will examine variation in the genome as a whole. The available environmental and expanded phenotype information will allow such genomic variation to be incorporated in autism etiology models in a way that has not previously been possible.

In addition, we also anticipate examination of particular candidate genotypes. Candidate genotypes to be examined include those with functional implications for the mechanisms of *a*

priori interest in this study (i.e., immune function, hormone regulation, gastrointestinal function, etc.) as well as additional candidate genotypes emerging from continuing work on linkage studies, including the planned collaboration between major multiplex family studies. Examination of candidate genotypes emerging from these efforts in the SEED sample of families unselected for genetic burden will be important in gauging the overall import of genotypes identified through family-based linkage studies.

6. Sociodemographic Features

Historically, ASD was linked with high social class (Kanner, 1943; Lotter, 1967; Rutter & Lockyer, 1967; Treffert, 1970) but studies carried out since the 1980s typically reported the occurrence of ASD across a wider socioeconomic spectrum. (Gillberg & Schaumann, 1982; Schopler et al., 1979; Tsai et al., 1982; Wing, 1980) More recently, studies carried out by California CADDRE investigators and CDC identified sociodemographic risk factors (e.g., maternal education, age, and race, and household income status) associated with ASD (Croen et al, 2002; Bhasin and Schendel, 2007; Durkin et al. 2010). Few epidemiologic studies have looked at multiple sociodemographic factors and ASD or considered their role as risk factors for the disorder or as markers for differential identification and diagnosis.

The goals for the sociodemographic area are to determine whether, compared to the other DD and POP comparison groups, children with ASD and their families have different sociodemographic characteristics and whether there is variation in ASD characteristics across sociodemographic subtypes.

7. Other Research Domains of Interest

In addition to the high priority research domains described above, SEED seeks additional information on substance use during pregnancy, the history of hospitalizations and injuries of the child, sleep disorders in the child and biologic parents, and information related to parental occupation and select mercury exposures. Specifically, the goals are to determine whether, compared to the other DD comparison group and general population:

Substance Use:

Is alcohol use, cigarette smoking and illegal substance use during pregnancy more common among mothers of children with ASD?

Hospitalizations and Injuries (SEED Phase 1 only):

1.) Do children with ASD experience more frequent physical injuries/ hospitalizations?
2.) Are there specific injuries/hospitalizations for which children with ASD are at increased risk?

Sleep Disorders:

Many components of the Sleep area based on detailed parent-report of child symptoms will be addressed by SEED 1 data only, whereas analyses based on more general parent-report on sleep problems will be possible with all three phases of SEED

Compared to children without ASD:

1.) Do children with ASD have higher frequencies of sleep disorders (overall, by type of sleep disorder)?
2.) Do children with ASD and sleep disorders have more impairment in cognition, social interaction or regulatory function?
    a. Also, more impairment in these areas than ASD children without sleep disorders?
3.) Do children with ASD and sleep disorders have more abnormalities in "clock genes" and biomarker levels such as serotonin and melatonin?
    a. Also, more abnormalities in these areas than ASD children without sleep disorders?
4.) Are children with ASD and sleep disorders more likely to have a personal or family history of anxiety disorders or sleep disorders?
    a. Also, more likely to have a history of these problems than ASD children without sleep disorders?

Select Mercury Exposures:

1) Is prenatal mercury exposure from a) RhoGAM, b) influenza vaccines higher among mothers of ASD children?
2) Are levels of mercury in hair samples higher among ASD children (SEED Phase 1 only)? The hair samples collected from study children at 2-5 years of age for the most part will represent recent exposures, unless the hair is sufficiently long to do serial measurements down the length of the hair shaft. Hair growth is estimated at

approximately 1cm/month. Depending on hair length, we may be able to estimate Hg exposure from early postnatal life especially for the youngest SEED children.

Parental occupation:

1) Are mothers and fathers (SEED Phase 1 only) of children with ASD more likely to have occupations requiring highly technical and analytic skills?
2) Are mothers and fathers of children with ASD more likely to have occupations with potential exposure to hazardous substances, e.g., pesticides, around the time of the index pregnancy?

II. Collaborators & Roles, SEED 3

SEED is a collaborative effort between the CDC, NCBDDD and the extramural CADDRE Centers. NCBDDD is also supported through contracts with Carter Consulting, Inc. and Chickasaw Nation Industries (CNI). The Carter Consulting, Inc. and CNI staff members are integrated with CDC staff and work on this and other projects on site at CDC in the Developmental Disabilities Branch (DD Branch), NCBDDD.

Karen Pazol (Team Lead, Epidemiology Team, Developmental Disabilities Branch) is the Science Lead for the CADDRE Network (also known as the "CADDRE PI") and Seema Gupta serves as Project Coordinator of CADDRE activities. Dr. Pazol is responsible for scientific oversight of the CADDRE sites and SEED project activities overall, providing direction to the collaboration. Dr. Pazol also serves as the co-PI for the CDC site -- GA SEED. Other personnel currently involved in SEED at NCBDDD include Daisy Christensen, (CDC Science Lead for the CADDRE-SEED Data Coordinating Center [located at Michigan

State University and funded through a separate contract] and GA SEED co-investigator), Aimee Alexander, (CADDRE-SEED Data Coordinator, Biomarker Analysis and Laboratory Coordinator, and GA SEED PI), Shericka Harris (CADDRE-SEED co-Data Coordinator), Lisa Wiggins (CADDRE lead clinician and GA SEED co-PI and supervising clinician), Charmaine McKenzie, (GA SEED project coordinator), Stuart Shapira, (GA SEED co-investigator), and Lin Tian (GA SEED co-investigator).

All CADDRE principal investigators worked collaboratively to develop the original SEED protocol. Multiple working groups were established by CADDRE in SEED 1 to organize and develop this large and multi-faceted case-control study. These working groups were retained in SEED 2.

In addition to the working groups, some of the original CADDRE sites (California, Pennsylvania, and Maryland) had advisory boards, including parents of children with autism, to review the original study materials and the study design. The CADDRE sites also conducted focus groups during the original planning phase and planning for SEED 2 with parents of children with and without developmental disabilities. The purpose of the focus groups was to obtain additional feedback on the study design and feasibility of the study.

Additionally, CDC, along with the CADDRE partners, established a five person peer review panel during the original study design phase. This panel consisted of experts in clinical research, epidemiology, genetics, immunology, and advocacy. Each of the panel members reviewed the SEED protocol and appendices and provided feedback to the CADDRE group.

The CADDRE PI's identified changes that were required of the protocol based on the panel's feedback and these changes were incorporated into the protocol prior to submission to the IRB. Also, in preparation for SEED Phase 2, CDC invited a panel of outside experts in epidemiology and population-based research field methods to review the SEED Phase 1 recruitment and enrollment methods and provide recommendations for improvement. These recommendations were incorporated into the SEED Phase 2 methods.

During the SEED 2 planning period (1st eight months of SEED 2 funding), all data collection instruments used in SEED 1 were carefully reviewed and revisions were made for SEED 2. While revisions were not major such that data consistency between SEED 1 and SEED 2 was maintained, many instruments underwent minor revisions whereby: 1) individual data items found to be ambiguous or otherwise uninformative during SEED 1 data collection were revised or deleted from SEED 2 instruments; 2) some new data items were added to various instruments; 3) some new instruments were created -- each combined subsets of the most important data items from multiple SEED 1 instruments; 4) one entirely new SEED 2 instrument was created; and 5) several SEED 1 instruments and examinations were eliminated entirely from the SEED 2 protocol.

The working group structure used in SEED 1 and SEED 2 and will be retained in SEED 3 to continue to facilitate collaboration. However, because the data collection protocol for SEED 3 has already been decided upon before the sites were funded, the specific working groups and functions of the working groups will change somewhat between SEED 2 and SEED 3.

Because of the need to harmonize and concatenate SEED 3 data with the data collected in SEED 1 and SEED 2, reduced CDC resources, and the need to keep the study startup burden on the SEED Data Coordinating Center (DCC) low to avoid long delays or cost overruns, CDC has stipulated that: 1) some data collection activities and instruments that were included in SEED 1 and SEED 2 will not be included in SEED 3; 2) data collection will be further reduced for some children included in the "other" DD group; and 3) the content of all interviews, examinations, and other study data collection instruments retained in SEED 3 will NOT be revised. Because DCC systems have already been established for SEED data entry and export into analytic files, in considering data collection instruments and forms for SEED 3 and the available resources for funding both CADDRE sites and the DCC, CDC decided a priori, that each data collection instrument included in SEED 2 would either be included in SEED 3 in its entirety or excluded from SEED 3 in its entirety. However, because some study SEED 2 instruments will be excluded from SEED 3, some of the ancillary SEED documents -- invitation letters and brochures and non-data collection enrollment materials – have been revised to reflect those changes.

With the above in mind, SEED 3 working groups and their responsibilities will likely include:

- **PI Working Group:** To provide general oversight of all aspects of SEED with a particular focus on establishing SEED scientific priorities that guide study design, implementation, analysis, and publication.
- **Project Coordinator Working Group:** To consider the study implementation plan and how to most practically implement the plan in consideration of the burden on the

participants. This WG will also hold regular calls to review and discuss quality control results from key instruments such as the maternal interview.

- **Data Sharing Committee:** Review all letters of intent for SEED analyses and manuscripts for publications to ensure they accurately and objectively reflect SEED and are scientifically sound.
- **Biomarker Committee**: A subgroup within the Data Sharing Committee that specifically reviews letters of intent for analyses proposing to use SEED bio specimens.
- **Clinician Working Group:** To oversee the use and quality control of psychometric assessments included in SEED. This includes considering whether, when, and how newer (updated) versions of standardized instruments used in SEED should be incorporated. This working group also reviews all of the human subjects issues related to these instruments (i.e., developmental feedback letters, consent forms, etc.).
- **Outreach Coordinators Working Group:** To promote visibility of SEED by the public, community partners and potential participants to enhance general study awareness and recruitment efforts specifically and consider all documents or messages which are provided to the participants for readability, accuracy of content, and clarity, including items such as fact sheets, public notices, invitation letter, consent forms, scripted communications, etc. This group also provides the participants with accurate information regarding the study (such as brochures, newsletters, etc. once participant has completed the study).

NCBDDD also funded a Data Coordinating Center (DCC) and a Central Laboratory and Biosample Repository (CLBR) for SEED. Michigan State University established and manages the Data Coordinating Center and Johns Hopkins University houses the CLBR for SEED 1 and SEED 2. Both Michigan State University and Johns Hopkins University are funded to manage the DCC and CLBR, respectively for SEED 3 activities.

In addition to their collaborative role in SEED development, the five extramural CADDRE principal investigators and co-investigators are also responsible for all aspects of SEED implementation at their individual sites, including obtaining all necessary IRB or other approvals from their own institution and local research partners. The PIs and co-PIs at each of the extramural sites are as follows:

- University of Colorado Denver/Anschutz Medical Campus – PI: Cordelia Rosenberg; co-PI: Carolyn DiGuiseppi; Co-investigators: Susan Hepburn, Kelly Kast, Eric Moody, Ann Reynolds, Steven Rosenberg
- Johns Hopkins University – PI: Margaret Fallin; Co-investigators: Rebecca Landa, Craig Newschaffer, Li-Ching Lee, Chris Ladd-Acosta, Heather Volk, Kelly Benke
- University of North Carolina at Chapel Hill – PI: Julie Daniels; Co-investigators: Kathleen Thomas, Rebecca Pretzel, Amy Herring
- Washington University in St. Louis – PI: John Constantino; co-PI: Robert Fitzgerald; Co-investigators: Paul Glaser, Stephen Kanne, Cy Nadler
- University of Wisconsin System, Board of Regents – PI: Maureen Durkin; Co-investigators: Maria Stanley, Traci Swink

III. SEED PHASE 3 METHODS

A. General Description/Overview

The overall purpose of SEED is to investigate risk factors for ASD and symptom subgroups of ASD, using a case-control study design that includes ascertainment of case and comparison groups that represent diverse population subgroups. As in SEED 1 and SEED 2, all six study sites for SEED 3 will implement the collaborative protocol, and common data elements across all sites will be pooled for analysis. All study participants will be drawn from children born in and residing in six study areas (SEED Phase 3 study areas are detailed in Appendix C). Study area children identified with ASD will be compared to a sample of children identified with other developmental problems, as well as a random sample of all study area children. Data collection will consist of five main components: 1) maternal telephone interview, 2) parent-completed questionnaires, 3) child developmental evaluation, 4) anthropometry, and 5) biosampling from biological parents and child.

B. Study Design

1. Study Design

An overview of the SEED 3 design is presented in Figure 1. The SEED 3 study cohort will consist of children:

- Born from January 1, 2014– December 31, 2017 (i.e., a 48-month birth interval; actual cohort dates subject to change depending on study implementation considerations) and

- Born in and currently reside in the catchment area of each site during the study enrollment and data collection period, which is 2017 – 2021 (actual dates subject to change due to study implementation considerations).

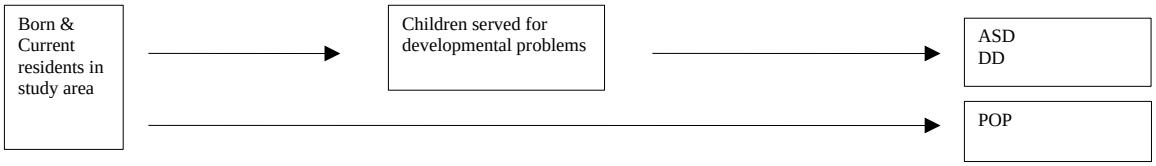


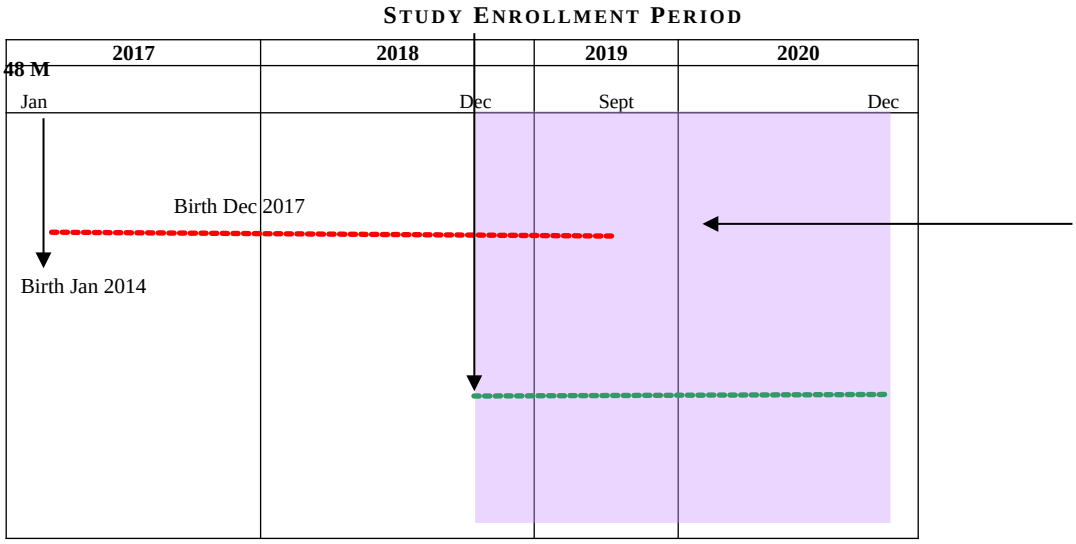
Figure 1: Overview of SEED Case-Control Study Design (Phases 1 and 2)

Case children will include children with ASD, as defined below. A sample of children who have developmental problems other than ASD will be called the other developmental delay or disorder comparison group (DD). Both ASD and DD groups will be drawn from children in the study cohort who are evaluated or receive services for developmental problems. A second comparison group of children will be drawn randomly from the entire study cohort population (POP). Comparisons between ASD and the POP comparison group are designed to identify risk factors for children with ASD relative to children from the general population, most of who are typically developing. Comparisons between children with ASD and DD may provide the opportunity to distinguish risk factors for ASD independent of factors common to other developmental problems. The DD group will also provide a means of controlling for recall bias that may be associated with having a child with developmental problems.

2. Timeline

We expect SEED 3 study enrollment to begin in 2017 and be completed in 2020 (see Figure 2 below). Due to the COVID-19 pandemic, as of July 2020, enrollment has been suspended indefinitely. Data collection might last until early 2021 in order to ensure complete collection on all enrolled subjects.

Figure 2: Timeline for SEED Phase 3 Enrollment



Assumptions

- Birth Cohort will cover up to 48 months of births: **January 1, 2014-December 31, 2017**
- Enrollment begins (i.e., contact for first enrollment) **January 2017** and will be completed by **December 2020**
- Children are typically enrolled between 36-60 months of age – this allows for data collection to be completed between 36-68 months of age (the valid age range for developmental instruments used in this study). NOTE: depending on site-specific recruitment protocols, it is allowable to enroll children prior to 36 months of age – i.e. between 24-36 months of age – with the understanding that much of the data collection could not take place until the child reached 36 months; this is expected to be a rare occurrence based on our experience in SEED 2.
- Data collection of all enrolled subjects is completed by **March 2021**

C. Study Population /Eligibility Criteria

The following criteria describe which children are eligible for SEED 3 and justifications for these criteria:

- Child is born between January 2014 and December 2017
- Child is 24-68 months old at time of enrollment (This age range, was chosen in order to limit recall bias for events in pregnancy and early life as much as possible while still allowing diagnostic accuracy for ASD and is similar to SEED Phases 1 and 2.
- Child was born in and currently resides in the study catchment area. A child is determined to have been "born in" the study catchment area based on maternal residence at birth of the child, not location of the birth hospital. The defined cohort is to be ascertained from birth certificate data; current residence is required for ascertainment purposes and to allow for examinations and other in-person assessments of enrolled subjects.
- Child currently resides with the knowledgeable biologic mother. For the purposes of this study a knowledgeable biologic mother resides with and has consistently been caring for the child since the child was 6 months of age or younger (based on self-report). This criterion is necessary in order to collect accurate information on pregnancy and early life events that may be risk factors for the development of autism. Although this criterion is somewhat more restrictive compared to SEED Phase 1 in which we required a "knowledgeable" caregiver (biological mother or other), as it turned out 97% of "knowledgeable" caregivers in SEED Phase 1 were biologic mothers.

- Legal consent is obtainable from the biologic mother. This criterion is necessary to obtain informed consent for the child to participate in the study.
- The biologic mother is competent to communicate orally in English (all sites) or Spanish (selected sites). This limitation is necessary since the clinical study instruments to be administered are only available and validated in English, although the majority of the instruments are also validated in Spanish. Sites may exclude Spanish-only speaking participants based on the percentage of Spanish-only speaking residents and other site specific factors.
- The child is not identified by the mother as deaf or blind, or unable to walk independently. Children who are reported to be deaf or blind or unable to walk independently will be excluded since they would be unable to complete the standardized instruments administered during the developmental evaluation.
- Child is not a sibling of a previously enrolled SEED 1, SEED 2, or SEED 3 participant. This limitation is necessary to avoid multiple observations with duplicate family data in the main case-control analyses.

Children must meet all of the above criteria to be eligible to participate.

Should SEED 3 selection processes simultaneously identify two or more otherwise eligible siblings, one of the siblings will be randomly selected for recruitment and enrollment. Only the child who was randomly selected will be enrolled as a research participant in the study.

If a mother with a child that is already enrolled in SEED (Phases 1, 2, or 3) later asks that a sibling also be enrolled in order to receive the developmental evaluation, sites will have the option of evaluating the sibling (using the same developmental assessment protocol as for all enrolled subjects) and providing a developmental feedback letter to the mother as a benefit to the family. However, the sibling will not be included in the SEED 3 data files as a study subject.

D. Case and Comparison Group Definitions

The birth date and birth residence of all children identified will be confirmed by birth certificate data. Current residence within the study catchment area will be confirmed prior to enrollment.

If CADDRE sites do not have accurate information about a potential participant's current residence, and therefore are unable to determine where to send invitation materials, sites will use tracing procedures to locate individuals. These procedures will primarily be the same as those used in SEED 2; they are described in Appendix D. Additionally, sites might consider other approaches such as use of multiple people search data bases and social networking sites that were not previously available to them for various reasons. (Appendix D).

In SEED 1 and SEED 2, one of the greatest barriers to recruitment has been accurately locating families so that we can send the study invitation packet. The process of tracing families at some sites currently relies heavily on identifying a phone number at which parents can be reached. Since SEED's inception, the presence of telephone land lines has decreased

and the sole use of cell phones has increased. Cell phones are currently much harder to trace than land lines because directories are not readily available. Another trend in communication has been the increasing use of e-mail and social networking sites. In a separate study being carried out by the NCBDDD, tracing staff for one site have documented several instances in which a mother's e-mail address was the only form of contact information or when a mother could be plausibly identified on a social networking site, but no other contact information was available. These instances were more common among younger mothers, a group who have traditionally been harder to trace for research purposes. Therefore in SEED 3 (as in SEED 2), we propose using e-mail and/or telephone as a means of inviting families to participate in SEED at sites where that mode of first contact is permitted. E-mail communication will be treated as analogous to telephone communication (similar script) when leaving messages on an answering machine.

1. Identification of Potential ASD and DD Children

Potential ASD and DD children are cohort children identified by the study as having a suspected ASD or other selected developmental delay or disorder (please refer to Appendix E for list of diagnoses). For the purposes of this study, ASD includes any DSM-IV-TR diagnoses of Autistic Disorder, Pervasive Developmental Disorder-Not Otherwise Specified, and Asperger's Syndrome, or a DSM-V diagnosis of Autism Spectrum Disorder. Potential ASD and DD children will be identified through sources serving or evaluating children with developmental problems; final case status (i.e., confirmed cases) will be determined from a clinical evaluation using standardized developmental measures conducted as part of the study. Sources for potential ASD and DD children may include Part C and Part B agencies, special education programs, state autism registries, hospitals, and clinics. Each SEED site will obtain

IRB approvals from their appropriate institutions and, if necessary, written agreements from their local sources. One of the SEED Phase 3 enhancements will be for sites to work more closely with their data sources to develop mechanisms to increase response to the initial invitation mailing, i.e. through active follow-up via telephone calls, email messages, or both. Site-specific case-ascertainment procedures are detailed in Appendix C.

The first step will be to identify potential ASD and DD children who meet any of the following criteria. This selection includes children with suspected ASD or other select developmental impairments based on:

- At early intervention/special educational facilities:
    - Child meets specific ASD or ASD-related exceptionality criteria from early intervention/special education services (see Appendix E for specific eligibilities)
    - Child has received an ASD or ASD-related diagnosis (if available at this educational site).
- At clinical facilities/state developmental service facilities:
    - Child has received an ASD diagnosis or
    - Child has received a diagnosis of one or more select conditions associated with ASD (Appendix E).

In addition to having at least one of the criteria above, the child must meet all eligibility criteria listed in Section II. C.

Note that the criteria described above for children with suspected ASD are quite broad and are not limited to children with a previous ASD diagnosis or autism exceptionality for early intervention/special education. This broad diagnostic net for possible cases ensures that young children with suspected ASD (i.e., young children without a formal diagnosis of ASD) are identified. However, it will also identify many children who may not have ASD. Therefore, a selection and ASD screening process will be implemented to identify potential ASD and DD children.

The purpose of the ASD screen is to identify those children in the broad net who are most likely to have ASD. The screening process will consist of administering the Social Communication Questionnaire (SCQ, Appendix F). The SCQ is a 5 to 10 minute screening interview completed by the child's mother. It is normed as a screening instrument for ASD in children 4 years and older.

Any child who has an ASD diagnosis or is receiving autism services from a public school (indicated in general by a code for autism special education services) will receive a full diagnostic assessment and be included as a case if they meet study criteria. The main purpose for the SCQ screen in this study is to identify children who might meet case criteria for SEED but do not have an ASD diagnosis and are not receiving services specific to autism from a public school. This group of children will principally be drawn from a pool of children receiving special education services for developmental delays and conditions other than autism/ASD and children with neurodevelopmental ICD-9 diagnoses other than ASD (see Appendix E).

The SCQ is considered to be the best parental screening instrument for this study. Although the published SCQ is validated on children 4 years of age and older, there has been considerable work investigating the performance of the SCQ in children younger than age 4 (Wiggins, L., Bakeman, R., Adamson, L., & Robins, D., 2007; Hanson, Sullivan, Ware, Lord, & Thurm, 2002; Corsello, Cook, & Leventhal, 2003; Eaves, Wingert, Ho, & Mickelson, 2006; Eaves, Wingert, & Ho, 2006; Baird, et al., 2006), There has also been a study conducted by the MD CADDRE site in preparation for SEED 1 (Lee, David, Rusyniak, Landa, & Newschaffer, 2007). Lee et al (2007) found that the sensitivity and specificity of the SCQ is improved by using a lower cut-off score in younger age cohorts. Likewise, Wiggins et al. (2007) found that in children younger than 4 years of age, sensitivity was maximized at a cut-off score of 11. Finally, Allen et al. (Allen, Silove, Williams, & Hutchins, 2007) reported that when using a cut-off score of 11, sensitivity and specificity of the SCQ was 93% and 58%, respectively, for children aged 2-6 years, and 100% and 62%, respectively, for children aged 3-5 years. Moreover, SCQ analyses of Phase 1 SEED data revealed that the cut-off score of 11 yielded more confirmed ASD children without increasing the rate of false positives compared to higher cut-off scores that yielded fewer confirmed ASD children.

Thus, a cutoff score of 11 will continue to be employed for SEED Phase 3. If the child scores $\geq$ 11, ASD will be considered indicated (i.e., positive screen). If the child scores <11, ASD will be considered not indicated (i.e., negative screen).

In order to obtain a systematic sample representing a range of diagnoses (for the DD group), and a range of birth cohorts (for both the DD and ASD group), it is important to try to obtain the diagnosis of each potential participant prior to contact. However, some local sources may not be able to release the diagnosis to SEED investigators without individual consent. Thus, the recruitment process and contact of potential ASD and DD children will likely vary among sources within each study site. Two recruitment scenarios have been developed to address these differences prior to first participant contact: one for sources that are able to release the diagnosis/exceptionalities of potential participants and one for sources not able to release diagnoses/exceptionalities without consent. The recruitment scenarios are discussed below and shown in Figures 3 and 4.

a. Sources with Diagnosis Known to Investigators Prior to Initial Subject Contact

Figure 3 describes the recruitment process for local sources that will release information about the diagnosis/exceptionality of the children to SEED investigators without consent (i.e., prior to first contact with a potential participant). Based on the released information from these sources, subjects are identified for recruitment following the ASD screening process:

- Children with a previous ASD diagnosis/exceptionality will be contacted, screened for eligibility, administered the SCQ, and then enrolled in the study. In previous phases of SEED, all children identified with a previous ASD diagnosis/exceptionality were contacted in order to meet study enrollment and completion targets; that is, this group of potential participants was not randomly sampled because the total number of children identified who met these criteria was fairly small. We expect this to also be true for SEED Phase 3.

- Children identified by the broad diagnostic net who do not have a previous autism/ASD diagnosis/exceptionality will be invited to participate following the eligibility and autism screening steps. At most sites, more children will be identified by the broad diagnostic net than are needed to meet enrollment and completion targets. Therefore this group might be randomly sampled in advance of sending invitation letters and/or at some point in the study when DD group targets are met, enrollment for this group might be limited to case-finding only (i.e. children identified in the "broad net" group will invited, screened for eligibility and if eligible, screened for possible ASD via the SCQ and only those who screen as positive for potential ASD will be enrolled). The proportion of the children identified by the broad diagnostic net criteria and subsequently sampled will vary among sources and within sites (see Appendix C).

As an illustration of this approach, the GA SEED site (CDC) recruitment process, is summarized as follows (details in Appendix C): Autism and developmental disability surveillance health and education sources are requested to provide lists of children who meet the SEED birth cohort eligibility and broad diagnostic net criteria. Once obtained, the various lists are compared to each other and de-duplicated and they are likewise compared to previously invited and enrolled children in SEED 1 and SEED 2 to avoid sending a duplicate invitation or an invitation to a previously enrolled sibling. The final de-duplicated list is linked to the GA birth certificate file which has been previously subset to select those births with maternal residency indicated to be in the GA SEED catchment area. Children without a birth certificate link are removed from the study invitation list because they do not meet the eligibility criterion of "born in the catchment area". All children who link to a birth record

AND who have a previous autism/ASD diagnosis/exceptionality are sent invitation letters. All remaining children who link to a birth record but do NOT have previous ASD diagnosis/exceptionality are subdivided into random samples stratified by child's birth date. Invitations are sent out in batches of these random samples (with older children being invited first) until enrollment and completion targets for both ASD and DD groups are met.

Once contact is made with the child's mother, an eligibility screen and ASD screen are administered: 1) Mothers of children with a previous autism/ASD diagnosis/exceptionality are screened for eligibility, and if eligible administered the SCQ, and enrolled in the Possible ASD Group. NOTE: all children with a previous ASD diagnosis/exceptionality are placed in the Possible ASD Group at this point of the study even if the SCQ score is <11; 2) Mothers of children ***without*** a previous autism/ASD diagnosis/exceptionality are contacted, screened for eligibility, and if eligible, they are administered the SCQ and asked whether a doctor or healthcare provider ever told her the child had autism or an autism spectrum disorder (this question is intended to capture any ASD diagnoses not already provided from study data sources, for example ASD diagnoses made by another health care provider); mothers and their children are then enrolled in either the Possible ASD group (if SCQ =>11 OR mother responds affirmatively to the question on previous ASD diagnosis) or in the DD group (SCQ <11 AND mother responds NO to question on previous ASD diagnosis).

Figure 3.

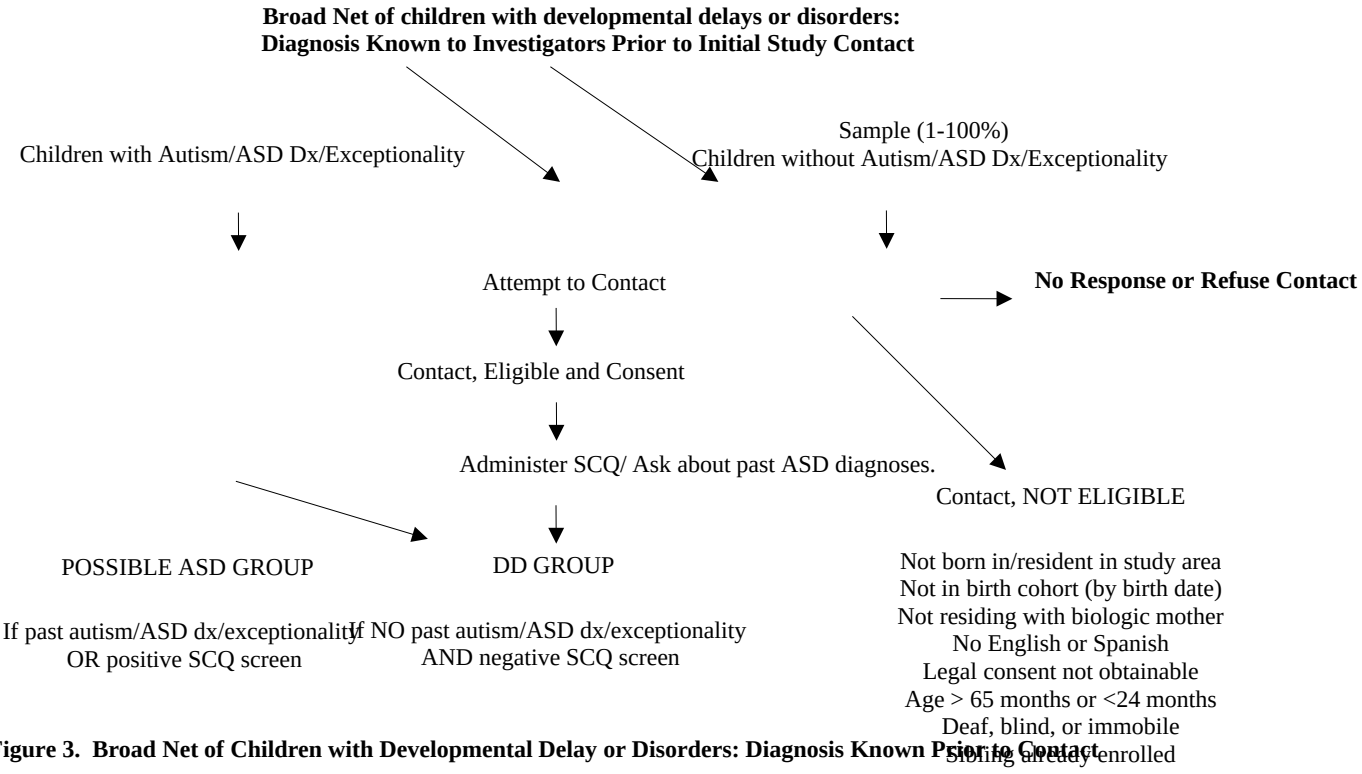


Figure 3. Broad Net of Children with Developmental Delay or Disorders: Diagnosis Known Prior to Contact

b. Sources with Diagnosis Unknown to Investigators Prior to Initial Subject Contact

Figure 4 describes the recruitment process for local sources that are unable to release the specific diagnosis type/exceptionality to the SEED investigators without consent from the potential participant. These sources have a slightly different process for determining their potential ASD and DD groups. Essentially, all children contacted are in the "broad net" until enrollment because at the time of the invitation mailing, the investigators do not know whether or not the child has a previous autism/ASD diagnosis/exceptionality. Once a child is determined to be eligible, the mother is administered the SCQ screener AND she is also asked whether a doctor or other health care provider ever told her that the (index) child has autism or an autism spectrum disorder. If the SCQ score is =>11 OR the mother answers

affirmatively to the question on previous diagnosis, the child is enrolled in the Possible ASD group. If the SCQ score is <11 AND the mother indicates there was no previous diagnosis, the child is enrolled in the DD group.

Figure 4.

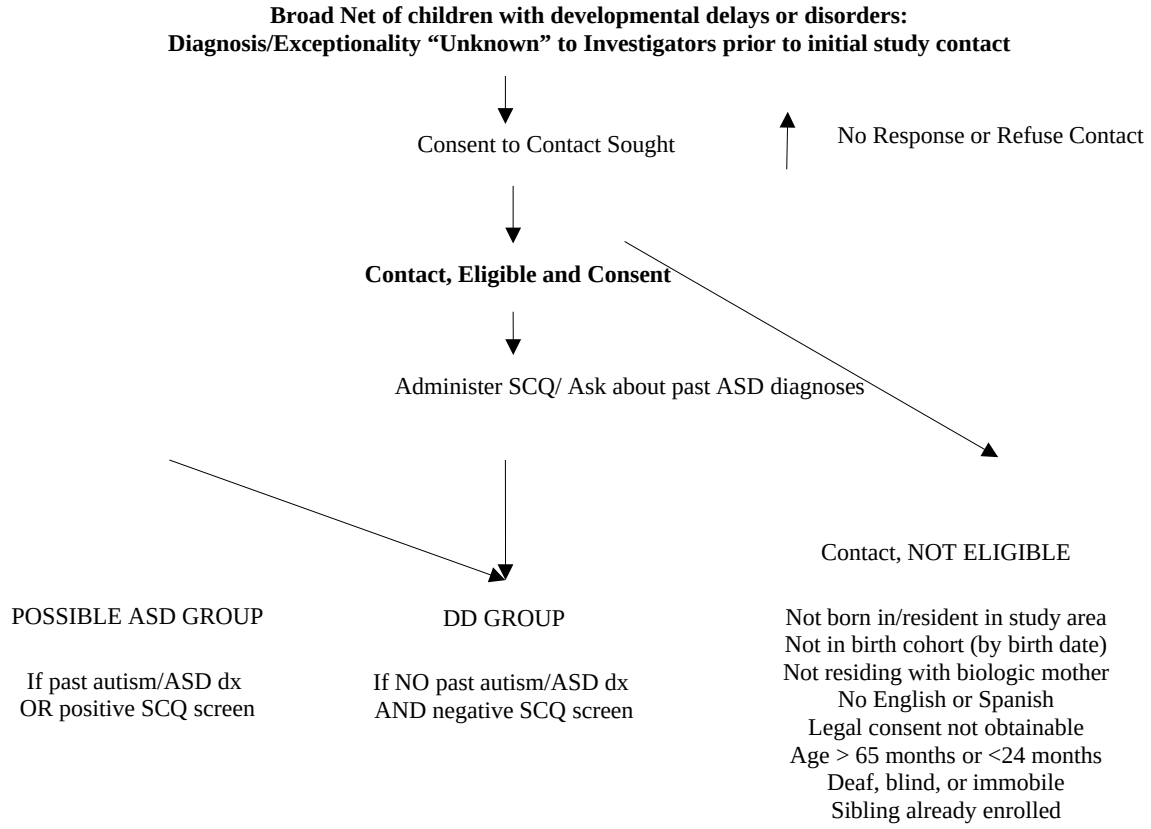


Figure 4. Broad Net of Children with Developmental Delay or Disorders: Diagnosis Unknown

c. Self Referrals

Although self referrals were accepted by some sites in SEED 1 and SEED 2, in SEED 3, most self referrals will not be allowed in order to improve the internal validity of the final SEED 3 sample. The exception to this new requirement is that if a mother contacts the study about

enrolling her child AND the study investigators are able to subsequently identify the child on the list of children already identified from health and education sources to be sent invitation letters, the child can be enrolled right away rather than going through the full invitation letter process.

2. Identification of Potential POP Children

POP children will be identified from birth certificates on the basis of birth date range and maternal residence in the catchment area at the time of birth. In addition to IRB approval, most sites will also require approval from the State Registrar or Vital Records Department to obtain files containing personal identifiers. Once approvals are obtained, potential POP member children will be randomly selected from among all cohort children to achieve a final enrollment of approximately 1,002 children (approximately 167 children per site). When possible, sites will link birth records to state death certificate files to remove any children from the contact list who are deceased (site-specific information on POP ascertainment is provided in Appendix C). The current residence in the study catchment area will also need to be established (procedures for tracing current address and contact information for potential participants are provided in Appendix D).

Based on these steps, contact will take place with each child's identified biologic mother as outlined in Section II. G, Contact and Enrollment, to determine if the child meets the study eligibility criteria and to obtain consent.

If the biologic mother consents to have the child participate in the study and the child meets the eligibility criteria, the SCQ screener will be administered, and the mother will be asked whether a health care provided ever told her the child has autism or an autism spectrum disorder. The child will be enrolled in either the Possible ASD or POP groups based on the results of the SCQ screener and ASD previous diagnosis question as shown in Figure 5.

Figure 5.

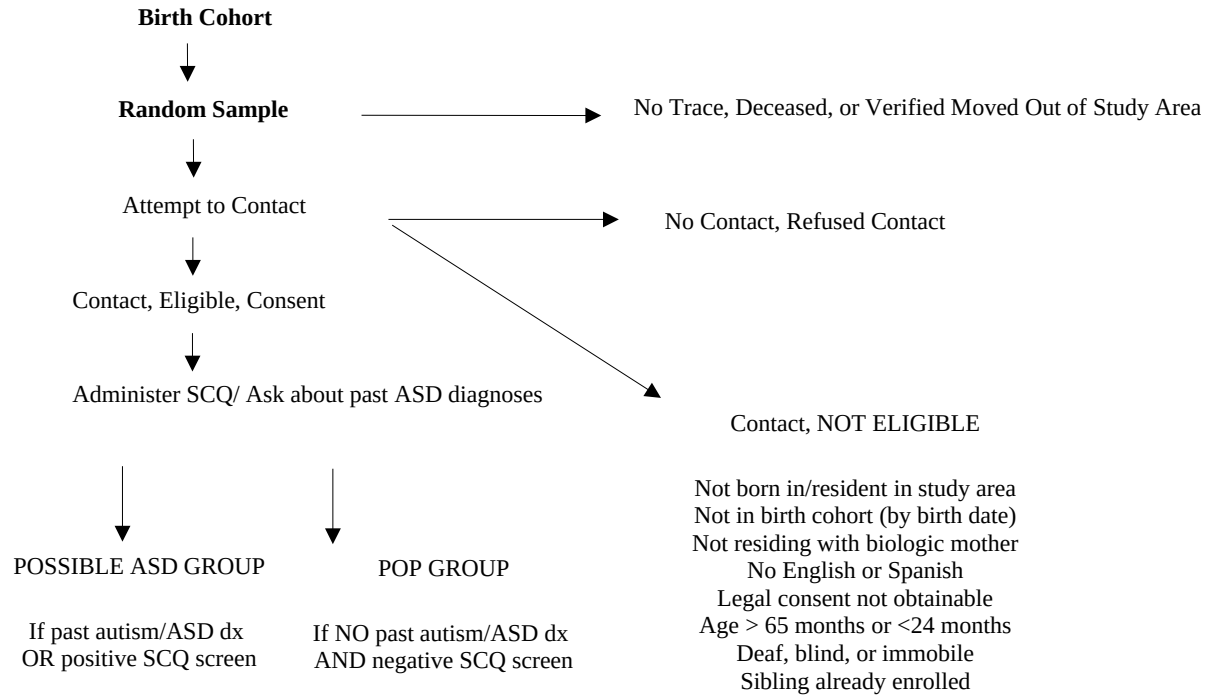


Figure 5. POP Comparison Group

3. SEED Phase 3 Site Support of Participant Identification and Invitation by Sources

One of the SEED Phase 3 enhancements is that all sites will be strongly encouraged to actively follow-up with potential participants sent mail invitations who do not respond to those invitations. This active follow-up might be via a series of phone calls, email contacts or both. In previous SEED phases some sites were only able to contact potential participants via mailings. A focus of SEED 3 will be to substantively increase response rates in all groups in order to better ensure the final sample is internally valid.

Thus, all sites are asked to work with each of their data sources -- including clinical and education sources that serve as the source for the ASD and DD groups and the state health agency that supplies the birth certificate files for identification of POP children -- to request permissions and develop mutually-agreeable strategies for some type of active follow-up. As in previous phases of SEED, the exact recruitment strategies will vary across sites. For example at some sites, sources might provide the site with the complete lists of children to include on the invitation lists and the site staff will conduct the tracing, mail the invitation letters and conduct the active follow-up onsite without the source's participation in recruitment activities. Other sites might provide staff or other support to sources to allow the tracing, invitation mailings and active follow-up to take place at the local sources.

In order to obtain permissions to conduct active follow-up via invitation phone calls and/or emails, some SEED sites might have to focus more closely on working with selected sources for ASD and DD children, rather than trying to enroll children from every source they included in SEED 1 and SEED 2. It is acceptable to limit enrollment sources; however, sites

must choose data sources that will allow enrollment of a study population that is diverse in terms of key demographics – race-ethnicity and socioeconomic status (as measured by maternal education and/or income). Thus, even though ASD and DD groups will not be selected to be completely population-based, they will still include diverse participants from key subgroups in the target population.

Another enhancement in SEED 3 is that sites are strongly encouraged to develop a mechanism and obtain necessary permissions to allow for linkage of limited birth certificate data on ***all*** participants (ASD, DD, and POP groups) invited including those who are never successfully contacted, those who refuse contact, and those contacted but found to be ineligible such that SEED investigators will be able to directly assess characteristics of respondents vs. non-respondents. Birth certificate data of interest include key socio-demographic factors – child sex, maternal race-ethnicity, maternal age at birth, maternal education at birth; of note: the latter three factors have been found to be associated with participant non-response in a recent analysis of SEED 1 data (unpublished data). The exact plan for data linkage might vary across sites. Some sites have access to the full birth records and are thus able to perform linkages between children invited from the ASD and DD source lists directly and retain all of the data onsite. Other sites might obtain more limited access. While individual-level data is preferred, aggregate data with distributions of aforementioned key variables will also be helpful in assessing potential response bias.

E. Determining Final Case and Comparison Group Status

We will perform a developmental evaluation on all children enrolled in the study and classified at enrollment as either POSSIBLE ASD or POP (See Figures 3 to 5).

Children initially identified as potential POP, who have a score of less than 11 points on the ASD (SCQ) screen and whose mothers respond negatively to the question on previous ASD diagnoses will remain in the POP group; they will undergo a limited developmental evaluation that consists of a cognitive measure only. POP children who score 11 or more points on the ASD screen, as well as all children who have a previous diagnosis of ASD or have ASD behaviors witnessed during the limited developmental evaluation, will receive a full developmental evaluation that includes a cognitive measure, an adaptive measure, and two autism diagnostic measures (full details of this evaluation are provided in section II. H., "Study Instruments and Data Collection Methods"). The developmental evaluation is crucial to ensure uniform case and comparison group designation across all study sites, thereby promoting greater confidence in multi-site pooling of data. Based on the outcome of the developmental evaluation, final status of the study subjects will be defined as outlined in Figures 6-8.

Children identified through ASD or DD sources who have documentation of a previous ASD diagnosis or autism special education exceptionality AND/OR whose mothers respond affirmatively to the question on past ASD diagnosis AND/OR who have a score of 11 or higher on the SCQ screen, will also receive a full developmental evaluation that includes a cognitive measure, an adaptive measure, and two autism diagnostic measures.

Children identified as potential DD members who score less than 11 points on the ASD screen and have no indication of a previous ASD diagnosis or autism special education exceptionality will undergo a reduced data collection protocol in SEED 3 because of resource limitations and because a large sample of DD children in SEED 1 and SEED 2 underwent the full SEED protocol and we have already gleaned much valuable information on this group. Because the DD group is such an important source for case finding, they will not be eliminated in SEED 3; rather, those who score less than 11 points on the ASD screen and have no indication of a previous ASD diagnosis or autism special education exceptionality (and thus remain classified as DD rather than being re-classified as POSSIBLE ASD) will not undergo the most costly aspects of SEED data collection – the in person developmental assessment, biosample collection, or anthropometry measurements. They will still continue to be included in the more resource efficient data collection processes – the maternal interview and the self-administered forms included in a mailing that are focused on maternal, paternal, and child health. Thus, they will still contribute to many analyses, and be particularly helpful in assessing potential recall bias.

1. Final ASD

ASD case confirmation will be established by administering the Autism Diagnostic Interview (ADI-R, Appendix G.1) to the primary caregiver, and the appropriate module of the Autism Diagnostic Observation Schedule (ADOS, Appendix G.2) to the child. Results from previous administration of the ADOS, ADI-R, or Mullen will not be accepted in lieu of testing for SEED.

- An ASD case is defined as any child who receives a complete developmental evaluation and meets the cutoff for autism on the ADI-R and ASD on the ADOS (see Figures 6-8). Note: Final ASD cases might have originated from ASD, DD, or POP samples.
  - To fall within the spectrum cutoff on the ADI-R, a child must have a social domain score ≥ 10, communication score ≥ 7 (nonverbal subject) or communication score ≥ 8 (verbal subject), and behavior score ≥ 3.
  - To fall within the spectrum cutoff on the ADOS, a child must have the minimum number of points as indicated on the most recent diagnostic algorithms (Gotham et al., 2007) according to the module administered, chronological age of the child, and language level of the child. For example, a child administered Module 1 who does not have any words must score 11 or more points in order to be considered within the ASD range.

ADOS cutoff scores to fall within the spectrum by module; ADOS WPS 2002; Gotham et al., 2007

	Minimum number of points for ASD classification
Module 1, no words	≥ 11
Module 1, some words	≥ 8
Module 2, younger than age 5	≥ 7
Module 2, aged 5 years or older	≥ 8
Module 3	≥ 7

If results are discordant on the ADI-R and ADOS, an algorithm was developed to determine final ASD status. The algorithm was developed by SEED epidemiologists, pediatricians, and psychologists who have extensive experience with evaluation and research involving children with ASD and was based on previous algorithms developed by the Collaborative Programs for Excellence in Autism, discussions with instrument authors, and an extensive review of the

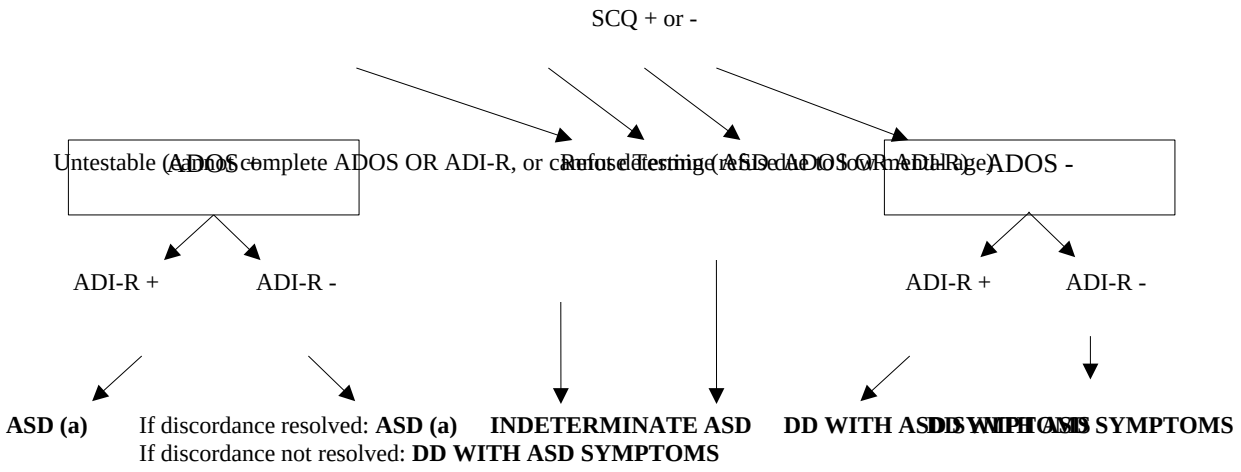
literature. The algorithm states that a child must meet ADOS criteria and one of 3 “relaxed” ADI-R criteria. The relaxed ADI-R criteria are: 1) child meets the cutoff for the social domain AND is within 2 points of the cutoff for the communication domain, 2) child meets the cutoff for the communication domain AND is within 2 points of the cutoff of the social domain, or 3) child meets the cutoff for the social domain and is within 1 point of the cutoff for the behavioral domain. If a child meets any of these discordance criteria, he or she will be classified as an ASD case. More information on SEED final classifications can be found in Wiggins, Reynolds, Rice, et al. 2015.

2. Indeterminate ASD

The Indeterminate ASD group will include any child who is targeted for a full (POSSIBLE ASD) developmental evaluation but does not complete the ADOS or ADI-R or refuses testing with the ADOS or ADI-R (rendering final classification impossible). See Figures 6-8 for final classification diagrams.

Figure 6:

Final Classification for Children with a Previous ASD Diagnosis or Autism Exceptionality



3. DD with ASD Symptoms

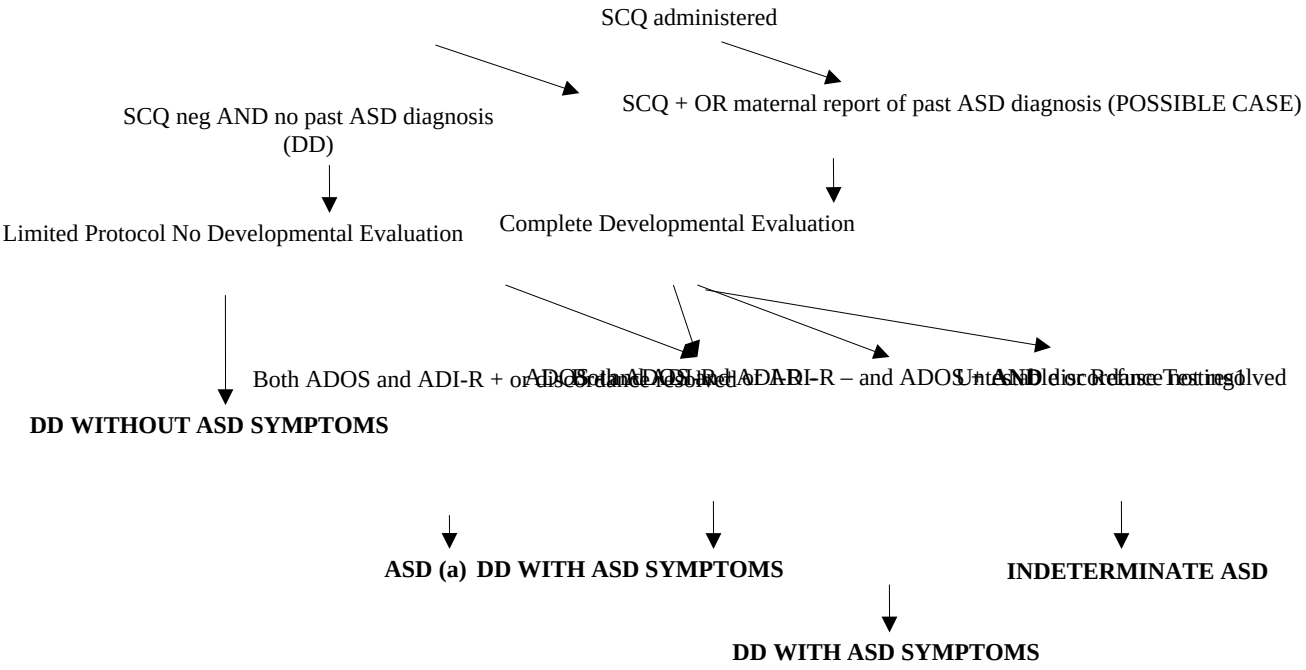
Children who are initially identified as a member of the ASD group because of a previous autism/ASD diagnosis/exceptionality and children who are initially identified as a member of the DD group but have either a positive ASD screening results (i.e., a score of 11 or more points on the SCQ) or maternal report of past ASD diagnosis will undergo a full developmental evaluation. They will be placed in either the ASD, DD, or Indeterminate ASD study groups (See Figure 6 and Figure 7) based on the results of the ADOS and ADI-R. If placed in the DD study group they will be classified as "DD with ASD symptoms" since ASD symptoms were noted on the SCQ, ADOS, and/or the ADI-R or through a past diagnosis. Analyses of Phase 1 data demonstrate that children classified as "DD with ASD Symptoms" are more phenotypically similar to children with ASD than children with other DD, which may indicate common etiologic risks.

4. DD without ASD Symptoms

Children who are initially identified as a member of the DD group who have a score of <11 on the SCQ screen and have no report of a past ASD diagnosis will not receive a developmental assessment. They will be classified as "DD without ASD Symptoms" (See Figure 7).

Figure 7.

Final Classification for Broad Net of Children with Developmental Delays or Disorders



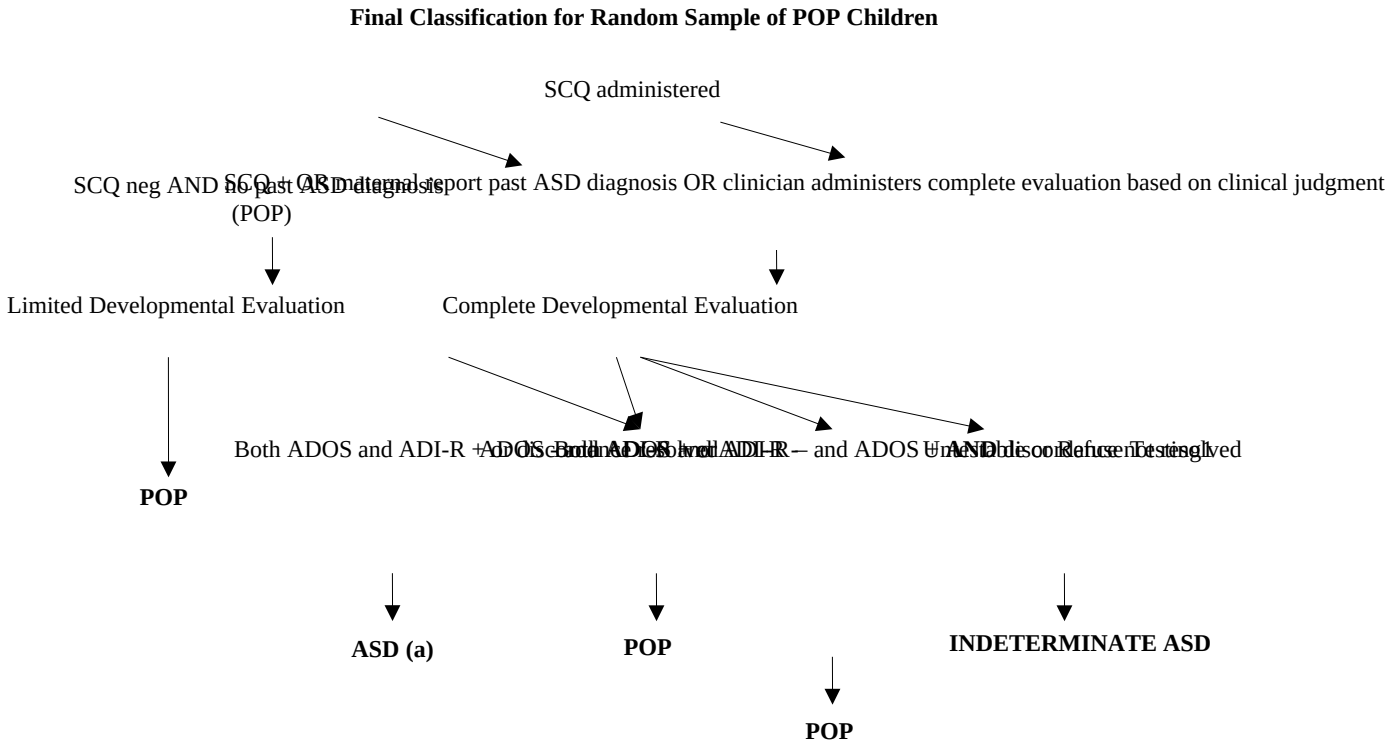
[1] Untestable and refuse testing are defined in Figure 6.

5. Final POP

A child that is sampled from the birth cohort, meets study eligibility criteria, has no maternal report of a past ASD diagnosis, and has a score of <11 on the SCQ will remain a member in the POP comparison group and undergo a limited developmental evaluation (cognitive assessment only). A child that is initially sampled as a member of the POP group who meets eligibility criteria and has a score of 11 or more points on the SCQ or has a maternal report of a past ASD diagnosis will complete the full developmental evaluation. If they meet case criteria based on the developmental evaluation they will be included in the Final ASD group (See Figure 8). Additionally, if a POP child is slated to undergo a limited developmental evaluation, but during that evaluation the developmental clinician witnesses behaviors

consistent with ASD, the protocol may be changed such that the child is instead administered the full developmental assessment and classified as an ASD case if they meet the established study criteria on the ADOS and ADI-R. All children who are originally classified as POP children and undergo the full developmental evaluation but do NOT meet ASD case criteria based on the ADOS and ADI-R, will receive a final classification of POP (See Figure 8).

Figure 8.



[1] Untestable and refuse testing are defined in Figure 6.

F. Sample Size and Statistical Power

1. Background: SEED Phases 1 and 2

The final sample sizes in each of the study groups (ASD, DD and POP) for SEED 1 and the expected SEED 2 sample sizes for those same groups are presented in Table 1. Altogether over 1400 children with ASD are expected to be included in the combined SEED 1 and SEED 2 dataset. The numbers in the DD and POP comparison groups will be slightly higher.

Table 1. Sample sizes from SEED 1 and expected from SEED 2

Final Classification	SEED 1	SEED 2 (preliminary estimate)	Total SEED 1 + 2
ASD Group	707	776	1,483
DD Group	995	813	1,808
POP Group	898	760	1,658

Note: All sample sizes pertain to the number of children who enrolled ***and*** completed the study protocol.

We expect this sample to allow us to address many important research questions within our primary and secondary research domains. However, the sample size and corresponding statistical power will not be adequate for many analyses of rare exposures, ASD subtypes, and genetic associations.

Table 2 provides data on the range of prevalence estimates we expect to observe among the SEED POP group for the types of exposures and child health conditions that will be assessed in SEED; these data are based on past population prevalence estimates. The examples highlighted in the Table illustrate that both rare and common maternal exposures and child health conditions will be examined in SEED.

Table 3 provides sample size estimates for various types of ASD subgroups of interest. These data highlight that even though we expect to achieve a sample size of close to 1,500 ASD cases by the end of SEED 2, our sample sizes for some important ASD subgroups might be as low as 300 or less.

Table 4 presents estimates of sample size needed for analyses based on various scenarios of exposure/health condition prevalence and the strength of the association between ASD (or ASD subtype) and exposure/health condition. All calculations assume 80% power, 5% alpha error, and a 1:1 ratio cases and controls. The odds ratio values used in these calculations – 1.50, 1.75, 2.50, and 3.00 -- are based on typical findings from previous studies of various risk factor-ASD associations. The exposure prevalence values used in these calculations match those expected for many maternal exposures and child health conditions being assessed in SEED (see Table 2). The findings indicate that the combined SEED 1 and SEED 2 sample with 1,483 expected ASD cases will likely be sufficient to examine associations between ASD and very prevalent exposures/health conditions (10% or higher), even when the magnitude of the association is fairly low (OR=1.5). Additionally, the SEED 1+2 sample will be sufficient to examine associations between ASD and exposures/health conditions that are strongly associated with ASD (OR=2.5 or 3.0), even when the prevalence of the exposure/health condition is low (1%). However, the sample size from the first two SEED phases might not be sufficient to examine all exposures with a prevalence of 1–5%, that are more modestly associated with ASD (OR<2.0). Moreover, analyses of associations with ASD subgroups, with much smaller sample sizes than the ASD group overall, will be fairly limited.

Table 2. Prevalence of select SEED candidate exposures and child health conditions and expected sample sizes.

POP exposure	Expected number in SEED 1+ 2	Example exposures and child health conditions

prevalence estimates	ASD group (N=1,483) with exposure or health condition if NO association (null hypothesis is correct)	included in SEED
1%	15	Assisted reproductive technology use for index pregnancy; Select maternal chronic conditions such as pre-existing diabetes; Select child conditions such as epilepsy
2%	30	Select maternal infections in pregnancy such as pyelonephritis or sexually transmitted diseases; Select maternal chronic conditions such as thyroid deficiency.
5%	74	Select past infertility conditions such as polycystic ovary syndrome (PCOS); Use of hormonal infertility treatments (all types combined) for index pregnancy; Select child health conditions such as asthma
10%	148	Any maternal autoimmune condition/allergy (all types combined); Select child conditions such as ADHD; GI dysfunction in children
20% or more	297+	Maternal labor induction or stimulation with Pitocin; Cesarean delivery; Maternal fever in pregnancy; Any maternal infection in pregnancy (all types combined)

Table 3. Sample size estimates for example ASD case subgroups of etiologic interest.

Proportion of full ASD group	Expected number in ASD subgroup for SEED 1 + 2 (based on expected total sample of 1,483 in ASD group)	Subgroup
20%	297	Complex autism
30%	445	Nonverbal; with regression
40%	593	With intellectual disability
60%	890	No intellectual disability
70%	1,038	Verbal; without regression
80%	1,186	Essential autism

Table 4. Sample size calculations for analyses under various assumptions of exposure prevalence and odds ratio.

Exposure Prevalence (%)	Odds Ratio	Sample Size Needed in ASD Group or ASD subgroup
1	1.50	7,964
2	1.50	4,041
5	1.50	1,690
10	1.50	912
1	1.75	3,900
2	1.75	1,983
5	1.75	834
10	1.75	454
1	2.50	1,247
2	2.50	637
5	2.50	272
10	2.50	151
1	3.00	804
2	3.00	413
5	3.00	177
10	3.00	100

All calculations assume 80% power, 5% alpha error, 1:1 ratio cases and controls.

2. SEED Phase 3

The goal of SEED 3 is to increase the sample sizes of the ASD, DD, and POP groups. The a priori targets for SEED 3 are that each of the six sites will add:

- 117 children with a final classification of ASD and complete study data (defined below)
- 117 children with a final classification of POP and complete study data
- 117 children with a final classification of DD and complete study data (defined below – of note: since the study protocol for the DD group will be streamlined, the definition of complete study data for this group will be different from the other two groups).

Overall, the expectation is that SEED 3 will add 702 children to the ASD group and each of the two comparison groups. Thus, after SEED 3 we expect to have a sample of 2,100 or more

children in the ASD group. This will expand our ability to analyze rare exposures and/or modest (yet scientifically important) associations. For example with the SEED 1+2+3 sample we expect to be better able to assess associations between ASD and maternal exposures/child health conditions in the following scenarios:

- Exposure/health condition prevalence 5% and expected odds ratio 1.50;
- Exposure/health condition prevalence 2% and expected odds ratio 1.75.

Also, we will substantially increase our ability to examine ASD subgroups in depth and to explore potential effect modifications between various ASD risk factors.

In setting enrollment and completion targets for all phases of SEED, we focused on the ASD group, since this is the group for which the number of children available in each site's target population is most limited. For SEED Phase 3 planning we assumed the following:

1) A similar or slightly larger birth cohort size for each site's catchment area as in SEED Phase 2. For SEED 3, CDC specified that each site's catchment area include a minimum of 40,000 births per year;
2) 15% out-migration of children born in the study areas (assumes an annual out-of-county move rate of 3% to 4% each year from birth to age 4);
3) An underlying population prevalence of ASD near 1% (a fairly conservative estimate based on recent data from Autism and Developmental Disabilities Monitoring Network [ADDM] for both 4 year olds and 8 year olds [NOTE this estimate does not assume that all cases of ASD in children will have been diagnosed by the time of SEED recruitment, but does assume that most children with an ASD will have come to the attention of a school or health care provider because of some type of developmental delay]);

4) Each site recruits SEED case participants from multiple data sources that cover at least 50% of the children with ASDs residing in the site study catchment area (again including children not yet identified as having an ASD diagnosis but being served by a clinical or education source for some DD code included in the broad net);
5) 50% of invited children will be successfully contacted (higher than observed at some sites in SEED 1 and SEED 2 but expected based on enhancements added to SEED 3 recruitment methods);
6) 70% of those contacted will be eligible to be included in SEED (estimate in line with our experience for potential ASD/DD participants in SEED 1 and 2);
7) 85% of those determined to be eligible will be enrolled in SEED (conservative estimate based on our experience for potential ASD/DD participants in SEED 1 and 2);
8) 70% of those enrolled will complete the SEED data collection protocol (estimate in line with our experience for ASD/DD participants in SEED 1and 2).

Based on those assumptions, we estimate that each site is able to enroll 203 case children and 142 of these children will complete the study protocol (See Table 5); thus based on this scenario, sites should be able to meet the a priori target of 117 ASD with complete study data. Of note: this calculation assumes a fairly low number of births per year (40,000) and that the site would only be working with data sources that cover half of the children from this birth cohort who remain in the area and have ASD. If a site had more births in their target area, but worked with sources that covered a lower percent of ASD cases in the population or a site had fewer births in their target area, but worked with sources that covered a higher percent of ASD cases in the population, it is possible to identify and enroll a similar number of children.

For example, if a site has a birth population of 50,000 per year but only works with sources that cover 40% of the population and all other assumptions remained the same, the site would be able to enroll the same number of children as detailed in Table 5. Likewise if a site only has a birth population of 35,000 per year but works with sources that cover 60% of the population and all other assumptions remained the same, the site would be able to enroll slightly more children than under the scenario presented in Table 5.

Table 5. Likely sample available for enrollment in ASD group per site.

Parameter	Assumption	Number of children in sample after applying assumption
Total Birth Cohort / site	Minimum 40,000 / year; 160,416 for full 4-year period	160,416
% out-migration by time of study contact	15% out-migration/85% remain in study area from birth to age 4 years	136,353
ASD prevalence (including not yet identified with ASD but being served for other DD)	1%	1,364
% potential ASD sources in community included in SEED case-finding	50%	681
Contact rate among invited children	50%	341
Eligibility rate among those contacted	70%	239
Enrollment rate among those eligible	85%	203
Study completion rate among those enrolled	70%	142

G. Contact and Enrollment

Recruitment and enrollment contact with study participants will consist of: 1) mailed invitation packet via United States Postal Service (USPS), 2) follow-up recruitment phone calls and/or emails, and 3) mailed enrollment packet. Please see Appendix H for a study flow chart. This section describes the initial contact through the Invitation Packet, the invitation

response mode options (i.e., options for indicating interest in or declining participation), the Telephone Screening for invitees who gave a positive response to the initial invitation or gave no response, and the subsequent enrollment packet following the Telephone Screening. The consent process is a multi-step process that begins with the Telephone Screening call, and is described in detail in Section II. I., Consent Process.

Due to the COVID-19 pandemic, as of July 2020, all further invitations and enrollments have been suspended indefinitely due to the inability to maintain safety and infection control measures during in-person developmental assessments. Families who have been sent invitations but have not yet been enrolled will be notified of this change. Because of the inability to maintain safety standards, materials have been developed to inform invitees who have already returned a response card expressing potential interest, as well as invitees who have not yet responded (Appendix I.5).

1. Invitation Packet

Following identification of a potential child participant by a local source, initial contact will be an invitation packet that is mailed to the primary residence of the prospective participant and will be directed to the child's biologic mother. The invitation packet will be sent from the study site (e.g., CDC) or collaborating local sources (e.g., Fulton County Department of Education) depending on the agreements each site has with their various sources. All invitation packets will be sent in English, or English and Spanish (selected sites only) For sites including Spanish speakers, the materials will be sent in both English and Spanish and the response card included in this packet will ask the mother to indicate language preference,

so that all future materials can be mailed in the appropriate language. The invitation packet (Appendix I) sent via USPS will include:

- A letter introducing the study and including (Appendix I.1):
    - Brief statement of the study purpose
    - A brief description of the what the participant will be asked to do
    - Description of monetary incentives
    - Brief statements regarding the voluntary nature of the different study parts and protection of participant confidentiality
    - A description of the ways the invitee can respond to the invitation: return the enclosed pre-paid response card, telephone, send a text message, or email the study site.
    - A statement informing the prospective participant about potential follow-up contact (this will be site- and source-specific)
- Study brochure (Appendix I.2). A SEED brochure will be included in all packets (which may include an optional site-specific fact page). The brochure provides a brief overview of the purpose of SEED and children being enrolled in SEED, and specific information about the participant activities and incentives. Because the study protocol will vary for children in the ASD, DD, and POP workflow groups (see below), three versions of the brochure have been developed; each includes the same language about the overall purpose of SEED and types of children being enrolled, but the language regarding study steps and incentives is tailored to reflect the source group from which a given child was originally identified (potential ASD, DD, or POP).

- A response card and pre-paid envelope that the invitee can return indicating interest or non-interest in future contact to learn more about the study (Appendix I.3).
- Some sites may also use additional invitation materials:
    - A site-specific incentive (e.g., magnet, key chain, small toy) valued at approximately $1.00 (current federal guidelines for use of federal funds will be followed in deciding whether to include this type of incentive in the invitation mailing).
    - Postcard sent in advance of the Invitation Packet. Some sites may send a brief postcard style mailing to potential participants in advance of the Invitation Packet to serve as a "heads up" to the recipient to look for a future mailing that includes the full Invitation packet materials. This option might increase participant response if the larger Invitation Packet is discarded without being read. The postcard may also be used as an alternative to the full second or third Invitation Packet mailings. Each site that uses this option will need to get approval from local IRB.

If a valid mailing address for a potential invitee is not obtainable or is for some reason questionable (e.g., initial mailing to last known address is returned to SEED site as undeliverable; or the answering machine message on follow-up telephone calls indicates a business or a family name other than that of the invitee; or contact with persons at the most recent address indicates the invitee no longer lives at that location) and an email address is obtained through tracing, then a brief invitation email will be sent to the biologic mother. The email message will only include text inviting the woman to email, text message, or telephone the SEED site in order to be provided with more information about the study and an invitation to participate (Appendix I.4). If the woman responds indicating interest then follow-up by

study staff with study information and an invitation to participate (equivalent in content to the invitation packet) will occur via her preferred mode of contact (email, telephone, regular mail).

a. Response to Invitation Not Received

If none of the response mode options are used by an invitee within two to six weeks of the invitation mailing, sites (in concert with their agreements with their data sources) will follow-up with the potential participant. The exact protocol for non-response follow-up will vary by site. At some sites, repeat mailings might be included as part of the follow-up protocol. However, all sites are asked to negotiate with their data sources to allow for a more active follow-up protocol (via phone calls and/or emails) as well.

One of the SEED Phase 3 enhancements is that all sites are strongly encouraged to actively follow-up with potential participants sent mail invitations who do not respond to those invitations. This active follow-up might be via a series of phone calls, email contacts or both. In previous SEED phases some sites were only able to contact potential participants via mailings. A focus of SEED 3 will be to substantively increase response rates in all groups in order to better ensure the final sample is internally valid.

Thus, all sites are asked to work with each of their data sources -- including clinical and education sources that serve as the source for the ASD and DD groups and the state health agency that supplies the birth certificate files for identification of POP children -- to request permissions and develop mutually-agreeable strategies for some type of active follow-up. As in previous phases of SEED, the exact recruitment strategies will vary across sites. For example at some sites, sources might provide the site with the complete lists of children to include on the invitation lists and the site staff will conduct the tracing, mail the invitation letters and conduct the active follow-up onsite without the source's participation in recruitment activities. Other sites might provide staff or other support to sources to allow the tracing, invitation mailings and active follow-up to take place at the local sources.

If the participant is contacted and indicates interest during this follow-up phone call or email, study staff will follow the same procedures as if the participant had returned a response to the invitation packet indicating interest in the study (see Screening and Invitation Phone Call, section 2. below). A maximum of 9 phone calls/emails will be made in an attempt to follow-up with the potential participant. At some sites the number of follow-up attempts might be less, depending on their agreements with data sources. Telephone calls will be attempted at

various times during the day and different days of the week. If telephone is the only mode of contact, up to seven telephone messages will be left for the potential participant. While only 3 telephone messages were left as part of SEED Phase 1, this was changed to seven telephone messages in SEED Phase 2 because caller-ID had become so pervasive on both landlines and cell phones, we believed that NOT leaving messages might have been concerning to some participants who repeatedly saw a call from the same number. Thus, beginning in SEED 2 we leave a short, friendly telephone message after most call attempts in order to tell the participant who we are, and after several call attempts, to reinforce that the study is voluntary, and that even if they don't want to participate in the study, we would appreciate their participation in the short eligibility screener. If after 9 attempts, the potential participant is not contacted, no further attempts to contact will be made. If a current phone number cannot be located, but an email address is available, email will be used instead. A total of 4 emails will be sent at intervals of one week or more. If both a phone number and an email address are located for a participant, a hybrid follow-up approach might be used, that will include an initial phone call and a follow-up email sent within a week of the phone call, and up to 4 additional phone call attempts and 3 additional email attempts at various intervals.

b. Negative Response Received

If the potential participant indicates she is not interested in further contact by responding negatively to the invitation letter or other contact, no further contact will be attempted. A negative response includes: returning the response card, with "No, I am not interested in

learning more about SEED" checked; sending a text or email to the study site indicating that she does not want to be further contacted; or calling the study site number and leaving a similar message. Likewise, if the potential participant indicates she does not wanted to be contacted again during a follow-up phone call or email, no further contact will be attempted.

c. Response Received

If the invitee returns a response indicating interest, recruitment will proceed with the screening and invitation telephone call.

To the extent possible, each site will cross check all lists of potential participants against each other in advance of the invitation mailing, such that individuals are not sent duplicate invitations. In circumstances where cross-checking of invitees across sources is not possible, individual sources or study staff may not know which potential participants have been contacted previously by other sources and whether or not they indicated interest in participating. Consequently, some invitees may be sent more than one invitation packet before study staff have received a response that they have declined to participate in the study; sites will attempt to limit this occurrence when possible.

2. Screening and Invitation Phone Call

The time between sending the invitation packet and the screening and invitation phone call may vary by site and by data sources within site. However, in the absence of a response it will occur no sooner than 14 days after the invitation packet was sent.

The appropriate respondent on the screening and invitation call is the specific invitee (i.e., the person presumed to be the biological mother as listed on a birth certificate or the child's health/education records). If the person contacted on the call is not the invitee, the interviewer will ask for the invitee's contact information and simply explain that we would like to talk to the woman in question about a health study we are conducting. Care will be taken in how this is approached in order to avoid a breach of confidentiality (e.g., indicating the invitee's relation to the child). The recruiter will not discuss any aspect of the study with any person other than the biological mother.

For participants who do not respond to the full invitation packet and are reached via a follow-up phone call (see Section 1a above), the initial portion of the call will be to determine the invitee's interest and if interest is expressed then the remainder of the call proceeds as described below for the screening and invitation call.

For invitees who do return a response to the full invitation packet indicating interest, the next contact is the screening and invitation phone call. This will consist of either one or two telephone contacts and will include an eligibility screen and ASD screen (SCQ). The phone call will include administration of the eligibility screen, verbal consent for the SCQ and administration of the SCQ. Scoring of the SCQ can occur during the call. As described in detail in Section II.D above and Figures 3-5, regardless of ascertainment source (i.e., "broad diagnostic net" or birth certificate), any eligible child who has a positive SCQ screen or a previous ASD diagnosis/special education exceptionality, will be enrolled as a POSSIBLE ASD. Eligible children ascertained via either the "broad diagnostic net" with no previous

autism/ASD diagnosis/special education exceptionality and a negative SCQ screen will be enrolled as DD participants. Likewise, eligible children ascertained via random sampling of the birth certificates with no previous ASD diagnosis and a negative SCQ screen will be enrolled as POP participants.

Some sites might only invite a sample of DD participants with negative SCQ screens and no previous autism/ASD diagnosis/special education exceptionality to enroll into the study. Depending on the site sampling process, individuals who are selected may need to be re-contacted with a second phone call to be invited into the study.

The flow of the screening and invitation phone call will vary somewhat depending on the results of the initial eligibility screen and, in some circumstances, on the results of the SCQ screen. Specific text for the different versions of the screening and invitation call and their corresponding circumstances are provided in Appendix J.1. Nevertheless, all calls will begin with the following:

1. Interviewer introduces themselves and confirms that respondent is the invitee in question (invitee is presumed to be the biological mother as listed on a birth certificate or the child's health/education records)
2. Brief introduction to the study
3. Conduct the eligibility screen

The eligibility screen includes questions covering the following topics:

- Confirming that the respondent is the biological mother of the child
- Birth date of the child

- Maternal residence at birth of child
- Current residence of child
- Determining if mother is a "knowledgeable" caregiver (i.e. has cared for child since at least 6 months of age)
- Determining if mother has legal guardianship of child
- Main language spoken in the home
- Determining if child is deaf or blind, or unable to walk independently
- Determining whether any siblings of the child have ever, or are currently, participating in SEED (actually, this question is to confirm a previously enrolled sibling since in most instances this can be determined by sites before the call is made)

During the eligibility screen, the participant will also be asked whether the child has had a previous ASD diagnosis. Although this is included in the eligibility screen for ease of the call flow, this question is not actually part of the eligibility screen; rather it is used to help determine the assignment of study workflow. To varying degrees across sites, data on past ASD diagnoses might be limited. Thus children ascertained initially as potential DD or even POP study participants might actually have had a previous ASD diagnosis. Any child determined to have a previous ASD diagnosis will be assigned to the POSSIBLE ASD workflow, regardless of initial source classification or his/her SCQ score.

In SEED 2 one of the SEED sites used an additional approach to the eligibility screen -- a brief on-line questionnaire accessible through the site website. The link to the website included the site's invitation letter and general source communications or flyers describing the

study. Participants could choose to click on an “Interest Form” link within the website. They were explicitly asked for their consent to provide personal contact information over the web and reassured that their information is submitted to a secure server maintained by the site IT department and only accessible by site staff. If they indicated they were willing to provide personal information they were taken to a new, secure screen asking them some of the eligibility questions above (child’s DOB, gender and race, the availability of biological mother, residence at time of child’s birth and current, mother & child’s ability to communicate in English, previous diagnosis of ASD in the child and then name, address, phone and relationship to child of person completing the form). All individuals submitting the web form were contacted by site staff to confirm information given via the web form and complete the remainder of the Invitation Call to determine their full eligibility and enrollment status. This alternative eligibility screen approach might also be used by one or more sites in SEED 3.

If the respondent is determined to be ineligible based on answers to the eligibility screen questions, then the call will end at this point. The person will be thanked for their time. For other invitees, the information obtained thus far on the call will be sufficient to tell most of them that they are eligible and the remainder of the call can proceed as described below. As stated previously, some DD children might not be enrolled in the study, even if eligible, if they have a negative SCQ screen and the DD sample size target has already been met.

Although the order of the call components may vary, the call components after the eligibility screen consist of the following:

- Providing further information on the study, an overview of the study steps, and corresponding incentives, and answering questions.
- If the invitee indicates interest, overall consent for the study and consent for the SCQ
- Administration of the SCQ
- Ascertaining additional information about the family
    - Mother's date of birth
    - Determining whether the biological father lives with the child or is otherwise available and his contact information
    - Information on another individual who can be contacted if needed during the in person visit
- Explaining the next steps of the study
- Scheduling two calls to complete: 1) the Pregnancy Reference Form (PRF) and 2) the Maternal Interview (MI). At some sites, the participant might be given the option of completing the PRF during the invitation call (see PRF details below in section H.1. Follow-up Phone Call 1). Otherwise, the call to complete the PRF will be scheduled for a date estimated to be at least 7 days after the participant has received the enrollment packet (or at least two weeks from the current date to allow for the participant to receive and review the enrollment packet). The Maternal Interview call will be scheduled for a date at least 1 week after the PRF call (or at least four weeks from the current date to allow for previous appointments and mailings to occur).

After the Invitation and Screening call, the site will send the participant the enrollment packet.

If at any time the invitee expresses concerns about the study, she will be referred to the site's principal investigator or his/her designee. Invitees who decide not to participate will receive no additional contact.

At the conclusion of the Invitation Call, all participants enrolled in the study will be classified in one of 3 workflows that will determine the data collection study protocol:

1. ASD workflow (also known as POSSIBLE ASD workflow):
    a. All children who are identified from data sources as having a previous ASD diagnosis and/or autism special education eligibility;
    b. All children identified as having a previous ASD diagnosis during the invitation call (including children originally ascertained from lists of potential DD or POP participants);
    c. All children who screen "positive" on the SCQ screen – defined as SCQ score =>11 (including children originally ascertained from lists of potential DD or POP participants).

    Based on experience from SEED 1 and 2 at most sites, less than half of the children included in the ASD workflow will have been identified by sources as being a potential ASD case; a majority (or sizable proportion) of the children assigned to the ASD workflow are likely to have been initially identified as potential DD without an indication of previous ASD diagnosis of autism special education eligibility.

2. DD workflow: children identified from sources as having one of the (non-ASD) diagnoses or special education eligibilities included in the "broad net" list who:
    a. Do NOT have a positive ASD screen: SCQ score <11; and
    b. Do NOT have a maternal report of a previous ASD diagnosis.

3. POP workflow: children randomly sampled from birth records who:
    a. Do NOT have a positive ASD screen: SCQ score <11; and
    b. Do NOT have a maternal report of a previous ASD diagnosis.

An overview of the data collection by workflow is presented in Appendix H. Differences by workflow are noted. In general, the DD workflow is much shorter than the ASD or POP workflows. The data collection protocol for children in the DD workflow does not include an in person clinic visit, parent-administered forms pertaining to child development, or biosample collection. The data collection protocols for the ASD and POP workflows are more similar; however children in the ASD workflow undergo more developmental evaluations than children in the POP workflow and the mothers of children in the ASD workflow are asked to complete a few more interviews and forms than mothers of children in the POP workflow.

3. Enrollment Packet

The Enrollment Packet (Appendix K) will be mailed to participants who verbally agree to be enrolled in the study during the screening and invitation call. This packet will include a cover

letter (Appendix K.1), a participant-friendly study flow diagram (Appendix K.2), consent document to review (Appendix K.3), and the study "Bill of Rights" (Appendix K.4).

As part of the subject recruitment and retention process, study sites will maintain a record of written and telephone contacts with invitees and subjects. Once a family is enrolled, each site will track the family's completion of data collection steps and receipt of incentives using tracking software created by the DCC and used at each of the sites to maintain site-specific data. The types of information that will be tracked include: dates of letters or calls, response received ('yes', 'no', 'no reply'), consents obtained for each data collection component, dates of scheduled study visits, incentives received, etc. As with all study data, these will be maintained in password protected files, and each person will be tracked using a unique identifying number assigned by the CADDRE Information System (CIS). For all study contact attempts with a participant, study staff will make no less than 8 attempts (over a 6 month period) to contact the participant via telephone at various times during a day and different days during the week. If after multiple attempts to contact the participant, they are unable to be reached, a letter will be sent indicating that they will be dropped from the study unless they contact study staff to express interest in continuing. If they do not respond, no more contact will be attempted.

H. Study Instruments and Data Collection Methods

Please refer to Appendix H for a study flow diagram for a data collection instruments summary table. Ideally, data collection will proceed in the order indicated in the study flow

diagram. However, study staff will try to accommodate participants' schedules and consider requests to change the order.

1. Follow-Up Phone Call 1

At least one week after the Enrollment Packet has been mailed, study staff will phone the participant to answer any initial questions about the study and the enrollment packet materials. Study staff will then complete the Pregnancy Reference Form (PRF) questionnaire (Appendix L.1-2), if not already completed during the invitation call. The PRF is a short questionnaire (<5 minutes) administered to the mother in advance of the Maternal Interview and designed to estimate dates pertaining to the index pregnancy (3 months prior to pregnancy and the first, second, and third trimesters of pregnancy) and the breastfeeding period to create a pregnancy calendar. We provide this calendar to the mother prior to the Maternal Interview such that she can refer to it when asked about the timing of various conditions and treatments just before, during or just after the index pregnancy.

Once the PRF is completed, the appointment for the Maternal Interview is confirmed and the participant is informed that the pregnancy calendar, the Maternal Interview Preparatory Guide (Appendix L.5) and a $30 money order or cash card for completing this follow-up call will be mailed in advance of the interview. The Maternal Interview Preparatory Guide includes reference lists of select medical conditions, medications, and other medical treatments for the biological mother and child.

An outline for Follow-up Call 1 can be found in Appendix J.2.

2. Maternal Interview

Key variables sought from the maternal interview include: pregnancy and reproductive history; obstetric and delivery complications; demographics; occupational history and lifestyle factors during pregnancy. It is expected that the interview will take approximately 45 minutes to complete. The telephone script for initiating the telephone call, including obtaining verbal consent for the interview, can be found in Appendix L.3, interviewer instructions on questions about medical conditions/procedures during the maternal interview can be found in Appendix L.4, and the complete Maternal Interview instrument can be found in Appendix L.6. Detailed interviewer manuals have also been developed. These manuals are provided to all interviewers and interviewers receive training on the content of the manuals and interview procedures to ensure standardization.

The interview will primarily be conducted via telephone using an electronic version of the interview (CATI: Computer Assisted Telephone Interview); however if the study participant does not have access to a telephone for this component, the participant will be asked to schedule an appointment to have the interview conducted face-to-face. Interviews may also be conducted face-to-face for Spanish-only speaking participants. In these cases, a hard copy of the interview will be used initially, and information will be transferred to the electronic database.

Once the interview is complete, the participant will be informed that a packet of checklist-type forms plus a $30 money order or cash card for completing the maternal interview will be mailed to her next. The participant has the option of completing these forms independently. If

the participant prefers that the forms be completed over the phone with study staff, then a telephone call will be scheduled. While we prefer that these forms are completed (either in a self-administered format or via a phone call) in advance of the clinical in-person visit, if for some reason this is not feasible or desirable to the participant, we will offer to help participants in the ASD or POP workflows complete these forms during the clinical visit. (Participants in the DD workflow will not have a clinic visit in keeping with their streamlined data collection protocol.)

Finally, for participants in the ASD and POP workflows, the in-person clinic visit might also be scheduled during this Maternal Interview phone contact. The entire phone call will take 60 minutes (45 minute Maternal Interview and 15 minute discussion of next steps and scheduling).

As mentioned above, participants in the DD workflow will not have a clinic visit and thus forms can be completed in a self/parent administered format and returned via mail to the site, or completed with assistance from the site over the phone. In addition to this difference between the DD workflow and other two workflows, DD workflow participants will receive fewer forms to complete in their packets. Specifically DD workflow participants will NOT be asked to complete two child development assessments – the Child Behavior Checklist and the Child Social Responsiveness Scale – that the other two workflow groups will be asked to complete. Because children in the DD workflow will not have a developmental assessment (conducted at the in person visit) to determine their cognitive scores, we will also not collect

other developmental assessments included in the self-administered packet; these types of assessments are not as meaningful without the accompanying cognitive data.

3. Clinic Visit Preparation, Medical History and Child Development Forms Packet

The materials contained in this packet include six parent/self-administered forms for the ASD and POP workflow children: Maternal Medical History; Paternal Medical and Occupational History; Maternal and Child Residential History; Child Health History; Child Social Responsiveness Scale; and Child Behavior Checklist. Some sites might also include the Services and Treatment Questionnaire (ASD workflow only) in this packet (other sites will help the mother complete this form during the clinic visit).

The packets for the DD workflow will include only four parent/self-administered forms: Maternal Medical History; Paternal Medical and Occupational History; Maternal and Child Residence History; and Child Health History.

In addition to the forms to complete, a $40 money order (or cash card) will be included in the Medical History and Child Development Forms Packet for all workflows. This money order is given to participants for completing the Maternal Interview in the previous step. Additionally for children in the ASD and POP workflows, the packet will include information about the upcoming clinic visit. Finally, the packet will include a cover letter explaining the materials in the packet (Appendix M.1); a cover sheet (Appendix M.2) for the materials to keep (Yellow Folder [ASD and POP workflows only] -- Biologic Sampling FAQ, Picture Story, Clinic Visit Prep Guide); a cover sheet (Appendix M.3a-c) for the questionnaire materials;

and a glossary to accompany the questionnaire materials which includes brief definitions for the medical conditions listed on the Maternal and Paternal Medical History forms and the Child Health History Form (Appendix M.3.d). All of the forms to complete will be included in a Green Folder (Appendix M.4).

It will take approximately 95 minutes to complete all of the forms. Participants may complete these forms on their own or with the help of study staff over the telephone or at a home or clinic visit.

Specific information included in the medical history and child development forms include:

- Maternal Medical History Form (Appendix M.5)

  The Maternal Medical History Form was designed to capture self-reported information on medical conditions and occupational history in the year before the child's birth for the biological mother. The conditions reported must have been diagnosed by a physician. In addition to capturing the presence or absence of the condition, information will be obtained on type of condition (if applicable), age of onset, and presence of the condition during pregnancy. The instrument will take approximately 10 minutes to complete.

- Paternal Medical and Occupational History Form (Appendix M.6)

  The Paternal Medical History Form was designed to capture self-reported medical history information and occupational history in the year before the child's birth for the biological father. Similar to the form for the biological mother, the conditions must have been diagnosed by a physician. This form also captures information on type of

condition (if applicable) and age of onset. The instrument will take approximately 10 minutes to complete. (Of note: Occupational history is only included on the father's medical history form because the mother's occupational history will be collected during the Maternal Interview.)

- Child Health History Form (Appendix M.7)

  The Child Health History Form was designed to capture information on the child's chronic conditions (including whether full and half siblings had similar conditions), allergies and infections, gastrointestinal and sleep characteristics, health insurance and health care. The instrument will take approximately 30 minutes to complete.

- Maternal and Child Residential History Form (Appendix M.8)

  This form was designed to capture information on the mother's residence history beginning one year prior to pregnancy up to the child's birth and the child's residence history from birth up to the present. This instrument will take about 5 minutes to complete.

- Child Behavior Checklist (CBCL) (Appendix M.9).

  The CBCL (Achenbach & Rescorla, 2000; Achenbach & Ruffle, 2000)) is designed to assess, in a standardized format, the behavior problems and social competencies of children as reported by parents. The CBCL takes approximately 20 minutes to complete.

- Social Responsiveness Scale (SRS) (Appendix M.10)

  The SRS is a 65-item rating scale that measures the severity of ASD symptoms. The SRS provides a clear quantitative score for an individual's social impairments including social awareness, social information processing, capacity for reciprocal

social communication, social anxiety/avoidance, and autistic preoccupations and traits. The mother will be asked to complete an age-appropriate version of the SRS for the child. It will take approximately 20 minutes to complete the SRS.

- Optional: Services and Treatments Questionnaire (Appendix G.6)

  This questionnaire is described in section H.6 below.

In addition to the foregoing forms, this packet will also include materials to assist the mothers and children in the ASD and POP workflows to prepare for the upcoming clinic visit:

- Frequently Asked Questions on Biologic Sampling (Appendix M.11)
- Picture Story (Appendix M.12) The picture story will help prepare families for the clinical visit. The story includes pictures and descriptions of each event in which the child will participate. It is written in child-friendly language and is meant to be read to the child by the parent as a story.
- Clinic Visit Prep Guide (Appendix M.13)

4. Follow Up Telephone Call 2 and Clinic Visit Reminder call

Study staff will telephone the participant approximately one week after the Clinic Visit Preparation, Medical History and Child Development Forms Packet is sent to make sure it was received, answer any questions about the packet and ask if there is any need of assistance in completing the forms. Staff may use this call to assist the participant to complete the forms.

For ASD and POP workflow participants, staff will also schedule the clinic visit (if the visit has not already been scheduled), discuss preparation for the clinic visit and answer any questions about the visit.

Within 1-2 days of the scheduled visit, staff will make a reminder call to the participant, answer any questions about the visit, discuss preparation about the visit, and, as needed, assist with the completion of any remaining forms.

5. Follow-up mailing for DD workflow.

Upon receipt of the completed forms in the Forms Packet or in some cases upon completion of the Forms Packet with assistance from study staff over the telephone, DD workflow participants will receive a letter thanking them for their participation in the study (Appendix Q.4). This mailing will also include their final $40 incentive.

6. Clinic Visit Overview (ASD and POP workflows)

The clinic visit will include four main components that can be split into different clinic visits or combined into one visit. Ideally, POP subjects will be scheduled for one clinic visit with

all components lasting approximately 1 hour and 50 minutes. The ASD subjects will be scheduled for one clinic visit with all components lasting approximately 5.5 hours (depending on study staffing, some components can be carried out simultaneously thereby reducing the overall length of the visit). The ASD workflow visit might also be divided into 2 separate visits depending on family preference or response of the child. The clinic visit components are:

a. Answer questions and obtain written informed consent. Complete Medical History and Child Development, forms if needed. (20 minutes)

b. Conduct developmental evaluation (50 minutes for POP workflow; 270 minutes for ASD workflow)

c. Anthropometry and collect saliva samples (20 minutes)

d. Collect blood samples (20 minutes)

The clinic visits will take place at predetermined locations at each of the study sites. Study staff will be responsible for coordinating visits with participants.

a. Answer questions and obtain informed consent

Regardless of the group assignment, the following items will be covered at the beginning of the clinic visit for both study groups (ASD and POP).

1. Project staff will first discuss the written informed consent and have the participant sign the consent form.
2. After the consent process is completed, the process of the developmental evaluation will be thoroughly explained in order to increase the comfort level among study participants.

This component is expected to take approximately 20 minutes to complete.

Also, this time serves as an opportunity to review and complete with the participant any parent/self-administered forms still pending.

b. Conduct developmental evaluation

It is expected the entire developmental evaluation will take between 50 minutes and 4 hours, 30 minutes to complete. The duration of the evaluation varies based on subject group assignment. Each family that participates in the clinical component of the study will be asked to come to a study clinic in order to complete the evaluation. Due to limited clinic space in some sites, if the family is not able to come to a study clinic, project staff will offer to conduct the evaluation in the participant's home.

The developmental assessment will begin once the child appears comfortable with the diagnostician and staff. Parents will be encouraged initially to sit with the child during the assessment in order to make him/her more comfortable if the child has a difficult time

separating from the parent. Children will be able to take a break if needed. If for any reason the child cannot complete the assessment, the family will be encouraged to schedule another appointment to complete the evaluation.

The measures administered during the child developmental assessment are listed below. Copies of the score sheets for each of the measures can be found in Appendix G.

Developmental Assessment Battery

The developmental battery includes measures administered to the child and, for ASD families, interviews and questionnaires administered to the parent. It is preferred that these measures be conducted in a clinical setting, however, if necessary study staff will conduct them in the home. Child measures cannot be conducted over the telephone because they include interactive assessment through structured activities. The developmental assessments administered to the child include:

- Mullen Scales of Early Learning (Appendix G.3): The Mullen will be administered to all children in the ASD and POP workflows. The Mullen assesses cognition in five developmental domains, although only four domains are appropriate for the age-range of children in this study: visual reception (e.g, recognition of patterns and shapes), fine motor (e.g., manipulation of small objects with the fingers), receptive language (e.g., understanding of language), and expressive language (e.g., use of language). These four domains are combined to yield an "early learning composite" score. The Mullen is a theory based neuro-developmental assessment designed to be interactive and is proven to be an effective tool for children with social

communication difficulties. It has adequate standardization and norms, is appropriate for the targeted age group (through 68 months), and has documented reliability and validity. It is expected to take an average of 50 minutes to administer.

Vineland Follow-Up Phone Call (if indicated)

Vineland Adaptive Behavior Scales (VABS) (Appendix G.4) The Vineland will be administered to all children in the ASD workflow. Additionally, if a child in the POP group tests 1.5 standard deviations below the mean on the Mullen, the Vineland (Appendix G.4) will be administered. Administration of the Vineland is only indicated for the POP workflow in this instance due to the importance of obtaining adaptive and non-verbal mental age scores (measured by the Vineland; see the description of the Vineland below) for children with low cognitive functioning. This event is expected to occur in about 3% of POP children. Study personnel will explain the results of the Mullen and subsequently administer items on the Vineland communication, daily living skills, socialization, and motor domains (during the clinic visit or on the phone; a telephone script for this phone call can be found in Appendix G.5); items on the maladaptive behavior index and maladaptive behavior critical items will not be administered due to overlap with other SEED instruments and the inappropriate nature of some of these questions for children in our target age range (e.g., does your child use alcohol or illegal drugs during the school day). Total administration time of the Vineland is expected to average 45 minutes.

- ADOS (Appendix G.2): The ADOS will be administered to children assigned to the ASD workflow, i.e. those with a previous autism/ASD diagnosis/exceptionality and/or

those who screen positive on the SCQ. The ADOS (Lord et al 2012) is a standardized instrument in which the researcher observes the child and tries to illicit social interaction and communication through the use of structured play activities. The examiner implements the module that best corresponds to the child's expressive language level in order to prevent language aptitude from impeding accurate diagnosis. This measure takes an average of 45 minutes to complete and generates observational data that are scored to determine ASD classification.

- Mothers of children assigned to the ASD workflow will be asked to complete three interviews asking questions about their child's development and services and treatments. These interviews will either be conducted in the clinic, on the telephone, or at the participant's home by the local study staff although telephone interviews are discouraged (especially for the ADI-R). The interviews include:
  - ADI-R (Appendix G.1) The ADI-R (Lord et al., 1994) is a semi-structured, investigator-based interview for caregivers. It is intended to provide historical information about social and communicative behaviors in order to assess whether a child has behaviors consistent with an ASD. This measure takes an average of 120 minutes to complete and generates interview data that are scored to determine ASD classification.
  - Vineland Adaptive Behavior Scales (VABS) (Appendix G.4) The VABS measures personal and social sufficiency across 4 domains of adaptive behavior: communication, daily living skills, socialization, and motor skills. It is administered to a respondent (parent or primary caregiver). The VABS is expected to take an average of 45 minutes to complete. NOTE: As described above, the

Vineland will also be administered to about 3% of POP children who score 1.5 standard deviations below the mean on the cognitive measure (Mullen).

- Services and Treatment Questionnaire (Appendix G.6) The purpose of this instrument is to capture information related to the child's current use of commonly used services among children with developmental difficulty. This tool also captures information on treatments, including complementary alternative medicines and biological treatments. This is expected to take approximately 10 minutes to complete.

With appropriate consent, at least 10% of the overall sample of children who receive the full developmental evaluation will have the ADI-R and ADOS and/or Mullen and Vineland videotaped for quality control purposes. Written permission for videotaping will be obtained in the comprehensive consent form. Participants will be told that allowing project staff to videotape the assessment is completely voluntary and that choosing not to have the assessment videotaped will not affect study participation in any way. They will also be told that the videotapes will be used for several purposes in the research study including 1) assisting the developmental evaluators in making case determination, 2) facilitating quality assurance between developmental evaluations, and 3) measuring inter-rater reliability (see section II.J "Training and Quality Control" of this protocol).

All data from the developmental assessment is recorded on hard-copy record forms that will subsequently be transferred into an electronic database.

c. Anthropometry and collect saliva samples

After completion of the developmental evaluation, a qualified examiner (as described in section II.J) will take height, weight and head circumference measurements of the child and height and head circumference measurement of the mother. The examiner will also ask the mother if the child has ever been diagnosed or been examined for a major birth defect or genetic syndrome. (Appendix N.1)

We will attempt to obtain saliva samples from all children and biological mothers and fathers in the ASD and POP workflows. Because genetic analysis is an important component of SEED, we include saliva collection to increase the completeness of the biologic samples. We anticipate compliance using this approach will be higher than blood sampling, because of the greater convenience and less discomfort to the participant. Also, in some instances, a blood draw attempt may fail. Unfortunately, saliva samples potentially yield less DNA, or a lower quality sample of DNA than blood. Collecting saliva samples will ensure that genetic material is obtained on a greater number of participants, while blood samples will ensure a higher quality sample of DNA.

In advance of the clinic visit, each saliva collection kit will be assembled based on the availability of targeted participants in the household (e.g., a household including the biologic mother and child, but not the biologic father, would need kits for mother and child only). A saliva collection kit will be mailed to the biologic father, as needed, if the biological father is willing to provide a saliva sample and lives in a separate household from the index child. If the biological father lives in the same household but is unable to attend the clinic visit, a

saliva collection kit for the father's specimen will be given to the mother to take home with her. Contents of the saliva collection kit are described below and in Appendix N.

Each participant will be asked to provide one tube of a saliva sample. This requires directly spitting into the tube. If the participant cannot or chooses not to spit into the tube, they can use a small sponge to soak up as much saliva for at least 30 seconds by moving the sponge in the cheek pouch. Parents will be asked to self-administer their own samples and to attempt the saliva collection with the child. In order to minimize anxiety that may be associated with the collection, it will be suggested that the child first watch the parent self-administer the sample before attempting to collect saliva from the child.

As mentioned above, if the biological father lives in the child's household and is not available for the clinic visit, but is willing to provide a saliva sample, then the father's saliva collection kit will be given to the biological mother at the clinic visit. Likewise, biological fathers not living with the mother and child may be mailed a saliva collection kit if they are willing to provide a specimen. In these cases, the father can mail the sample back to the SEED site. The saliva collection kit (Appendix N) will include the following:

- Oragene saliva kit – contains 1 collection tube, 1 collection tube cap, collection sponge and directions.
    - It will take an average of 5-10 minutes to complete the contents of the saliva collection for a participant.
- Saliva Collection Instruction Sheet (Appendix N.2)
- Frequently Asked Questions on Biosampling (Appendix M.11)

- Saliva Collection Written Consent Form including copies to sign & copies to keep (Appendix N.3)
- Saliva Transmittal Form (Appendix N.4)
- Self- addressed, stamped mailer to return the kit and consent form to the SEED site.

SEED sites will send saliva samples to the SEED Central Lab where they will be stored at -80ºC until DNA is extracted.

- In addition, as discussed below under the quality control section (section II. J. "Training and Quality Control") there is also a very small chance that we may ask mothers or fathers to provide an additional sample of saliva. They will be provided with another kit to do so. This can happen if DNA extraction fails on the first samples that were provided. Subjects providing saliva samples that do not yield a sufficient amount of DNA and for whom there is no blood sample might be asked to supply another saliva sample via a self-administered kit mailed to the participant.

d. Blood sample

Due to funding restrictions, SEED sites will discontinue blood sample collection on both children and mothers in 2018. We anticipate that blood collection will be discontinued in July, but there might be some variability by site. We have revised all participant materials that mention the blood collection. (See Appendices). We will continue to collect saliva specimens and thus, will continue to get samples that allow us to conduct genetic analyses. If funding becomes available at a future date, we may re-instate blood collection; if that happens we will notify the IRB via amendment. We are not further revising the protocol to

delete the verbiage about venous blood sampling because of the aforementioned possibility that we may resume blood sampling at a future point in the study.

Venous blood sampling will be conducted on all children and biologic mothers in the ASD and POP workflows. Because genetic analysis is an important component of SEED, we are requesting venous blood samples from participants to ensure a high quality sample. This biosampling will take approximately 20 minutes to complete. Mothers will be asked to complete for herself and for the child a brief checklist of recent exposures that might affect blood analyses (e.g., medications, illnesses) (Appendix O.1-2). Please refer to section II.J for a detailed description of the hiring requirements for the phlebotomist, as well as a detailed description of the training regimen.

Venous blood sampling will be standardized to the greatest extent possible across all study sites. Trained study staff will discuss the child blood draw with the mother prior to the attempt. Staff will attempt to assess social, communication, sensory and behavioral strengths and limitations of child as they relate to the child's ability to comprehend and cope with the blood draw. Staff will seek parent input on approaches the parent has used for handling blood draw or other similar situations in the past. The mother may be asked to bring relevant materials from home to facilitate the procedure (e.g., favored toys, activities, books, videos, or other communication aids) and staff will obtain other materials (e.g., food, suckers, stickers) as needed. In certain instances a pre-visit education plan may be indicated, although for most families a standard approach consisting of gentle restraint, distraction, and a reward will be adequate. In all instances, decisions about the blood draw administration will be made based on mother's input and the team's assessment.

All blood draws on children will be conducted at the end of the clinic visit by protocol-trained study phlebotomists having prior pediatric blood draw experience. Other staff and the mother will be present to assist with restraint, distraction, and reinforcement. No more than two attempts will be made to complete the draw during any clinic visit. Study staff may offer mothers the option that children receive a local topical anesthetic prior to venipuncture. However, the mother will make the final decision as to whether the anesthesia cream is used. The agent used will be a non-prescription strength topical cream of (4% lidocaine in a liposomal delivery system). The cream is applied 30 minutes prior to blood draw. If the child will tolerate an adhesive bandage, a hypoallergenic latex free bandage will be applied to cover the area to prevent accidental removal or ingestion. Alternatively, in some clinic settings if the child can tolerate long-sleeved shirts, individual sites can recommend during the pre-visit contact that a long-sleeve shirt be worn. Potential side-effects of the anesthetic cream are rare, transient, and localized including irritation, redness, or rash.

The pediatric venipuncture itself will be done using the butterfly system. The butterfly system uses shorter needle (appears less threatening) and, because of the flexible tubing, there is allowance for movement on the part of both the subject and the phlebotomist. A maximum of 13mL of blood will be collected using vacutainers appropriate for pediatric draw volumes. Optional residual blood in the butterfly tube (less than 1mL) will be used for the filter paper blood spots. Federal guidelines prohibit the draw of more than 3mL/kg body weight per eight-week period. Consequently, a child would only have to weigh about 4.33 kg (9.5 lbs) to be eligible to provide this amount of blood.

Blood tubes will be processed at an appropriate temperature and shipped per lab specifications in a specialized dual temperature shipper via Federal Express or other commercial carrier to the Central Lab. Filter paper cards will be shipped to the Central Lab after drying (24 hours) and with a subsequent shipment of biosamples. Details on the handling, storage, and shipment of these biosamples are outlined for study staff in the Biologics Procedures Manual.

Biological mothers in ASD and POP workflow groups will be asked to provide venous blood samples. Because genetic analysis is an important component of SEED, we are requesting venous blood samples from participants to ensure a high quality sample with enough volume to complete analysis. Venous blood sampling will be standardized to the greatest extent possible across all study sites. Protocol-trained phlebotomists will conduct blood draws on biological mothers during the clinic visit. Mothers will be asked to donate a maximum of 7mL of blood with residual blood in the butterfly tube (less than 1mL) will be used for the filter paper blood spots. The number of blood draw attempts made will be determined on an individual basis, based on discussion between the participant and the phlebotomist.

Blood tubes will be processed at an appropriate temperature and shipped per lab specifications in a specialized dual temperature shipper via Federal Express or other commercial carrier to the Central Lab the same day they are drawn. Filter paper cards will be shipped to the Central Lab after drying (24 hours) and with a subsequent shipment of biosamples. For quality control purposes (discussed further below in Section II.J, Training and Quality Control) a small (2-3%), randomly chosen sample of mothers providing blood samples will be asked to

provide double the amount of blood so that duplicate aliquots can be established and shipped to the repository in a blinded fashion.

7. Follow-up Call 3

Following the clinic visit, if there are any self-administered forms that have not been completed or need further information, study staff will make a follow-up call to finalize this data collection.

8. Passive Refusal Letter

Efforts will be made to complete full data collection on each study element in which the family consented to participate. In the event that one or more of the above data collection elements or study instruments is not returned to the study, study staff will attempt to contact the family up to 8 times by telephone (or email if that is the optimal mode of contact) over a six month period to complete data collection. After the 8 contacts have been made, the study will send a Passive Refusal letter (Appendix P, via regular mail or email) that indicates we will no longer attempt to make contact with the family to collect the remaining items. The contact schedule is as follows:

- Over a 1-2 month period: 1 morning call, 1 afternoon call, 1 evening call, 1 weekend call
- No call attempts will be made for the following 1-2 months.
- Over the next 1-2 month period: 1 morning call, 1 afternoon call1, 1 evening call, and 1 weekend call.

Separate from the 8 phone attempts, a participant may receive a passive refusal letter if any of the following are true: 4 cancellations (for study appointments or calls), even if the appointment/call is re-scheduled, or if 2 appointments (Maternal Interview, Clinic Visit, or other scheduled phone calls) have been unattended without notification to study staff. Once these minimum numbers of failed attempts at contact have occurred, sites may decide on a case-by-case basis when to send a passive refusal letter to a participant.

9. Vital Records Data

a. Use of Data for Study Recruitment and Related Purposes

Sites will obtain approval from the State Registrar or Vital Records Department to obtain birth records containing personal identifiers. These identifiers are needed to identify potential participants (POP group) and verify study eligibility (all study groups). Identifiers collected from vital records data, including names of children, mothers and fathers; county of residence at child's birth; mailing/ e-mail addresses (at some sites), and phone numbers (at some sites) will not be included in SEED analytic data files. Only limited SEED study staff have access to these data and all SEED staff are held to confidentiality standards and are bound by the CDC Certificate of Confidentiality (see below).

b. Analytic Data Files on Enrolled Participants

At all sites, individual-level vital records data on children enrolled in SEED 3 will be obtained. These data will be de-identified and compiled in analytic data files for use in various risk factor analyses.

Personally identifiable information (PII) obtained from each site's state vital records department will not be included in SEED analytic data files. Access to these data will be limited to the SEED site to which they pertain. However, while PII on SEED participants will not be shared across SEED sites, some vital records data on enrolled participants that are potentially identifiable will be used to create new non-identifiable data items that will contribute to SEED analytic data files. For example mother's and father's dates of birth might be systematically shifted up to +/- 7 days to de-identify the specific dates while maintaining accurate age information. The date shifting procedure and algorithm were developed by the SEED Data Coordinating Center (DCC) to de-identify dates from various study instruments that are included in the SEED analytic data files (see below, K. Data Management & Analyses) DCC previously QC'd their date shifting algorithm to ensure that the newly created shifted dates are not able to be linked back to the original dates. While the basic aspects of the general date shifting procedure is known to the SEED sites, the DCC does not share the exact algorithm, such that study staff from one SEED site are not able to determine the original (correct) dates from other SEED sites' data that are included the analytic files.

Other non-identifiable vital records data items that will be included in the SEED analytic data files are: maternal education, paternal education, maternal marital status, maternal place of birth, paternal place of birth, insurance status, maternal race/ ethnicity, paternal race/ ethnicity, child birth weight, clinical estimation of child's gestational age in weeks, plurality of child, birth order of child, child Apgar score at 5 minutes, month of pregnancy prenatal

care began, number of prenatal visits, maternal smoking, method of delivery, induction of labor, stimulation of labor, tocolysis, complications with delivery, number of live births now living, number of live births now dead, and other similar variables that describe demographic and pregnancy characteristics.

These analytic data will only be presented in aggregate.

c. *Use of Data to Understand Characteristics of Non-Participants*

Vital records data will also be used to understand characteristics of non-participants. Again, the findings from all analyses will only be presented in aggregate.

One of the enhancements of SEED 3 is that sites are strongly encouraged to develop a mechanism and obtain necessary permissions to allow for linkage of limited birth certificate data on ***all*** participants (ASD, DD, and POP groups) invited including those who are never successfully contacted, those who refuse contact, and those contacted but found to be ineligible, such that SEED investigators will be able to directly assess characteristics of respondents vs. non-respondents. Birth certificate data of interest include key socio-demographic factors – child sex, maternal race-ethnicity, maternal age at birth, maternal education at birth; of note: the latter three factors have been found to be associated with participant non-response in a recent analysis of SEED 1 data (unpublished data). The exact plan for data linkage might vary across sites. Some sites have access to the full birth records and are thus able to perform linkages between children invited from the ASD and DD source

lists directly and retain all of the data onsite. Other sites might obtain more limited access. While individual-level data is preferred, aggregate data with distributions of aforementioned key variables will also be helpful in assessing potential response bias.

I. Consent Process

Consent for the data collection described above will be obtained on four separate occasions for mother-child pairs included in the ASD and POP workflows. First, initial verbal (oral) consent to conduct an ASD screen and then to participate in the study will be obtained during the first telephone contact after the individual has received the invitation letter. Second, verbal consent for the maternal interview will be obtained during the interview phone call. Third, staff will obtain written, oral, or implied consent for the Medical History and Child Development Forms, depending on whether the forms are completed face-to-face, over the telephone, or self-administered. Some sites will use Appendix M.4 to obtain written consent if the participants complete the packets either face-to-face or self-administered. Fourth, during the clinic visit, study staff will obtain informed consent for the overall study including blood and saliva swabs collection.

Given the streamlined protocol for mother-child pairs included in the DD workflow, consent for the data collection will be obtained on three separate occasions. First, initial verbal (oral) consent to conduct an ASD screen and then to participate in the study will be obtained during the first telephone contact after the individual has received the invitation letter. Second, verbal consent for the maternal interview will be obtained during the interview phone call. Third, staff will obtain written, oral, or implied consent for the Medical History and Child

Development Forms, depending on whether the forms are completed over the telephone, or self-administered. Some sites will use Appendix M.4 to obtain written consent if the participants complete the packets in a self-administered mode.

1. Verbal Consent

After the invitation letter has been sent, individuals who provide an affirmative response to the invitation or during the follow-up phone call for invitees who do not respond will be administered the Screening and Invitation phone call. This phone call includes the eligibility and autism screening, as well as the introduction to the study. These screener and invitation phone calls will consist of one or two telephone contacts.

The initial phone call will be made to administer the eligibility screen and the SCQ. The SCQ will be administered to all participants (POP, DD, and ASD). This phone call will include verbal consent for the SCQ. Scoring of the SCQ can occur during the call. Verbal consent for participation in the study will be obtained at this point. We are requesting the IRB waive the requirement to obtain a signed consent form (45 CFR 46.117(c) 2) for the invitation phone call for all participants. SEED is requesting this waiver because the investigators feel this research presents no more than minimal risk of harm to the participants and involves no procedures for which written consent is normally required outside the research context. It is imperative to the study for the investigators to obtain minimal information about the participant during this call to assign the potential participant to the appropriate group.

A script for this call can be found in Appendix J. Specifically the process is described below:

- Project staff will introduce themselves, review the purpose of the study, and answer any questions the participant may have;
- Project staff will review the expected roles and responsibilities of the biologic mother and child, which components of the study they are being asked to participate in, risks and benefits involved with the study, and confidentiality procedures;
- All study participants will be referred to the principal investigator at each site (if they wish) in order to address any other particular questions or concerns; and
- If the participant agrees to take part in the study, his/her verbal consent will be documented on a study contact log.

2. Verbal Consent for the Maternal Interview

Study staff will schedule a time to call the participant for the conduct of the maternal interview. The interviewer will obtain verbal consent of the participant. We are requesting the IRB waive the requirement to obtain a signed consent form (45 CFR 46.117(c) 2) for the maternal interview for all participants. SEED is requesting this waiver because this research presents no more than minimal risk of harm to the participants and involves no procedures for which written consent is normally required outside the research context. This verbal consent is obtained orally, and will be documented on study call logs.

3. Medical History and Child Development Forms

Self/Parent Administered Questionnaires Consent

As described in the contact and enrollment section (II G) above, the Medical History and Child Development forms will be mailed to the participant if they have verbally agreed to

participate in this component of the study. The forms will be bundled in one packet with a cover sheet or letter that itemizes the questionnaires and includes a statement about consent, if required by an individual site's IRB. A site-specific consent form will be included if required (Appendix M.4). (See Appendix M for Medical History and Child Development Forms Packet)

4. Written Consent for the Study

A written informed consent for the entire study will be included in the Enrollment Packet for ASD and POP workflows (Appendix K) such that the participant may review prior to the clinic visit. The consent document will include a description of the entire study and expected roles and responsibilities of the mother and child, risks and benefits to participants, and confidentiality procedures. The consent document will also clearly state that participation in each component of the study is completely voluntary, and participants can drop out of the study without penalty at any time. A separate consent section will also be included for the participant to sign if they consent to their genetic information to be contributed to the National Database for Autism Research (NDAR) at the end of the study. A description of NDAR and how the participant's information will be contributed is explained in this section. Study staff will review this consent form at the clinic in-person contact. At this time, study staff will answer questions and obtain written consent from the participant. A copy of this consent form can be found in Appendix K.

All consent forms will follow the guidelines outlined by the Office of Human Subjects Protection of the Centers for Disease Control and Prevention and ethical guidelines set forth

by the state and federal governments. The composition of the consent forms will be adapted to conform to site-specific IRB requirements. Documentation of subject consent will be stored in a locked file in the PI or designee's office at each study site.

c. Written Consent for Saliva Collection

A written standalone consent form is available and given to the family for use by the father who is not available during the clinic visit. The saliva collection kit (Appendix N) includes two copies of each consent form requested from the participant; one for the participant to keep, and one for him to sign and return. The consent documents are returned to the site and the saliva collection kit is sent to the Central Laboratory and Biorepository (CLBR) in separate postage-paid mailers provided by the site.

J. Training and Quality Control Methods for Data Collection

All sites will begin training with a standardized training protocol and instruction manual that provides general training to all study staff as follows:

1. Oral presentation by a senior investigator or Project Coordinator describing the overall background, purpose and approach of the study;
2. Overview of each component of the data collection;
3. Detailed discussion of general instructions for conducting field work; and
4. Detailed discussion of procedures concerning data security and safeguards for protecting privacy and confidentiality of personal information.

Each staff member will receive additional training with respect to the specific data collection components for which they will be responsible:

1. Classroom instruction for specific data collection components
2. Classroom practice administering the specific data collection component.
3. Initial and subsequent periodic observation of staff by the data collection supervisor, followed by constructive review or correction of techniques to ensure adherence to the protocol. The specific focus of the initial and subsequent quality assurance monitoring by the supervisor will include:
    - Standardized and complete administration of instruments.
    - Accurate recording of data
    - Ongoing checks for data range, consistency, and logic
    - Inter-rater reliability of recording responses during interviews.

Additional quality assurance measures and specific staff qualifications are described below and a summary of the quality control requirements for each study instrument is presented in Table 6.

Table 6. SEED quality control procedures summary

Study contacts and instruments	**Type of QC assessment(s) and requirements**	**Specific QC training requirements[1]**	**Ongoing QC requirements (frequency of QC)**
Invitation phone call, including eligibility screener and Social Communication Questionnaire	*Intra-site:* Semi-qualitative call rating form -- a priori criteria established for acceptable score.[2]	Acceptable scores on 3 role playing (mock) calls and first 2 "live" calls.	5% per interviewer
Follow-up call, including structured Pregnancy Reference Form interview	*Intra-site:* Semi-qualitative call rating form -- a priori criteria established for acceptable score.[2]	Acceptable scores on 3 role playing (mock) calls.	5% per interviewer
Maternal Interview	*Intra-site:* 1) Semi-qualitative call rating form -- a priori criteria established for acceptable score.[2]	Acceptable scores for both assessments on 3 role playing (mock) interviews	5% per interviewer

	2) Quantitative inter-rater reliability assessment of selected interview items. Acceptable score is ≥95% concordance.	and first 2 "live" calls.	
Parent/self-administered forms[3] (forms sometimes administered by SEED study staff via phone or in person)	No specific systematic QC requirements, but continual supervisor oversight and all forms reviewed for missing or illegible data or contradictory entries. Participants are re-contacted as needed.	None. General training provided on forms and appropriate responses to participant queries.	NA
Autism Diagnostic Observation Schedule (ADOS)	*Inter-site:* supervising clinicians establish reliability by scoring the same ADOS exam videotapes. Acceptable score is ≥80% concordance on algorithm items. *Intra-site:* all clinicians establish reliability with supervising clinician. Acceptable score is ≥80% concordance on algorithm items.	Both inter-site and intra-site reliability established in advance of study start.	quarterly inter-site and intra-site reliability exercises
Autism Diagnostic Interview - revised (ADI-r)	*Inter-site:* supervising clinicians establish reliability by scoring the same ADI-r interview videotapes. Acceptable score is ≥90% concordance on algorithm items. *Intra-site:* all clinicians establish reliability with supervising clinician. Acceptable score is ≥90% concordance on algorithm items.	Both inter-site and intra-site reliability established in advance of study start.	quarterly inter-site and intra-site reliability exercises.
Mullen Scales of Early Learning	No specific systematic QC requirements, but continual supervisor oversight and all forms reviewed for missing or illegible data or contradictory entries.	None. Supervising site clinicians monitor initial assessments until competency determined.	NA
Vineland Adaptive Behavioral Scales	No specific systematic QC requirements, but continual supervisor oversight and all forms reviewed for missing or illegible data or contradictory entries.	None. Supervising site clinicians monitor initial assessments until competency determined.	NA
Anthropometrics: Height, weight, and head circumference	*Intra-site:* All examiners establish reliability with project coordinator or other qualified examiner on age appropriate individuals. Acceptable reliability defined as agreement in 95% of instances (agreement within 0.5 cm for height and head circumference and 0.1kg in weight). Equipment calibrated periodically as needed.	Acceptable measurements on 5 role play (mock) individuals and on first 2 participants.	5% per examiner
Biologic specimens: Saliva specimens (child, mother and	*All*: central lab staff processes specimens upon receipt and performs preliminary QC (gross	None. Extensive staff training on study protocol for	2% sample of duplicate blood

father if available) and blood specimens (child and mother) (ASD and POP workflow only)	visual inspection). *Sample of mothers:* second blood specimen obtained for duplicate processing and analysis	obtaining and processing biologic specimens	specimens

[1] Training QC requirements include requirement for staff to pass formal reliability or other QC assessment on mock exercises in advance of "live" field work and initial QC requirement on first instruments/exams once in the field. For each instrument, if a study staff member does not meet criteria for acceptable score during ongoing QC, retraining and training QC requirements are instituted.

[2] Semi-qualitative call rating forms for invitation, follow-up, and maternal interview calls include items such as use of call script, coverage of essential points, ability to respond to participant questions, probing on unclear or neutral responses, professionalism, and delivery and response recording for applicable study instruments (Social Communications Questionnaire, Pregnancy Reference Form, or Maternal Interview). For each item, QC supervisor rates interviewer as "good", "fair" or "poor". Criteria for acceptable score include: no item rated as "poor" and 20% or less rated as "fair"; and mandatory ratings of "good" for select items (dependent on call type).

[3] Parent/self-administered forms include Maternal Medical History Form, Paternal Medical and Occupational History Form, Child Health History Form, Maternal and Child Residence History Form, Child Behavior Checklist (ASD and POP workflows only), Child Social Responsiveness Scale (ASD and POP workflows only), Child Services and Treatments Questionnaire (ASD workflow only).

1. Invitation and Screening Call and Follow-up Call 1 (Telephone)

The project coordinator at each site will train all individuals conducting calls with the study participants. Each staff member conducting such calls must also successfully complete several mock call role playing exercises in advance of beginning field work. The assessment of both the mock calls and ongoing calls with participants once the staff member begins field work consists of a semi-quantitative rating form which includes items related to the interviewer's professionalism, ability to answer questions from participants, success in keeping the participant focused, and accuracy in administering various components of the call correctly (such as the eligibility screen, SCQ, and participant consent in the Invitation and Screening Call and the Pregnancy Reference Form in the Follow-up Call).

Once the staff member completes the mock calls and begins calling SEED participants, the PC will listen to 5% of the calls; sites will either tape calls for QC at a later time or have the PC listen to a live call. The PC will complete the quality control form described above.

2. Maternal Interview Personnel (Telephone)

Interview personnel should hold a Bachelor's degree or have interviewing experience. An interviewer supervisor (project coordinator or her/his designee) will oversee the additional training and ongoing review of interviewers. Detailed interviewer manuals have also been developed. These manuals will be provided to all interviewers and interviewers receive training on the content of the manuals and interview procedures to ensure standardization within and across sites.

Each interviewer must successfully complete several mock interview role playing exercises in advance of beginning field work Assessments of both the mock interviews and the ongoing interviews with participants once field work is initiated consist of two components: 1) a semi-quantitative rating form which includes items related to the interviewer's professionalism, ability to answer questions from participants, ability to keep the participant focused and administer the interview correctly (as written); and 2) a quantitative inter-rater reliability assessment of selected interview items.

Once the interviewer completes the mock interviews and begins interviewing SEED participants, the interviewer supervisor will listen to 5% of the Maternal Interviews; sites will either tape the interview for QC at a later time or have the supervisor listen to a live interview. The supervisor will complete two quality control forms described above. The supervisor will calculate inter-rater reliability rates and produce a report to share with the interviewer; inter-rater reliability should be no less than 95%.

3. Developmental Evaluation/Assessment and Parent Interview

Supervising Site Clinicians

Supervising site clinicians (SSC) are defined as clinicians with a doctorate degree in medicine, psychology, education, or related field, experience with the assessment and diagnosis of children with autism spectrum disorders, relevant core competencies as defined by the American Psychological Association (APA) and test administration experience, and licensure according to regulations of the site center or facility. SEED SSC clinicians will have received general protocol training and administered and scored at least three Mullen, Vineland, ADOS, and ADI-R assessments within the year prior to study enrollment (to ensure familiarity with administration guidelines and coding rules). In addition, SSC will have established general research reliability on the ADOS and ADI-R through the UMACC, WPS, or other certified trainer and inter-site reliability prior to assessing the first study participant (in each phase of the study).

Inter-site reliability will be established by reviewing and scoring the same ADOS and ADI-R videotape (as other SSC) and forwarding the scoring algorithm to a CWG representative prior to a scheduled consensus call or in-person meeting. The consensus call or meeting will be held to establish the final scoring criteria for which each site will be evaluated against. Reliability is established with 80% agreement on ADOS algorithm items and 90% agreement on ADI-R algorithm items. A CWG representative will maintain reliability records. Inter-site scoring reliability will be maintained at least quarterly thereafter, with scoring algorithms sent to a CWG representative before the scheduled call or in-person meeting at least once per year.

Sites will rotate responsibility for providing tapes for reliability review at least two weeks before the scheduled call or in-person meeting.

ADOS and ADI-R administration competency will be demonstrated via once-per-phase submissions of an ADOS and ADI-R administration to another SSC. The other SSC will review each ADOS tape and at least one hour of each ADI-R tape and provide feedback to the SSC. The other SSC will also provide notes on the SEED QC ongoing data collection form and forward to the Clinical Workgroup Chair to maintain throughout the study period.

Remediation in SSC administration procedures and/or scoring discrepancies will proceed as outlined in the SEED Clinical Workgroup Quality Control Manual.

Field Clinicians

To administer the Mullen, Vineland, ADOS, and ADI-R for SEED, qualifying staff will be defined as clinicians with at least a Master's degree or higher in related field and relevant core competencies as defined by the APA and test administration experience. To administer all other instruments (e.g., Social Responsiveness Scale, Child Behavior Check List) qualifying staff will be persons who have completed protocol training as describe above and related interview or questionnaire administration experience. These staff members will be supervised by a qualifying degree clinician (i.e., PhD, MD).

Additionally, SEED clinicians will have administered and scored at least three of each of these protocols within the year prior to study enrollment (to ensure familiarity with administration guidelines and coding rules). Administration competency in these particular instruments will be demonstrated via videotape or live observation prior to SEED data

collection and at least once per year thereafter. The supervising site clinician is responsible for judging competency and will note any major administration deviancies and/or scoring discrepancies. The supervising site clinician is responsible for determining when each clinician is competent to begin administration of SEED assessments; if competency is determined acceptable after review of the first example (either video or in-person observation) no additional review is needed for the clinician to initiate actual study assessments.

In addition to the above criteria, clinicians responsible for ADOS and ADI-R administration must establish intra-site reliably before completing a SEED developmental evaluation. The field clinician will either 1) administer and score an ADOS evaluation shadowed by the SSC or 2) watch a videotape scored by the SSC and score the same assessments. The SSC's results will be compared with the field clinician's results. A field clinician will be considered "research reliable" when they establish and maintain intra-site reliability of .80 on the ADOS and .90 on the ADI-R. Intra-site reliability exercises will be completed by each field clinician at least quarterly thereafter.

All other instruments (e.g., Social Responsiveness Scale and the Child Behavior Checklist) can be administered by study personnel with protocol training as described above and relevant interview or questionnaire administration experience. All study personnel who attend clinic visits will review the instruments included in the Clinic Visit Preparation, Medical History and Child Development Forms Packet and the Services and Treatments Questionnaire and proper administration guidelines with the site supervising clinician. The site supervising clinician will determine when each person is competent to administer these SEED assessments.

Remediation in field clinician administration deviancies and/or scoring discrepancies will proceed as outlined in the SEED Clinical Workgroup Quality Control Manual.

4. Anthropometry

Personnel administering the anthropometry must be a trained research assistant. Training will involve the following elements:

1. Basic didactic instruction on the anthropometry measures (with a manual with illustrations); including review of principles of examining children with autism and developmental disabilities, including the special considerations required for this population.
2. Practice anthropometry sessions

Inter-rater reliability for measurements will be examined periodically with agreement in 95% of instances (within 0.5 cm for height and head circumference and 0.1kg in weight). Periodic reliability checks will be carried out to maintain the quality of the collected data, to refresh data collection skills, and reinforce the initial training instructions. Inter-rater reliability will be conducted by having a second qualified examiner perform the anthropometry measurements at the time of the exam.

5. Saliva Sample

Within 30 days of receipt, DNA will be extracted from the saliva collection tubes at the Central Laboratory and Biosample Repository (CLBR) and test amplified for a moderate-length fragment. Failure rate for moderate-length fragments is expected to be around 2.5% (Garcia-Closas et al., 2001).

If the saliva DNA yield is less than 5 micrograms, but a blood sample has been drawn on the participant, then no additional saliva collection is attempted. If the saliva yield is less than 5 micrograms and no blood sample has been drawn, then the site is notified and a second saliva sample collection from the participant is attempted. This re-contact could be done at the time of the Clinic Visit or through a separate mailing.

6. Venipuncture

Personnel completing phlebotomy will be trained clinician(s) (medical doctor, nurse or nurse practitioner) or a certified laboratory technician or phlebotomist who is trained in phlebotomy techniques and precautions according to JCAHO standards.

Adult venipuncture will be performed as outlined in the Biologic Specimen Manual. Staff performing venipuncture during home visits will be trained to follow a home phlebotomy protocol. Study coordinators at each site will distribute a Biologic Specimen Manual to staff and ensure that staff learn and follow study protocol, occasionally observing study sessions.

Child venipuncture will be performed by a phlebotomist or nurse having prior experience conducting blood draws in 2-5 year old children. Study coordinators at each site will distribute a Biologic Specimen Manual to these staff and will be responsible for assuring that staff learn and follow study protocol. To familiarize staff with venipuncture in children with developmental disabilities (unless the staff member comes with prior experience with this population), study coordinators, working with study site supervising clinicians, will arrange for staff to observe, in-person, clinical blood draws in this population on-site. Training

material specific to the conduct of blood draws on children with developmental disabilities, including guidelines for pre-visit interview, application of ELA-max or equivalent topical anesthetic, and restraint option guidelines (including visuals), will be developed and circulated to all sites. A training teleconference will be held across sites to review and discuss training materials with the phlebotomist or nurse, and any other study staff, who might be conducting the pre-visit assessment and/or assisting during blood draw. All staff assisting in the blood draw will review the Biologic Specimen Manual and the training material.

Study coordinators across sites also discuss during their weekly calls any issues that may have emerged regarding protocol implementation and together learn from these experiences.

7. Quality Assurance for Venous Samples

For quality-control purposes a small (2-3%), randomly chosen sample of parents providing biosamples will provide sample duplicates (venous blood). These samples will be sent to the Central Repository with unique ID numbers (the Data Center will maintain the link between these identifiers and the participants' original ID numbers). The extra blood volume will be aliquoted. The duplicate samples will allow additional volume on a subset of participants facilitating assessment of laboratory assay variation by creating duplicate aliquots from the same individuals. These samples will also allow evaluation of other quality control concerns (e.g. the effect of thaw/refreeze on assay performance). If a participant is asked to provide a second sample because the first sample was not adequate, they will NOT be asked to provide an additional sample for quality control.

K. Data Management & Analysis

The data analysis and management plan for a complex, multi-site study such as this involves planning at the level of the Principal Investigators, coordination, management and implementation at the Data Coordinating Center level, and implementation at the specific site, specific investigator level.

1. Data Analysis Planning

During the SEED 1 implementation period, the CADDRE PIs developed a process whereby SEED research analyses work groups were established, each led by one to two CADDRE PIs or co-PIs, to review each of the primary and secondary research domains (See Section I.D), further research and review the scientific literature on each topic included in these domains, review the specific data collected in SEED related to the primary and secondary research domains, and develop specific "principal" analyses within each primary research domain. In this way the CADDRE PIs were intentional in ensuring that the most important analyses -- those that directly pertained to the primary research domains SEED was funded to address -- were developed thoughtfully with input from a group of subject matter experts and that these analyses were given priority status in terms of data preparation. At the end of the work group process, all the CADDRE PIs convened to discuss the work group recommendations for principal analyses and collectively made the final decisions on which analyses would be considered "principal". In all, 28 principal analyses were developed through this process and lead authors were assigned.

Of NOTE: these SEED analytic work groups purposefully met and completed their work before the SEED 1 analytic data files were prepared; in fact these work groups met during the time period in which SEED 1 data were still being collected and entered. This was done such that principal analyses could be further developed and initiated as soon as the analytic data files became available. With the exception of a few ancillary data files containing genetic data, SEED 1 analytic data files were finalized in February, 2015. Work has begun on all principal analyses and the first two analyses have been completed – one has already been published (Wiggins, Levy, Daniels, et al., 2015) and the second has been submitted to a journal and is currently under review.

In addition to the work of the SEED research analyses work groups, a more permanent Data Sharing Committee (DSC) was also established during SEED 1. The DSC is chaired by the CDC Science Lead for CADDRE and includes 2 members from each SEED site. The purpose of the DSC is

1. To assure and expedite orderly and timely presentation to the scientific community of all pertinent data resulting from the collaborative Study to Explore Early Development (SEED);

2. To promote accurate and scientifically sound presentations and papers from the CADDRE Network and its collaborating investigators conducting SEED and analyzing SEED data;

3. To promote collaboration among CADDRE investigators and to assure that all participating investigators have the opportunity to be involved in data analysis and the preparation of CADDRE papers and presentations;

4. To assure that press releases, presentations, and publications related to CADDRE are accurate and objective, and do not compromise the collaborative study and the acceptance of its results;

5. To establish guidelines for authorship, acknowledgments, and funding citations for any presentations and publications of CADDRE.

The DSC has thus developed a SEED Analysis and Publications Guidelines manual that outlines procedures by which SEED investigators 1) propose and gain CADDRE Network approval for new analyses beyond the principal analyses (which were already approved by a separate process – See above); 2) form authorship groups that include interested SEED investigators across SEED sites for approved research analysis topics; 3) complete analyses in a timely manner; 4) submit all planned presentations and publications resulting from analyses of SEED data to the CADDRE Network for review prior to presentation or publication; 5) comply with all CDC and site-specific requirements (e.g. CDC clearance, site specific requirements for acknowledgments of certain data sources, etc.); 6) notify the CADDRE Network of single-site data analyses and ensure that these analyses do not violate the CADDRE requirement that SEED analyses shall not be limited to a single site's data if a multi-site data analysis is possible; 7) and obtain approval from the CADDRE Network to pursue funding from sources other than CDC (NIH, HRSA, etc.) to conduct SEED analyses including bio-specimen analyses.

The Data Sharing Committee has also formed a subgroup, the SEED Biomonitoring Committee (BMC). The BMC is comprised of one investigator from each CADDRE site, two additional scientists from CDC with specific expertise in genetic epidemiology and laboratory methods, and the CADDRE laboratory coordinator. Additionally, CDC or other sites might request ad hoc consultation for particular proposals from experts at their institutions. The Data Sharing Committee will refer any proposals from SEED investigators involving genotyping or other biosample analyses to the BMC for their review and recommendations.

Similarly the BMC will review requests for ancillary genetic or other biosample analyses from outside researchers sponsored by a CADDRE site PI (note such outside researchers must take all confidentiality training and adhere to all confidentiality procedures as core SEED investigators). The final recommendation from the BMC is then forwarded to the Data Sharing Committee for final review and formal approval. Once specific analyses are approved, study IRBs will be informed and addendums sent through review processes as needed. The BMC will give priority to candidate genotypes emerging from family-based linkage studies and genotypes that influence pathways also potentially affected by the environmental exposures for which the study has collected data. DNA is a depletable resource and the BMC may recommend in some instances that sequential analysis procedures (e.g., Kaaks et al., 1994) are used in order to preserve sample. These approaches involve the analysis of sample in small sets until there is sufficient evidence to either accept or reject a null hypothesis.

To date over 80 analytic topics (including the original 28 principal analyses) have been proposed and reviewed by the DSC; most have been approved and approval is expected for the others shortly (See Appendix R). Currently all approved analyses pertain to SEED 1 data. SEED 2 data collection is still underway; however, the same procedures and oversight by the DSC will be used to guide SEED 2 analyses, most of which will likely be analyses of the pooled SEED 1 and SEED 2 analytic data files. Likewise, at the completion of SEED 3, analytic data files will be prepared and the SEED 3 files will be combined with SEED 1 and SEED 2 files. The SEED DSC will continue to oversee all SEED analyses proposals and publications.

Additionally, planning is currently underway for the eventual development of SEED public use datasets. The development of public use datasets will adhere to the CDC and National Center on Birth Defects and Developmental Disabilities (NCBDDD) policies and procedures for de-identification (currently under development). All public use data files will be reviewed by the NCBDDD prior to release to ensure compliance with these de-identification policies.

2. Data Coordinating Center Responsibilities

a. Coordination

The SEED Data Coordinating Center has responsibility for coordinating information and maintaining a database for study data analyses. The DCC worked with CDC and the SEED sites during SEED 1 and SEED 2 to develop a centrally installed CADDRE Information System (CIS) to track participants, schedule visits, manage data entry, and to maintain the link to identifying information. The DCC also contracts with the Internet System for Assessing Autistic Children (ISAAC) for some of the data entry tools. ISAAC is a web-based application for administering and managing health research projects/studies. ISAAC is a third-party internet portal scoring service for copyrighted assessment instruments. ISAAC access requires a username and password and all communication with the ISAAC servers uses 128-bit SSL encryption, the highest available.

In SEED 1, ISAAC was employed as a functional subsystem of the CADDRE Information System (CIS) to provide data entry forms and scores for the copyrighted instruments used as part of the SEED battery of developmental assessments (e.g. ADOS, ADI-R). To save costs

in SEED 2 participant raw scores were data entered and scored directly in ISAAC. Participant date of birth was entered into ISAAC to calculate age at the time of assessment, and gender was also entered. No other participant personal identifying information was entered in ISAAC. Once all SEED data are transferred from ISAAC to DCC, the data in ISAAC will be expunged.

Since no data collection instruments are being revised between SEED Phase 2 and SEED Phase 3, but rather most instruments are being retained from SEED 2 in their entirety and a few instruments are being discontinued altogether, it is expected that the data collection system used in SEED 3 will largely be developed from the existing systems used in SEED Phase 2.

b. Data Preparation, Cleaning, Recoding, and New Variable Creation

The DCC works with the CDC to organize and prioritize the preparation of analytic data files. In summary: 1) at each site all data from the various SEED data collection instruments are entered in electronic capture engines administered by DCC and ISAAC with data transfer to DCC as indicated above); 2) DCC runs limited checks (developed by DCC and CDC with input from SEED sites) on both the dynamic data set (before all data are entered) and the static data set (once all data are entered) to assess major errors in data entry (e.g. an extreme proportion of missing data or out-of-range values). Based on these checks, sites are asked to review certain data against the paper forms that are available at the site and make corrections if needed. (NOTE: these types of corrections are limited to the most important data items and instruments – such as those used in defining cases status. Also, these types of checks can only

be conducted for those instruments for which the site has a paper copy); 3) once the aforementioned checks and corrections are complete, the DCC exports all of the data needed for analyses (i.e. Nearly all data collected other than participant tracking data) into analytic data files; 4) as part of the export process the DCC de-identifies all date data using a date shifting algorithm they developed and QC'd previously and they also remove any open string text field data elements that are inherently identifiable or deemed very likely to be identifiable (such as name of child's hospital of delivery); 5) the DCC runs various quality control checks on the data export process to ensure that the data items were correctly exported; 6) the DCC prepares documentation – including quality control report, data dictionaries, etc. – to accompany the data files; 7) the DCC uploads the analytic data files (nearly all de-identified) and accompanying documentation onto a remote data access server (RDA) for access by the site investigators. Note: the RDA meets CDC security requirements and is located in a HIPAA-compliant data center with full redundant power and security measures; 8) the DCC also delivers a complete set of the analytic data files to the CDC on an encrypted hard drive. The CDC stores these data in a secure location on SQL servers with limited access to the databases and identifiers are encrypted. CDC performs additional quality control checks to assess the data export process and ensure data harmonization from various SEED phases was correctly performed.

The DCC date shifting procedure and algorithm have been previously QC'd to ensure that the newly created shifted dates are not able to be linked back to the original dates. While the basic aspects of the general date shifting procedure is known to the SEED sites, the DCC does not share the exact algorithm, such that study staff from one SEED site are not able to

determine the original (correct) dates from other SEED sites' data that are included the analytic files.

Overall, DCC is responsible for full data preparation which includes: exporting data from CIS into SQL tables; shifting all dates using aforementioned algorithm that ensures that participants cannot be identified using various combinations of their date data; removing text field data elements, such as name, address, hospital of birth, etc., that are identifying or likely identifying; and preparing data dictionaries and other documentation. However, beyond the initial cleaning whereby sites are asked to go back to their original paper records and confirm or correct certain key data, the DCC does not conduct additional data cleaning activities. Rather, the raw data are included in the analytic data tables and analysts for various research analyses review the data for discrepancies, implausible values, etc. Analysts often create, new "cleaned" data variables but they do not write over the original raw data. New created (derived) variables are compiled into new data tables with appropriate documentation and shared on the RDA such that other SEED analysts are able to benefit from the data cleaning already undertaken.

As described above, all analytic data files will be shared with SEED analysts through the RDA maintained by the DCC with oversight from CDC. Specific personally identifiable information (PII) (such as names, address, phone numbers, etc.) will not be included in the data files on the RDA, with one very limited exception (see below). The DCC will work with CDC to ensure the data available through the RDA are de-identified to the extent possible– e.g. by using aforementioned date shifting procedures and removing text variables pertaining

to specific PII (such as name) and variables likely to be identifiable by reference to other data that might be available (such as child's hospital of delivery) from all analytic files. However, the analytic data files will not (and cannot) be considered completely de-identified for several reasons:

- Some genetic data derived from analyses of biosamples will be compiled into analytic files and will be shared with SEED investigators through a second RDA server maintained by the DCC. Because genetic data are unique to individuals, they are not considered de-identified. Nonetheless, CDC wants to ensure that these important data are available to the SEED investigators to maximize their use in research analyses. The RDA approach offers a secure platform for these analyses. The second server will include all security provisions as the original RDA. Additionally, access to the genetic data files will be more restrictive. Only investigators with a clearly identified need to use the genetic data in their analyses will be provided access.
- The address data collected on the Maternal and Child Residence History form will not be included directly on the RDA. However, each site will be asked to allow DCC to derive various geocode variables from the data collected on this form and/or to themselves derive various geocode variables. If sites derive geocoded variables themselves, these data will be sent to DCC following an encryption protocol and the DCC will compile the information. The DCC will provide access to the geocode data on a designated restricted folder on one of the DCC RDA servers to individual SEED investigators who require access for data linkage purposes. The investigator will have to demonstrate a specific need for these sensitive data and that proposed analysis must be approved by the CADDRE/SEED Data Sharing Committee. DCC will provide

such investigators time-limited access to the geocoded data such that they can link the data with other data files such as environmental monitoring data or Census data. No data will be downloaded from the RDA. Once the linkage occurs, the geocodes will be stripped from the linked data file. This de-identified linked data file will contain participant study ID. DCC will move this de-identified file to the main section of the RDA whereby all SEED investigators needing such data will have access and can combine these data with other SEED data.

- Beyond genetic and geocode data, the composite SEED data files will contain up to 20,000 variables per participant for participant families. Although the major issues related to identifiability will be addressed, with this volume of data it is not possible to ensure that combination of data about a research participant will not result indirectly, by reference to other information to identification of study participants.

3. Study Analysis Categories and Basic Analytic Approaches

SEED will provide an unprecedented array of data on ASD cases and two comparison groups. Analyses can be crudely classified in the following categories: 1) characterization of phenotype (which includes case-only analyses and case-comparison group contrasts); 2) estimation of risk factor associations (includes evaluations of heritable and non-heritable risk factors, assessment of specificity of associations, and assessments of interactions); 3) comparison of biomarkers across ASD, DD, and POP groups; and 4) comparison of child health conditions across ASD, DD, and POP groups.

Recall that the final status of the ASD group subjects is based on the results of the children's developmental evaluation administered as part of the study data collection. In some instances, (e.g. refusal, loss to follow-up) these subjects may not complete the study's developmental evaluation and, thus, will be classified as Indeterminate ASD for analysis purposes. Consequently, analyses may be performed separately on subjects classified on the basis of the developmental evaluation and subjects classified on the basis of the ASD screen.

Both case-only analyses and case-comparison group contrasts will be conducted. Case-only analyses are primarily designed to identify novel, specific phenotypic subgroups in ASD, while the case-comparison group contrasts assess the specificity of an independent factor of interest with ASD – overall or by phenotypic subgroup – relative to the DD and POP. A priori, we may consider the following ASD subgroups for analysis:

- with (30%)/without (70%) regression
- with (40%)/without (60%) intellectual disability
- complex (20%)/essential (80%) autism
- verbal (70%)/nonverbal (30%)

These are not mutually exclusive categories, however, and one of our goals is to explore the utility of more complex combinations that include multiple features and may be potentially etiologically distinct, phenotypic subgroups.

Further, analyses may consider stratification on common variables (e.g., gender, gestational age) across all 3 subject groups.

Although the primary unit of analysis will be the index child, for some analyses classification of affected/unaffected status (or exposed/unexposed) may include criteria that consider

diagnosed or reported medical, neurologic, and developmental conditions in parents and/or siblings.

4. Data Sharing

As described above under Data Analysis Planning, both prioritized principal analyses and additional analyses of primary and secondary research domain topics have been developed for SEED 1 data and many are underway. Affiliated investigators (SEED site investigators or their designated colleagues/ students/ collaborators) can apply to the Data Sharing Committee (DSC) for permission to use multi-site pooled SEED data to complete analyses. The Data Sharing Committee receives and tracks these letters of intent (LOIs). Each LOI is reviewed for approval by the Data Sharing Committee. Detailed analysis and publication guidelines for CADDRE have been developed and approved by the CADDRE Data Sharing Committee. For each phase of SEED (1, 2, and 3), each site who contributes data to the pooled analytic dataset will have representation on the committee and will be included in decision-making for projects that use their sites' data. The use of the collaborative study data will initially be limited to the SEED investigators.

Affiliated SEED investigators and their collaborators who are given access to clinical, interview and biologic data from SEED must sign a confidentiality and data use oath that describes how the data should be used and stored. SEED data analysts will be required to follow the CDC-approved mechanism for data access, which will ensure that all CDC-required data security standards are maintained. Currently, the approved mechanism for SEED data access is through the RDA maintained by the DCC. All analyses, with the

exception of limited genetic analyses, are conducted on the RDA and aggregate results are downloaded but individual-level data are not. Previously, certain genetic analyses were conducted on a secure server outside of the RDA at Johns Hopkins University because of data space and analytic requirements. As described above, DCC will now maintain a second RDA server that is able to accommodate the capacity of the genetic data files. Also as described above, the DCC and CDC have developed a mechanism for limited access to geocoded data derived from participant residence address data. Of note: access to both genetic data and geocoded data will be through a mechanism as secure as the more general RDA but will be limited to much fewer individuals with a demonstrated need for the data. Access to the geocoded data for the purpose of data linkages will be time limited and once the linkage is complete the linked data will be de-identified by removing the geocodes.

The Principal Investigator of each SEED site has full and direct responsibility for tracking the use of SEED data at their site and assuring that each person who has access to the data has read and signed the confidentiality and data use oath. Each site should maintain files of the signed confidentiality and data use oaths. Signed statements will also be kept on file at the DCC; CDC will have access to these files as needed. It will be left to the discretion of the individual sites to determine when the statements should be renewed for specific individuals or projects.

The data will become available to outside researchers in line with CDC/ATSDR policies for the release of de-identified data to outside investigators. CDC is required to make data generated using CDC funding available for interested researchers as a de-identified public use

dataset or restricted access dataset. The specific policy and de-identification implementation guidelines will be developed by CDC and shared with the DCC and CADDRE DSC. A data release plan that adheres to the CDC policy will be developed with input from the CADDRE DSC and the DCC.

For SEED 2, specific consent language was included in the written consent forms to permit, if signed by the participant, the participant's de-identified data from SEED 2 to be placed in the National Database for Autism Research (NDAR) and the Database for Genotypes and Phenotypes (dbGaP). Both databases are run by the National Institutes of Health (NIH) that allows researchers studying autism to easily share and pool information with each other. To protect participant privacy, all identifying information will be removed and replaced with an NDAR specific code number.

Consent for participant's genetic data to be placed in NIH data sharing repositories is also included in SEED 3 consent forms for the ASD and POP groups. However, the language was revised in accordance with the evolving understanding of the identifiability of genetic data. The latest guidance from the National Human Genome Research Institute at NIH states:

> "Sharing individual and even summary-level genomic data carries some degree of privacy risk to study participants. When data will be shared, researchers should explain how privacy and confidentiality will be protected.
>
> A primary privacy concern in genomics research is re-identification. Researchers have shown that subjects can be re-identified by combining de-identified genomic information with other information types that are publicly available, and that individual subjects sometimes can be distinguished even in summary-level genomic data. Current federal regulations do not classify genomic information as "identifiable," but a proposal to revise the Common Rule released in 2011 discussed the identifiability of genomic data.

> Data repositories may protect against the possibility of re-identification by controlling access to the data and requiring data users to agree not to attempt to re-identify research participants. However, it is not possible to eliminate completely the risk of re-identification. Researchers should explain this risk to participants' privacy and confidentiality and note whether there may be related unanticipated risks in the future."
>
> http://www.genome.gov/27559024

We carefully reviewed our SEED 2 consent form language and revised it in accordance with this guidance for SEED 3 such that parents are fully informed of risks and benefits when providing consent for genetic data sharing.

L. Provisions for Protecting Privacy and Confidentiality of Human Subjects

All data on an individual participating in the study will remain confidential at all times. A Federal Certificate of Confidentiality will be obtained from the Department of Health and Human Services which, in most circumstances, affords the study and the participant additional protection from involuntary disclosure of information collected in the study.

Participants will be interviewed on multiple occasions and will be asked some questions that are potentially sensitive, including items on alcohol use during pregnancy, adverse pregnancy outcomes, family medical history, family income questions, and other lifestyle questions. The interviews have some risk of psychological discomfort, but women will be told at the beginning of each interview that they may choose to skip any question at any time during the interview. In addition, we will accommodate the participants' wishes with regards to the timing of the interviews and will hire interviewers who are sensitive to the well-being of participants who are emotionally vulnerable. Extensive training will be held with all interviewers to address these issues.

The exposure of the identity of study participants will be avoided wherever feasible in the study workflows. A 12- digit identification number will be employed to encode the participant identity on data collection forms, specimens, and various other study materials. The linkage of the identification number and the participant personal identifiers is provided by the CADDRE Information System (CIS). All efforts will be taken to ensure that all proposed research methods comply with human subjects requirements.

The role-based security system of the CIS ensures that only those personnel who have prior authorization to access personal identifiers have the capability of associating an identification number with a participant's identity. The entire CIS database will be encrypted. The encryption implementation method will be approved by the CDC. Encrypting the database will preclude access to personally identifying information in the event that an intruder penetrates the server security. No information concerning the operation of the encryption system will be stored on the CIS servers. Proper firewall and network security configurations will be maintained. All applicable security updates of comprising commercial software components will be applied to the servers rapidly after release by the software vendors. The CIS complies with all current (and future) guidelines required by the CDC for such web-based applications. All data in the CIS data repository will be maintained on dedicated secured servers. CIS production servers will be housed in physically secure professional server rooms. The CIS incorporates all appropriate system security practices and methodologies for the CIS applications and database servers to protect the data and maintain the confidentiality of the participant information.

Biologics samples will be stored in two ways or destroyed at the end of the study, based on a choice by the study participants. The first way of storing the samples would keep them linked to personal information (through a study ID). This will allow study investigators, or other researchers approved by the study team, to contact participants again in the future. Participants who agree to have a sample stored with the study ID link intact are informed that they are only agreeing to potentially being contacted for future studies (which will require additional consent from participant). They will also be told they have the option to request this link be broken in the future, and are requested to do this by sending a written, signed letter to the study staff. Study participants will also have the option to store their samples without a link to personal identifiers. Under this approach the link between the participant's study ID and their biologic samples will be destroyed at the end of the study. This way their samples and the information given for other parts of the study could be used for future analyses of child development, but researchers would not be able to add any new information. Participants can also request to have their biologics samples destroyed at the end of the study. Under this approach, the sample would not be stored for future studies.

As described under the data sharing section of this protocol, all additional studies and investigations involving any data collected under SEED would need to first be approved by the Data Sharing Committee. After this approval, additional IRB review will be required for any research involving collaborators outside the CDC/SEED group and also for any studies which fall outside the scope of the current protocol.

In the future, the data will become available to outside researchers in line with CDC/ATSDR policies for the release of de-identified data to outside investigators. CDC is required to make data generated using CDC funding available for interested researchers as a de-identified public use dataset or restricted access dataset. The specific policy and de-identification implementation guidelines will be developed by CDC and shared with the DCC and CADDRE DSC. A data release plan that adheres to the CDC policy will be developed with input from the CADDRE DSC and the DCC.

M. Procedures for Avoiding Bias

To avoid measurement bias, an attempt will be made to minimize knowledge of case or comparison group status of all subjects by interview and data collection staff (i.e., no intentional disclosure of group status to staff), although, it may become apparent to study staff during data collection. Interviewers will receive training in the proper conduct of an interview, emphasizing the importance of standardizing the means by which questions are asked and responses are recorded. An additional important consideration is the potential for recall bias in the interview; the DD group was designed to help mitigate this problem.

For the in-person clinical visit, developmental examiners will administer all assessments before assessing degree of certainty the child has an ASD. Due to the nature of the information collected on each participant, it will be impossible to blind the developmental examiner to the case status of the child. However, all developmental examiners will be trained to research-reliable standards in standardized instrument and clinical assessment procedures.

Selection bias (including referral bias) is a potential issue, since the POP children most likely to participate might be those whose parents have concerns about development. This bias can be assessed by examining differences between participants and non-participants using variables available on the birth certificate. Selection bias could also be introduced since those children most severely affected by ASD are more likely to have been diagnosed previously. To mitigate this problem, the broad diagnostic net will be cast to identify previously unidentified cases, and cases will be obtained from multiple sources.

Misclassification bias is potentially an issue, but will be largely avoided for cases by the use of a stringent definition of ASD using standardized instruments and training examiners using standardized methods.

IV. Dissemination of Results

A. Notifying Parents of Individual Results and Study Findings and Other Communication with Parents

Mothers of all children in the study will receive feedback letters that explain the results of relevant developmental assessments that were administered to her and her child (parental developmental feedback letter can be found in Appendix Q.1). The feedback letter will describe each assessment that was administered, will provide scores for each assessment, and will offer a general interpretation of the scores. The letter will also include contact information for a study clinician if the parent has additional questions that are not answered in the feedback letter. A list of community referral resources will also be provided so that parents can schedule further developmental evaluation if desired. In addition to the

developmental feedback letter sites may send parents of children in the study additional letters and cards: specifically, thank you letters or cards (See Appendix Q.2 (Thank you letter for participants who didn't complete the study), Q.3 (Thank you letter for participants who completed the study) and Q.5 (general thank you card or acknowledgement of other occasion [e.g. Mother's Day]); and an age-out letter (Appendix Q.6) in the event that the child ages past the target age range for the developmental evaluation instruments before completing the clinic visit.

Individual laboratory results will not be provided since little is known about which genes and other biologic substances are associated with autism and child development and most of the findings from this study will be initial leads (hypothesis generating). Similar findings will have to be seen again in other studies before the information could be considered useful to any individual child or their family. If by the end of the study, one or more of the measures becomes clinically significant, the participants will be contacted and will be given the opportunity to receive these results, and the protocol will be revised to reflect this change.

B. Dissemination of Results to the Public

The Project Coordinators at each SEED site will themselves participate or will delegate someone at their site to participate in the Outreach Coordinators Working Group (OCWG) of SEED. The OCWG will have responsibility for reviewing the content and process of the information distributed to the subjects and public. Announcements will be made to the general public at key junctures in the project, for example, when recruitment begins, when recruitment ends, and when study results are available. Subjects will be kept informed of the progress of the study on a regular basis.

1. Dissemination to the Public

When the study has been approved by the IRB's of CDC and all sites a press release will be prepared by the OCWG and reviewed by the CDC and all investigators to be distributed to broadsheet and broadcast news agencies, host institutions, and sources. For example, recipients will include but not be limited to local and regional newspapers, source newsletters, parent support groups, autism advocacy groups, and public service announcers. The press release will report general information about the project such as the name of the study, the sites involved, the study purpose, and general timeline. Additional news briefs will be prepared over the course of the study by the investigators if they determine there is information that may be of interest to and can be reported to the general public.

2. Dissemination to Subjects and Other Stakeholders

Subjects, sources, and other stakeholders will be offered the opportunity to receive study updates. Those who agree or request this opportunity will be placed on a post/electronic mailing list prepared and maintained by the Project Coordinators at each site. The mailing list will be composed of names and addresses, only. These individuals will be sent an All States SEED newsletter two times a year that contains information of interest to subjects and sources in particular. The newsletter may include a tally of subjects recruited at all sites, descriptions of public relations activities at all sites, descriptions of educational activities at all sites, or a description of activities conducted at one site that has relevance for all sites. The sites will share responsibility for producing at least one issue of a four page newsletter that reports on general activities related to the project as it is conducted in all states. The newsletter will have the title of the study as a header and list all participating sites. Prior to sending the newsletter

each site (including CDC) will receive approval from their respective IRB, Public Relations Department, or other clearance process as outlined at the site.

V. PROTECTION OF HUMAN RESEARCH PARTICIPANTS

A. Participant Risks and Methods to Minimize Risks

The risks to study participants are expected to be minimal. Provisions for protecting privacy and confidentiality of study subjects will be strictly maintained. Study results will always be presented in aggregate form, thereby further preventing identification of individual subjects.

Participants may be uncomfortable answering certain questions during the interviews (i.e., questions relating to sexually transmitted diseases or alcohol use). However, respondents will be reminded that they maintain the option of not answering any individual question. Respondents are assured that all information that they provide will be analyzed to identify characteristics that are common to many or all cases, no information will be published in any manner that would personally identify them, and all information is regarded as medically privileged information.

Risks involved with the developmental evaluation are also expected to be minimal. There is a chance that families may learn for the first time that their child is displaying developmental delays, which could be a stressful experience for the family. Thus, parents of children with no previous ASD diagnosis or special education classification will receive a call from a study clinician (prior to being sent the study feedback letter) if the results of the developmental evaluation suggest behaviors consistent with an ASD. The feedback letter, sent to all

participants who complete a developmental evaluation, encourages parents to share the results with their child's education and medical providers. The feedback letter also states that the evaluation was conducted for research and should not be used to determine clinical diagnoses. A list of community resources will also be provided with the letter. All recipients are given contact information for a study clinician if they have any further questions. We will take all necessary precautions to prevent undue distress. We believe the plan described here, based on past experiences with this type of research, will adequately minimize any negative psychological effects families may experience.

Participants may experience minor discomfort when having their blood drawn for biosampling procedures. In order to minimize discomfort, all blood draw procedures will only be performed by trained personnel and topical anesthetic will be provided for the children. All blood will be drawn using universal precautions, and safety measures will be taken in order to prevent infection and bruising at the puncture site for participants donating blood samples. Child participants may experience minor discomfort when having the saliva collected. Stress from this procedure will be reduced by having the parent obtain the samples at the optimal time for the child. Parents will be provided with clear instructions to assure safe and appropriate saliva collection techniques.

B. Handling of Unexpected Adverse Events

As stated above, the risks to potential participants are expected to be minimal and precautions will be taken to minimize all risks. All unexpected adverse events will be reported by the sites to the Project Officer at CDC and the IRB at individual institutions at each site. A log of

all unexpected adverse events will be kept at CDC and at each site for tracking these events and the potential impact on the study.

C. Injury Compensation

Although this study was determined to pose no greater than minimal risk to participants, clinical research often involves the chance that something unexpected may occur. This may include the risk of personal injury. In spite of all safety measures implemented, a participant might develop a reaction or injury from the various procedures (i.e. blood and/ or saliva collection) that are a part of this study. Some of the participants may experience minor discomfort when having their blood drawn for biosampling procedures. In order to minimize discomfort, all blood draw procedures will only be performed by trained personnel and topical anesthetic will be provided for the children. All blood will be drawn using universal precautions, and safety measures will be taken in order to prevent infection and bruising at the puncture site for participants donating blood samples. These are the physical injuries that could occur. If such issues occur, the researchers will refer participants to a medical provider, but any costs for the medical care will be billed to the participant and/or the participant's insurance company. Neither the Centers for Disease Control and Prevention nor any of the collaborating institutions in this study have set aside funds to pay participants for any such reactions or injuries, or for the related medical care. However, by participating in this study, participants are not asked to relinquish any of their legal rights. Some sites are specifically informing participants that should any of the aforementioned discomforts or injuries occur (that are stated above) then the project could <u>not</u> provide any compensation for emergency

medical expenses, or treatment for the injury without payment from the participant or their insurer.

D. Participant Benefits

Participants will be periodically updated on the overall progress of the study and results through a SEED newsletter will be mailed to their home if they choose to be included on the mailing list. Science and society in general will benefit from our improved understanding of the causes of developmental problems. This information may lead to possible intervention and prevention strategies for developmental problems. These benefits are believed to outweigh the inconvenience and minimal risk to individual participants.

E. Vulnerable Populations

Since childhood developmental disabilities is the focus of the research, it is necessary to include children, including those with mental disabilities, as study participants. Because of the young age of the enrolled children (e.g., less than five years of age), their assent will not be sought. The consent of the child's mother will be sought, and no child will be enrolled without such consent. In addition, the research is minimal risk to the children. Provisions for protecting privacy and confidentiality of study subjects will be strictly maintained. Study results will always be presented in aggregate form, thereby further preventing identification of individual subjects.

F. Costs and Incentives to Participants

Successful completion of this study depends on our ability to recruit and retain the majority of invited subjects. Mothers will be asked to complete multiple data collection steps including questionnaires, a telephone interview, one or more clinic or home visits, and biosampling including blood and saliva. Children participate in a set of evaluations, videotaping of behaviors, and biosampling including blood and saliva. We estimate time burden to be 3.75 hours for data collection from mother-child pairs in the DD workflow, 5.5 hours for data collection from mother-child pairs in the POP workflow and an average of up to 9 hours for data collection from mother-child pairs in the ASD (parent and child) workflow. Due to the multiple components of data collection, an additional consideration is the potential difficulty we may have retaining subjects once the study has begun.

Time Incentives to Participants

Individuals in the ASD and DD, and POP workflow groups will be provided incentives as motivation to participate in the study while completing specific data collection steps. The incentive process is shown in Tables 7, 8, and 9:

Table 7: Incentives for ASD workflow mother-child pairs

GROUP: ASD			
	Data Collection Step*	**Time to Complete**	**Time Incentives**
1	Invitation Letter		
2	Screening and Invitation Phone Call (with the SCQ)	30 min.	
3.	Enrollment Packet		
4.	Follow Up Phone Call 1 (includes completing Pregnancy Reference Form)	15 min	$30
5.	Maternal Interview Call	60 min	$30
6.	Clinic Visit Preparation, Medical History and Child Development Forms Packet	95 min	$40
7.	Follow Up Phone Call 2 (includes answering questions and providing help completing Medical History and Child Development Forms)	20 min	
8.	Clinic/Home Visit	330 min	$200
	TOTAL	**9 hours, 10 minutes**	**$300**

***refer to Appendix H for detailed study flow diagram*

Table 8: Incentives for POP workflow mother-child pairs

GROUP: POP			
	Data Collection Step*	**Time to Complete**	**Time Incentives**
1	Invitation Letter		
2	Screening and Invitation Phone Call (with the SCQ)	30 min.	
3.	Enrollment Packet		
4.	Follow Up Phone Call 1(includes completing the Pregnancy Reference Form)	15 min	$30
5.	Maternal Interview Call	60 min	$30
6.	Clinic Visit Preparation, Medical History and Child Development Forms Packet	95 min	$40
7.	Follow Up Phone Call 2 (includes answering questions and providing help completing Medical History and Child Development Form)	20 min	
8.	Clinic/Home Visit	110 min	$75
	TOTAL	**5 hours, 30 minutes**	**$175**

***refer to Appendix H for detailed study flow diagram*

Table 9: Incentives for DD workflow mother-child pairs

GROUP: DD			
	Data Collection Step*	**Time to Complete**	**Time Incentives**
1	Invitation Letter		
2	Screening and Invitation Phone Call (with the SCQ)	30 min.	
3.	Enrollment Packet		
4.	Follow Up Phone Call 1(includes completing the Pregnancy Reference Form)	15 min	$30
5.	Maternal Interview Call	60 min	$30
6.	Medical History Forms Packet (4 forms only)	60 min	$40
	TOTAL	**2 hours, 45 minutes**	**$100**

***refer to Appendix H for detailed study flow diagram*

The investigators recognize that all subjects may not participate in all phases of data collection. Subjects may choose to complete all or some of the measures. We have created a schedule of incentives for subjects to simplify the administrative tasks associated with providing incentives. Subjects will be provided incentives primarily in a step-wise fashion in which completion of one phase on data collection is followed by an incentive; the incentive for the clinic visit is provided at the beginning of the visit.

While incentives have been applied as monetary value it may take a variety of forms such as cash, or store voucher, or donation to a local community center or autism organization. All methods of disbursement will be negotiated with business departments at each site. To protect subject confidentiality during check disbursement, the respective business offices will not have information that links subject's name or social security number to the study ID number.

The final incentive will be given to subjects prior to the last data collection component. There will be no other incentives beyond this date and the above amounts.

Each site will track phases of data collection and incentives by recording the time, purpose, and the amount disbursed to each subject. Each site will record information required for tax exemption by the federal government and the institution and all disbursements will be reconciled at the end of the study period. Tracking information will be stored in the project coordinator's files at each site.

G. Certificate of Confidentiality

Due to the sensitive nature of components of the maternal interview, SEED has obtained additional confidentiality protections. Discussions were held with the CDC Confidentiality Officer, who reviewed a draft of the SEED protocol. Judgment was that formal confidentiality protection would increase the likelihood of valid responses to the sensitive information areas being explored. A consolidated application for 301(d) Certificates of Confidentiality for the six sites conducting SEED Phase 2 was requested and received by NCBDDD. A consolidated application for 301(d) Certificates of Confidentiality for the six sites conducting SEED Phase 3 will also be requested.

H. Internal and External Reviews and Approvals

Upon approval by the CDC IRB panel, this protocol will be approved by the IRB of each of the other five collaborating centers as well as any additional IRBs as required through the collaboration (for example, schools, universities, departments of health). Once these approvals have been obtained, all documents supporting these approvals will be included as an addendum to this protocol.

References

Achenbach, T. M., & Edelbrock, C. (1981). Child behavior checklist. *Burlington, VT*.

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the Social Communication Questionnaire in Assessing Risk of Autism in Preschool Children with Developmental Problems. *Journal of Autism and Developmental Disorders*, *37*(7), 1272–1278. http://doi.org/10.1007/s10803-006-0279-7

Atladóttir, H. Ó., Thorsen, P., Østergaard, L., Schendel, D. E., Lemcke, S., Abdallah, M., & Parner, E. T. (2010). Maternal Infection Requiring Hospitalization During Pregnancy and Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *40*(12), 1423–1430. http://doi.org/10.1007/s10803-010-1006-y

Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveillance Summaries*, *63*(2).

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine*, *25*(01), 63–77.

Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: The phenotype in relatives. *Journal of Autism and Developmental Disorders*, *28*(5), 369–392.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *The Lancet*, *368*(9531), 210–215.

Barton, M., & Volkmar, F. (1998). How commonly are known medical conditions associated with autism? *Journal of Autism and Developmental Disorders*, *28*(4), 273–278.

Bhasin, T. K., & Schendel, D. (2007). Sociodemographic Risk Factors for Autism in a US Metropolitan Area. *Journal of Autism and Developmental Disorders*, *37*(4), 667–677. http://doi.org/10.1007/s10803-006-0194-y

Buie, T., Fuchs, G. J., Furuta, G. T., Kooros, K., Levy, J., Lewis, J. D., … Winter, H. (2010). Recommendations for Evaluation and Treatment of Common Gastrointestinal Problems in Children With ASDs. *PEDIATRICS*, *125*(Supplement), S19–S29. http://doi.org/10.1542/peds.2009-1878D

Chaidez, V., Hansen, R. L., & Hertz-Picciotto, I. (2014). Gastrointestinal Problems in Children with Autism, Developmental Delays or Typical Development. *Journal of Autism and Developmental Disorders*, *44*(5), 1117–1127. http://doi.org/10.1007/s10803-013-1973-x

Chess, S. (1971). Autism in children with congenital rubella. *Journal of Autism and Childhood Schizophrenia*, *1*(1), 33–47.

Close, H. A., Lee, L.-C., Kaufmann, C. N., & Zimmerman, A. W. (2012). Co-occurring Conditions and Change in Diagnosis in Autism Spectrum Disorders. *PEDIATRICS*, *129*(2), e305–e316. http://doi.org/10.1542/peds.2011-1717

Comi, A. M., Zimmerman, A. W., Frye, V. H., Law, P. A., & Peeden, J. N. (1999). Familial clustering of autoimmune disorders and evaluation of medical risk factors in autism. *Journal of Child Neurology*, *14*(6), 388–394.

Constantino, J. N., & Gruber, C. P. (2002a). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

Constantino, J. N., & Gruber, C. P. (2002b). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

Constantino, J. N., & Todd, R. D. (2014). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*. Retrieved from http://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.157.12.2043

Cook, E. H. (1990). Autism: review of neurochemical investigation. *Synapse*, *6*(3), 292–308.

Corsello, C., Cook, E., & Leventhal, B. (2003). A screening instrument for autistic spectrum disorders. In *Biennial Meeting of the Society for Research in Child Development, Miami, FL*.

Couteur, A., Bailey, A., Goode, S., Pickles, A., Gottesman, I., Robertson, S., & Rutter, M. (1996). A broader phenotype of autism: the clinical spectrum in twins. *Journal of Child Psychology and Psychiatry*, *37*(7), 785–801.

Croen, L. A., Grether, J. K., & Selvin, S. (2002). Descriptive epidemiology of autism in a California population: who is at risk? *Journal of Autism and Developmental Disorders*, *32*(3), 217–224.

Daniels, W., Warren, R., Odell, J., Maciulis, A., Burger, R., Warren, W., & Torres, A. (1995). Increased Frequency of the Extended or Ancestral Haplotype B44-SC3O-DR4 in Autism. *Neuropsychobiology*, *32*(3), 120–123.

Deykin, E. Y., & MacMahon, B. (1979). The incidence of seizures among children with autistic symptoms. *The American Journal of Psychiatry*.

Durkin, M. S., Maenner, M. J., Meaney, F. J., Levy, S. E., DiGuiseppi, C., Nicholas, J. S., … Schieve, L. A. (2010). Socioeconomic inequality in the prevalence of autism spectrum disorder: evidence from a U.S. cross-sectional study. *PLoS One*, *5*(7), e11551. http://doi.org/10.1371/journal.pone.0011551

Eaton, W. W., Mortensen, P. B., Thomsen, P. H., & Frydenberg, M. (2001). Obstetric

complications and risk for severe psychopathology in childhood. *Journal of Autism and Developmental Disorders*, *31*(3), 279–285.

Eaves, L. C. (2006). Screening for autism: Agreement with diagnosis. *Autism*, *10*(3), 229–242. http://doi.org/10.1177/1362361306063288

Eaves, L. C., Wingert, H. D., Ho, H. H., & MICKELSON, E. C. (2006). Screening for autism spectrum disorders with the social communication questionnaire. *Journal of Developmental & Behavioral Pediatrics*, *27*(2), S95–S103.

Efficient_Use_of_Biological_Banks_for_Biochemical.9.pdf. (n.d.).

Eufemia, P. d', Celli, M., Finocchiaro, R., Pacifico, L., Viozzi, L., Zaccagnini, M., … Giardini, O. (1996). Abnormal intestinal permeability in children with autism. *Acta Paediatrica*, *85*(9), 1076–1079.

Funderburk, S. J., Carter, J., Tanguay, P., Freeman, B. J., & Westlake, J. R. (1983). Parental reproductive problems and gestational hormonal exposure in autistic and schizophrenic children. *Journal of Autism and Developmental Disorders*, *13*(3), 325–332.

Ganea, D., & Delgado, M. (2002). Vasoactive intestinal peptide (VIP) and pituitary adenylate cyclase-activating polypeptide (PACAP) as modulators of both innate and adaptive immunity. *Critical Reviews in Oral Biology & Medicine*, *13*(3), 229–237.

García-Closas, M., Egan, K. M., Abruzzo, J., Newcomb, P. A., Titus-Ernstoff, L., Franklin, T., … others. (2001). Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiology Biomarkers & Prevention*, *10*(6), 687–696.

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., … Pathways

in ASD Study Team. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach: ASD factor mixture model. *Journal of Child Psychology and Psychiatry*, *54*(2), 206–215. http://doi.org/10.1111/j.1469-7610.2012.02588.x

Geschwind, D. H. (2011). Genetics of autism spectrum disorders. *Trends in Cognitive Sciences*, *15*(9), 409–416. http://doi.org/10.1016/j.tics.2011.07.003

Geschwind, N., & Behan, P. O. (1984). Laterality, hormones, and immunity. *Cerebral Dominance: The Biological Foundations*, 211–224.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised Algorithms for Improved Diagnostic Validity. *Journal of Autism and Developmental Disorders*, *37*(4), 613–627. http://doi.org/10.1007/s10803-006-0280-1

Gupta, S., Aggarwal, S., Rashanravan, B., & Lee, T. (1998). Th1-and Th2-like cytokines in CD4+ and CD8+ T cells in autism. *Journal of Neuroimmunology*, *85*(1), 106–109.

Hallmayer, J. (2011). Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Archives of General Psychiatry*, *68*(11), 1095. http://doi.org/10.1001/archgenpsychiatry.2011.76

Hanson, E., Sullivan, N., Ware, J., Lord, C., & Thum, A. (2002). Social communication questionnaire (SCQ). In *International Meeting for Autism Research*.

Höglund Carlsson, L., Norrelgen, F., Kjellmer, L., Westerlund, J., Gillberg, C., & Fernell, E. (2013). Coexisting Disorders and Problems in Preschool Children with Autism Spectrum Disorders. *The Scientific World Journal*, *2013*, 1–6. http://doi.org/10.1155/2013/213979

Hönekopp, J. (2012). Digit Ratio 2D:4D in Relation to Autism Spectrum Disorders, Empathizing, and Systemizing: A Quantitative Review: Digit ratio 2D:4D and autism. *Autism Research*, *5*(4), 221–230. http://doi.org/10.1002/aur.1230

Horvath, K., Papadimitriou, J. C., Rabsztyn, A., Drachenberg, C., & Tildon, J. T. (1999). Gastrointestinal abnormalities in children with autistic disorder. *The Journal of Pediatrics*, *135*(5), 559–563.

Insel, T. R., O'Brien, D. J., & Leckman, J. F. (1999). Oxytocin, vasopressin, and autism: is there a connection? *Biological Psychiatry*, *45*(2), 145–157.

Ivarsson, S., Bjerre, I., Vegfors, P., & Ahlfors, K. (1990). Autism as one of several disabilities in two children with congenital cytomegalovirus infection. *Neuropediatrics*, *21*(2), 102–103.

Johnson, W. G., Buyske, S., Mars, A. E., Sreenath, M., Stenroos, E. S., Williams, T. A., … Lambert, G. H. (2009). HLA-DR4 as a risk allele for autism acting in mothers of probands possibly during pregnancy. *Archives of Pediatrics & Adolescent Medicine*, *163*(6), 542–546.

Jyonouchi, H., Sun, S., & Le, H. (2001). Proinflammatory and regulatory cytokine production associated with innate and adaptive immune responses in children with autism spectrum disorders and developmental regression. *Journal of Neuroimmunology*, *120*(1), 170–179.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*.

Korvatska, E., Van de Water, J., Anders, T. F., & Gershwin, M. E. (2002). Genetic and Immunologic Considerations in Autism. *Neurobiology of Disease*, *9*(2), 107–125. http://doi.org/10.1006/nbdi.2002.0479

Krause, I., He, X.-S., Gershwin, M. E., & Shoenfeld, Y. (2002). Brief report: immune factors in autism: a critical review. *Journal of Autism and Developmental Disorders*, *32*(4), 337–345.

Le Couteur, A., Lord, C., & Rutter, M. (2003). The autism diagnostic interview-revised (ADI-R). *Los Angeles, CA: Western Psychological Services*.

Lee, L.-C., David, A. B., Rusyniak, J., Landa, R., & Newschaffer, C. J. (2007). Performance of the Social Communication Questionnaire in children receiving preschool special education services. *Research in Autism Spectrum Disorders*, *1*(2), 126–138. http://doi.org/10.1016/j.rasd.2006.08.004

Lee, L.-C., Zachary, A. A., Leffell, M. S., Newschaffer, C. J., Matteson, K. J., Tyler, J. D., & Zimmerman, A. W. (2006). HLA-DR4 in Families With Autism. *Pediatric Neurology*, *35*(5), 303–307. http://doi.org/10.1016/j.pediatrneurol.2006.06.006

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., … Rice, C. E. (2010). Autism Spectrum Disorder and Co-occurring Developmental, Psychiatric, and Medical Conditions Among Children in Multiple Populations of the United States: *Journal of Developmental & Behavioral Pediatrics*, *31*(4), 267–275. http://doi.org/10.1097/DBP.0b013e3181d5d03b

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2002). *Autism diagnostic observation schedule: ADOS*. Western Psychological Services Los Angeles, CA.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism Diagnostic Observation Schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a

revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685.

Lotter, V. (1967). Epidemiology of autistic conditions in young children. *Social Psychiatry*, *1*(4), 163–173.

Manning, J., & Bundred, P. (2000). The ratio of 2nd to 4th digit length: a new predictor of disease predisposition? *Medical Hypotheses*, *54*(5), 855–857.

Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behavior, and health*. Rutgers University Press.

Mazzocchi, G., Malendowicz, L. K., Rebuffat, P., Gottardo, L., & Nussdorfer, G. G. (2002). Expression and function of vasoactive intestinal peptide, pituitary adenylate cyclase-activating polypeptide, and their receptors in the human adrenal gland. *The Journal of Clinical Endocrinology & Metabolism*, *87*(6), 2575–2580.

Miles, J. H. (2011). Autism spectrum disorders—A genetics review. *Genetics in Medicine*, *13*(4), 278–294. http://doi.org/10.1097/GIM.0b013e3181ff67ba

Minderaa, R. B., Anderson, G. M., Volkmar, F. R., Akkerhuis, G. W., & Cohen, D. J. (1987). Urinary 5-hydroxyindoleacetic acid and whole blood serotonin and tryptophan in autistic and normal subjects. *Biological Psychiatry*, *22*(8), 933–940.

Modahl, C., Green, L. A., Fein, D., Morris, M., Waterhouse, L., Feinstein, C., & Levin, H. (1998). Plasma oxytocin levels in autistic children. *Biological Psychiatry*, *43*(4), 270–277.

Mullen, E. M., & others. (1995). *Mullen scales of early learning*. AGS Circle Pines, MN.

Nelson, K. B., Grether, J. K., Croen, L. A., Dambrosia, J. M., Dickens, B. F., Jelliffe, L. L.,

… Phillips, T. M. (2001). Neuropeptides and neurotrophins in neonatal blood of children with autism or mental retardation. *Annals of Neurology*, *49*(5), 597–606.

Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., … Windham, G. C. (2007). The Epidemiology of Autism Spectrum Disorders *. *Annual Review of Public Health*, *28*(1), 235–258. http://doi.org/10.1146/annurev.publhealth.28.021406.144007

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry*, *154*(2), 185–190.

Piven, J., Simon, J., Chase, G. A., Wzorek, M., Landa, R., Gayle, J., & Folstein, S. (1993). The etiology of autism: pre-, peri-and neonatal factors. *Journal of the American Academy of Child & Adolescent Psychiatry*, *32*(6), 1256–1263.

Ronalds, G., Phillips, D. I. W., Godfrey, K. M., & Manning, J. T. (2002). The ratio of second to fourth digit lengths: a marker of impaired fetal growth? *Early Human Development*, *68*(1), 21–26.

Rutter, M., & Lockyer, L. (1967). A five to fifteen year follow-up study of infantile psychosis I. Description of sample. *The British Journal of Psychiatry*, *113*(504), 1169–1182.

Sasson, N. J., Lam, K. S., Parlier, M., Daniels, J. L., & Piven, J. (2013). Autism and the broad autism phenotype: familial patterns and intergenerational transmission. *J Neurodev Disord*, *5*(1), 11.

Schopler, E., Andrews, C. E., & Strupp, K. (1979). Do autistic children come from upper-middle-class parents? *Journal of Autism and Developmental Disorders*, *9*(2), 139–152.

Singh, V. K., Warren, R. P., Odell, J. D., Warren, W. L., & Cole, P. (1993). Antibodies to

myelin basic protein in children with autistic behavior. *Brain, Behavior, and Immunity*, *7*(1), 97–103.

Sparrow, S. S., Balla, D. A., Cicchetti, D. V., Harrison, P. L., & Doll, E. A. (1984). Vineland adaptive behavior scales.

Tordjman, S., McBride, P. A., Hertzig, M. E., Snow, M. E., Anderson, G. M., Hall, L. M., … Cohen, D. J. (1997). Plasma β-endorphin, adrenocorticotropin hormone, and cortisol in autism. *Journal of Child Psychology and Psychiatry*, *38*(6), 705–715.

Treffert, D. A. (1970). Epidemiology of infantile autism. *Archives of General Psychiatry*, *22*(5), 431–438.

Tsai, L., Stewart, M. A., Faust, M., & Shook, S. (1982). Social class distribution of fathers of children enrolled in the Iowa Autism Program. *Journal of Autism and Developmental Disorders*, *12*(3), 211–221.

Warren, R. P., Singh, V. K., Cole, P., Odell, J. D., Pingree, C. B., Warren, W. L., & White, E. (1991). Increased frequency of the null allele at the complement C4b locus in autism. *Clinical & Experimental Immunology*, *83*(3), 438–440.

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities*, *22*(1), 33–38.

Wiggins, L. D., Levy, S. E., Daniels, J., Schieve, L., Croen, L. A., DiGuiseppi, C. Schendel, D. (2015). Autism Spectrum Disorder Symptoms Among Children Enrolled in the Study to Explore Early Development (SEED). *Journal of Autism and Developmental Disorders*. http://doi.org/10.1007/s10803-015-2476-8

Wiggins, L. D., Reynolds, A., Rice, C. E., Moody, E. J., Bernal, P., Blaskey, L., … Levy, S. E. (2015). Using Standardized Diagnostic Instruments to Classify Children with Autism in the Study to Explore Early Development. *Journal of Autism and Developmental Disorders*, *45*(5), 1271–1280. http://doi.org/10.1007/s10803-014-2287-3

Wing, L. (1980). Childhood autism and social class: a question of selection? *The British Journal of Psychiatry*, *137*(5), 410–417.