

**Supporting Statement Part A**

**The Study to Explore Early Development (SEED)  
SEED Phase 3**

**OMB # 0920-16PA**

**New**

**Laura Schieve, PhD**  
Epidemiologist  
Centers for Disease Control and Prevention  
Email: [LSchieve@cdc.gov](mailto:LSchieve@cdc.gov)  
Phone: 404.498.3888  
Fax: 770.498.1541

**March 8, 2017**

## Table of Contents

Abstract.....	4
A. Justification.....	5
A.1 Circumstances Making the Collection of Information Necessary .....	5
A.2 Purpose and Use of Information Collection .....	5
A.3. Use of Improved Information Technology and Burden Reduction.....	13
Participant Burden Reduction.....	15
A.4. Efforts to Identify Duplication and Use of Similar Information.....	16
A.5. Impact on Small Businesses or Other Small Entities.....	19
A.6. Consequences of Collecting the Information Less Frequently.....	19
A.7. Special Circumstances Relating to the Guidelines of 5 CFR 1320.5.....	20
A.8. Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency .....	20
A.9. Explanation of Any Payment or Gift to Respondents.....	22
A.10. Protection of the Privacy and Confidentiality of Information Provided by Respondents.....	25
A.11 Institutional Review Board Approval and Justification for Sensitive Questions.....	30
Justification for Sensitive Questions.....	30
A.12. Estimates of Annualized Burden Hours and Costs.....	32
A.13. Estimates of Other Total Annual Cost Burden to Respondents or Record Keepers.....	33
A.14. Annualized Cost to the Government.....	33
A.15. Explanation for Program Changes or Adjustments.....	34
A.16. Plans for Tabulation and Publication and Project Time Schedule.....	34
A.17. Reason(s) Display of OMB Expiration Date Is Inappropriate.....	35
A.18. Exceptions to Certification for Paperwork Reduction Act Submissions.....	36

## Attachments

<b>1</b>	<b>Authorizing Legislation</b>
1.a	Public Health Service Act
1.b.	The Combating Autism Act of 2006
1.c	Children’s Health Care Act of 2000
<b>2</b>	<b>60-Day Federal Register Notice</b>
<b>3</b>	<b>Site Specific Ascertainment SEED</b>
<b>4</b>	<b>Eligibility</b>
4.a.	SEED 3 Eligibility Criteria

- 4.b. ICD-9 Codes & Part B/School Eligibility Codes & List of Variable needed from data request
- 5 Social Community Questionnaire (SCQ)**
- 6 Maternal Interview**
  - 6.a. Maternal Interview Call Script
  - 6.b. Maternal Interview Definitions
  - 6.c. Maternal Interview Preparatory Guide
  - 6.d. Maternal Interview
- 7 Pregnancy Reference Questionnaire Packet**
  - 7.a. Pregnancy Reference Form Script
  - 7.b. Pregnancy Reference Form
- 8 Self/Parent Administered Forms**
  - 8.a. Maternal Medical History
  - 8.b. Paternal Medical and Occupational History
  - 8.c. Child Health History
  - 8.d. Maternal and Child Residential History
  - 8.e. Child Behavior Checklist (CBCL)
  - 8.f. Social Responsiveness Scale (Preschool)
  - 8.g. Social Responsiveness Scale (Child)
  - 8.h. Services and Treatment Questionnaire
  - 8.i. Cover Letter ASD and POP
  - 8.j. Cover Letter DD
  - 8.k. Instructions Yellow Folder – Materials to Keep
  - 8.l. Instructions and Check list Green Folder – Things to Return ASD Version
  - 8.m. Instructions and Check list Green Folder – Things to Return DD Version
  - 8.n. Instructions and Check list Green Folder – Things to Return POP Version
  - 8.o. Medical Glossary for Maternal Paternal and Child Health Forms
  - 8.p. Self-Administered Consent Forms SEED 3 ASD, POP, and DD
- 9 Developmental Assessment Score Sheets**
  - 9.a. Mullen Scales of Early Learning Score Sheet
  - 9.b. Vineland Adaptive Behavior Scales (VABS) Score Sheet
  - 9.c. Vineland Phone Script
  - 9.d. ADOS Score Sheet Module-1
  - 9.e. ADOS Score Sheet Module -2
  - 9.f. ADOS Score Sheet Module -3
  - 9.g. ADI-R Score Sheet
- 10 IRB Approval Letter**
- 11 Consent Forms**
  - 11.a. Overall Consent Document – ASD
  - 11.b. Overall Consent Document – DD
  - 11.c. Overall Consent Document – POP
- 12 SEED Semi-annual Newsletter Sample**
- 13 List of SEED 2 PIs**
- 14 Review Panel for SEED – 2015**
- 15 Clinic/ Home Visit Biosamples Packet**
  - 15.a. Frequently Asked Questions about Biologic Sampling

15.b.	Picture Story
15.c.	Clinic Visit Prep Guide - ASD
15.d.	Clinic Visit Prep Guide – POP
<b>16</b>	<b>Self-Administered Saliva Collection Kit</b>
16.a.	Anthropometric Exam Form
16.b.	Saliva Collection Instructions
16.c.	Saliva Collection Written Consent Form
16.d.	Saliva Transmittal Form
<b>17</b>	<b>Blood Draw Packet</b>
17.a.	Child Blood Draw Form
17.b.	Adult Blood Draw Form
<b>18</b>	<b>Study Flow for Participant SEED 3</b>
<b>19</b>	<b>Invitation Packet</b>
19.a.	Invitation Letter – ASD
19.b.	Invitation Letter – DD
19.c.	Invitation Letter – POP
19.d.	Study Brochure – ASD
19.e.	Study Brochure – DD
19.f.	Study Brochure – POP
19.g.	Response Card
19.h.	E-mail Notice
<b>20</b>	<b>Enrollment Packet</b>
20.a.	Enrollment Packet Cover Letter ASD and POP SEED 3
20.b.	Enrollment Packet Cover Letter DD SEED 3
20.c.	Study Bill of Rights
<b>21</b>	<b>Telephone Schedules and Scripts</b>
21.a.	Invitation Call Script - Previous ASD Diagnosis SEED 3
21.b.	Invitation Call Script - DD source selection SEED 3
21.c.	Invitation Call Script - POP source selection SEED 3
21.d.	Follow-up Call Items Checklist SEED 3
21.e.	Follow-up Call 2 Items Checklist SEED 3
<b>22</b>	<b>Passive Refusal Letter</b>
<b>23</b>	<b>Parental Feedback Letter and Other Communication with Parents</b>
23.a.	Parental Feedback Letter
23.b.	Thank You Open - Ended Refusal Letter
23.c.	Thank You Letter ASD POP
23.d.	Thank You Letter DD Completion Letter
23.e.	Thank you or General Greeting Card SEED 3
23.f.	Age Out Letter
<b>24</b>	<b>Public Comments on 60-Day Federal Register Notice</b>
<b>25</b>	<b>References</b>

- **Goal of the study:** SEED is a multi-site case-control study designed to investigate risk factors for Autism Spectrum Disorders (ASDs) and the health and behavioral characteristics of children with ASDs. SEED is specifically focused on preconception, prenatal, perinatal and early postnatal risk factors with primary emphasis on genetic, infectious, immunological, hormonal, and obstetric risk factors. Additional SEED focus areas include characterization of distinct ASD behavioral phenotypes and assessment of ASD health impacts, particularly gastrointestinal disorders and symptoms.
- **Intended use of the resulting data:** The data from SEED 3 will be combined with data from the first two SEED phases to enable investigators to conduct in depth analyses of ASD risk factors including assessment of potential etiologic subgroups and gene-environment interactions. SEED findings will inform the public about potential causes of ASD with implications for primary prevention, and the health of children with ASD with implications for secondary prevention of associated sequelae.
- **Methods to be used to collect:** Data will be collected once from participant families using multiple methods including: 1) maternal telephone interview with questions about maternal reproductive history and pregnancy with the index child, 2) parent-completed questionnaires about parental and child health and child development, 3) in-person child developmental evaluation, 4) maternal and child anthropometry measurements, and 5) biosampling from biological parents and child.
- **The subpopulation to be studied:** SEED 3 will enroll children born between 2014 and 2017 and aged 2-5 at enrollment and their parents. Children will be identified at 6 participating sites (5 sites funded through a competitive funding process and 1 site in metropolitan Atlanta managed by CDC). Three groups of children will be included: children with ASDs (ASD case group – defined based on SEED developmental assessment), children with other developmental (non-ASD) conditions (DD comparison group), and children from the general population (POP comparison group). Potential ASD and DD children will be identified from multiple health and education sources at each site. POP children will be sampled from site birth certificates.
- **How data will be analyzed:** Children with ASDs will be compared to the two control groups (DD and POP). Odds ratios with 95% confidence intervals will be calculated for associations between ASD and various risk factors: overall; for etiologic ASD subgroups defined empirically based on analyses of the detailed behavioral and other phenotypic data; within strata defined by key demographic characteristics, such as child sex, race-ethnicity and maternal age; and after adjustment for demographic and perinatal characteristics. Interactions between genetic and non-genetic risk factors will be assessed through stratification and modelling. Child health characteristics will be similarly assessed through comparison of children with ASD to children in the DD and POP comparison groups.

## **A. Justification**

### **A.1. Circumstances Making the Collection of Information Necessary**

This Information Collection Request (ICR) is submitted under the classification “New.” The length of data collection requested for Office of Management and Budget (OMB) approval is 3 years. ‘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. NCBDDD at CDC is making this request as authorized by Section 301(a) [42 U.S.C. Section 241(a)] and 317(c) of the Public Health Service Act [42 U.S.C. 247b-4], as amended (**Attachment 1.a.**); the *Combating Autism Act of 2006, Pub. Law No. 109-416* (**Attachment 1.b.**); and the Children’s Health Care Act of 2000 (**Attachment 1.c.**).

### **Background**

The Children’s Health Care Act of 2000 mandated CDC to establish autism surveillance and research programs to address the number, incidence, correlates, and causes of autism. Under the provisions of this act, NCBDDD has previously implemented and completed data collection to this end. OMB first approved “The Study to Explore Early Development’ (OMB 0920-0741) in October 2007. The second phase of the effort (SEED 2) was inappropriately granted an OMB-PRA clinical research exemption. Therefore, for this new project (SEED 3) we seek OMB-PRA clearance under a new ICR versus a reinstatement of the previously approved collection. The study protocols for SEED phases 1, 2, and 3 are very similar. The major difference between the phases is that each subsequent phase includes a more recent birth cohort. While all SEED phases have the same research goals and the same recruitment protocol and study design, data collection has been streamlined between SEED 1, SEED 2, and SEED 3 such that 1) many study instruments and data collection components included in the SEED 1 protocol were not included in the SEED 2 protocol, and are similarly not part of the SEED 3 protocol; 2) two instruments

included in both SEED 2 and SEED 3 protocols were developed subsequent to SEED 1 to capture an abbreviated version of information that had been included on some of the discontinued SEED 1 forms and to capture some additional information overlooked in the SEED 1 protocol; 3) instruments included in SEED 1 underwent review and minor revision following SEED 1 to address ambiguities and difficulties experienced during SEED 1 data collection; 4) collection of buccal swabs was replaced with collection of saliva specimens in SEED 2 and SEED 3 to increase the DNA yield; and 5) upon review of data needed to address study research questions, it was decided that for SEED 3, only minimal data collection will be needed for the group with non-ASD developmental disabilities. Implementing this third phase of SEED will increase the total SEED pooled sample size for investigation of high priority hypotheses. Maintaining the same basic study design and general protocol integrity will ensure that data pooling can be achieved across SEED phases.

The overall purpose of SEED is to investigate risk factors for autism spectrum disorders ASD and symptom subgroups of ASD, using a case-control study design that includes ascertainment of case and comparison groups that represent diverse population subgroups.

The U. S. prevalence of ASDs is estimated at 1-2% (Zablotsky et al., 2015; Blumberg et al., 2013, Autism and Developmental Disabilities Monitoring [ADDM] Principal Investigators, 2014). Apart from the identification of some rare genetic conditions that are commonly associated with autism (Miles, 2011), causal mechanisms for the disorder largely remain unknown. While numerous genetic factors have been implicated in the etiology of ASDs and sibling and twin studies suggest high heritability (Risch et al., 2014; Hallmayer et al., 2011), the specific genetic mechanisms appear complex and research gaps remain. Moreover, select prenatal environmental factors and adverse perinatal outcomes have also been associated with

ASDs (Schieve et al., 2015; Schieve et al., 2011; Talbott et al., 2015). The composite evidence supports the likelihood of gene-environment interactions (Miles, 2011; Hallmayer et al., 2011).

In the face of these considerable research gaps, large population-based epidemiologic studies of ASD etiology are lacking. The proposed data collection for a third phase of SEED addresses this critical need. While the composite sample from SEED 1 and 2 allows investigators to address many important research questions, it is insufficient for many analyses of interest including examination of the many important risk factors with a prevalence of 1-5%, that are modestly associated with ASD (OR<2.5) and associations with ASD subgroups. SEED 3 will increase the composite sample to  $\geq 2,000$  children in all groups – ASD case and two comparison groups (children with other developmental disabilities [DD] and children from the general population [POP]) -- thus, expanding our ability to analyze rare exposures and/or modest (yet scientifically important) associations, assess associations within key ASD subgroups, and explore potential effect modifications between various ASD risk factors, including the potential for gene-environment interactions.

## **A.2. Purpose and Use of Information Collection**

### ***I. How this information will be used and for what purpose:***

The information collected in SEED is used to conduct epidemiologic analyses to 1) characterize the ASD phenotype according to current research standards; 2) assess risk factors for ASD, how these risk factors might vary according to distinct ASD subtypes, and potential effect modifications between risk factors; and 3) assess associations between ASD and health conditions such as gastrointestinal disorders.

The prevalence of ASDs in the U.S. is 1-2% (Zablotsky et al., 2015; Blumberg et al., 2013, ADDM Principal Investigators, 2014). In addition to the profound, lifelong impacts on individuals' functioning given the core deficits in social-communication abilities, a high

proportion of children with ASD also have one or more other developmental impairments such as intellectual disability or attention-deficit-hyperactivity-disorder (ADDM Principal Investigators, 2014; Schieve et al., 2012; Levy et al., 2010) and children with ASDs have higher than expected prevalences of health conditions such as obesity, asthma and respiratory disorders, eczema and skin allergies, migraine headaches, and gastrointestinal symptoms and disorders (Phillips et al., 2014; Schieve et al., 2012). Apart from the identification of some rare genetic conditions that are commonly associated with autism (Miles 2011), causal mechanisms for the disorder largely remain unknown. The composite evidence supports the likelihood of gene-environment interactions (Miles, 201; Hallmayer, 2011); however, considerable research gaps remain in our understanding of specific genetic and environmental risk factors and whether risk factors are differential for ASD subtypes. SEED was designed to address this critical need.

In planning and designing SEED, study investigators conducted an extensive review of the literature (Newschaffer et al., 2007) and based on their findings, designated 6 primary research domains: 1) investigation of the ASD phenotype; 2) assessment of genetic risk factors and genetic differences in children with and without ASD; 3) assessment of prenatal infection and immunologic risk factors; 4) assessment of reproductive and hormonal risk factors (subsequently split into two domains – hormonal and obstetric risk factors); 5) assessment of child health with a focus on gastrointestinal symptoms; 6) assessment of sociodemographic features associated with ASD. Additionally, several areas were designated as secondary research domains: assessment of substance use during pregnancy; assessment of maternal and paternal occupational exposures before and during pregnancy; assessment of maternal environmental exposures before and during pregnancy; assessment of hospitalizations and injuries of the child; and assessment of sleep disorders in the child. Secondary research domains were considered important areas with notable research gaps; yet they were prioritized below the primary research

domains. The data collection protocol was designed to ascertain data on both primary and secondary research domains; however, more extensive data are collected to answer research questions pertaining to the primary research domains. The original primary and secondary scientific interests of SEED were retained in SEED Phase 2 and SEED Phase 3.

The data from SEED 3 will be combined with data from the first two SEED phases to enable investigators to conduct in depth analyses of ASD risk factors including assessment of potential etiologic subgroups and gene-environment interactions. Altogether over 1,400 children with ASD are expected to be included in the combined SEED 1 and SEED 2 dataset with approximately equal numbers in the two control groups. While this existing sample will allow us to address many important research questions within our primary and secondary research domains, the sample size and corresponding statistical power will not be adequate for many analyses of infrequent exposures or exposure subtypes (such as looking at type of maternal infection rather than broadly assessing maternal infection as a risk factor), ASD subtypes, and genetic associations.

The findings from analyses of SEED 1, 2, and 3 data will be published in peer-reviewed journals and presented at national scientific and public health meetings and at local community meetings at each site. CDC will also prepare summaries of key findings from these studies written in plain language so as to be accessible to the general public and we will make these available on our website. We will also prepare webinars and reports detailing SEED findings for partner organizations and stakeholders. The personally identifying information collected will be used to maintain contact with the participants throughout the course of the study and (if the participant consents), this information may be retained for future contact for a follow-up study.

***II. Justification for data collection in terms of positive needs and the negative consequences of not having the information:***

The purpose of this case-controlled study is to ascertain a case and comparison groups that represent diverse population subgroups. All SEED study sites will implement the collaborative protocol, and common data elements across all sites will be pooled for analysis. Three groups of children will be enrolled at each site: children with ASD; children with other developmental disabilities and delays (DD comparison group); and a random sample of study area children from same birth cohorts (POP group). Data will be collected once from participant families using multiple methods including: 1) maternal telephone interview with questions about maternal reproductive history and pregnancy with the index child (**Attachment 6.a-n**), 2) parent-completed questionnaires about parental and child health and child development (**Attachment 8.a-p**), 3) in-person child developmental evaluation (**Attachment 9.a-g**), 4) maternal and child anthropometry measurements (**Attachment 16.a**), and 5) biosampling from biological parents and child (**Attachments 15-17**).

The same methods for identification and recruitment of cases and controls and data collection are used in all phases of SEED implementation. However, the SEED 3 data collection protocol has been reduced in comparison to SEED 1 and 2 in order to:

- o Eliminate some of the most time-intensive data collection instruments; and
- o Reduce processing and storage costs for the SEED Biorepository.

In reducing the data collection protocol, CDC carefully considered which instruments were no longer necessary because sufficient sample size had already been achieved in earlier phases of SEED to answer the specific relevant research questions pertaining to the instrument. In addition to eliminating 2 instruments altogether between SEED 2 and SEED 3, the SEED 3 data collection protocol has been greatly streamlined for one group – the DD comparison group; the rationale for this reduction is that the data already collected in SEED 1 and 2 will be sufficient to address many important research questions needing this particular comparison group.

Specific data collection instruments that will be retained in SEED 3 and justifications for their need are included below:

- o **Social Communication Questionnaire (SCQ)**, administered to mothers of all children upon entry into the study (Attachment 5). This instrument is key in determining the data collection workflow for SEED participants upon enrollment in the study. Additionally data collected on the SCQ inform many analyses subsumed under the ASD Phenotype research domain; these data analyses are important for informing the scientific and clinical community on the range of characteristics and behaviors exhibited among children on the autism spectrum and in defining potential etiologic ASD subtypes to be used in other SEED analyses. **(Attachment 5)**
- o **Maternal Interview (MI) and accompanying Pregnancy Reference Form (PRF)**, administered to mothers of all enrolled children – The MI is a telephone-assisted interview that includes questions on maternal reproductive history, maternal health and behaviors during the index pregnancy, maternal occupational history, and family demographics. The PRF is a short questionnaire administered to the mother via phone in advance of the MI; it is designed to estimate dates pertaining to the index pregnancy and the breastfeeding periods **(Attachment 6 and 7 – both with multiple instruments – 6.a-d, 7.a-b)**. The MI and PRF instruments provide risk factor data for several primary and secondary research domains: Prenatal Infection and Immunologic Factors; Prenatal Hormonal Risk Factors; Obstetric Risk Factors; Sociodemographic Factors; Pregnancy Substance Use; and Maternal Occupational Exposures.

**Self/Parent-Administered Forms:**

- o **Maternal Medical History Form**, collected from mothers of all enrolled children **(Attachment 8.a)**. This instrument provides risk factor data for several primary research domains: Genetics (family history data on parental developmental and psychiatric conditions); Preconception and Prenatal Infection and Immunologic Factors; Preconception and Prenatal Hormonal Risk Factors; and Obstetric Risk Factors.
- o **Paternal Medical and Occupational History Form**, collected from fathers (or mothers serving as respondent) of all enrolled children **(Attachment 8.b)**. This instrument provides risk factor data for several primary and secondary research domains: Genetics (family history data on parental developmental and psychiatric conditions); Preconception Immunologic Factors; and Paternal Occupational Exposures.
- o **Child Health History Form**, collected from mothers of all enrolled children **(Attachment 8.c)**. This instrument provides risk factor data for several primary and secondary research domains: Genetics (sibling history collected in addition to index child for various health conditions); Preconception, Prenatal, and Postnatal Immunologic Factors; Sociodemographic Factors (specifically child health insurance which is not captured on the MI); Child Gastrointestinal Symptoms; Child Hospitalizations and Injuries; Child Sleep Disorders.
- o **Maternal and Child Residential History Form**, collected from mothers of all enrolled children, **(Attachment 8.d)**. This instrument provides data that can be geocoded and linked to other data sources such as environmental monitoring data and Census data to examine risk factors related to Maternal Environmental Exposures and Maternal and Child Sociodemographic Features.

- o **Child Behavior Checklist (CBCL)**, collected from mothers of children in ASD and POP groups (**Attachment 8.e**). This instrument informs many analyses subsumed under the ASD Phenotype research domain. Additionally, this instrument provides important information on the POP comparison group – specifically, information about the degree to which this group is typically developing versus experiencing symptoms consistent with developmental impairments.
- o **Child Social Responsiveness Scale**, collected from mothers of children in ASD and POP groups (**Attachment 8.f-g**). This instrument informs many analyses subsumed under the ASD Phenotype research domain. Additionally, this instrument provides important information on the POP comparison group – specifically, information about the degree to which this group is typically developing versus experiencing symptoms consistent with developmental impairments.
- o **Child Services and Treatment Questionnaire**, collected from mothers of children in ASD group (**Attachment 8.h**). This instrument informs many analyses subsumed under the ASD Phenotype research domain. Additionally, this instrument informs separate analyses of service needs and access to services among children with ASDs.

**Clinical in-person developmental evaluation (ASD and POP workflows only):**

- o **Mullen Scales of Early Learning**, collected from mothers of children in ASD and POP groups (**Attachment 9.a**). This instrument informs many analyses subsumed under the ASD Phenotype research domain. Additionally, this instrument provides important information on the POP comparison group – specifically, information about the degree to which this group is typically developing versus experiencing symptoms consistent with developmental impairments.
- o **Vineland Adaptive Behavioral Scales (VABS)**, collected from mothers of children in ASD and POP groups (**Attachment 9.b-c**). This instrument informs many analyses subsumed under the ASD Phenotype research domain. Additionally, this instrument provides important information on the POP comparison group – specifically, information about the degree to which this group is typically developing versus experiencing symptoms consistent with developmental impairments.
- o **Autism Diagnostic Observation Schedule (ADOS)**, collected from mothers of children in ASD group (**Attachment 9.d-f**). This instrument is used along with the ADI-R to determine the final study group classification for children in the ASD work flow. Additionally data collected on the ADOS inform many analyses subsumed under the ASD Phenotype research domain.
- o **Autism Diagnostic Interview-Revised (ADI-R)**, collected from mothers of children in ASD group (**Attachment 9.g**). This instrument is used along with the ADOS to determine the final study group classification for children in the ASD work flow. Additionally data collected on the ADI-R inform many analyses included within the ASD Phenotype research domain.
- o **Saliva Specimens**, collected from mothers and children in ASD and POP groups (**Attachment 16**). This biospecimen collection provides risk factor data for the Genetics research domain and also informs gene-environment interaction studies.
- o **Blood Specimens**, collected from mothers and children in ASD and POP groups (**Attachment 17**). This biospecimen collection provides risk factor data for the Genetics research domain and also informs gene-environment interaction studies and child health studies.

In addition to the data collected from the SEED study instruments, each site will obtain limited birth certificate data from all enrolled participants. These data on maternal and child demographic and pregnancy factors will be used in various analyses (both as adjustment factors and in some analyses as the primary risk factor of interest, e.g. preterm delivery). Sites are also asked to obtain to the extent possible, birth certificate data on all invited participants to enable comparison of responders and non-responders on key demographic and pregnancy factors. Some additional publicly available data might also be used in SEED analyses. For example, data collected on the Maternal and Child Residence History Form will be geocoded such that they can be linked to other data files such as various environmental monitoring databases such as those maintained by the EPA to study hazardous air pollutants as potential risk factor for ASD.

Overall the various types of information collected in SEED collection are necessary to fulfill the study objectives. Collecting medical and health information will allow investigators to identify risk factors that might be associated with ASD. Without this information SEED will not be able to answer the scientific questions proposed as part of the purpose for this study and consequently, CDC will not be able to develop recommendations about primary prevention of ASDs or secondary prevention of associated sequelae.

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

NCBDDD will fund a Data Coordinating Center (DCC) and a Central Biosample Repository (Central Lab) for SEED. The DCC, at Michigan State University has developed an electronic data collection system to centrally store (100%) of the data. The Central Biosample Repository, is where all biosamples from the study will be shipped, processed, and stored. The DCC and Central Biosample Repository will work on an ongoing basis with the SEED investigators to implement the study.

SEED will apply information technology broadly to collect data efficiently, to assure both the quality of the collected data and the privacy and security of the collected data, and to minimize the burden to the study participants. The DCC will be responsible for the information technology aspects of the study. The DCC has previously created and hosted a custom web-based information system, called the CADDRE Information System (CIS), which was used for the first two phases of the study. The same system will be used in SEED 3. CIS was 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 the CIS will be used by the study personnel only, and not by the participants themselves. Relevant services provided by CIS include:

- Generation of customized task lists specific to the role of each authenticated user.
- Role-based security that restricts user access privileges to the minimum required.
- Automated tracking of participant progress.
- Generation of bar code labels to identify all study documents and biologic samples.
- Computer-assisted-telephone-interviews (CATI) for MI and PRF instruments **(Attachments 6-7)**.
- Double data entry for data collected on paper forms.
- Support for data entry and coding of copyrighted clinical assessment instruments by interfacing with other approved electronic systems.
- Ongoing data quality assurance checks
- Automated tracking and quality assurance reports.
- Comprehensive audit logging functions
- User support services.

Upon completion of data collection activities, the DCC works with the CDC to organize the preparation of pooled analytic data files from the data entered into CIS. This process includes quality control checks of the data collected in CIS before its exported into analytic files; quality control of the export process itself; and de-identification of most analytic files (other than limited data needed for genetic analyses or data linkage based on geocodes) by applying a date shifting algorithm previously developed and tested and removing open string text field data elements that

are inherently or likely identifiable. Upon completion of the export and QC process, DCC uploads the analytic data files and accompanying data dictionaries and other documentation onto a remote access server (RDA) for access by the site investigators. The RDA meets CDC security requirements and is located in a HIPAA-compliant data-center with full redundant power and security measures. DCC also delivers a complete set of the analytic data files to the CDC on an encrypted hard drive. The CDC stores these data in a secure location on SQL servers with limited access to the databases and identifiers are encrypted.

### Participant Burden Reduction

The CIS will facilitate computer-assisted-telephone-interviews (CATIs) for the *Maternal Interview (MI)* and *Pregnancy Reference Form (Attachments 6-7)*. The required logical branching is automatically provided by the CIS guidance to the interviewer during the interview. This implementation improves data quality and reduces errors to preclude the burden of follow-up calls to participants. During the MI and PRF interview and during all other calls with the participants, study staff also employ the CIS system to aid the interviewer in tracking completeness of response and thus reducing time for the participant.

Additionally, the CIS proactively tracks all aspects of participant's needs, requests, scheduled activities, and study protocol requirements. All contacts with each participant are tracked to ensure the efficient execution of the study. Staff are alerted automatically at login to all pending actions/tasks. Special care is given to preparation for the clinical visits. 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, and any pending paper forms that still need to be completed.

Electronic data collection systems are used as much as possible in SEED both to minimize participant burden and reduce data entry errors. Nonetheless, for several data collection instruments participants are provided with paper forms and are given the option of completing the forms on their own or of having a staff member complete the forms for them over the phone. These forms include medical history checklist type forms (**Attachments 8.a-c**) similar to those individuals are asked to complete at their doctor's offices and standardized checklist child development forms that cannot be reformatted into a CATI because of copyright constraints. In SEED 1 and 2, we found that many participants preferred to complete these forms on their own rather than with a staff member. To ensure high data quality, SEED study staff members carefully review all forms upon receipt and follow-up with participants about inconsistent or ambiguous information; additionally, all paper forms are entered into CIS using a duplicate data entry function to minimize data entry errors.

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

No data collection activities currently supported by the Department of Health and Human Services (DHHS), other government institutions, or other private agencies, are comparable to the SEED 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 2000 children in each of the three study groups – ASD, DD, and POP – at the end of SEED 3. Not only does the large SEED sample size increase study power and statistical precision overall, but it also enhances the capability for stratified analyses of phenotypic subtypes within the ASD case group as well as stratification on other factors across all subject groups.

Another recent autism epidemiologic project is the California Childhood Autism Risks from Genetics and the Environment (CHARGE) study. The CHARGE Study was 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 utilized data collection methods similar to SEED, there are multiple differences between CHARGE and SEED. The CHARGE sample is only 25% that of SEED. Moreover, CHARGE is collecting data only in the state of California, and therefore is less generalizable to a national population; CHARGE relied on a single source (Department of Developmental Services) for case ascertainment while SEED uses multiple source case ascertainment to achieve both a large and demographically diverse

sample of study participants; the CHARGE Study case and developmental delay comparison groups are more narrowly defined than those for SEED; thus the SEED ASD case group includes a more complete representation of children across the autism spectrum (which allows for more complete phenotype analyses and more accurate ASD subtypes for etiologic analyses); the research goals and corresponding data collection batteries differ somewhat between the SEED and CHARGE studies (while CHARGE collects more data on some environmental exposures than SEED, SEED collects much more detailed information on child health including GI function and sleep features; and child behavioral phenotype).

A literature review conducted for SEED protocol development identified 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 as detailed exposure data collection methods as SEED. This comprehensive literature review helped detect gaps in our current understanding of ASD which, in turn, led to identification of high priority research domains.

The SEED sample size and unique data being collected in SEED allow for study of research domains not covered or not covered as fully by CHARGE or other autism studies. Additionally, while there is some overlap in data collection between SEED and other studies, this will permit replication of many smaller analyses published by other studies. 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. Autism is a complex neurological disorder that is difficult to diagnose because there is no clear biologic marker; early studies suggest a multi-factorial

etiology with the likelihood of both gene-gene and gene-environment interactions. Thus, duplication of findings in multiple independent studies is a key component of assessing the causality of identified risk factors.

Finally, the Director of the Division of Birth Defects and Developmental Disabilities at CDC 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 DHHS 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.

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

This data collection will not involve small businesses.

#### **A.6. Consequences of Collecting the Information Less Frequently**

The information collected from each SEED participant will only be collected once and has not been collected previously. The SEED case-control study is the first and largest multi-site, population based study on ASD planned and implemented to date, and the findings from this study will be essential to advancing the understanding of the causes of autism and ASDs. Since this data collection is in response to a mandate for research into the causes of ASD in the Children's Health Act of 2000 and the Combating Autism Act of 2006, the consequence of not collecting the information would be to severely limit information on autism causes from studies of US children. If these data were not collected researchers' ability to provide timely and important information related to the risk factors and causes of ASD and the characteristics and health of children with ASD would be greatly impacted.

There are no legal obstacles to reduce the burden.

#### **A.7. Special Circumstances Relating to the Guidelines of 5 CFR 1320.5**

This request fully complies with the guidelines of 5 CFR 1320.5.

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

A 60-day Federal Register Notice was published in the Federal Registry on February 24, 2016, Vol. 81, No. 36, pp 9200-9201 (**Attachment 2**). CDC received 3 non-substantive comment(s) (**Attachment 24**) and replied with a standard CDC response.

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 1 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 same sites were funded in SEED 2 and the PIs from these sites worked together throughout study implementation to discuss and resolve any issues related to implementing the study protocol, maintaining data security, and analyzing the data. Please see **Attachment 13** for a list of SEED 2 PIs.

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 protocol was revised based on the panel's feedback.

After protocol approval, CDC assembled periodic peer review panels to assess the portfolio of research conducted in the Developmental Disabilities Branch, including the SEED project. The most recent of these was in January 2015. Please see **Attachment 14** for the list of the external review panel. Although the purpose of that review was not specifically to suggest revisions to the SEED protocol (since the protocol necessarily needs to be consistent across phases to enable data harmonization), the CDC presented the panel members with a comprehensive overview of the SEED program and progress to date and asked panel members to consider SEED in the context of some specific issues related to SEED data collection, such as maximizing the use of the biospecimens already collected) and to consider any research gaps not covered by the current SEED protocol. Based on their review, we are continuing another round of data collection (SEED 3) to ensure we have an appropriate sample size to answer the critical research questions that SEED was designed to address.

In addition to the more formal peer review panels that have been assembled, we have periodically sought consultation from various individuals at academic institutions with particular areas of expertise (such as sampling, statistical analyses, genetics, etc.) needed to address various challenges we faced throughout the course of SEED implementation.

Finally, the Associate Director for Science for the National Center on Birth Defects and Developmental Disabilities at CDC, Stuart Shapira, 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. Dr. Shapira provides updates on SEED progress at all annual IACC meetings.

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

All SEED families have young children and two-thirds of SEED families 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 families in all 3 study groups without providing incentives (Dunn and Gordon, 2005). Thus, we propose providing incentives to participants for completing each study step in SEED 3 to ensure a more representative study sample. These incentives are in keeping with those provided to participants in SEED 1 and SEED 2, although the incentive for the DD group has been reduced commensurate with the reduction of the data collection protocol for this group. Description and specific justification of incentives in SEED 3 are as follows:

1. Incentives of \$30-40 at multiple times during the study period for data collection via phone interview or completion of self-administered forms.

Given that SEED includes an intensive data collection protocol with multiple data collection components occurring at varying points in time over a period of several months, it is important to include strategies to retain participants who are successfully enrolled. It can be difficult for SEED families to find the time to participate.

All of the families enrolled in SEED have at least 1 pre-school aged child; two-thirds of SEED families have at least 1 pre-schooled aged child with a developmental disability; many families have multiple young children and multiple children with developmental disabilities. Thus, even for phone interviews and completion of self-administered forms, finding sufficient periods of uninterrupted time can be challenging for mothers. This problem is often exacerbated for many SEED families who are of lower SES – groups SEED was specifically designed to include.

In considering these incentives it is also very important to bear in mind that while these incentives are linked to certain data collection components, those particular components do not account for the total time

we ask a participant to spend on SEED study activities. For example, we have several follow-up calls with subjects to discuss study steps, schedule visits, etc. We send several packets of materials to participants that we ask them to review (for example a packet of materials to help them prepare their child for the lengthy in person visit that takes place at the end of the data collection protocol). We also offer participants help in completing self-administered forms via phone and offer to go over study materials we send them. This adds time to the protocol for some participants, but we find it is an important option that many families take. Given the diversity of our study population, we need to be mindful that some participants have low literacy level and thus need more support from SEED staff. These participants might also have the largest time constraints.

2. Incentives of \$75 POP group and \$200 ASD group for in person assessment.

This is a particularly intensive and burdensome visit in terms of both participant time and acceptability data collection components.

***Time:*** Actual visit time is nearly 2 hours for the POP group and 5 and a half hours for the ASD group, In addition to visit time, these visits most often take place in a clinic which can be many miles from the participant's home. At some sites, some participants must travel an hour or more one way to reach the clinic location.

As discussed above, for most SEED families, time to participate in research studies can be very limited. Thus, incentives allow participants to purchase meals, child care, or even take time off work without pay. The incentive for this data collection component must also cover travel expenses,

***Acceptability of Data Collection Components:*** This visit includes simultaneous data collection activities for both mothers and their children. Mothers undergo intensive developmental interviews while children undergo a developmental assessment. While all SEED study staff are professionally trained with past experience conducting assessments with children, some children nonetheless have difficulty working with a stranger. Some mothers will bring along another trusted caregiver to help the child get comfortable with the visit. In other cases, the mother must help the child get started and thus start her own data collection components later, which adds to the total time burden.

During these visits we also ask mothers to allow us to collect biosamples, including blood, from themselves and their children. This can be a particularly difficult component for the child. Yet, blood specimens are important as the genetic yield is usually greater and the genetic material obtained is of higher quality than that for saliva specimens. Thus, while families can refuse any data collection component and we do not put undue pressure on them to participate in blood collections, we do include this as a possible data collection component we ask them to consider.

Given the issues with both time burden and acceptability it is important to include a generous incentive to motivate families to complete the SEED study data collection protocol and to offset the time demands of the study.

The following are the specific incentive structures according to data collection workflow for all enrolled participants.

**Incentives for ASD workflow mother-child pairs (NOTE: only enrolled mother-child pairs receive incentives)**

<b>GROUP: ASD</b>			
	<b>Data Collection Step*</b>	<b>Time to Complete</b>	<b>Time Incentives</b>
1	Invitation Letter		
2	Screening and Invitation Phone Call (Includes eligibility screen, description of the study, consent [ <b>Appendix 21.a</b> ], and administration of the SCQ [ <b>Appendix 5</b> ])	30 min.	
3.	Enrollment Packet (includes incentive for the eligibility screening)		\$30
4.	Follow Up Phone Call (includes administration of Pregnancy Reference Form ( <b>Appendix 7.a, 7.b</b> ) and discussion of maternal interview) ( <b>Appendix 21.d</b> )	15 min	
5.	Maternal Interview Call (includes administration of interview and discussion of next steps in study) ( <b>Attachment 6.b, 6.d</b> )	60 min	\$30
6.	Mailing of Self/Parent Administered Forms Packet (includes 7 forms to complete on parental and child health and child development, and materials about clinic visit preparation)	95 min	\$40
7.	Follow Up Phone Call 2 (includes answering questions, help with self-administered forms as needed, and preparation for clinic visit) ( <b>Appendix 21.e</b> )	20 min	
8.	Clinic/Home Visit (includes in depth developmental assessments, anthropometry, collection of biosamples, signing consents, and completing any remaining forms) ( <b>Appendix 11.a</b> )	330 min	\$200
	<b>TOTAL</b>	<b>9 hours, 10 minutes</b>	<b>\$300</b>

Note: Steps 1 through 7 involve contact with and data collection from the biological mother. During the clinic/home visit, data are collected from both the mother and the index child.

**Incentives for POP workflow mother-child pairs (NOTE: only enrolled mother-child pairs receive incentives)**

<b>GROUP: POP</b>			
	<b>Data Collection Step*</b>	<b>Time to Complete</b>	<b>Time Incentives</b>
1	Invitation Letter		
2	Screening and Invitation Phone Call (Includes eligibility screen, description of the study, consent [ <b>Appendix 21.c</b> ], and administration of the SCQ [ <b>Appendix 5</b> ])	30 min.	
3.	Enrollment Packet (includes incentive for the eligibility screening)		\$30
4.	Follow Up Phone Call (includes administration of Pregnancy Reference Form ( <b>Appendix 7.a, 7.b</b> ) and discussion of maternal interview) ( <b>Appendix 21.d</b> )	15 min	
5.	Maternal Interview Call (includes administration of interview and discussion of next steps in study) ( <b>Attachment 6.b, 6.d</b> )	60 min	\$30
6.	Mailing of Self/Parent Administered Forms Packet (includes 6 forms to complete on parental and child health and child development, and materials about clinic visit preparation)	95 min	\$40
7.	Follow Up Phone Call 2 (includes answering questions, help with self-administered forms as needed, and preparation for clinic visit) ( <b>Appendix 21.e</b> )	20 min	
8.	Clinic/Home Visit (includes developmental assessments, anthropometry, collection of biosamples, signing consents and completing any remaining forms) ( <b>Appendix 11.c</b> )	110 min	\$75
	<b>TOTAL</b>	<b>5 hours, 30 minutes</b>	<b>\$175</b>

Note: Steps 1 through 7 involve contact with and data collection from the biological mother. During the clinic/home visit, data are collected from both the mother and the index child.

**Incentives for DD workflow mother-child pairs (NOTE: only enrolled mother-child pairs receive incentives)**

<b>GROUP: DD</b>			
	<b>Data Collection Step*</b>	<b>Time to Complete</b>	<b>Time Incentives</b>
1	Invitation Letter		
2	Screening and Invitation Phone Call (Includes eligibility screen, description of the study, consent [ <b>Appendix 11.b, 21.b</b> ], and administration of the SCQ [ <b>Appendix 5</b> ])	30 min.	
3.	Enrollment Packet (includes incentive for the eligibility screening)		\$30
4.	Follow Up Phone Call (includes administration of Pregnancy Reference Form ( <b>Appendix 7.a, 7.b</b> ) and discussion of maternal interview) ( <b>Appendix 21.d</b> )	15 min	
5.	Maternal Interview Call (includes administration of interview and discussion of next step in study) ( <b>Attachment 6.b, 6.d</b> )	60 min	\$30
6.	Mailing of Self/Parent Administered Forms Packet (includes 4 forms to complete)	60 min	\$40
	<b>TOTAL</b>	<b>2 hours, 45 minutes</b>	<b>\$100</b>

**A.10. Protection of the Privacy and Confidentiality of Information Provided by**

**Respondents**

This submission has been reviewed by the NCBDDD Privacy Officer, who has determined that the Privacy Act does apply. The Privacy Act is applicable to data collection activities at the Georgia SEED site (involving a few contractors - currently, Research Triangle Institute (RTI) International, Acentia, Carter Consulting, Inc., and McNeal Professional Services). 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 encrypted. The data collected is jointly owned by the CDC and the participating research sites.

Due to the sensitive nature of certain data collection components, SEED will obtain additional confidentiality protections. As in SEED 1 and SEED 2, an application for a 301(d) Certificate of Confidentiality for protection of the individual participants at all six sites conducting the study will be submitted and once approved will be provided.

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

The role-based security system of the CIS ensures that only those personnel who have prior authorization to access personal identifiers have the capability of associating an identification number with a participant’s identity. The entire CIS database will be encrypted. The encryption implementation method will be approved by the CDC. Encrypting the database will preclude access to personally identifying information in the event that an intruder penetrates the server security. No information concerning the operation of the encryption system will be stored on the CIS servers. Proper firewall and network security configurations will be maintained. All applicable security updates of comprising commercial software components will

be applied to the servers rapidly after release by the software vendors. The CIS complies with all current (and future) guidelines required by the CDC for such web-based applications. All data in the CIS data repository will be maintained on dedicated secured servers. CIS production servers will be housed in physically secure professional server rooms. The CIS incorporates all appropriate system security practices and methodologies for the CIS applications and database servers to protect the data and maintain the confidentiality of the participant information.

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

All analytic data files will be shared with SEED investigators through a Remote Data Access (RDA) platform maintained by the DCC with oversight from CDC. Specific personally identifiable information (PII) (such as names, address, phone numbers, etc.) will not be included

in the data files on the RDA, with one very limited exception (see below). The DCC will work with CDC to ensure the data available through the RDA are de-identified to the extent possible—e.g. by using date shifting procedures and removing text variables pertaining to specific PII (such as name) and variables likely to be identifiable by reference to other data that might be available (such as child’s hospital of delivery) from all analytic files. However, the analytic data files will not (and cannot) be considered completely de-identified for several reasons:

- Some genetic data derived from analyses of biosamples will be compiled into analytic files and will be shared with SEED investigators through a second RDA server maintained by the DCC. Because genetic data are unique to individuals, they are not considered de-identified. Nonetheless, CDC wants to ensure that these important data are available to the SEED investigators to maximize their use in research analyses. The RDA approach offers a secure platform for these analyses. The second server will include all security provisions as the original RDA. Additionally, access to the genetic data files will be more restrictive. Only investigators with a clearly identified need to use the genetic data in their analyses will be provided access.
- The address data collected on the Maternal and Child Residence History form will not be included directly on the RDA. However, each site will be asked to allow DCC to derive various geocode variables from the data collected on this form and/or to themselves derive various geocode variables. If sites derive geocoded variables themselves, these data will be sent to DCC following an encryption protocol and the DCC will compile the information. The DCC will provide access to the geocode data on a designated restricted folder on one of the DCC RDA servers to individual SEED investigators who require access for data linkage purposes. The investigator will have to demonstrate a specific need for these sensitive data and that proposed analysis must be approved by the CDC and the CADDRE/SEED Data Sharing Committee (comprised of site principal investigators). DCC will provide such investigators time-limited access to the geocoded data such that they can link the data with other data files such as environmental monitoring data or Census data. No data will be downloaded from the RDA. Once the linkage occurs, the geocodes will be stripped from the linked data file. This de-identified linked data file will contain participant study ID. DCC will move this de-identified file to the main section of the RDA whereby all SEED investigators needing such data will have access and can combine these data with other SEED data.
- Beyond genetic and geocode data, the composite SEED datafiles will contain up to 20,000 variables per participant for participant families. Although the major issues related to identifiability will be addressed, with this volume of data it is not possible to ensure that combination of data about a research participant will not result indirectly, by reference to other information to identification of study participants.

Scientists, colleagues, and collaborators who are given access to clinical, interview and biologic data from SEED must sign a confidentiality and data use oath that describes how the

data should be used and stored. The only approved mechanism for SEED data access is through the RDA maintained by the DCC. All analyses are conducted on the RDA and aggregate results are downloaded but individual-level data are not. DCC actively monitors the RDA and ensures that data are not downloaded.

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

Specific consent language was included in the written consent forms to permit, if signed by the participant, the participant's genetic data to be placed in the National Database for Autism Research (NDAR) and the Database for Genotypes and Phenotypes (dbGaP) (**Attachment 11.a, 11.c**). Both databases are run by the National Institutes of Health (NIH) that allows researchers studying autism to easily share and pool information with each other. To protect participant privacy, all identifying information such as name and address will be removed and replaced with an NDAR specific code number. The consent form language was developed in accordance with the evolving understanding of the identifiability of genetic data such that parents are fully informed of risks and benefits when providing consent for genetic data sharing. The latest guidance from the National Human Genome Research Institute at NIH states:

- “Sharing individual and even summary-level genomic data carries some degree of privacy risk to study participants. When data will be shared, researchers should explain how privacy and confidentiality will be protected.

- A primary privacy concern in genomics research is re-identification. Researchers have shown that subjects can be re-identified by combining de-identified genomic information with other information types that are publicly available, and that individual subjects sometimes can be distinguished even in summary-level genomic data. Current federal regulations do not classify genomic information as "identifiable," but a proposal to revise the Common Rule released in 2011 discussed the identifiability of genomic data.
- Data repositories may protect against the possibility of re-identification by controlling access to the data and requiring data users to agree not to attempt to re-identify research participants. However, it is not possible to eliminate completely the risk of re-identification. Researchers should explain this risk to participants' privacy and confidentiality and note whether there may be related unanticipated risks in the future."
- <http://www.genome.gov/27559024>

### **A.11 Institutional Review Board Approval and Justification for Sensitive Questions**

IRB approval was granted on 12/7/2015 and will expire on 12/7/2016. The current IRB approval letter is included as **Attachment 10**.

The consent forms for the parents (see *Written Informed Consent Document*, **Attachment 11.a-c**) 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* (**Attachment 12**) 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 started, no such publications have yet occurred).

### **Justification for Sensitive Questions**

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

Questions of particular sensitivity can be found in the following:

1. Maternal Interview (**Attachment 6.a-d**)
2. Self/Parent Administered Forms (**Attachment 8**)
3. Child developmental evaluation (**Attachment 9**)

We have included these items despite their potential sensitivity because research suggests that they are 1) potential risk factors for ASDs and the associations need further clarification; 2) important health conditions potentially related to ASDs in need of further study; 3) behaviors and symptoms important in further characterizing the spectrum of autism.

Specifically, some of 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. Prior to beginning the *Maternal Interview (Attachment 6)* 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.’ Participants in ASD and POP workflows sign a *written informed consent (Attachment 11)* at the initial clinic visit. It informs participants that: ‘You can refuse any task and still participate in the study.’

## **A.12. Estimates of Annualized Burden Hours and Costs**

Although children are initially identified as potentially being eligible for a given group – ASD, DD, or POP – the final study group classification is determined from standardized research developmental assessments. Upon enrollment, all children are screened for possible autism characteristics through their mother’s completion of the Social Communication Questionnaire (SCQ). Children with SCQ scores above a predetermined threshold (>11) are designated as potential ASD cases regardless of how they were initially identified. Additionally, all children who had a previous ASD diagnosis or autism special education classification are designated as potential ASD cases regardless of their SCQ scores. Potential ASD or DD or POP participants are designated into specific protocols, and these protocols are called workflows.

Children in the potential ASD workflow will have a more comprehensive assessment than children in the other groups. The clinic visit includes a developmental assessment, anthropometric measurements, and biologic sample collection. The child and mother parts of the clinic visit can happen simultaneously, and fathers are mailed the saliva kits. In addition to a general developmental assessment, children are administered the Autism Diagnostic Observation Schedule (ADOS) and their mothers are administered the Autism Diagnostic Interview revised (ADI-R). Final ASD case classification are based on the ADOS and ADI-r scores.

Those in the potential POP workflow will have a general developmental assessment resulting in a shorter clinic visit, fewer paper forms to fill out, and less burden than the potential ASD workflow group. Participants in the potential DD workflow will not have a clinic visit and will have even fewer paper forms to fill out than the ASD or POP workflow groups.

To estimate annualized burden hours, we have made the following assumptions based on preliminary SEED2 data:

**Mother-child pairs sampled from birth records for potential POP workflow group**

- Of *potentially* eligible participants sent invitation mailings, study staff will have contact with 50%.
- Of those with contact, 24% will be ineligible.
- Of those with contact who are eligible, 60% will enroll.
- Of those eligible, consented, and enrolled, 90% of mothers will complete the first follow-up phone call and pregnancy reference form and the maternal interview.
- Of those eligible, consented, and enrolled, 80% of mothers will complete the self-administered forms and second follow-up phone call to review the forms.
- Of those eligible, consented, and enrolled, 70% will complete the clinic visit.
- Of those who complete the clinic visit, 50% of fathers will provide saliva specimens.

**Mother-child pairs identified from health and school sources for potential ASD or DD**

**workflow groups**

- Of *potentially* eligible participants sent invitation mailings, study staff will have contact with 50%.
- Of those with contact, 22% will be ineligible.
- Of those with contact who are eligible, 70% will enroll.
- Of those who do not have a previous ASD diagnosis and are initially identified as potential DD workflow, 40% will screen positive on the ASD screen and will be evaluated in the ASD workflow instead of the DD workflow.
- Of those eligible, consented, and enrolled, 90% of mothers will complete the first follow-up phone call and pregnancy reference form and the maternal interview.

- Of those eligible, consented, and enrolled, 80% of mothers will complete the self-administered forms and second follow-up phone call to review the forms.
- Of those eligible, consented, and enrolled, 70% will complete the clinic visit (ASD workflow only)
- Of those who complete the clinic visit, 50% of fathers will provide saliva specimens (ASD workflow only).

Given the complexity of a study with participants in 3 different workflow groups the table below presents the estimates by pre-enrollment, enrollment, and post-enrollment of each workflow group (ASD, POP, and DD) to give the most accurate estimate of participant burden. The total participant burden for this study is estimated at 21,354 hours. The estimated annual burden is 7,118 hours.

#### A.12.A Estimated Annualized Burden Hours

Type of Respondents	Form Name	No. of Respondents	No. Responses per Respondent	Average Burden per Response (in hours)	Total Burden Hours
<b>ASD Workflow</b>					
<b>PRE-ENROLLMENT</b>					
<b>Mother</b> <i>All potential participants sent mailing</i>	<b>Invitation Packet/Response Card</b> (Attachments 19.a, 19.d, and 19.g)	1,718	1	10/60	286
<b>ENROLLMENT</b>					
<b>Mother</b> <i>Potentially eligible with contact by study staff</i>	<b>Invitation Call Script</b> (Attachment 21.a) <b>and SCQ</b> (Attachment 5)	859	1	30/60	430
<b>Mother</b> <i>Eligible, consented, and enrolled; assigned to</i>	<b>Enrollment Packet</b> (Attachments 20.a, 18, and 11.a) – Reference materials only; could be read	469	1	20/60	156

<i>the ASD workflow based on enrollment intake</i>	on own or reviewed during subsequent calls.				
<b>POST-ENROLLMENT</b>					
<b>Mother Completed this study step</b>	<b>Follow-up Phone Call Script (Attachment 21.d) and Pregnancy Reference Form (Attachments 7.a and 7.b)</b>	422	1	15/60	106
<b>Mother Completed this study step</b>	<b>Maternal Interview Call (Attachments 6.a and 6.d)</b>	422	1	1	422
<b>Mother Completed this study step</b>	<b>Self-Administered Forms (Attachments 8.a-p, excluding 8.j, 8.m. and 8.n.)</b>	375	1	105/60	656
<b>Mother Completed this study step</b>	<b>Follow-up Call 2</b> answer questions, provide help with self-administered forms as needed, and prepare for clinic visit (Attachment 21.e)	375	1	20/60	125
<b>Mother Completed this study step</b>	<b>Clinic / Home Visit – Developmental Assessment (Attachments 11.a; 9.b; 9.c; 9.g; 16.a-d; 17.b)</b>	328	1	225/60	1,230
<b>Father Completed this study step</b>	<b>Clinic / Home Visit – Saliva Collection (optional - on own) (Attachments 16.b-d)</b>	164	1	15/60	41
<b>Child Completed this study step</b>	<b>Clinic / Home Visit – Developmental Assessment (Attachments 9.a; 9.d or 9.e or 9.f; 16.a-d;</b>	328	1	135/60	738

	17.a)				
<b>ASD Workflow SUBTOTAL</b>					<b>4,190</b>
<b>POP Workflow</b>					
<b>PRE-ENROLLMENT</b>					
<b>Mother</b> <i>All potential participants sent mailing</i>	<b>Invitation Packet/Response Card</b> (Attachments 19.c, 19.f, and 19.g)	1,466	1	10/60	244
<b>ENROLLMENT</b>					
<b>Mother</b> <i>Potentially eligible with contact by study staff</i>	<b>Invitation Call Script</b> (Attachment 21.c) <b>and SCQ</b> (Attachment 5)	733	1	30/60	367
<b>Mother</b> <i>Eligible, consented, and enrolled; assigned to the POP workflow based on enrollment intake</i>	<b>Enrollment Packet</b> (Attachments 20.a, 20.c, 18, and 11.c) Reference materials only; could be read on own or reviewed during subsequent calls.	334	1	20/60	111
<b>POST-ENROLLMENT</b>					
<b>Mother</b> <i>Completed this study step</i>	<b>Follow-up Phone Call Script</b> (Attachment 21.d) <b>and Pregnancy Reference Form</b> (Attachments 7.a and 7.b)	301	1	15/60	75
<b>Mother</b> <i>Completed this study step</i>	<b>Maternal Interview Call</b> (Attachments 6.a and 6.d)	301	1	1	301
<b>Mother</b> <i>Completed this study step</i>	<b>Self-Administered Forms</b> (Attachments 8.a-p, excluding 8.j, 8.l, and 8.m.)	267	1	105/60	467
<b>Mother</b> <i>Completed</i>	<b>Follow-up Call 2</b> answer questions,	267	1	20/60	89

<i>this study step</i>	provide help with self-administered forms as needed, and prepare for clinic visit (Attachment 21.e)				
<b>Mother</b> <i>Completed this study step</i>	<b>Clinic / Home Visit – Developmental Assessment</b> (Attachments 11.c; 16.a-d; 17.b)	234	1	50/60	195
<b>Father</b> <i>Completed this study step</i>	<b>Clinic / Home Visit – Saliva Collection (optional - on own)</b> (Attachments 16.b-d)	117	1	15/60	29
<b>Child</b> <i>Completed this study step</i>	<b>Clinic / Home Visit – Developmental Assessment</b> (Attachments 9.a; 16.a-d; 17.a)	234	1	90/60	351
<b>POP Workflow SUBTOTAL</b>					<b>2,229</b>
<b>DD Workflow</b>					
<b>PRE-ENROLLMENT</b>					
<b>Mother</b> <i>All potential participants sent mailing</i>	<b>Invitation Packet/Response Card</b> (Attachments 19.b, 19.e, and 19.g)	641	1	10/60	107
<b>ENROLLMENT</b>					
<b>Mother</b> <i>Potentially eligible with contact by study staff</i>	<b>Invitation Call Script</b> (Attachment 21.b) <b>and SCQ</b> (Attachment 5)	321	1	30/60	161
<b>Mother</b> <i>Eligible, consented, and enrolled; assigned to the DD workflow</i>	<b>Enrollment Packet</b> (Attachment 20.b, 18, and 11.b) Reference materials only; could be read on own or reviewed during subsequent	175	1	20/60	58

<i>based on enrollment intake</i>	calls.				
<b>POST-ENROLLMENT</b>					
<b>Mother Completed this study step</b>	<b>Follow-up Phone Call Script</b> (Attachment 21.d) <b>and Pregnancy Reference Form</b> (Attachments 7.a and 7.b)	158	1	15/60	40
<b>Mother Completed this study step</b>	<b>Maternal Interview Call</b> (Attachments 6.a and 6.d)	158	1	1	158
<b>Mother Completed this study step</b>	<b>Self-Administered Forms</b> (Attachments 8.a-d, 8.j, 8.k, 8.m, 8.o, 8.p)	140	1	55/60	128
<b>Mother Completed this study step</b>	<b>Follow-up Call 2</b> answer questions, provide help with self-administered forms as needed (Attachment 21.e)	140	1	20/60	47
<b>DD Workflow SUBTOTAL</b>					<b>699</b>
<b>GRAND TOTAL</b>					<b>7.118</b>

**A.12.B. Estimated Annualized Burden Costs (beginning at ENROLLMENT)**

<b>Type of Respondents</b>	<b>Form Name</b>	<b>Total Burden Hours</b>	<b>Average Hourly Wage Rate (\$)</b>	<b>Total Respondent Costs (\$)</b>
<b>ASD Workflow</b>				
<b>Mother</b> N=859	<b>Invitation Call Script and SCQ</b> (30 minutes)	430	\$20.00	\$8,600
<b>Mother</b> N=422	<b>Enrollment Packet, Follow-up Phone Call Script</b>	262	\$34.29	\$8,984

	<b>and Pregnancy Reference Form</b> (Total: 35 minutes)			
<b>Mother</b> N=422	<b>Maternal Interview Call</b> (1 hour)	422	\$45.00	\$18,990
<b>Mother</b> N=375	<b>Self- Administered Forms</b> (1.6 hours) and <b>Follow-up call 2</b> answer questions, provide help with self-administered <b>forms as needed</b> , and prepare for clinic visit	781	\$26.09	\$20,376
<b>Mother and Child</b> N=328	<b>Clinic / Home Visit – Developmental Assessment</b> (5.7 hours)	1,968	\$ 35.09	\$69,057
<b>ASD SUBTOTAL</b>				<b>\$126,007</b>
<b>POP Workflow</b>				
<b>Mother</b> N=733	<b>Invitation Call Script and SCQ</b> (30 minutes)	367	\$20.00	\$7,340
<b>Mother</b> N=334 (Enrollment) + N=301 (Call Script and PRF)	<b>Enrollment Packet, Follow-up Phone Call Script and Pregnancy Reference Form</b> (Total: 35 minutes)	186	\$34.29	\$6,378
<b>Mother</b> N=301	<b>Maternal Interview Call</b> (1 hour)	301	\$45.00	\$13,545
<b>Mother</b> N=267	<b>Self- Administered Forms</b> (1.6 hours) and <b>Follow-up call 2</b> (20 minutes) answer questions, provide help with self-	556	\$26.09	\$14,506

	administered <b>forms as needed</b> , and prepare for clinic visit			
<b>Mother and Child</b> N=234	<b>Clinic / Home Visit – Developmental Assessment</b> (2.1 hours)	499	\$35.16	\$17,545
<b>POP SUBTOTAL</b>				<b>\$59,314</b>
<b>DD Workflow</b>				
<b>Mother</b> N=321	<b>Invitation Call Script and SCQ</b> (30 minutes)	161	\$20	\$3,220
<b>Mother</b> N=175 (Enrollment) + N=158 (Call Script and PRF)	<b>Enrollment Packet, Follow-up Phone Call Script and Pregnancy Reference Form</b> (Total: 35 minutes)	98	\$34.29	\$3,360
<b>Mother</b> N=158	<b>Maternal Interview Call</b> (1 hour)	158	\$45	\$7,110
<b>Mother</b> N=140	<b>Self- Administered Forms</b> (55 minutes) and <b>Follow-up call 2</b> (20 minutes) answer questions, provide help with self-administered <b>forms as needed</b>	175	\$40	\$7,000
<b>DD SUBTOTAL</b>				<b>\$20,690</b>
<b>GRAND TOTAL</b>				<b>\$206,011</b>

### **A.13. Estimates of Other Total Annual Cost Burden to Respondents or Record Keepers**

There are no costs to respondents associated with either capital and startup efforts or operation and maintenance of services for this project.

### **A.14. Annualized Cost to the Government**

The average annualized cost to the Government to collect this information is:

<b>Federal Government Personnel costs</b>		
	CDC Site Principal Investigator	\$90,000
	CDC Project Officer	\$60,000
	CDC Site Co-Principal Investigator	\$80,000
	CDC Site Co-Principal Investigator	\$80,000
	CDC Site Co-Principal Investigator	\$80,000
	CDC Health Scientist	\$75,000
	CDC Public Health Analyst	\$65,000
	CDC Collaborator	\$50,000
	CDC Collaborator	\$50,000
	<b>Subtotal, For Government Costs</b>	<b>\$630,000</b>
<b>Contractor and Grantee Costs</b>	GA SEED (CDC)	800,000
	Awardee #1	800,000
	Awardee #2	800,000
	Awardee #3	800,000
	Awardee #4	800,000
	Awardee #5	800,000
	Awardee #6 Biorepository	300,000
	Awardee #7 DCC	1,200,000
		<b>Total (\$) 6,130,000</b>

### **A.15. Explanation for Program Changes or Adjustments**

This is a new data collection.

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

Data collection will commence 1 month after OMB approval is obtained and is expected to 3 years. Data cleaning and analytic preparation and QC of all SEED analytic data files

(including harmonization of SEED 3 data with data from SEED 1 and 2) will take up to 1 year following data collection. Data analysis will begin as soon as the analytic files are finalized.

In most risk factor and child health analyses, children with a final classification of ASD will be compared to children in the two control groups (DD and POP). Odds ratios with 95% confidence intervals will be calculated for associations between ASD and various risk factors: overall; for etiologic ASD subgroups defined empirically based on analyses of the detailed behavioral and other phenotypic data; within strata defined by key demographic characteristics, such as child sex, race-ethnicity and maternal age; and after adjustment for demographic and perinatal characteristics. Interactions between genetic and non-genetic risk factors will be assessed through stratification and modelling. Child health characteristics will be similarly assessed through comparison of children with ASD to children in the DD and POP comparison groups. Separate in depth analyses of the ASD group will also be performed that examine the proportions of children with various behavioral and developmental symptoms and thus characterize ASD subgroups and the extent to which various ASD symptoms and other health characteristics cluster together.

<b>Project Time Schedule</b>	
<b>Activity</b>	<b>Time Schedule*</b>
Letters of invitation sent to potential participants	Immediately after OMB approval
Data collection begins	1 month after OMB approval
Complete data collection	3 years after OMB approval
Finalize data cleaning and entry	3.5 years after OMB approval
Prepare analytic data files and harmonize SEED 1 and 2 data files	4 years after OMB approval

Begin to analyze data	4 years after OMB approval
Prepare first manuscript	4.5 years after OMB approval
Publication of first manuscript	5 years after OMB approval

**Section A.17. Reason(s) Display of OMB Expiration Date Is Inappropriate**

The display of the OMB expiration date is appropriate, no exception is sought.

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

There are no exceptions to the certification.