Title: The Study to Explore Early Development: Child Development and Autism

Project Officer: Diana Schendel, Ph.D.

Address:

National Center on Birth Defects and Developmental Disabilities, CDC,

1600 Clifton Road

(MS E-86),

Atlanta, GA,

30333

Telephone: 404-498-3845
Fax: 404-498-3550
Email: dcs6@cdc.gov

April 2010

Supporting Statement

for

The Study to Explore Early Development (SEED)

OMB 0920-0741

Revision of a Previously Approved Collection

Table of Contents

Appendices

Appendix A: Authorizing legislation & other relevant laws

Appendix B.1: 60- day Federal Registry Notice

Appendix C: Primary Caregiver Interview
C.1 Telephone Script
C.2 Primary Caregiver Interview

Appendix D: Study Flow*
D.1 Data Collection Flow Diagram
D.3 Data Collection Instruments Summary Table
D.4 Research Domains by Data Collection Activity

Appendix E: Enrollment Packet
E.1 Enrollment Packet Cover Letter
E.1.1 Enrollment Packet Study Flow
E.2 Yellow Folder Instructions
E.2.1 Blue Folder Instructions
E.2.2 Green Folder Instructions
E.2.3 Questionnaire Checklist
E.3 Written Informed Consent Document
E.4 Clinic Visit Prep Guide
E.5 Caregiver Interview Prep Guide
E.6 Medical Glossary
E.7 Frequently Asked Questions on Biosampling
E.8 Rights of Research Subjects Fact Sheet
E.9 Autoimmune Disease Survey
E.10 Carey Temperament Scales
E.10.1 Carey Coversheet
E.11 Child Behavior Checklist
E.11.1 Child Behavior Checklist Coversheet
E.12 Survey of Gastrointestinal Function
E.13 Maternal Medical History
E.14 Paternal Medical History
E.15 Paternal Occupational Questionnaire
E.16 Child's Sleep Habits Questionnaire
E.16.1 Child's Sleep Habits Questionnaire Coversheet
E.17 Social Responsiveness Scale (Adult, Child & pre-school versions)
E.17.1 Social Responsiveness Scale (Child) Coversheet
E.17.2 Social Responsiveness Scale (Adult) Coversheet
E.17.3 Social Responsiveness Scale (Pre-school) Coversheet
E.18 HIPAA Provider Checklist and Medical Records Release Form
E.18.1 HIPAA Letter Provider Worksheet
E.18.2 HIPAA Medical Records Release coversheets
E.19 How to Collect Cheek Cell Samples
E.20 Informed Consent Document – Cheek Swab (Mother & Father)
E.21 Cheek Swab Sample Record Sheet
E.22 Social Story Example: Trip to the JFK Center

*Appendix D.2 has been deleted

Appendix T: Study Hypotheses and Data Collection Tools

Appendix U: Data Sharing Approval Process

Appendix V: List of Requested changes to ICR

Appendix W: 301(d) Certificates of Confidentiality
W.1. 301(d) (Batelle)
W.2. 301(d) (RTI)
W.3. 301(d) (California)
W.4. 301(d) (Colorado)
W.5. 301(d) (North Carolina)
W.6. 301(d) (Maryland: Johns Hopkins)
W.7. 301(d) (Pennsylvania: University of PA)

Appendix X. Confidentiality Agreement

Appendix Y. Participant Newsletter

Appendix Z. Detailed breakdown of Participant Burden Hours

A. Justification

A.1. Circumstances Making the Collection of Information Necessary

Background:

'The Study to Explore Early Development (SEED): Child Development and Autism' was developed under the National Center on Birth Defects and Developmental Disabilities (NCBDDD) at CDC. Authorization for the center's activity is granted under *SEC 317C of the Public Health Service Act (42 U.S.C. 241) and (42 U.S.C. 257b-4)*. (Appendix A) This authorization has been further supported by the *Combating Autism Act of 2006, Pub. Law No. 109-416* (Appendix A) which authorizes the Center's activities through 2012.

In addition, 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 previsions 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): Kaiser Research Foundation in California, Colorado Department of Public Health and Environment, Johns Hopkins University, University of Pennsylvania, and University of North Carolina at Chapel Hill. CDC participates as the sixth CADDRE site.

The Study to Explore Early Development (OMB 0920-0741) was initiated in December 2007 (following OMB approval in October 2007 of a name change required prior to study start). It is expected that SEED enrollment will continue through March 2011 and data

collection will continue until July 2012. Therefore, OMB approval is requested for three additional years of data collection.

This application classification is for "revision". This revision is requested due to two proposed changes to SEED. First, minor data collection changes are requested, including modifications of some of the self administered instruments and the *Primary Caregiver Interview* (Appendix C). These changes consist primarily of clarification of questionnaire instructions and clarifying the text for specific questions to make the instruments easier to complete and further improve data quality. None of the instrument changes have a measureable impact on participant burden. A complete list of the modifications to the instruments and appendices is provided in *List of Requested Changes to the ICR* (Appendix V). Second, a single study design change is proposed. We propose expanding the eligible study participant birth date range from September 1, 2003-August 31, 2005 to September 1, 2003-August 31, 2006. The expansion of the eligible birth date range will be used by sites if they expect that they are unable to achieve their target sample size for any of the participant groups (case or comparison groups) based on the original cohort. The potential shortfall has arisen primarily due to the larger than expected proportion of potential participants from whom we never get a response to the mailed study invitation (indicating if they are interested in being contacted for further study information) or for whom we are unable to achieve initial contact for invitation into the study (whereas rates of enrollment among contacted participants are good). The pace and rates of contact and enrollment are being closely monitored as the study progresses and if the cohort expansion appears not to be needed then it will not be pursued. We have calculated the burden estimate for the upcoming OMB approval period to accommodate the latter

change, however the total burden for the upcoming 3-year OMB approval period is less than the original 3-year period. This lower burden reflects that fact that the study will be winding down during the upcoming OMB approval period and thus the fewer total participant contacts to be made and averaged over the three year period lowers the annualized burden rate.

Despite significant advances in our understanding of the clinical features of autism spectrum disorders (ASD) and substantial progress in establishing ASD prevalence studies across multiple populations (Rice, et al, 2003), 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 (Bacchelli, 2006; Klauck, 2006).

In the face of these considerable gaps in our understanding of the causes of and risk factors for ASDs, large population-based epidemiologic studies of ASD etiology are lacking. The proposed data collection is designed to address this critical need.

Privacy Impact Assessment:

The PIA for this data collection was assessed previously in 2006 during the first OMB-PRA approval process.   During that assessment, the Privacy Officer for CDC recommended that a *301(d) Certificate of Confidentiality* be obtained for the CDC/Georgia site, and that Certificate was signed in July 2007 with Batelle as the contractor for the data collection in Georgia (see Appendix W.1). When a contractor modification was sought to replace Batelle with Research Triangle Institute, a revised

301(d) Certificate of Confidentiality was provided on January 27, 2010 (see Appendix W.2). Similar *301(d) Certificates of Confidentiality* were also issued for the other five sites (see Appendices W.3 through W.7).

Overview of the Data Collection System:

The overall goal of this data collection system is to collect possible risk factors and possible causes of ASD among selected children born between September 1, 2003 and August 31, 2006. This assessment is being done in order to identify risk factors and potential causes associated with ASD. The individual assessments are performed at six participating sites in the Unites States (Kaiser Research Foundation in California, Colorado Department of Public Health and Environment, Johns Hopkins University, University of Pennsylvania, and University of North Carolina at Chapel Hill, and CDC in Georgia). The assessments are done using a series of standardized assessment tools and SEED-specific data instruments that are listed in *Research Domains by Data Collection Activity* (Appendix D4).

Items of Information to be Collected:

The following six "domain areas" will be assessed: 1) investigation of the ASD phenotype, 2) infection and immune function, 3) reproductive and hormonal features, 4) gastrointestinal features, 5) sociodemographic features, and 6) genetic features. These domains are described in more detail in Part A.2 below.

Identification of websites and Website Content Directed at Children under 13 years of Age:

The websites for SEED consist of the following three sites list below. None of the

websites are directed at children who are under 13 years of age:

1) http://www.cdc.gov/ncbddd/autism/seed.html.
2) http://www.cdc.gov/ncbddd/autism/seed-ga.html
3) http://www.cdc.gov/ncbddd/autism/caddre.html

A.2. Purpose and Use of the Information Collection

As mentioned above, the Study to Explore Early Development (OMB 0920-0741) was initiated in December 2007. Given the size and scope of the study, it was expected that enrollment and data collection would not be completed within the initial 3-year OMB approval period. To date, we have enrolled about two thirds of our target number of participants. Our experience with the processes of participant ascertainment, enrollment and data collection thus far have been extremely informative. Fortunately our experiences with the data collection instruments themselves and associated quality controls have been generally positive and we have had to make only very minor changes to study instruments, primarily to clarify instructions or a few specific questions. We have assessed the results of our initial participant ASD screening process to determine if the cut-off point is optimal and have concluded that no change was needed. We have evaluated the results of our ASD case confirmation algorithm used by study clinical staff to assign final case status based on results of the clinical evaluations and, based on recent published data on the same standardized instruments adopted for SEED, have made a few minor modifications for final case classification (with no change in the clinical evaluation process with participants). Our quality control activities designed to assure a certain training standard is achieved prior to staff engagement with study participants have identified opportunities for improvement of initial staff training and also steps for maintaining standards of ongoing performance. Based on our experience, we also

continue to monitor and adapt our staffing and study resources to improve the pace and efficiency of study implementation. We have created a variety of mechanisms to monitor study flow and to gather data on study implementation processes. One example is a data export from our web-based centralized tracking system that study staff can generate at any time that provides a list of data collection items still be collected from individual enrolled participants. This data export has improved our ability to target staff time allocation and improve data collection completeness. Although enrollment and data collection is ongoing, we have begun prioritization of our analytic plans and initiated data coding, data cleaning and preparing algorithms for analytic variable creation. We are also more well informed as to the pace of participant study completion (once a family is enrolled) and have accommodated our study timelines accordingly. It is expected that SEED enrollment will continue through March 2011 and data collection will continue until August 2012. This application for revision is to allow SEED enrollment and data collection to be completed.

The overall purpose of this study is to identify risk factors and potential causes associated with ASD. This will be accomplished through the investigation of six high priority research domains concerning potential etiologic factors. Study investigators selected the domains after an extensive review of the literature (Newschaffer CJ, 2006). Investigators designated each of the domains as high priority 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 design and feasibility of obtaining relevant data. The specific research goals of each domain are as follows: investigation of the ASD phenotype, infection and immune function, reproductive and hormonal features,

gastrointestinal features, sociodemographic features, and genetic features.

The goals of the ASD phenotype domain are to identify:

- the distinctive features of children with ASD compared to children in the control groups[1] related to: physical traits, medical conditions, developmental problems, and behavior difficulties,

- the distinctive features of parents or siblings of ASD children compared to parents or siblings in the control groups related to: parental psychiatric/affective problems, medical conditions, developmental problems, behavior difficulties, and

- discriminating features of children with ASD, with and without regression, related to: language skills, cognitive or adaptive delays, medical, physical or genetic traits.

The study goals in the area of infection are to identify whether, compared to the neurodevelopmentally impaired comparison group and subcohort:

- Mothers of children with ASD are more likely to experience, during pregnancy or

_____

[1] Study participants include children age 2-5 years and their parents or primary caregivers. All study children will be drawn from the cohort of children born among residents in the CADDRE site study areas in select birth years. Three groups of children will be selected: cohort children identified with autism spectrum disorders will be compared to 1) a sample of children identified with other developmental problems (neurodevelopmentally impaired comparison group or NIC), and 2) a random sample of all cohort children (most of whom are typically developing).

through the end of breastfeeding: a) clinical illness from infections (e.g., STDs, Group B strep), b) clinical illness from viral infections specifically, c) other infection-related exposures such as vaccines (e.g., influenza), or d) different treatment histories for infectious illness during pregnancy (e.g., prescription medications such as antibiotics).

- Children with ASD, from birth up to the 3rd birthday, are more likely to experience: a) clinical illness from infections, b) clinical illness from viral illnesses, c) clinical illness from ear infections, d) different treatment histories for infectious illness (e.g., antibiotics), or e) different vaccine histories or reactions to vaccines.

The goals in the immune function area are to identify whether, compared to the neurodevelopmentally impaired comparison group and subcohort:

- Children with ASD: a) are more likely to have a nuclear family history of autoimmune disorders, b) if occurrence of autoimmune disorder in mother is time related to pregnancy, c) if nuclear family history is present, is associated with specific ASD subgroups;

- Children with ASD: a) are more likely to have an autoimmune disorder, b) have abnormal levels of specific biomarkers of autoimmune disease, (e.g. auto antibodies to CNS proteins), c) if present, associated with specific ASD subgroups;

- Children with ASD: a) have abnormal levels of specific chemical messengers involved in CNS, immune, and endocrine development and regulation (e.g., cytokines, neuropeptides, neurotrophins, neurotransmitters), b) if present,

associated with specific ASD subgroups.

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

- Assess whether mothers of children with ASD have, compared to the neurodevelopmentally impaired comparison group and subcohort: 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 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)

- Postnatal hormone features – Assess whether children with ASD, compared to the neurodevelopmentally impaired comparison group and subcohort, have different levels of serotonin, melatonin, oxytocin, vasopressin.

The goals in the GI area are to determine identify whether:

- Children with ASD are more likely to have GI symptoms than children in the neurodevelopmentally impaired comparison group and subcohort;

- 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

neurodevelopmentally impaired comparison group and subcohort;

- GI symptoms are associated with dietary patterns,

    o children with ASD are more likely to have restricted diets than children in the neurodevelopmentally impaired comparison group and subcohort,

    o restricted diets in ASD children are associated with specific measures of abnormal nutrient intake or behavior (e.g., temperament)

- GI symptoms in ASD children are associated with candidate biologic markers or genes for ASD.

The goals in the genetics area are to:

- Investigate genetic main effects,

- Investigate interactions between genetic and environmental effects, and

- Determine whether genetic effects are offspring or parentally mediated.

The goals in the sociodemographics domain are to determine whether, compared to the neurodevelopmentally impaired comparison group and subcohort, children with ASD and their families have different sociodemographic characteristics.

Many of the domains are linked by different theoretical causal pathways leading to ASD. Each of the six research domains requires comprehensive and standardized case ascertainment and/or confirmation of previously diagnosed cases. The table in Appendix D.4 summarizes explicitly which data collection instruments address each specific research domain. The last row of the table in Appendix D.4 lists which instruments are used for the case ascertainment and confirmation process. Use of these same instruments

also allows cases and controls to be subdivided into potentially etiologically distinct subtypes, according to dysmorphology, cognitive ability, various genetic markers, and case presentation with or without regression.

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.

In addition to the high priority research domains described above, SEED seeks additional information on substance use during pregnancy, maternal and paternal occupational exposures before and during pregnancy, the history of hospitalizations and injuries of the child, sleep disorders in the child and biologic parents, and information related to select mercury exposures. This additional information will be used to test secondary

hypotheses. Specific hypotheses in each domain are found in *Study Hypothesis and Data Collection Tools* (Appendix T).

All of the secondary hypotheses are related to, and limited by, the data collected to support the primary hypotheses. Take, for example, the select mercury exposure hypothesis. SEED captures information (i.e., through interviews, questionnaires, medical record review, and biologic sampling) related to prenatal influenza vaccine, RhoGAM exposures, prenatal thimerosol exposures, mercury exposure related to maternal and

paternal occupational histories, and child mercury levels measured in hair since this information is already collected to address the primary hypotheses. However, other sources of prenatal, perinatal and early postnatal mercury exposure, such as maternal diet and non-occupational environmental mercury exposures are not captured in SEED. As a result, the select mercury exposure hypothesis addresses certain medical and occupational mercury exposures, but does not include a complete mercury exposure history. Given the retrospective nature of exposure ascertainment for SEED and the lag time of three or more years between exposure and ascertainment, collecting valid and complete data on dietary and environmental histories throughout this interval, especially on common exposures, was not deemed practicable. As such, the select mercury exposure hypothesis is a secondary hypothesis.

In summary, SEED permits investigators to estimate for each specific causal factor the prevalence of the factor, the magnitude of the risk associated with that factor, and the proportion of individuals with ASDs that is attributable to the factor across sites. This knowledge will ultimately assist CDC to develop recommendations concerning identification of individuals with ASDs, identify interventions, and design more effective programs for prevention of ASDs. Without these data, CDC would be limited in its ability to identify interventions that are likely to have the greatest effect on the prevention of ASDs.

Privacy Impact Assessment Information:

    i)   Why the information is being collected:

    Personally identifiable information is collected on the consent forms, the HIPAA

forms, and in the medical records of the child that are collected as a part of the data collection. The information collected includes name, address, date of birth, and personal medical information about the child and even of the parents as reflected in the child's medical record. This information collection is necessary to fulfill the SEED objective of collecting medical and health information factors that might be associated with ASD, and the need for continuing contact with the participants makes necessary the collection and continued use of the locating information (name, address) throughout the course of the SEED project.

ii)   <u>The intended use of the information</u>:

Since the goal of the project is to identify risk factors and potential causes associated with ASD, the identifying and personal health information collected will be used to maintain contact with the participants throughout the course of the study, and to use the health data collected to perform epidemiologic analysis to study possible risk factors associated with ASD.

The data collected will be shared only with other research personnel and contractors assigned to the SEED project. The identifying information linked to each participant will not be shared with the CDC personnel, as the CDC will only receive data that is linked to the unique study ID number assigned to each participant (see below). The linkage key of the identifying information and the study ID number are maintained by the Data Coordination Center (DCC) and not the CDC. The data elements that will be shared among the research personnel are the name and locating information of the participants, as well as the other health information collected as part of the SEED project.

iii)   <u>Impact on privacy and safeguards in place for the protection of information</u>:

None of the data forms (except for initial consent forms and HIPAA Release Forms

which are stored in a locked cabinet within a locked office at each site with access

limited to staff personnel only) will record any identifiable information on the form

itself except for the unique study ID number that is placed on each form. Since the

data collected on each form includes sensitive health information, *301(d) Certificates*

of Confidentiality have been obtained for the project (Appendices W2 through W7).

Furthermore, the data collection staff will each sign a *Confidentiality Agreement*

(Appendix X) that states that they will not share the information collected for this

project with any person or entity outside of the project.

A.3. Use of Improved Information Technology and Burden Reduction

<u>Application of Information Technology</u>
In addition to the CADDRE centers, NCBDDD funded a Data Coordinating Center

(DCC) and a Central Biosample Repository (Central Lab) for SEED. Michigan State

University established and manages the DCC. The DCC developed an electronic data

collection system to centrally store the data.   Johns Hopkins University houses the

Central Lab, where all biosamples from the study are shipped, processed, and stored. The

DCC and Central Lab work on an ongoing basis with the SEED investigators to

implement the study.

SEED applies information technology broadly to collect data efficiently, to assure the

quality of the collected data, to assure the privacy and security of the collected data, and

to minimize the burden to the study participants. As stated previously, the DCC is

responsible for the information technology aspects of the study. The DCC created and hosts a custom web-based information system, called the CADDRE Information System (CIS), which is carefully designed to directly support all of SEED data collection workflows, data quality assurance processes, and provide secure database and Internet transaction services. Please note that this CIS system is used by the study personnel only, and not by the participants themselves.

A sampling of relevant services of the CIS includes:

- Upon login, the CIS automatically presents the user a list of tasks that are currently open items required to be performed, or alerts to exceptional issues. The task list is customized for the specific organizational role of that authenticated user.

- Employ role-based security that restricts user access privileges to the minimum required for that specific staff person's organizational functions

- Automated tracking of all aspects of a participant as s/he proceeds through the SEED protocol

- Bar code labels will identify all study documents exchanged with participants, as well as all biologic samples. For efficient processing of all documents and biosamples, bar codes will be scanned into the CIS to drive automated processing.

- Facilitate efficient computer assisted telephone interviews (CATI) for the telephone interviews occurring in the study

- Support electronic versions of copyrighted clinical assessment instruments whenever possible

- Support double data entry (QA) operations whenever data collection is performed using paper forms

- Support direct entry of medical record abstractions

- Data quality assurance processes are provided via:

  o The application of logic rules to all data entry fields in forms—checking of range, data type, consistency with data contained in other fields

  o Prevention and detection of the duplicate participants and other records in the database

  o Extensive set of automated reports to support the detection and evaluation of data quality and completeness

- Provide a broad range of automated reports to enable careful monitoring of data quality and operations, and data cleaning, etc.

- Provide comprehensive audit logging facilities capturing the relevant details of all updates to the database, user login and logouts, and user accesses to personally identifiable data

- Provide a secure method to distribute cleaned, SAS/Microsoft Access-ready tables of site and pooled analytic data sets for analysis at each SEED research site

- Provide exported data in standard interchangeable file formats accessible by various analytical software applications (e.g., SAS, SPSS, S Plus)

- Provide 8 am to 10 pm EST Monday – Saturday user support for study staff using the CIS to facilitate efficient operations and improve the availability of CIS services

Participant Burden Reduction

The CIS facilitates one computer assisted telephone interview (CATI) that is given to the study participant parents, the *Primary Caregiver Interview* (Appendix C.2). In addition, three *Follow-up Phone Calls* (Appendix O.1 repeated at three time points) also employ the CIS system to aid the interviewer in tracking completeness of response and reducing time for the participant. These interviews are scheduled per the convenience of the participant parents. The *Primary Caregiver Interview* (Appendix C.2), has a particularly complex structure involving branching and looping dependent on responses to prior questions. The required logical branching is automatically provided by the CIS guidance to the interviewer during the interview. This implementation will improve data quality and reduce errors to preclude the burden of follow-up calls to participants. Since the Primary Caregiver Interview and the follow-up phone calls comprise a total of 980 hours of burden per year, out of a total burden of 5,238 burden hours annually to the participants, about 19% of the participant burden hours are reduced via information technology tools.

Much of the data collection of the SEED protocol takes place in a clinical setting and

involves the participant children who are 2-5 years of age. Many of the data collection instruments are filled out by clinicians working in real time with the children. Most of the data collection instruments are copyrighted which limits our options for direct response entry. The copyrighted instruments are standardized developmental measures, including the *Autism Diagnostic Observation Scale (ADOS)* (Appendix G.1), the *Carey Temperament Scales* (Appendix E.10), and the *Mullen Scales of Early Learning* (Appendix G.2). Together these aspects limit the opportunities for direct computer entry of responses into data collection instruments in the conduct of the study.

Nevertheless, the following list describes various technical actions taken to reduce SEED participant burden:

- The CIS proactively tracks all aspects of participant's needs, requests, scheduled activities, and study protocol requirements. Study staff is alerted automatically when they login to all currently required actions/tasks to do. The aim is to preclude oversights and errors to avoid inconvenience and inefficient use of participant's time.
- The CIS implementation of the CATI workflow reduces burden:
  - Automated navigation through the interview logic speeds the interview process and prevents errors by the interviewer--precludes the need for a follow-up call to collect correct data.
  - Support for suspending the interview whenever the participant requests. Rescheduling the follow-on call at the participant's convenience automatically begins when a call must be suspended. The follow-on

interview script resumes automatically at the ending location of the prior

call.

- Automation for scheduling or rescheduling any call of any type with a participant to maximize the participant's convenience. The study staff is automatically alerted on the day when calls are promised with a scheduled task.

- Careful automated tracking of all study items and tasks to preclude errors requiring the re-collection of data, or other accidental oversights

  o Special care is given to preparation for the clinical visits. Staff is automatically notified of every task and all data items required in that visit, including any exceptional elements that are required for this visit (items which are usually already accomplished by that time)

  o Automatic alerts for the clinical visit are provided to the staff about a participant's special needs, prior special requests, allergies, sibling child care, incentives, etc.

- In the tracing (recruitment) workflow, the CIS implementation assures that only intended participants are contacted for invitation.

- All contacts with each participant are tracked and processed to determine the next required action per the study protocol and rules. This ensures the efficient execution of the study and reduces the chances of wasting participant's time.

- For the participant family situation where a biological parent does not live with the participant child, the study workflow is implemented in an entirely independent thread. All study communications and actions proceed independently from those related to the participant child. No communication is required between the two parts of the family or caregiver(s) not living together.

A.4. Efforts to Identify Duplication and Use of Similar Information

No data collection activities currently supported by HHS, other government institutions, or other private agencies, are comparable to the proposed data collection. The Collaborative Programs of Excellence in Autism (CPEA) network, co-funded by the National Institute for Child Health and Development (NICHD), the National Institute of Deafness and Other Communication Disorders, and the National Center for Complementary and Alternative Medicine, were investigating the cause of autism at 25 sites in the United States, Canada, Great Britain, France and Germany. The National Institutes of Health (NIH) and the Interagency Autism Coordinating Committee (IACC), established Studies to Advance Autism Research and Treatment (STAART) Network to conduct basic and clinical research in autism at eight centers in the United States. In 2007, NIH initiated the Autism Centers of Excellence (ACE) program to support studies covering a broad range of autism research areas, including early brain development and functioning, social interactions in infants, rare genetic variants and mutations, associations between autism-related genes and physical traits, possible environmental risk factors and biomarkers, and a potential new medication treatment. Although some CPEA, STAART and ACE grantees have in the past, or are currently, investigating research domains similar to those in SEED, the CPEA, STAART or ACE sites do not all adhere to a common protocol. Use of a common protocol will allow SEED sites to pool data, resulting in a sample of 2,700 children and their families. Not only does the large SEED sample size increase study power and statistical precision overall, but also enhances our

capability for stratified analysis of phenotypic subtypes within the ASD case group as well as stratification across all subject groups.

Another large, multi-site, collaborative project that is currently ongoing is the California Childhood Autism Risks from Genetics and the Environment (CHARGE) study. The CHARGE Study is funded by the National Institute of Environmental Health Sciences, the United States Environmental Protection Agency, and the University of California Davis – Medical Investigation of Neurodevelopmental Disorders (M.I.N.D.) Institute, and is investigating factors in the environment that are associated with autism in some children and families.

Although the CHARGE study is population-based and utilizes data collection methods similar to SEED, there are multiple differences between CHARGE and SEED. CHARGE will enroll 1,600 children and their families, significantly less than the 2,700 SEED will enroll. Moreover, CHARGE is collecting data only in the state of California and data, therefore, is less generalizable to a national population.

In addition, CHARGE relies on a single source (Department of Developmental Services) for case ascertainment and the CHARGE Study case and developmental delay comparison groups are more narrowly defined: children who meet criteria for autism services and children who meet criteria for Mental Retardation /Developmental Disability services by the California Department of Developmental Services. Finally, the research goals and corresponding data collection batteries differ somewhat between the SEED and CHARGE studies. CHARGE collects more data on environmental exposures than SEED, while SEED will be collecting more detailed data on GI function, including diet, sleep

features, and child/parent behavioral phenotype than CHARGE. In addition to extending the domains being studied by the CHARGE and Early Markers for Autism Study (EMA) studies, the overlap in data collection in SEED will permit replication of many of the CHARGE and EMA analyses. In fact, many aspects of SEED data collection were explicitly set up to enable this kind of replication which will allow for comparison of results.

A literature review conducted for SEED protocol development did identify other case-control population-based studies on the pre- and perinatal etiological risk factors for autism; although none have utilized comparable data collection procedures (Burd et al, 1999; Croen et al 2002; Hultman et al 2002; Juul Dam et al 2001; Glasson, 2004; Larsson, 2005). For instance, previous investigations have used relatively small sample sizes, did not verify autism case status, and did not employ diverse data collection methods (i.e., maternal interview in addition to medical record review). This

comprehensive literature review helped detect gaps in our current understanding of ASD which, in turn, led to identification of high priority research domains.

In addition, the Director of NCBDDD is a member of the Interagency Autism Coordinating Committee (IACC). The IACC was established in accordance with the Combating Autism Act of 2006 (P.L. 109-416) and coordinates all efforts within the Department of Health and Human Services (HHS) concerning autism spectrum disorder (ASD). Through its inclusion of both Federal and public members, the IACC helps to ensure that a wide range of ideas and perspectives are represented and discussed in a

public forum. Several SEED investigators and collaborators also attended the 2004 Autism Summit Conference held in Washington, D.C. and were also involved in the more recent development of the IACC Strategic Plan for ASD research. The proposed data collection contributes to multiple high priority autism research questions described in the IACC's Strategic Plan released in January 2009.

A.5. Impact on Small Businesses or Other Small Entities

No small businesses or other small entities will be involved in this data collection.

A.6. Consequences of Collecting the Information Less Frequently

As the protocol is currently written, SEED proposes *one-time data collection* in response to a mandate for research into the causes of ASD in the Children's Health Act of 2000. If this data was not able to be collected, it would impact the ability of the researchers to provide timely and important information related to the risk factors and causes of ASD. Some of the outcomes the investigators hope to achieve include having greater knowledge of the etiology of autism, improving the phenotypic descriptions of children with ASDs, allowing for better and earlier screening for ASDs in younger children, and possibly improving the services and treatments for these children. Without this collection, or if information was collected less frequently, these data would be delayed or never reported. The SEED case control study will be the first and largest multi-site, population based study on ASD planned to date, and the findings from this study will be essential to advancing the understanding of autism and ASDs.

There are no legal obstacles to reduce the burden.

A.7. Special Circumstances Relating to the Guidelines of 5CFR1320.5

This request fully complies with the regulation 5 CFR 1320.5.

A.8. Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency

A. The 60 Day Federal Register Notice was published in the *Federal Register* on September 8th, 2009, pages 46200-46201, Volume 74, Number 172. No substantive public comments were received. See *60 Day Federal Register Notice* (Appendix B).

B. We have consulted a number of persons outside CDC to ensure that this data collection is not duplicative and that the study design, data elements, and instruments are appropriate.

The principal investigators (PIs) at each of the SEED sites played an integral role in the design and the development of SEED. They conducted an extensive review of the literature, identified the research domains, selected the study design and data collection instruments, and developed the study protocol. The PIs are:

Diana Schendel, Ph.D.
Centers for Disease Control and Prevention
(404)498-3845
Email: dcs6@cdc.gov

Lisa Croen, Ph.D.
Kaiser Permanente Division of Research
Phone: (510) 891-3463
Email: Lisa.A.Croen@kp.org**M. Daniele Fallin, Ph.D.**
Johns Hopkins University, Bloomberg School of Public Health
Phone: (410) 955-3463
Email: dfallin@jhsph.edu

Jennifer Pinto-Martin, Ph.D., MPH,
University of Pennsylvania, School of Nursing
Phone: (215) 898-4726
Email: pinto@nursing.upenn.edu

Lisa Miller, M.D., MSPH
Colorado Department of Public Health and Environment
Phone: (303) 692-2663
Email: lisa.miller@state.co.us

Julie Daniels, Ph.D.
University of North Carolina at Chapel Hill
School of Public Health
Phone: (919) 966-7096
Email: Julie_daniels@unc.edu

Co-investigators and other collaborators include:

Gayle C. Windham, PhD
Division of Environmental and Occupational
  Disease Control
CA Department of Public Health
Phone: 510-620-3638
Email: Gayle.Windham@cdph.ca.gov

Craig Newschaffer, Ph.D.
Drexel University
School of Public Health
Phone: (215) 762-7152
Email: cnewscha@drexel.edu

Rebecca Landa, Ph.D.
Center for Autism and Related Disorders
Kennedy Krieger Research Institute
Phone: (443) 923-7680
Email: landa@kennedykrieger.org

Susan Levy, M.D.
Division of Child Development and Rehabilitation
Children's Seashore House of The Children's Hospital of Philadelphia
Telephone 215 590-7528
levys@email.chop.edu

Cordelia Robinson (Corry), Ph.D., R.N.
JFK Partners/UCHSC

Phone: (303) 864-5261
Email: robinson.cordelia@tchden.org

Laura Schieve, Ph.D.
Centers for Disease Control and Prevention
(404) 498-3888
Email: Ljs9@cdc.gov

In December 2003, prior to submission to CDC IRB, the SEED group established a five person external peer review panel. This panel consisted of experts in clinical research, epidemiology, genetics, immunology, and advocacy, who were chosen on the basis of their expertise, balance, independence, and lack of conflicts of interest. Each of the panel members reviewed the SEED protocol and appendices with regard to several factors, including:

- the relevance of the proposed research domains and associated hypotheses,

- the effectiveness and feasibility of the scientific study plan,

- the appropriateness of the study design, study population, eligibility criteria, and case determination,

- adequacy of the sample size and study power, and,

- appropriateness of the data collection instruments and methods.

The SEED PIs identified changes that were required of the protocol based on the panel's feedback and these changes were incorporated into the protocol. External peer reviewers were:

- Eric Fombonne, M.D., Professor/McGill University, Canada Research Chair in Child and Adolescent Psychiatry, Director of Psychiatry/Montreal Children's Hospital, 514/412-44 49, eric.fombonne@mcgill.ca

- Judy Van de Water, Ph.D., Associate Professor/University of California, Davis,

530/752-2154, javandewater@ucdavis.edu

- M. Anne Spence, Ph.D., Professor, Department of Pediatrics/University of California, Irvine, 530/752-2154, maspence@uci.edu

- Eric London, M.D., National Alliance for Autism Research, Co-founder; Boardmember, Clinical Assistant Professor in the Department of Psychiatry at the University of Medicine and Dentistry of New Jersey (UMDNJ), Consulting Psychiatrist to Hunterdon Developmental Center in New Jersey, drlondon@earthlink.net

- Susan Hyman, M.D., Assistant Professor of Pediatrics/Strong Children's Research Center, 585/275-2986, susan_hyman@urmc.rochester.edu

The Data Coordinating Center and the Central Lab Repository were also involved in the collaboration of the SEED project. The Principal Investigators for these two entities are:

Data Coordinating Center
Philip L. Reed, PhD (PI)
Michigan State University
Data Coordinating Center
Room 100, Conrad Hall
East Lansing, MI 48824
517.353.9445
Phil.Reed@hc.msu.edu

Central Biosample Repository
Homayoon Farzadegan
Bloomberg School of Public Health
Department of Epidemiology
East Baltimore Campus
615 N. Wolfe Street E7140
Baltimore, MD 21205
410.955-3786
hfarzade@jhsph.edu

In 2009, as a consequence of internal review of the data collected to date , this Information Collection Request underwent additional proposed revisions to the cohort birth date definition and the clarification of some of the question wording and questionnaire instructions. The result of that internal review is this revised ICR request.

A.9. Explanation of Any Payment or Gift to Respondents

In order to ensure the validity of the data, it is important that SEED has high response rates. As stated previously, all SEED families will have young children and two-thirds of SEED families will include children with autism or other developmental disabilities. These parents cope with challenges above and beyond what parents of typically developing children face. Also, since the burden is higher than many other studies, it will be difficult to enroll and retain all families without providing incentives (Dunn and Gordon, 2005). Thus, we propose the following incentive structure to ensure a more representative study sample:

Information Collection Step	Incentive Offered
Enrollment Packet Introductory Letter & Study Cover Letter (E.1) Cover Letter (E.2) Informed Consent (E.3) Rights of Research Subjects Fact Sheet (E.8) Prep Guides (E.4 & E.5) Social Story Example: Trip to the JFK Center (E.22) HIPAA Provider Checklist & Medical Record Release Forms (E.18) Buccal Swab Kit (E.19-E.21)	$25 included in packet
Primary Caregiver Interview (C.1 &C.2)	$30 mailed when scheduled
First Questionnaire Packet Paternal Medical History (E.14) Maternal Medical History (E.13) Autoimmune Disease Survey (E.9) GI questionnaire (E.12) Paternal Occupational Questionnaire (E.15) Services & Treatments (Cases only) (F.3) EDQ (Cases only) (F.2)	$30 mailed when scheduled

Second Questionnaire Packet CBCL (E.11) Carey Temperament Scales (E.10) Sleep Habits Questionnaire (E.16) Social Responsiveness Scale (E.17)	$30 mailed when scheduled
Case Child Clinic Visit ADOS (G.1) Mullen (G.2) Dysmorphology Exam (P) Biosampling (R.1 & R.2)	$80 at visit
Parent Child Clinic Visit ADI-R (F.1)(case only) Vineland (F.4)(primarily case only) Biosampling (R.1 & R.2)	$80 at visit
NIC/Subcohort Clinic Visit Mullen (G.2) Dysmorphology Exam (P) Biosampling (R.1 & R.2)	$80 at visit
Third Questionnaire Packet Three Day Diet Diary (H.1) Seven Day Stool Diary (H.2)	$40 when handed out

We propose this incentive structure for the following reasons:

- Enrollment Packet

The enrollment packet includes materials that further explain the study: (*Enrollment Packet Cover Letter* (Appendix E.1)*, Cover letter* (Appendix E.2)*, Informed Consent* (Appendix E.3)*, Rights Of Research Subjects Facts Sheet* (Appendix E.8)*, Social Story Example: Trip to the JFK Center* (Appendix E.22)*, Clinic Visit Prep Guide* (Appendix E.4 and E.5)*, HIPAA Medical Records Release Form* (Appendix E.18.), and *How to Collect Check Cell Samples* (Appendix E.19-E.21). We propose to include $25 in the enrollment packet. We believe this amount to be appropriate since we are requesting access to the participants' medical records and biologic samples; both of these activities are more intrusive than asking a participant to answer a questionnaire. Since we propose to employ a graduated incentive structure, this is the lowest amount offered.

- Primary Caregiver Interview

The *Primary Caregiver Interview* (Appendix C1 and C2) is a 90-minute long computer assisted telephone interview (CATI) asking the participant for information about the biological mother's pregnancy and reproductive history, postnatal medical and developmental history of the child, lifestyle factors during pregnancy, and demographics. We propose to mail the participant $30 when the caregiver interview is scheduled. We believe this amount is appropriate since we are requesting the participant share sensitive information that they may otherwise be reluctant to provide.

- First Questionnaire and Second Questionnaire Packets

The first and second questionnaire packets will include surveys about family medical history, occupational exposures, and standardized developmental tests comprised of the following appendices: *Paternal Medical History* (E.14), *Maternal Medical History* (E.13), *Autoimmune Disease Survey* (E.9), *GI questionnaire* (E.12), *Paternal Occupational Questionnaire* (E.15), *Services & Treatments (Cases only)* (F.3) *EDQ (Cases only)* (F.2), *CBCL* (E.11), *Carey Temperament Scales* (E.10), *Sleep Habits Questionnaire* (E.16), *Social Responsiveness Scale* (E.17). We propose to mail the participant $30 when each packet is scheduled. This incentive is in line with our plan to offer a graduated incentive structure. Although there is a high time burden for these surveys, they are not particularly invasive.

- Case Child Clinic Visit, Case Parent Clinic Visit, NIC/Subcohort Clinic Visit

During the case child clinic visits, study staff will administer standardized developmental evaluations to the child, perform a dysmorphology examination, and draw biologic

samples using the following appendices: *ADOS* (G.1), *Mullen* (G.2), *Dysmorphology Exam* (P), and *Biosampling* (R.1 & R.2). The case parent clinic visit will include standardized developmental evaluations and biologic samples from both biological parents using the following appendices: *ADI-R* (F.1), *Vineland* (F.4), and *Biosampling* (R.1 & R.2). During the NIC/Subcohort clinic visit, study staff will administer a standardized developmental evaluation to the child, perform a dysmorphology examination, and draw biologic samples from the child and the biological parents using the following appendices: *Mullen* (G.2), *Dysmorphology Exam* (P), and *Biosampling* (R.1 & R.2). We propose to give the participant $80 at the beginning of each of the three visits. This is the highest amount we propose to offer during the study because it is the most inconvenient and intrusive activity and it occurs relatively late in the study. We believe that this level is appropriate for the aforementioned reasons.

- Third Questionnaire Packet

The third questionnaire packet includes the use of the following appendices: *3-Day Diet Diary* (H.1) and a *7-Day Stool Diary* (H.2). We will give the diaries at the conclusion of the clinic visit and, we propose to give the participant $40 at that time. We believe this amount to be appropriate because although the diaries do not constitute a significant time burden, participants otherwise may be reluctant to provide detailed information requested in the diaries.

The investigators recognize that all subjects may not participate in all phases of data collection. Subjects may choose to drop out of the study at any time. Due to differing regulations, incentives will vary across SEED sites, but will include: gift cards, money

orders, checks, and cash.

A.10. Assurance of Confidentiality Provided to Respondents

The CDC Privacy Act Officer reviewed the initial SEED application in 2007 and determined that

the Privacy Act is not applicable to the data collection activities conducted by CDC-funded

grantees at the five sites outside of Georgia. However, the Privacy Act is applicable to data

collection activities at the Georgia SEED site (involving a contractor, Research Triangle Institute

(RTI). All employees associated with this project, including contractors, will continue to sign a

non-disclosure agreement. Where applicable, personally identifiable information will be collected

and maintained under Privacy Act System of Records 09-20-0136, "Epidemiologic Studies and

Surveillance of Disease Problems." Analytic datasets transmitted to CDC by the Data

Coordinating Center (DCC) will be in de-identified form. The data collected is jointly owned by

the CDC and the participating clinical sites, however the CDC does not own any of the

identifying data collected at the sites.

Due to the sensitive nature of certain data collection components, SEED has obtained additional

confidentiality protections. A 301(d) Certificate of Confidentiality for protection of the

individual participants at all six sites conducting the study were approved in July 2007

(Appendices W.1, W3 through W.7) and the Georgia site 301(d) was reapproved in January 2010

when the awardee for data collection efforts changed from Batelle to RTI (Appendix W.2)

The SEED project has been approved by the CDC IRB; please see the *IRB Approval Letter*

(Appendix I). The consent forms for parents or caregivers (see *Written Informed Consent*

Document, Appendix E.3) include the advisements required by the Privacy Act as well as the

advisements required by 45 CFR 46. Due to the age of the children involved in this study

(2-5 years), parental consent alone is sufficient and the explicit assent of the child is not

required. During the consent process, participants are fully informed about the potential uses of the information and the fact that their participation is completely voluntary. Participants are also assured that their decision about participating in the study will not affect their child's medical care. In addition, participants are given a chance to receive a semi-annual *Participant Newsletter* (Appendix Y) which keeps them informed about the study's progress and when the study results will be shared in general medical and public health journals (since study enrollment is not yet complete, no such publications have yet occurred).

Multiple steps are taken during the data collection process to ensure that the privacy and confidentiality of each participant is ensured only to the best of the researchers' ability within the extent of the law. Each study subject is given a unique identifier (study ID) upon entry into the study. The study ID is assigned by the CADDRE Information System (CIS); the DCC maintains the records that link the ID code to the respondent name. No data collection forms will have any personally identifying information; they will only include this unique identifier. Any forms with personally identifying information (e.g., consent forms, caregiver interview), photographs, and, videotapes are kept in a locked file cabinet in a locked room with limited access to these data. Study staff limit the amount of staff who have access to the identifiable information and all study staff, including the Data Coordinating Center and the Central Lab, are required to undergo confidentiality training as part of their orientation. All study staff also must sign a *Confidentiality Agreement* (Appendix X). All forms, photographs, and videotapes will be destroyed one year after analyses are completed.

The Data Coordinating Center (DCC) plans to provide a centralized web-based data collection system that holds all of the study data. Data, including some identifiable data, acquired at the sites is transmitted and stored at the DCC as it is obtained at each of the sites. All transactions across the Internet that involve individually identifiable health information are sent to/from the

DCC as encrypted data. Personal identifiers are transmitted in encrypted form, and then stored in the database in only an encrypted form. These identifiers allow us to maintain the accuracy and validity of the data. Each SEED site's is only allowed to view its own data in identifiable form. No means exists for one site to access the personal identifiable data stored by another site. The DCC does not release identifiable data from other sites to CDC.

The approved policies and procedures for safeguarding respondent privacy are documented in a Manual of Procedures. This ensures that adequate and uniform privacy safeguards are utilized at all data collection sites, the data coordinating center, and the central biologic sample processing laboratory.

Biologic samples are collected at study sites, labeled with the study participant's ID code, and transmitted to the Central Laboratory at Johns Hopkins University for analysis and storage. The Lab does not have access to participants' personal information. Samples are stored in one of 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 is to keep them linked to personal information (through a study ID). This allows study investigators, or other researchers approved by the SEED Principal Investigators, to contact participants again in the future. Future research studies would be conducted after obtaining any needed IRB or OMB approvals. 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 requires additional consent from participant). They are also 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 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; in other words, researchers would not

be able to contact respondents to request additional 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.

A.11. Justification for Sensitive Questions

SEED includes several items that could be considered sensitive: race and ethnicity information,

family medical history, including psychiatric conditions and history of suicide; history of sexually

transmitted diseases; reproductive history, including miscarriages, abortions, and fertility

treatments; drug and alcohol use during pregnancy; child diet and stool history; use of

contraceptives; and educational level and household income. Questions of particular sensitivity

can be found in:

1. _First and Second Questionnaire Packets_ (Appendix E). Questions concerning:

      a. Family medical history, including psychiatric conditions

      b. History of child development

2. _Caregiver interview_ (Appendix C). Questions concerning:

      a. Infections of reproductive organs, including sexually transmitted diseases

      b. Reproductive history, including miscarriages, abortions, and fertility treatments,

      c. Alcohol use during pregnancy

      d. Race and Ethnicity

3. *Parent interview* (related to the Child's development) (Appendix F). Questions concerning:

     a. History of child development

     b. Services and treatments questionnaire (for child participants)

4. *Child developmental evaluation* (Appendix G) Questions concerning:

     a. Child behavioral characteristics related to developmental delays.

5. Third Questionnaire Packet (Appendix H) Questions concerning:

     a. Child participant diet and stool history

We have included these items despite their potential sensitivity because research suggests that they are potential risk factors for ASDs and the associations need further clarification. Specifically, these questions explore risk factors that may be:

- Direct hazards to the developing fetus (e.g., recreational drugs use during pregnancy, infectious diseases of the genitourinary system, medications taken during pregnancy)

- Pathways of exposure to potentially harmful agents to the developing fetus (e.g., infectious disease transmission associated with sexual intercourse)

- Related to poor reproductive outcomes (e.g., abnormal menstrual patterns or indicators of abnormal hormonal patterns such as menstrual history and fertility treatments).

Throughout the data collection process, subjects repeatedly are reminded that they may choose to skip any question that causes them undue discomfort and that their answers are not divulged to anyone outside the research group. The *Invitation Telephone Script* (Appendix N) informs participants: 'You may feel uncomfortable answering sensitive questions about your child's development. You can also skip any questions you feel uncomfortable answering.' The *Self-Administered written Consent* (Appendix E.3) that is included in the Enrollment packet states:

'Some of the questions may make you feel uneasy. You can skip any question you do not want to answer.'

Participants sign a *written informed consent* (Appendix E.3) at the initial clinic visit. It informs participants that: 'You can refuse any task and still participate in the study.' Prior to beginning the *Primary Caregiver Interview* (Appendix C.2), interviewers notify participants 'You may find some of the questions sensitive in nature but you can choose not to answer any question you wish' and, again that 'You may feel uncomfortable answering sensitive questions or discussing your pregnancies. Again, you can choose not to answer any question that makes you feel uncomfortable.'

Additionally, we ask participants to provide the last four digits of their social security number on the *HIPAA Medical Records Release Form* (Appendix E.18). While we realize that OMB is reluctant to provide approval for collection of social security numbers, we believe that collection of a partial number is necessary to conduct the study. Many providers use social security numbers as a patient identification number and require it to release medical records. We do not know in advance the specific providers (and associated study participants) that have a SSN requirement for medical record release. Because we are requesting medical records from the time of birth (i.e. 3-5 years prior to study enrollment), and even earlier for certain maternal medical records, some of the records we are requesting might very well be archived offsite which presents barriers to convenient and timely retrieval by the provider's staff. We thus anticipate the possibility of the need for repeated records requests and long lag times between request and receipt of some medical records. As such it is desirable to collect the last 4 digits of the SSN on all participants during the enrollment phase of the study; in this way we hope to minimize medical records ascertainment difficulties because of lack of required data that might arise months after the participant has completed all other components of the study protocol and might be more difficult

to contact for additional data.

We only use the last four digits of the social security number to request medical records and, as such, will not enter the number into the CADDRE Information System (CIS). Instead, we store the HIPAA Medical Records Release Forms and any other medical records request forms that a provider requires in a locked file cabinet in a locked room with limited access to these data. We limit the number of staff who have access to the information and the paper forms will be destroyed after data collection is complete.

A.12. Estimates of Annualized Burden Hours and Costs

We estimate that we will be able to successfully trace and send an *Introductory Packet* (Appendix M) to a 2,458 potential participants during the remainder of the study implementation period (Table A.12.). Potential participants are identified through schools and clinics that serve children with developmental problems and through state birth certificate registries. After the *-Introductory Packet* is sent, sites conduct an *invitation phone call* (Appendix N) with any potential study participant who responds indicating interest in the study or, when possible, with potential participants who have not returned the *invitation response card* (part of Appendix M). The *invitation phone call* includes an eligibility screen and autism screen as well as an introduction to the study and a verbal consent for the study. We estimate that 1,008 (41%) participants will return a *invitation response card* (part of Appendix M) indicating interest and be called or be contacted by phone and be screened for the study (Table A.12).

Of the potential participants who receive the invitation phone call, we estimate 423 (42%)

potential participants will be eligible to participate, based on the criteria

defined/described in Section B-1. This is the number of potential participants who will

satisfy the autism screen and selection criteria for enrollment into one of 3 subject groups

and who also agree to continue in the study. These 423 participants will be sent the

enrollment packet (Appendix E).

The next step for the study participants will be to complete the *Primary Caregiver*

Interview (Appendix C).   We expect 402 of the participants will complete this interview

(Table A.12.).

The next steps are to complete two *questionnaire packets* (Appendices E and H). (Table

A.12.). The participant will be given the option to complete these in person with study

staff, over the telephone, or as a self-administered packet. We expect 347 participants

will complete these packets.

The final step of the study will be to complete a clinical visit, including a *child*

development exam, parent interviews, biosampling, and *dysmorphology exam*

(Appendices G1, G2, P, F1 through F4, S1).   We expect 76% (321) of all participants to

complete the components of the clinical visit. The burden for cases (5 hours, 50 minutes)

is longer than the burden for the NIC and Subcohort groups (2 hours, 5 minutes). See

Tables A.12.

For a more detailed breakdown of participant burden hours and costs, which describes the

burden for each individual form, please see Appendix Z entitled *Detailed Breakdown of*

Participant Burden Hours.

Table A.12.A. Estimated Annualized Burden Hours :

Type of Respondent	Form Name	Number of Respondents	Responses Per Respondent	Avg. Burden per Response (in hours)	Total Burden Hours
Parent	Appendix M: Response Card	2,458	1	10/60	410
Parent	Appendix J and N: Invitation packet	1,008	1	30/60	504
Parent	Appendix E and H: Questionnaire packet	347	1	3.5	1215
Parent	Appendix C: Caregiver Interview packet	402	1	1.5	603
Parent	Appendix O: Follow-up telephone call packet	347	3	1.0	347
Parent and Child	App E20 and E21: Biosample packet	1,041	1	40/60	694
Parent and Child	Appendix R: Blood Draw	966	1	15/60	242
Child	Appendix G2 and P: Clinic Visit- control children packet	214	1	1.0	214
Parent	Appendix F4: Clinic Visit-control parent packet	80	1	45/60	60
Parent	Appendix E3: control parent consent form	214	1	10/60	36
Child	Appendix G1, G2,and P: Clinic Visit— Case Children packet	107	1	1.5	161
Parent	Appendix F1-F4, E3: Clinic Visit— Case Parent packet	107	1	3.5	375
Parent	Appendix S1: Medical Record Abstraction	347	5	3/60	87
TOTAL					**4,948**

Table A.12.B. Estimated Annualized Burden Costs

Type of Respondent	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response in hrs (min if <60	Total Burden (in hours)	Cost per Hour	Respondent Cost
Parent: Response Card	2,458	1	10/60	410	$18.62	$7,634
Parent: Invitation	1008	1	30/60	504	$18.62	$9,384
Parent: Questionnaire Packet	347	1	3.5 (205/60)	1,215	$18.62	$22,623
Parent: Caregiver Int. Packet	402	1	1.5	603	$18.62	$11,228
Parent: Follow-up telephone call. Packet	347	3	1.0	347	$18.62	$6,461
Parent & Child: Biosample packet	1,041	1	40/60	694	$18.62	$12,922
Parent & Child: Blood Draw	966	1	15/60	242	$18.62	$4,506
Child: Clinic control packet	214	1	1.0 (75/60)	214	$18.62	$3,985
Parent: Clinic control packet	80	1	45/60	60	$18.62	$1,117
Parent: Control consent	214	1	10/60	36	$18.62	$670
Child: Case Packet	107	1	1.5 (100/60)	161	$18.62	$2,998
Parent: Case Packet	107	1	3.5 205/60	375	$18.62	$6,983
Parent: Medical Record Abstraction	347	5	3/60	87	$18.62	$1,620
TOTAL				4,948		$92,132

According to the Bureau of Labor Statistics the average wage in the United States in June 2005 was $18.62 per hour.

Additional Supplementary Documents
- Prenatal Chart Abstraction Form (S.2)
- Labor & Delivery Chart Abstraction Form (S.3)
- Neonatal Medical Record Abstraction Form (S.4)
- Pediatric Chart Abstraction form (S.5)

Supplementary Documents to the Supporting Statement
- Authorizing Legislation and Other Relevant Laws (A)
- 60 Day Federal Register Notice (B)
- Data Flow Diagram (D.1)
- Data Collection Instruments Summary Table (D.3)
- Research Domains by Data Collection Activity (D.4)
- CDC IRB Approval Letter (I)
- Case, Comparison, and Subcohort Ascertainment Methodology (K)
- ICD-9 Codes/Part B School Eligibility Criteria (L)
- Study Hypotheses and Data Collection Tools (T)
- Data Sharing Approval Process (U)

We developed the SEED data collection battery to be a careful balance between what we ideally would like to collect (e.g., neuroimaging and more detailed medical examinations) without undue burden on the participant. Although SEED has multiple hypotheses, they are in fact only a subset of potential etiologic hypotheses for autism. Autism etiologic research is still very much in the stages of delving more deeply into multiple avenues of inquiry. No one study can address all the open questions in autism and for SEED we selected a few main hypotheses that we believed were strongly supported in the literature and for which we could get good data based on the proposed study design, being always mindful of the associated participant burden. It would be premature to narrow the scope of SEED further. The proposed study protocol and data collection battery reflects a final

balance of many compromises.

A.13. Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers

There are no capital or maintenance costs to respondents.

A.14. Annualized Costs to the Federal Government

Estimated Annual Costs for SEED Fiscal Year 2009

Estimated Annual Costs for SEED Study

Fiscal Year 2009 Expense Type		
Direct Costs to the Federal Government	Expense Explanation	Annual Costs (Dollars)
	CDC Principal Investigator (GS-15, .55 FTE)	95,528
	CDC Co-Principal Investigator (GS-14, .25 FTE)	40,977
	CDC Health Scientist (GS-14, .20 FTE)	30,351
	CDC Health Scientist (GS-13, .50 FTE)	56,803
	CDC Medical Epidemiologist (GS-15, .05 FTE)	12,352
	CDC Public Health Analyst (GS-14, .06 FTE)	8,321
	CDC Public Health Analyst (GS-13, .10 FTE)	10,517
	CDC Public Health Analyst (GS-12, 1.0 FTE)	91,940
	CDC Public Health Analyst (GS-11, 1.0 FTE)	71,050
	CDC Public Health Analyst (GS-7, .5 FTE)	10,113
	GA CADDRE Supplies/Postage/Printing/Medical Records	21,000
	Travel	10,000
	Subtotal, Direct Costs to the Government	458,952
Contractor & Grantee Costs		
	GA CADDRE Site – Contract	912,398
	CADDRE Program – Contract	136,079
	California CADDRE Site – Grant	1,386,677
	Colorado CADDRE Site – Grant	1,192,664
	JHU CADDRE Site - Grant	1,937,600
	UNC CADDRE Site - Grant	1,209,900
	Univ of Pennsylvania CADDRE Site – Grant	1,565,617
	Data Coordinating Center – Grant	700,000
	Subtotal, Contracted/Grantee Services	9,040,935
	Total cost to the government	9,499,887

Grantee costs include training site staff, full data collection, data management, and data analysis and reporting.

Grantee costs include training site staff, data collection, data management, and data analysis and reporting. The estimate takes into account an expected 5% cost of living increases for the next 2 fiscal years and a $200,000 increase in laboratory responsibilities (due to increased study enrollment).

A.15. Explanation for Program Changes or Adjustments

This is a request for OMB revision to complete enrollment and data collection activities. This revision is requested due to two proposed changes to SEED. First, minor data collection changes are requested, including modifications of some of the self administered instruments and the *Primary Caregiver Interview* (Appendix C). These changes consist primarily of clarification of questionnaire instructions and clarifying the text for specific questions to make the instruments easier to complete and further improve data quality. None of the instrument changes have a measureable impact on participant burden and individual participant burden remains the same as originally proposed. A complete list of the modifications to the instruments and appendices is provided in *List of Requested Changes to the ICR* (Appendix V). Second, a single study design change is proposed. We propose expanding the eligible study participant birth date range from September 1, 2003-August 31, 2005 to September 1, 2003-August 31, 2006. The expansion of the eligible birth date range will be used by sites if they expect that they are unable to achieve their target sample size for any of the participant groups (case or comparison groups) based on the original cohort. The potential shortfall has arisen primarily due to the larger than expected proportion of potential participants from whom we never get a response to the mailed study invitation (indicating if they are interested in

being contacted for further study information) or for whom we are unable to achieve initial contact for invitation into the study (whereas rates of enrollment among contacted participants are good). The pace and rates of contact and enrollment are being closely monitored as the study progresses and if the cohort expansion appears not to be needed then it will not be pursued. We have calculated the burden estimate for the upcoming OMB approval period to accommodate the latter change, however the total burden for the upcoming 3-year OMB approval period is less than the original 3-year period. This lower burden reflects that fact that the study will be winding down during the upcoming OMB approval period and thus the fewer total participant contacts to be made and averaged over the three year period lowers the annualized burden rate.

A.16. Plans for Tabulation and Publication and Project Time Schedule

The following schedule is designed to reflect the data collection, preparation, analysis, and reporting for the study.

A. Time Schedule:

Task	Time Schedule
1. Letters of invitation sent to potential participants	Immediately after OMB approval of revision
2. Data collection begins	Immediately after OMB approval of revision
3. Complete data collection	25 months after OMB approval of revision
4. Prepare first analytic files	2 years, 6 months after OMB approval of revision
5. Begin to Analyze data	3 years after OMB approval of revision
6. Prepare first manuscripts	3 years, 6 months after OMB approval of revision
7. Publication of first manuscripts	4-5 years after OMB approval of revision

B. Analysis Plan

General Considerations

Multiple analyses will be conducted on the data gathered in SEED. The SEED PIs will set priorities for principal analyses involving multi-site pooled data that address primary study aims. The PIs will also make decisions on the composition of the analytic teams. Once these decisions have been made, the analyses will be registered with the DCC and the Data Sharing Committee, which is responsible for approving analyses of multi-site data. The lead analyst assumes responsibility for coordinating and implementing analysis and reporting back on progress to the Data Sharing Committee.

Once principal analyses are underway, other primary and secondary analyses involving multi-site data can begin. Affiliated investigators (SEED site investigators or their colleagues/ students/ collaborators who have registered as additional investigators) can apply to the Data Sharing Committee for permission to use multi-site SEED data to complete primary and secondary analyses. The Data Sharing Committee will receive and track these proposals. Each proposal will be reviewed for approval by the Data Sharing Committee. For more information about the Data Sharing approval process, please see *Data Sharing Approval Process* document (Appendix U).

Additionally, CDC Policy on Releasing and Sharing Data requires PIs to release data as a public use data set or to share data as a restricted access data set. Given the sensitivity of the topic, we intend to share a restricted access data set with the public so that only interested, well qualified researchers can access it. These researchers must adhere to the processes and procedures outlined in Appendix U and must sign a data sharing

agreement. We will use the HIPPA "safe harbor" method to de-identify the data in the restricted access data set (e.g., no level of geography lower than the State level will be shared in the restricted access data set). Also, at the time of informed consent, respondents will indicate (by checking a box) whether they grant permission for subsequent researchers to link their data to other data sets (see Appendix E.3 Informed Consent).

Each site can also conduct site-specific analyses on the subset of data from participants they recruited, without approval from the Data Sharing Committee (the exception being analyses involving biosamples, since these are an exhaustible resource). Such site-specific analyses may not address primary study aims. Analysis data sets for site-specific projects will be subsets from the main study database maintained by the Data Coordinating Center, not independent data sets generated locally. This will be done to assure consistency of data between subanalyses and overall analyses. It is recommended that each site PI establish a process for approving applications for analyses involving site-specific data only. While site-specific analyses do not require approval by the Data Sharing Committee, the site must report to the Data Sharing Committee the aims, data elements involved, and anticipated timetable for each local analysis.

As mentioned above, the Data Coordinating Center will have a responsibility for coordinating information and will also maintain a database on study data analyses. The DCC has worked with the sites 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 contracts with Internet System for Assessing

Autistic Children (ISAAC) for some of the data entry tools.

The DCC is responsible for all final checks and edits on data submitted from study sites

and ISAAC.   The DCC will also create a series of standard core recoded and new

variables based on input from the SEED investigators. This work could include

comparison of information on common exposure from two alternative data sources (e.g.,

maternal interview and maternal medical records) as well as creation of summary

variables (e.g., total scores from behavioral assessments, construction of summary indices

of obstetric suboptimality, etc.). Analysts working on approved projects who develop

additional recodes or who create new variables will be responsible for submitting the

code and rationale for development of these variables to the DCC. Even if these are not

adopted as additional core recoded variables or new variables, it is critical that there be a

central record of how any variable ultimately used in a disseminated analysis was

recoded or created.

The DCC is responsible for establishing central data file architecture which is expected to

include linkable core files organized by both data collection instrument, study group

(Case, NIC, subcohort), study subject (child, mother, father), and, in some cases, domains

cross-cutting several instruments (e.g., a behavioral phenotype summary file).

Codebooks will be developed for each file and a user's guide developed for the inter-

relationships between files. The DCC will also have responsibility for the assembly of

analysis files by linking variables from core files requested by investigators who have

had analyses approved by the Data Sharing Committee.

As a direct extension of the activities performed in cleaning, recoding, and new variable creation, the DCC may perform initial descriptive analyses on study variables. For dichotomous and categorical variables this would include frequency distributions and missing value counts. For continuous variables this would also include assessments of central tendency, spread, skewness, and recommended transformation for normality as well as missing value counts.

All other analyses will be performed by members of the analytic teams for each analysis registered with the SEED Data Sharing Committee. Primary 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 nonheritable risk factors, assessment of specificity of associations, and assessments of interactions), and 3) comparison of biomarkers across Case, NIC, and Subcohort groups.

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 NIC and subcohort. A priori, we may consider the following ASD subgroups for analysis:

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

These are not mutually exclusive categories, however, and one of our goals will be 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, cognitive status) across all 3 subject groups.

Although the primary unit of analysis will be the index child, for some analyses classification of affected/unaffected status may include criteria that consider diagnosed or reported medical, neurologic, and developmental conditions in parents and/or siblings.

Specific Analyses

The foregoing discussion provided an overview of general analytic features that apply to all SEED analyses. What follows are more specific examples concerning the analytic approach for select hypotheses under 5 of the 6 domains (including approaches to biomarker analyses): characterizing the autism phenotype (including gastrointestinal features), infection and immune function, reproductive and hormonal features, and genetic features. For reference, all SEED hypotheses are provided in *Study Hypotheses and Data Collection Tools* (Appendix T).

Characterization of Phenotype (including biomarkers)

Because of suspected etiologic heterogeneity in ASD one avenue of analyses will define novel, and potentially etiologically distinct, phenotypic ASD subgroups. Characterization of ASD phenotype includes analyses focused on the ASD case group and analyses

involving comparison groups. Analyses focused on the ASD group will include those using variables capturing behavioral characteristics known a priori to be associated with ASD (e.g., results from Mullen scales, ADOS scales, and ADI-R scales; indicators of regression; other indicators of core symptoms) to identify subgroups where particular traits tend to co-occur. Statistical analyses used here will include true multivariate techniques such as factor or principal components analyses. Examination of these behavioral data may facilitate characterization of intermediate traits to ASD, or endophenotypes.

In addition, phenotypic features strongly suspected, but not confirmed, to be associated with ASD will be considered. Phenotypic features considered include symptoms not currently considered in the realm of core characteristics (e.g., gastrointestinal disturbances, differences in gut-derived hormones, sleep disturbance, sensory dysfunction), anthropometrics (head circumference) and minor dysmorphology. Analyses here will focus first on determining whether these features do occur with greater frequency in the ASD population and then will explore whether adding these features to the known list of behavioral characteristics leads to different subgroup clustering.

<u>Hypothesis:</u> Children with ASD are more likely than children in the NIC or sub-cohort to have co-morbid medical or neurodevelopmental conditions including Tuberous sclerosis (TS), Neurofibromatosis (NF), Fragile X, seizure disorders/epilepsy, and attention /hyperactivity problems. (Note: these items were chosen as examples because they represent both diagnosed conditions - of varying expected prevalence - and measures of abnormal behavior, some of which may be indicators of core symptoms.)

Medical (e.g., Tuberous sclerosis, Neurofibromatosis, Fragile X, seizure disorders/epilepsy) and neurodevelopmental conditions (e.g., ADHD) that have been

diagnosed by a physician are captured in the Caregiver Interview and child medical record abstraction (neonatal/pediatric/specialty). The reported prevalence of these specific diagnoses in the general population and prior reported prevalence in ASD is provided below:

	General population	Children with ASD
Seizures	3-5%	20-25%
TS	0.0106% *	0.4-2.9%
NF	0.03% **	0.2-14%
ADHD	5-10%	25-30%
Fragile X	0.025% with MR ~ 3%	13%

*10.6/100,000
**3/10,000

Note: These estimates are summaries and do not take into consideration variation by child age. For example, prescribed stimulant use (as a proxy for ADHD) has been reported in different studies to range from 0.18% to 0.68% in 2-4 year olds and 2.4% to 7.8% in 5-9 year olds.

Other behavioral information [e.g., decreased ability to shift attention (also a possible core ASD symptom), inattention, hyperactivity, impulsivity] are captured through the Child Behavior Checklist 1 ½ to 5 (CBCL), Carey Temperament Scale (CTS), Vineland Adaptive Behavior Scale, and Social Responsiveness Scale (SRS). For discrete diagnoses, children are classified on the basis of presence/absence of the condition while outcomes derived from standardized test scores may be defined on the basis of score as a continuous measure or categorized/ dichotomized on the basis of score falling above or below a specified cutoff. Further, children may be classified on the basis of having any of the co-morbid or neurodevelopmental conditions in question, having one or more of a class of co-morbid or neurodevelopmental conditions, or having a specific co-morbid or neurodevelopmental condition (depending on prevalence). Apart from diagnosis or test score, other features of the condition in question are considered in characterizing affected individuals, such as: age of symptom onset or diagnosis, severity (where relevant, e.g.,

type, frequency of, and medication for seizures).

Potential confounders to be considered include family history of relevant neuropsychiatric and developmental conditions, severity of core deficits, presence of other co-morbid conditions, age, medication/treatment history, and measures of pre- and perinatal risk such as minor dysmorphic features, gestational age and abnormal fetal growth. Careful consideration will be given as to whether these factors are true confounders or, in fact, links in the causal pathway to ASD.

Analyses will focus on comparisons between children with ASD and both the NIC and sub-cohort groups to determine the specific association of a condition with ASD relative to children with other developmental problems or the general population. Analysis will begin with descriptive measures (for categorical or quantitative data) of the distribution of the condition(s) of interest and its associated features across subject groups. For these analyses, contrasts will employ standard analytic methods for case-control designs. Unadjusted and adjusted associations will be estimated with relative risk estimates. Unconditional logistic regression techniques will be used to adjust measures of association between ASD and the condition of interest for the effects of significant confounders identified in the descriptive analyses. Stratification on key factors, such as gender, cognitive status, and developmental regression (applicable to ASD cases only) will be performed in both descriptive and multivariable analyses.

Finally, conditions that are significantly associated with ASD in these analyses will be subsequently examined in case-only analyses - using multivariate techniques such as

principal components or factor analyses - for evidence of clustering with other

phenotypic features to identify ASD subgroups for consideration in etiologic analyses.

<u>Hypothesis:</u> The prevalence of GI dysfunction will be higher in children with ASD compared to children in the subcohort and NIC group

Using data from the seven-day stool diary and the "Survey of Gastrointestinal Function,"

gastrointestinal dysfunction is defined as the presence of one or more of the following:

- Four or more stools per day;
- Two or more hard (type 1) stools per week;
- Only one stool per week when that stool is hard (type 1) or loose/watery (type 6 or 7);
- More than one third of stools are loose/watery (type 6 or 7);
- Two or more stools per week are watery (type 7);
- Vomiting in any frequency.
- Laxative or stool softener use in the past 30 days.

Based on pilot data involving 47 children with ASD and 31 typically developing controls,

we expect 25 to 35% of the children in the ASD group to have GI dysfunction, compared

to 6-13% in the sub-cohort. We will first compare the ASD group to each of the two

control groups (NIC and sub-cohort) using simple chi-square tests.

Using unconditional logistic regression, we will separately model GI dysfunction in

children with ASD versus each of the control groups. Children with ASD are likely to

differ from children in the sub-cohort in other important ways that may influence

reported gastrointestinal function. Potential confounders to consider will be age, IQ level

(from the Mullen), whether the child wears diapers (from ''Survey of Gastrointestinal

Function'), and fiber and caloric intake (from 3 day diet record). Tests for interactions

between covariates and outcome will be performed. Relative risk estimates and 95%

confidence intervals will be presented. Specific ASD subgroup analyses will examine the

relationship between ASD and GI dysfunction among 1) children with and without

regression, and 2) children with and without a family history of autoimmune and

gastrointestinal dysfunction. Sweeten, 2003, reported 30% of parents of children with

PDD and 12% of parents of healthy controls reported an autoimmune disorder.

<u>Hypothesis:</u> *Children with ASD and GI dysfunction will have higher levels of serotonin compared to children with ASD without GI dysfunction*.

We will examine differences in serotonin levels between ASD cases with and without GI

symptoms based on analysis of blood samples collected on children during the clinic

visit.   We will evaluate the normality assumption by examining the distribution of the

data via histograms and/or by performing a normality test. The equality of variances

assumption will be verified with the *F* test. If assumptions are met, we will compare

means between groups using Student t tests and ANOVAs. If assumptions are not met,

we will evaluate differences in means between groups using a nonparametric alternative

to the Student t test.

Specific Analyses: Estimation of Risk Factor Associations, Including Biomarkers

Many analyses will focus on associations of potential risk factors with ASD. Risk factors

include data collected about family history, especially maternal medical history,

exposures during the windows of the preconception period and the index pregnancy (e.g.,

maternal medication use, presence of indicators of infection) and early life of the child

(e.g., frequency of otitis media). Construction of exposure variables themselves will

typically require sophisticated analysis – this being an area requiring close collaboration

between the analytic team and the DCC.

An additional class of analyses involves comparison of non-genetic biomarkers across study groups. Biomarkers of interest include, for example, cytokines, neuropeptides, neurotrophins, autoantibodies, antibodies, hormones, and immune cell counts. A Biomarker Studies Advisory Committee composed of one investigator from each SEED site, a representative from the Central Biosample Laboratory and Repository, and one outside expert will advise the Data Sharing Committee on technical issues related biosample management of and technical aspects of analyses involving biomarkers. Biomarker analyses may involve descriptive approaches (e.g., the comparison of assay levels across study groups) or incorporation of biomarker data into relative risk models (e.g., estimation of relative risks associated with second and third tertile levels compared to first tertile). As with genetic analyses, sequential analytic approaches may also be recommended for biomarker analyses involving stored sample.

All analyses will generally proceed through four phases: univariate descriptive analyses of hypothesized risk factors and potential confounders, multivariate examination of potentially related factors for possible collinearity, simple analysis of associations between the selected risk factor and outcome as well as the relationship of both to potential confounders (e.g. parental age, parity, gestational age, birthweight, other treatments or conditions, etc.). Unadjusted and adjusted associations will be estimated with relative risk estimates. Finally, multivariable unconditional logistic regression models will be used to adjust for possible confounders and to assess the relative contribution of different factors, including potential effect modifiers.

Infection/Immune Function

Converging evidence points towards an immunologic component in an unknown

proportion of children with autism, including exposure to maternal infection and

inflammation during pregnancy and immune function abnormalities, including

autoimmunity.

*Hypothesis: **Mothers of children with ASD are more likely to have infections during pregnancy compared to mothers of sub-cohort children.***

Infections during pregnancy are quite common (40-60%), with specific conditions

occurring in the range of 5-10%: 11% reported UTI's , 20% reported fever, 11.5%

reported influenza/pneumonia. In a national survey of OB/GYNs, respondents estimated

that 5% of their patients had URI symptoms at their office visits.

The number, type, and timing of maternal infections around the time of pregnancy

(including neonatal infection within 24 hours post-delivery), as reported by mothers

during the Caregiver interview, and as recorded in maternal and neonatal medical

records, will be compared between all ASD cases and sub-cohort controls in the dataset.

For the subset of physician diagnosed and documented infections abstracted from

prenatal records, data on confirmation of the diagnosis (e.g., lab, clinical), duration of

infection, and fever associated with infection will also be examined. Timing of infection

will be defined by trimester and by intervals between infection and labor onset/delivery

and between multiple infections. Several exposure definitions will be evaluated,

including dichotomous (any infection vs. no infections), categorical (e.g.,

chorioamnionitis, UTI, Renal, Vaginal, STD, GI, URI, perinatal, etc), and individual

infections, depending on frequency. Factors such as maternal age, gestational age,

maternal autoimmune disease in pregnancy, and treatment of infections or exposure to

other anti-inflammatory therapies, will be evaluated as potential confounders and

included in multivariable models when appropriate.

Hypothesis: ***Families of children with ASD are more likely to have a history of***
autoimmune disorders than families of subcohort or NIC children.

Reported prevalence of maternal history of autoimmune diseases:

- all autoimmune disorders as a group: 8% in 4-year period around date of delivery
  (-2yrs to +2 yrs), and 14% anytime before or after date of delivery
- specific autoimmune disorders in 4-year time period: alopecia (1.4%);
  autoimmune thyroid disease (3.2%); IBD (0.4%); psoriasis (1%); rheumatoid
  arthritis (0.3%); Type 1 diabetes (0.4%)

Family history of autoimmune disorders as a group, and for specific autoimmune

disorders, as reported by parents on the autoimmune survey, and as recorded in maternal

medical records, will be compared between all children with ASD and the NIC and

subcohort controls. Since autoimmune disease is collected on different sets of

individuals, exposure may be variously defined, for example: index children who have

any family history of autoimmune disease, index children who have a maternal history of

autoimmune disease (anytime, during pregnancy), and index children who have a

diagnosis of autoimmune disease. For the subset of physician diagnosed and documented

autoimmune disorders abstracted from the maternal medical records, data on date of

diagnosis, time period during pregnancy when condition was active, treatment during

pregnancy, and age at initial diagnosis will also be examined. Several exposure

definitions will be evaluated, including dichotomous (any family history vs. none),

categorical (e.g., by organ system), and individual autoimmune disorders, depending on

frequency. We will also investigate autism risk associated with family history of

autoimmune disorders for specific family members (e.g. mother, father) and for number

of affected family members. Factors such as maternal age, total number of family members, measures of infection in pregnancy, and treatment for autoimmune disorders and other inflammatory conditions (which may also be considered separately as exposures) will be evaluated as potential confounders and included in multivariable models when appropriate.

Hypothesis: ***Children with ASD will have different blood levels of markers of inflammation compared to subcohort or NIC children.***

To evaluate biomarkers of inflammation, we analyze the blood collected from the children during the clinic visit and measure levels of cytokines and chemokines (e.g., interleukin-1B, 6, 8, 10, tumor necrosis factor alpha (TNFa), Interleukin-1RA, Interferon g), immunoglobulins, autoantibodies (e.g., anti-myelin basic protein, anti-neuron-axon filament protein, anti-glial fibrillary acidic protein, antinuclear antibodies (ANA), and antibodies to infectious agents (e.g., maternal and child cytomegalovirus IgG), and leptin. We examine individual biomarker distributions for departures from normality, apply transformations (e.g., log transformation, square root transformation), and use non-parameteric techniques as needed. We will construct continuous, categorical (e.g., quartiles, quintiles), and dichotomous (e.g., present vs. absent; above 90% vs. below; above median vs. below median) measures of inflammation, depending on the actual distributions of the analytes measured. We will examine correlation matrices of all analytes, and employ statistical techniques to identify inflammatory biomarker 'clusters' that may be predictive of ASD risk. Assessments of relations between markers may indicate approaches to reducing the number of candidate markers - by identifying sentinel markers, or creating index variables - in multivariate analysis. These data reduction approaches will be used with caution, however, as we do not want to lose important

information.

Reproductive/Hormonal Features

The natural fluctuation of maternal hormones pre- and perinatally is important to allow

conception, maintain the pregnancy, and initiate birth. 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, including

fertility treatment or labor induction, contraceptives) also suggest an association between

ASD and prenatal hormonal features

Hypothesis: Mothers of children with ASD have different patterns of exogenous hormone exposures (hormonal medications including oral contraceptives, infertility treatments, treatments for conditions, and medications administered during the labor and delivery and perinatal period such as oxytocin and pitocin) during pregnancy and through the end of breastfeeding than mothers of children in the NIC group or subcohort.

Reported prevalence of select exogenous hormone exposures:

2-7% any hormone treatment for fertility
  1 % use of artificial reproductive techniques (ART)
  2-5% use of hormone in non-ART fertility treatment
2-8% oral contraceptives during pregnancy
  1-2% failure rate for prescribed use and up to 8% failure rate with typical use
21% exposure to pitocin/oxytocin during labor induction

Information on exogenous hormone exposure will be collected as a part of the maternal

obstetric history through the caregiver interview and through maternal medical records.

SEED will collect details about the type and timing of contraceptive use and consult

available drug dictionaries as needed to determine specific hormone exposure. The

timing of exposure with respect to the pregnancy period and during breastfeeding will be

determined to classify exposure status to specific hormones during critical periods during

fetal and infant development. While the details on the dose and duration may not be

available for all, we will likely have the ability to evaluate whether any exposure to

exogenous hormones occurred and the purpose for the hormones (infertility,

contraception, labor induction, etc).

We will initially describe the rate of each type of hormone use among all study groups

(ASD, NIC, sub-cohort). The relation between a particular type of hormone use and

ASD compared to the sub-cohort and compared to the NIC group will be estimated using

multivariable logistic regression, adjusting for potential confounding factors such as

maternal age, parity, and various indicators of socio-economic status. If phenotypic data

can be used to further refine ASD into more homogenous subgroups, we will investigate

whether any observed association between exogenous hormone exposure and ASD might

be stronger among specific case subgroups.

Genetic Analysis

Principal analytic goals involving genetic associations will be to identify genomic

variation associated with autism (genome-wide associations) and to test candidate gene

main effect associations and interactions between genotypes and environmental

exposures. Exploration of potential epigenetic influences will also be pursued.

Analytic strategies will focus on genetic main effects and particular hypotheses of gene-

by-environment interactions. This will be carried out in a generalized linear modeling

framework (Schaid et al. 2002; Lake et al. 2003), where genotype or diplotype (pair of haplotypes) information is included as an independent factor, along with other non-heritable factors thought to be important in the model. The GLM framework allows us to model etiology (logistic regression for ASD versus subcohort comparisons, or Poisson for time-to-ASD analyses), as well as to test for the importance of genetic factors on ASD phenotypic subgroups (e.g., head circumference or logistic regression for psychiatric phenotypes among cases). The latter set of analyses may help identify phenotypic subgroups that are more directly related to a particular etiologic class. We also intend to highlight the wealth of information to be collected in this study including extensive environmental exposure information, and phenotype characterization. We anticipate exploratory analyses to incorporate genetic and non-heritable risk factor information simultaneously. These exploratory analyses will include (and likely extend beyond) multiple dimensionality reduction methods (Hahn et al. 2003), classification and regression trees (Hizer et al. 2004), and logic regression (Kooperberg et al. 2001; Kooperberg and Ruczinski 2005). Each of these approaches aim to find sets of factors (whether multiple genes, non-heritable risk factors, or both) that act together as a risk set for ASD. These methods can be considered as "model searching", with the detected models then used to estimate effect sizes in a validation study, or through cross-validation within the same data set.

The fact that DNA will be available from children as well as their parents provides an opportunity to test genetic and environmental associations at the parent and child levels in several ways. Comparisons using the Case and sub-cohort children as the unit of analysis can test associations between the genes carried by the child and risk for ASD

(eg, using logistic regression as described above). Family-based tests (e.g., the TDT) can also be performed among the case-parent trios that are available, to further elucidate parent-of-origin effects and provide tests of genetic linkage. The availability of trios from the sub-cohort and NIC groups would then provide the opportunity to test assumptions such as absence of general transmission distortion. The loglinear modeling approach of Weinberg et al (1998) can be applied to case-parent trio data to estimate genetic main effects, interactions between genetic and environmental effects, and, also, make determinations of whether genetic effects are offspring or parentally mediated. This ability to consider parent-of-origin effects is a particularly attractive advantage of the available trio analyses.

Finally, because sampling will be population-based, we will be able to estimate allele frequencies as well as penetrances (risk estimates). With these data in hand, attributable risk estimates for particular genes or gene-environment combinations can be constructed.

The Data Sharing Committee will be advised on decisions about genome-wide genotyping and candidate gene genotypes to be explored by a Genetic Studies Advisory Committee (GSAC) comprised of one investigator from each SEED site and two outside genetic researchers. The Data Sharing Committee will refer any proposals for ancillary analyses involving genotyping received from others within SEED to this group for their opinion. Once specific analyses are approved, study IRBs will be informed and addendums sent through review processes as needed. The GSAC will give priority to candidate genotypes emerging from family-based linkage studies and genotypes that influence pathways also potentially affected by the environmental exposures on which the study has collected data. Until immortalized cell lines are established or other

techniques are available (i.e., whole genome amplification), DNA is a depletable resource (although there are several biosamples that can provide DNA), and the GSAC 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.

A.17. Reason(s) Display of OMB Expiration Date is Inappropriate

No such exemption is requested.

A.18. Exceptions to Certification for Paperwork Reduction Act Submissions

No exceptions apply to this data collection.