Title: The National CADDRE Study: 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

June 4, 2007

Supporting Statement

for

The National CADDRE Study: Child Development and Autism

New OMB Application

Table of Contents

A. Justification....1
A.1. Circumstances Making the Collection of Information Necessary....1
A.2. Purpose and Use of the Information Collection....2
A.3. Use of Improved Information Technology and Burden Reduction....9
A.4. Efforts to Identify Duplication and Use of Similar Information....14
A.5. Impact on Small Businesses or Other Small Entities....17
A.6. Consequences of Collecting the Information Less Frequently....17
A.7. Special Circumstances Relating to the Guidelines of 5CFR1320.5....17
A.8. Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency....18
A.9. Explanation of Any Payment or Gift to Respondents....22
A.10. Assurance of Confidentiality Provided to Respondents....25
A.11. Justification for Sensitive Questions....28
A.12. Estimates of Annualized Burden Hours and Costs....31
A.13. Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers....36
A.14. Annualized Costs to the Federal Government....37
A.15. Explanation for Program Changes or Adjustments....37
A.16. Plans for Tabulation and Publication and Project Time Schedule....38
A.17. Reason(s) Display of OMB Expiration Date is Inappropriate....58
A.18. Exceptions to Certification for Paperwork Reduction Act Submissions....58
B. Statistical Methods....58
B.1. Respondent Universe and Sampling Methods....58
B2.Procedures for the Collection of Information....72
B.3.Methods to Maximize Response Rates and Deal with Nonresponse....84
B.4. Tests of Procedures or Methods to be Undertaken....87
B.5. Individuals Consulted on Statistical Aspects and Individuals Collection and/or Analyzing Data....88
References....91

Appendices

Appendix A: Authorizing legislation & other relevant laws

Appendix B: Primary Caregiver Interview
B.1 –Telephone Script
B.2 – Primary Caregiver Interview

Appendix C.1: 60 day Federal Registry Notice
Appendix C.2: Public Comment Received

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 Cover Letter
E.2 Cover Sheet
E.3 Written Informed Consent Document
E.4 Clinic Visit Prep Guide
E.5 Caregiver Interview Prep Guide
E.6 Glossary of Medical Terms
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.11 Child Behavior Checklist
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.17 Social Responsiveness Scale (Adult & Child versions)
E.18 HIPAA Provider Checklist and Medical Records Release Form
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
E.23 Self-Administered Questionnaire Checklist

Appendix F: Clinic Visit – Parent Portion
F.1 ADI-R
F.2 Early Development Questionnaire
F.3 Services and Treatments Interview
F.4 Vineland Adaptive Behavior Scales

Appendix G: Clinic Visit – Child Development Evaluation
G.1 ADOS
G.2 Mullen Scales of Early Learning

Appendix H: Self-Administered Packet II
H.1 Three-day diet diary
H.2 Seven-day stool diary

Appendix I: CDC IRB Approval Letter

Appendix J: Social Communication Questionnaire
Appendix K: CADDRE Case, Comparison Group, and Subcohort Ascertainment Methodology

Appendix L: ICD-9 Codes/Part B School Eligibility Criteria

Appendix M: Introductory Packet (Letter of Introduction, Study Brochure, Response Card, Study Posters)

Appendix N: Invitation Phone Call

Appendix O: Follow-Up Phone Calls
O.1 Follow-Up Phone Call
O.2 Reminder Phone Call Script for Clinic Visits and Caregiver Interview

Appendix P: Dysmorphology Exam: Protocol and Data Collection Form

Appendix Q: Administration of the Vineland in the Sub-cohort Telephone Script

Appendix R: Biosamples
R.1 Summary of Biosample Shipping, Processing and Storage
R.2 Blood Draw Information Form

Appendix S: Medical Record Abstraction Forms
S.1 Medical Record Request Script and Fax Letter
S.2 Prenatal Chart Abstraction Form
S.3 Labor & Delivery Chart Abstraction Form
S.4 Neonatal Medical Record Abstraction Form
S.5 Pediatric Chart Abstraction Form

Appendix T: Study Hypotheses and Data Collection Tools

*Appendix D.2 has been deleted

I. A. Justification

A.1. Circumstances Making the Collection of Information Necessary

The National CADDRE Study: 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 amended by public law 108-154 "The Birth Defects and Developmental Disabilities Prevent Act of 2003", which has extended the center's activities until the year 2007. (Appendix A)

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.

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.

OMB approval is requested for 3 years.

A.2. Purpose and Use of the Information Collection

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 CADDRE Study 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, and 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 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

[1] Study participants include children age 3-5 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).

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

part of 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, the National CADDRE Study 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 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. CADDRE will capture 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 will already be 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 CADDRE. As a result, the select mercury exposure hypothesis will address certain medical and occupational mercury exposures, but will not include a complete mercury exposure history. Given the retrospective nature of exposure ascertainment for CADDRE 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, the National CADDRE Study will permit 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.

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

Application of Information Technology

In addition to the CADDRE centers, NCBDDD funded a Data Coordinating Center (DCC) and a Central Biosample Repository (Central Lab) for the National CADDRE Study. Michigan State University established and manages the DCC. The DCC is developing an electronic data collection system to centrally store the data. Johns Hopkins University houses the Central Lab, where all biosamples from the study will be shipped, processed, and stored. The DCC and Central Lab work on an ongoing basis with the CADDRE investigators to prepare for the implementation of the study.

The CADDRE study will apply 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 will create and host a custom web-based information system, called the CADDRE Information System (CIS), which is carefully designed to directly support all of the CADDRE study data collection workflows, data quality assurance processes, and provide secure database and Internet transaction services.

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 CADDRE Study 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 three 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:

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

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

  - 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 CADDRE research site

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

- Provide weekday (Monday – Friday) user support for study staff using the CIS to facilitate efficient operations and improve the availability of CIS services

Participant Burden Reduction

The CIS will facilitate administering three computer assisted telephone interviews given to the study participant parents. These interviews are scheduled per the convenience of the participant parents. One of the interviews, the Caregiver Interview (Appendix B), 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.

Much of the data collection of the CADDRE Study protocol will take place in a clinical setting and involves the participant children who will be 3-5 years of age. Many of the data collection instruments will be 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 Scale (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 CADDRE Study 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
    - 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)
    - Automatic alerts for the clinical visit are provided to the staff about a participants 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 of 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, is investigating the cause of autism at 25 sites in the United States, Canada, Great Britain, France and Germany. The National Institute 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. Although both CPEA and STAART are investigating research domains similar to those in CADDRE, neither CPEA nor STAART sites adhere to a common protocol. Use of a common protocol will allow CADDRE sites to pool data, resulting in a sample of 2,700 children and their families. Not only does the large CADDRE 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 CADDRE, there are multiple differences between CHARGE and CADDRE. CHARGE will enroll 1,600 children and their families, significantly less than the 2,700 CADDRE will enroll. Moreover, CHARGE is collecting data only in the state of California and data, therefore, is not 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 CADDRE and CHARGE studies. CHARGE collects more data on environmental exposures than CADDRE, while CADDRE 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 CADDRE will permit replication of many of the CHARGE and EMA analyses. In fact, many aspects of CADDRE data collection were explicitly set up to enable this kind of replication which will allow for comparison of results.

A literature review conducted for CADDRE 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, several CADDRE investigators and collaborators are members of the IACC. These memberships provide expertise into research priorities among the federal government and other autism authorities. Several CADDRE investigators and collaborators also attended the 2004 Autism Summit Conference held in Washington, D.C. Information gained from this conference led to the development of the IACC Research Agenda. The proposed data collection contributes to multiple high priority autism research questions raised by the IACC.

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

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

A.6. Consequences of Collecting the Information Less Frequently

As the protocol is currently written, The National CADDRE Study 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 CADDRE case cohort 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

There are no special circumstances.

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

8A.

A 60 Day Federal Register Notice was published in the *Federal* Register on October 5, 2006, page 58865-58866, Volume 71, Number 193. No substantive public comments were received. (See Appendix C).

8B.

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 CADDRE sites played an integral role in the design and the development of the National CADDRE Study. 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

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:

Judith Grether, Ph.D
California Department of Health and Human Services, Environmental Health Investigation Branch
Phone: (510)620-3626
Email: jgrether@dhs.ca.gov

Craig Newschaffer, Ph.D.
Johns Hopkins University
Bloomberg School of Public Health
Phone: (410) 955-1268
Email: cnewscha@jhsph.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

In December 2003, prior to submission to CDC IRB, the CADDRE 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 CADDRE 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 CADDRE PI's identified changes that were required of the protocol based on the panel's feedback and these changes were incorporated into the protocol. 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 CADDRE project. The Principal Investigators for these two entities are:

Data Coordinating Center
Mohammad Hossein Rahbar, PhD (PI)
Michigan State University
Data Coordinating Center
Room 100, Conrad Hall
East Lansing, MI 48824
517.432.4083
517.353.9445
rahbar@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 2006, this Information Collection Request underwent significant additional revisions as a consequence of both external and internal reviews.

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

In order to ensure the validity of the data, it is important that CADDRE has high response rates. As stated previously, all CADDRE study families will have young children and two-thirds of CADDRE study 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 strutcture to ensure a more representative study sample:

Information Collection Step	**Incentive Offered**
Enrollment Packet Introductory Letter & Study Flow (E.1) Forms to Keep Cover Letter (E.2) Informed Consent (E.3) Study Bill of Rights (E.8) Prep Guides (E.4 & E.5) Picture Story (E.22) HIPAA Checklist & Medical Record Release Forms (E.18) Buccal Swab Kit (E.19-E.21)	$25 included in packet
Caregiver Interview (B.1 & B.2)	$30 mailed when scheduled
First Questionnaire Packet Paternal Medical History (E.14) Maternal Medical History (E.13) Autoimmune History Worksheet (E.9) GI questionnaire (E.12) Paternal Occupational Exposure (E.15)	$30 mailed when scheduled

Services & Treatments (Cases only) (F.3) EDQ (Cases only) (F.2)	
Second Questionnaire Packet CBCL (E.11) Carey Temperament Scale (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) Vineland (F.4) 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.2) 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 (introductory letter, study flow sheet, study bill of rights, picture story, and prep guides), HIPAA medical release forms, and buccal swab kits. 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.

- Caregiver Interview

The caregiver interview is an hour long computer assisted telephone interview 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. 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. The case parent clinic visit will include standardized developmental evaluations and biologic samples from both biological parents. 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. 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 a 3-Day Diet and a 7-Day Stool Diary. 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 CADDRE sites, but will include: gift cards, money orders, checks, and cash.

A.10. Assurance of Confidentiality Provided to Respondents

The CDC Privacy Act Officer has reviewed this application and has determined that the Privacy Act is not applicable to the data collection activities conducted by CDC-funded grantees. The Privacy Act is applicable to data collection activities at the Georgia CADDRE site (involving a contractor, Battelle, and to some CADDRE program activities involving the contractor, Battelle). All employees associated with this project, including contractors, will 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.

Due to the sensitive nature of certain data collection components, the CADDRE Study will obtain additional confidentiality protections. A consolidated application for 301(d) Confidentiality Certificates for protection of the individual participants at the six sites conducting the study has been requested by NCBDDD and is currently under review at CDC. Consent forms and other materials will be modified as needed upon finalization of the Certificate of Confidentiality.

The CADDRE project has been approved by the CDC IRB (Appendix I). The consent forms for parents or caregivers 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 (3-5 years), parental consent alone is sufficient and the explicit assent of the child is not required.

Multiple steps will be taken during the data collection process to ensure that the privacy and confidentiality of each participant is ensured to the best of the researchers' ability. Each study subject will be given a unique identifier (study ID) upon entry into the study. The study ID will be assigned by the CADDRE Information System (CIS); the DCC will maintain 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 will be kept in a locked file cabinet in a locked room with limited access to these data. Study staff will 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, will be required to undergo confidentiality training as part of their orientation. 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 will be sent to/from the DCC as encrypted data. Personal identifiers will be 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 CADDRE site's is only allowed to view its own data in identifiable form. No means will exist for one site to access the personal identifiable data stored by another site. The DCC will not release identifiable data from other sites to CDC.

The approved policies and procedures for safeguarding respondent privacy will be documented in a Manual of Procedures. This will ensure 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 will be 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 will not have access to participants' personal information. Samples will be 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 would keep them linked to personal information (through a study ID). This will allow study investigators, or other researchers approved by the CADDRE 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 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; 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

The CADDRE Study 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 B). 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 will be reminded that they may choose to skip any question that causes them undue discomfort and that their answers will not be

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 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 will 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 Caregiver Interview (Appendix B.1), 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 will 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 number, we believe that collection 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 will 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 will 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 will 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 will 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

Similar to CHARGE study rates, we estimate that we will be able to successfully trace and send a recruitment letter to a total of 17,610 potential participants. Potential participants will be identified through schools and clinics that serve children with developmental problems and through state birth certificate registries. After the introductory letters are sent, sites will conduct an invitation phone call with any potential study participant who responds indicating interest in the study. This phone call will include an eligibility screen and autism screen as well as an introduction to the study and a verbal consent for the study. We estimate that 8,922 (50%) participants will return a response card indicating interest and be called and screened for the study (Table A.12.A).

Of the potential participants who receive the invitation phone call, we estimate 3,456 (38%) 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 3,456 participants will be sent the enrollment packet.

The next step for the study participants will be to complete a telephone interview. We expect 3,456 of the participants will complete this interview (Table A.12.D).

The next steps are to complete two questionnaire packets (Table A.12.C). 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 3,456 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. We expect 90% (3,114) of all participants to complete the components of the clinical visit. The burden for cases (5 hours, 55 minutes) is longer than the burden for the NIC and Subcohort groups (2 hours, 5 minutes). See Tables A.12.F-J.

Table A.12.A. Burden Hours and Cost for Respondents: Mailed Invitation Packet

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
Response Card (M.3)	8,922	1	10/60	1,487	$18.62	$27,688

We expect to mail the Invitation Packet (Appendix M), which contains the Introductory Letter, Study Brochure, and response card, to 17,610 people and we expect 8,922 people to respond.

Table A.12.B. Burden Hours and Cost for Respondents: Invitation Telephone Call

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
Social Communication Questionnaire (J)	8,922	1	20/60	2,974	$18.62	$55,376
Invitation Phone Call (N)	8,922	1	20/60	2,974	$18.62	$55,376

Table A.12.C. Burden Hours and Cost for Respondents: Questionnaire Packets

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
Self-Administered Consent	3,456	1	5/60	288	$18.62	$5,363
HIPAA Release forms (E.18)	3,456	1	15/60	864	$18.62	$16,088
Autoimmune Disease Survey (E.9)	3,456	1	20/60	1,152	$18.62	$21,450

Carey Temperament Scales (E.10)	3,456	1	10/60	576	$18.62	$10,725
Child Behavior Checklist (E.11)	3,456	1	15/60	864	$18.62	$16,088
Survey of Gastrointestinal Function (E.12)	3,456	1	10/60	576	$18.62	$10,725
Maternal Medical History (E.13)	3,456	1	10/60	576	$18.62	$10,725
Paternal Medical History (E.14)	3,456	1	10/60	576	$18.62	$10,725
Paternal Occupational Questionnaire (E.15)	3,456	1	5/60	288	$18.62	$5,363
Child's Sleep Habits Questionnaire (E.16)	3,456	1	10/60	576	$18.62	$10,725
Social Responsiveness Scale (Adult & Child versions) (E.17)	3,456	1	45/60	2,592	$18.62	$48,263
3-Day Diet Diary (H.1)	3,456	1	20/60	1,152	$18.62	$21,450
7-Day Stool Diary (H.2)	3,456	1	40/60	2,304	$18.62	$42,900

Table A.12.D. Burden Hours and Cost for Respondents: Caregiver Interview

	No. of Respondents	**No. of Responses per Respondent**	**Avg. Burden per Response (in hours)**	**Total Burden (in hours)**	**Cost per Hour**	**Respondent Cost**
Primary Caregiver Interview (B.2)	3,456	1	60/60	3,456	$18.62	$64,351

Table A.12.E. Burden Hours and Cost for Respondents: Follow-Up Telephone Calls

	No. of Respondents	**No. of Responses per Respondent**	**Avg. Burden per Response (in hours)**	**Total Burden (in hours)**	**Cost per Hour**	**Respondent Cost**
First Follow-Up Phone Call (O.1)	3,456	1	30/60	1,728	$18.62	$32,175
Second Follow-Up Phone Call	3,456	1	20/60	1,152	$18.62	$21,450
Third Follow-Up Phone Call	3,456	1	15/60	864	$18.62	$16,088

Table A.12.F. Burden Hours and Cost for Respondents: Biosamples

	No. of Respondents	**No. of Responses per Respondent**	**Avg. Burden per Response (in hours)**	**Total Burden (in hours)**	**Cost per Hour**	**Respondent Cost**
Biosample Informed Consent (E.20)	10,368	1	20/60	3,456	$18.62	$64,351
Cheek Swab Sample Record Sheet (E.21)	10,368	1	20/60	3,456	$18.62	$64,351
Blood Draw Information Form (R.2)	9,342	1	35/60	5,450	$18.62	$101,470

We expect 10,368 index children and their parents (or 3,456 families) to provide a cheek swab sample. We expect 9,342 index children and their parents (or 3,114 families) to provide a blood sample.

Table A.12.G. Burden Hours and Cost for Respondents: Clinic Visit – Control Children

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
Mullen Scales of Early Learning (G.2)	2,304	1	45/60	1,728	$18.62	$32,175
Dysmorphology Exam Data Collection Form (P.2)	2,304	1	15/60	576	$18.62	$10,725

Table A.12.H. Burden Hours and Cost for Respondents: Clinic Visit – Control Parent

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
Vineland Adaptive Behavior Scales (F.4)	10	1	45/60	8	$18.62	$140
Written Informed Consent Document	2,304	1	10/60	384	$18.62	$7,150

The Vineland will be administered to parents of subcohort and NIC children who score less than 1.5 standard deviations below the mean on the Mullen. This outcome is expected to be extremely rare, but if it occurs, the Vineland will be administered as a follow-up telephone interview.

Table A.12.I. Burden Hours and Cost for Respondents: Clinic Visit – Case Child

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
ADOS (G.1)	810	1	40/60	540	$18.62	$10,055
Mullen Scales of Early Learning (G.2)	810	1	45/60	608	$18.62	$11,312
Dysmorphology Exam Data Collection Form (P.2)	810	1	15/60	203	$18.62	$3,771

Table A.12.J. Burden Hours and Cost for Respondents: Clinic Visit – Case Parent

	No. of Respondents	No. of Responses per Respondent	Avg. Burden per Response (in hours)	Total Burden (in hours)	Cost per Hour	Respondent Cost
ADI-R (F.1)	810	1	120/60	1,620	$18.62	$30,164
Early Development Questionnaire (F.2)	810	1	20/60	270	$18.62	$5,027
Services and Treatments Interview (F.3)	810	1	5/60	68	$18.62	$1,257
Vineland Adaptive Behavior Scales (F.4)	810	1	45/60	608	$18.62	$11,312
Written Informed Consent Document	810	1	10/60	135	$18.62	$2,514

Table A.12.K. Burden Hours and Cost for Respondents: Medical Record Abstraction

	No. of	No. of	Avg. Burden	Total	Cost per	Respondent

	Respondents	Responses per Respondent	per Response (in hours)	Burden (in hours)	Hour	Cost
Medical Record Request Script and Fax Letter (S.1)	3,456	5	15	4,320	$18.62	$80,438

We developed the CADDRE 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 CADDRE 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 CADDRE 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 CADDRE further. The proposed study protocol and data collection battery reflects a final balance of many compromises.

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

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 CADDRE Study

Fiscal Year 2007

Expense Type	**Expense Explanation**	**Annual Costs (Dollars)**
Direct Costs to the Federal Government	CDC Principal Investigator (GS-14, .55 FTE)	$83,590
	CDC Co-Principal Investigator (GS-14, .70 FTE)	$100,096
	CDC Health Scientist (GS-14, .20 FTE)	$26,370
	CDC Epidemiologist (GS-14, .03 FTE)	$3,677
	CDC Medical Epidemiologist (GS-15, .05 FTE)	$10,707
	CDC Epidemiologist (CC, .10 FTE)	$8,531
	CDC Public Health Analyst (GS-14, .06 FTE)	$7,124
	GA CADDRE Supplies/Postage/Printing	$17,500
	Travel	$5,000
	Subtotal, Direct Costs to the Government	$262,595
Contractor & Grantee Costs	GA CADDRE Site – Contract	$841,641
	CADDRE Program – Contract	$129,598
	California CADDRE Site – Grant	$368,291
	Colorado CADDRE Site – Grant	$358,252
	JHU CADDRE Site - Grant	$557,643
	UNC CADDRE Site - Grant	$337,353
	Univ of Pennsylvania CADDRE Site – Grant	$358,252
	Data Coordinating Center – Grant	$166,667
	Subtotal, Contracted/Grantee	$3,117,6

	Services	97
	TOTAL COST TO THE GOVERNMENT	**$3,380,292**

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 new data collection.

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 first three years of the study.

A. Time Schedule:

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

*Because we do not expect data collection to conclude until 3 months after OMB approval expires, an extension request will be submitted prior to the OMB expiration date.

B. Analysis Plan

General Considerations

Multiple analyses will be conducted on the data gathered in the National CADDRE study. The CADDRE 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 (National CADDRE Study 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 National CADDRE Study 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 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 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 will also work 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 will contract with Internet System for Assessing Autistic Children (ISAAC) for some of the data entry tools.

The DCC will be 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 CADDRE Study 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 will be 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 and 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 CADDRE 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 CADDRE Study 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 CADDRE hypotheses are provided in 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, digit ratio) 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.

***Hypothesis:* 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 will be 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%	13%

	with MR ~ 3%	

*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] will be 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 will be 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 will be 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 will 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-cohort designs. Unadjusted associations will be estimated with relative risk estimates calculated according to the approach described by Sato (1992). 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 will be 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 CADDRE Study 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 associations will be estimated with relative risk estimates calculated according to the approach described by Sato (1992). 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.

A theoretical advantage of case-cohort designs is that time of disease onset can be considered in the analysis as is done in a true prospective design (for example, using the Cox regression approach described by Barlow et al., 1999). However, because there is no means of determining an accurate time of onset for ASD, these methods may have limited utility in this study. Nonetheless it is expected that risk estimates based on these approaches will be compared to those from simple logistic regressions.

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, respondants 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.

<u>Hypothesis:</u> 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.

<u>Hypothesis:</u> Children with ASD will have different blood levels of markers of inflammation in compared to subcohort or NIC children.

To evaluate biomarkers of inflammation, we will 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 will 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.

<u>Reproductive/Hormonal Features</u>

The natural fluctuation of maternal hormones pre- and peri-natally 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

<u>Hypothesis:</u> 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. CADDRE 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 the 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 CADDRE site and two outside genetic researchers. The Data Sharing Committee will refer any proposals for ancillary analyses involving genotyping received from others within the CADDRE Study 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.

B. Statistical Methods

B.1. Respondent Universe and Sampling Methods

CADDRE study participants will be drawn from children born in and residing in the six study areas: the San Francisco Bay area, Denver metropolitan area, Philadelphia metropolitan area, Maryland, central North Carolina, and the Atlanta metropolitan area. Each CADDRE site selected 2-10 counties for its catchment area based on the following criteria: proximity to the study site, at least 30,000 live births collectively across counties, and geographic adjacency (i.e., counties must be contiguous). The sample frame for the cases and the NIC group, for each site, consists of

all clinical and educational settings serving children with ASDs (both public and private) within the catchment areas. The sample frame for the subcohort is the birth certificates for all children born during the birth cohort, excluding those children who have died. Please see Appendix K (Ascertainment Methodology) for additional detail.

The study cohort will consist of children:

- Born from January 2003 – December 2004 (i.e., a 24-month birth interval; actual dates subject to change) and
- Born in and currently reside in the catchment area of each site during the study period, which is January 2007 - December 2009 (actual dates subject to change due to study implementation considerations).

The following criteria describe which children are potentially eligible for this study and justifications for these criteria:

- Child is 24-60 months old at time of eligibility (birth date range of January 2003 to December 2004) This age range, which is younger than the age range seen in some other studies, was chosen in order to limit recall bias for events in pregnancy and early life as much as possible while still allowing diagnostic accuracy for ASD.
- The child is between 30 – 60 months at completion of data collection components. This age range is appropriate for the instruments to be administered in the study.
- Child was born in and currently resides in the study catchment area. The defined cohort (required by the case-cohort design) is to be ascertained from birth

certificate data; current residence is required for ascertainment purposes and to allow for examinations and other in-person assessments of enrolled subjects.

- Child currently resides with a knowledgeable caregiver. For purposes of the study, a knowledgeable caregiver is defined as a family member or other caregiver of legal age who has resided with and consistently been caring for the child since the child was 6 months of age or younger (based on self-report). This criterion is necessary in order to collect accurate information on early life events that may be risk factors for the development of autism.
- Legal consent is obtainable.
    - Note. Any children with legal circumstances (e.g., legal adoption) preventing access to birth certificates/ legal consent are not eligible for the study. Foster children will be excluded because, as wards of the state, it is difficult to obtain informed consent for them. Children who have been legally adopted will not be enrolled because birth certificate records are typically sealed. Birth certificates are essential in determining place of birth, one of the eligibility criteria for the study.
- The knowledgeable caregiver is competent to communicate orally in English or Spanish. This limitation is necessary since the clinical study instruments to be administered are only available and validated in English, although the majority of the instruments are also validated in Spanish. Some sites may exclude Spanish-only speaking participants based on the percentage of Spanish-only speaking residents and other site specific factors.

Children who do not meet all of these criteria will not be eligible to participate.

Identification of Potential Case and NIC Children

Potential cases and NIC children are cohort children identified by the study as having a suspected ASD or other selected neurodevelopmental impairments (please refer to Appendix L for list of diagnoses). For the purposes of this study, ASD includes Autistic Disorder, Pervasive Developmental Disorder-Not Otherwise Specified, and Asperger's Syndrome. Potential cases and NIC children will be identified through sources serving or evaluating children with developmental problems.. Sources for potential cases and NIC children may include Part C and Part B agencies, special education programs, state autism registries, hospitals, and clinics. These types of data sources were chosen based on experience with CDC autism surveillance activities. Each National CADDRE Study site will obtain IRB approvals from their appropriate institutions and, if necessary, written agreements from their local sources. Site-specific case-ascertainment procedures, including definitions of Part C and Part B agencies, are detailed in Appendix K.

The first step will be to identify potential case and NIC children who: a) have received an ASD diagnosis, b) meet specific ASD or ASD-related exceptionality criteria, or c) have received a diagnosis of one or more select conditions associated with ASD. Please see Appendix L (ICD-9 Codes/Part B School Eligibility Criteria) for specific eligibilities. In addition to having at least one of the criteria above, the child must meet all eligibility criteria listed in Section B.1.

Note that the criteria described above for children with suspected ASD are quite broad and are not limited to children with a previous autism diagnosis or autism exceptionality for early intervention/special education. This broad diagnostic net for possible cases ensures that young children with suspected ASD (i.e., young children without a formal diagnosis of ASD) are identified. However, it will also identify many children who may not have ASD. Therefore, CADDRE sites will use an ASD screening instrument to assign participants into the potential case group or the NIC group. All of the respondents who screen into the potential case group will be invited to enroll in the study. However, since the study analysis plan requires a one-to-one case to control ratio, those respondents who screen into the NIC group may or may not be invited to enroll. At each CADDRE stie, a random sample of NIC children will be drawn from all NIC children across the catchment area.

The purpose of the ASD screen is to identify those children in the broad net who are most likely to have ASD. Due to variations in the recruitment process among sources within each study site, two options for screening were developed. The site will determine the most appropriate screening process for a given source population. The screening process will consist of either 1) a record review for descriptions of behaviors characteristic of ASD; this process is based on the autism surveillance methods conducted at CDC and elsewhere, or 2) parents will be administered the Social Communication Questionnaire (SCQ, Appendix J). California will use a record review; Georgia (CDC), University of North Carolina, Chapel Hill, University of Pennsylvania, Colorado, and Johns Hopkins University plan on using an SCQ. The SCQ is a 5 to 10 minute screening interview

completed by the child's primary caregiver. It is normed as a screening instrument for ASD in children 4 years and older.

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

The SCQ is the considered to be the best parental screening instrument for this study. Although the published SCQ is validated on children 4 years of age and older, there has been considerable work recently investigating the performance of the SCQ in children younger than age 4 (Wiggins, L., Bakeman, R., Adamson, L., & Robins, D., in press; Hanson, Sullivan, Ware, Lord, & Thurm, 2002; Corsello, Cook, & Leventhal, 2003; Corsello, Anderson, Qui, Risi, & Lord, 2004; Corsello, Lord, Hus, & Qui, 2005; Eaves, Wingert, Ho, & Mickelson, 2006; Eaves, Wingert, & Ho, 2006; Baird, et al., 2006), There has also been a study conducted by the MD CADDRE site in preparation for the CADDRE study (Lee, David, Rusyniak, Landa, & Newschaffer, in press). Lee et al (in press) found that the sensitivity and specificity of the SCQ is improved at a reduced cut-off score in younger age cohorts. Likewise, Wiggins et al. (in press) found that sensitivity

and specificity improved at a cut-off score of 13 in children younger than 4 years of age; sensitivity and specificity were maximized at a cut-off score of 11. Finally, Allen et al. (Allen, Silove, Williams, & Hutchins, 2006) reported that when using a cut-off score of 11, sensitivity and specificity of the SCQ was 93% and 58% for children aged 2-6 years, and 100% and 62% for children aged 3-5 years.

Thus, a cutoff score of 11 will be employed for this study. If the child scores ≥ 11, ASD will be considered indicated (i.e., positive screen). If the child scores <11, ASD will be considered not indicated (i.e., negative screen).

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

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

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

- Children with a previous ASD diagnosis/exceptionality will be contacted, screened for eligibility, and then enrolled in the study.
- Children identified by the broad diagnostic net who do not have a previous ASD diagnosis/exceptionality, a systematic sample representing a range of diagnoses and ages found in the database will be used to identify which children will be contacted for the secondary eligibility screening step. The proportion of the children identified by the broad diagnostic net criteria and subsequently sampled will vary among sources depending on the site-specific requirements for subject contact (see Appendix K). At sites where consent is required before eligibility screening, the sample may be up to 100% of those consenting to eligibility screening if the proportion of subjects who are contacted and consent to the eligibility screen is low. The sample proportion may be lower than 100% at sites where consent is not required before contacting for the eligibility screening. Thus, the sample proportion will be adjusted by each National CADDRE Study site to ensure the target enrollment is met. Each child's identified legal guardian will be contacted to determine if the child meets the study eligibility criteria.

As an illustration of this approach, the GA National CADDRE Study site (CDC) recruitment process is summarized as follows (details in Appendix K): 1) Following procedures established as part of CDC ASD surveillance, sources are requested to provide lists of children who meet the cohort eligibility and broad diagnostic net criteria. Based on the diagnostic/exceptionality information released by the sources children with a previous ASD diagnosis/exceptionality will be contacted, screened for eligibility, administered the SCQ, and enrolled (Possible Case Group), 2) children without a previous ASD diagnosis/exceptionality will be contacted, screened for eligibility, administered the SCQ, and enrolled in either the Possible Case or NIC groups based on the results of the SCQ screener. For the NIC group, the sample proportion to be contacted will be adjusted to produce an approximate 1:1 ratio of case-NIC children.

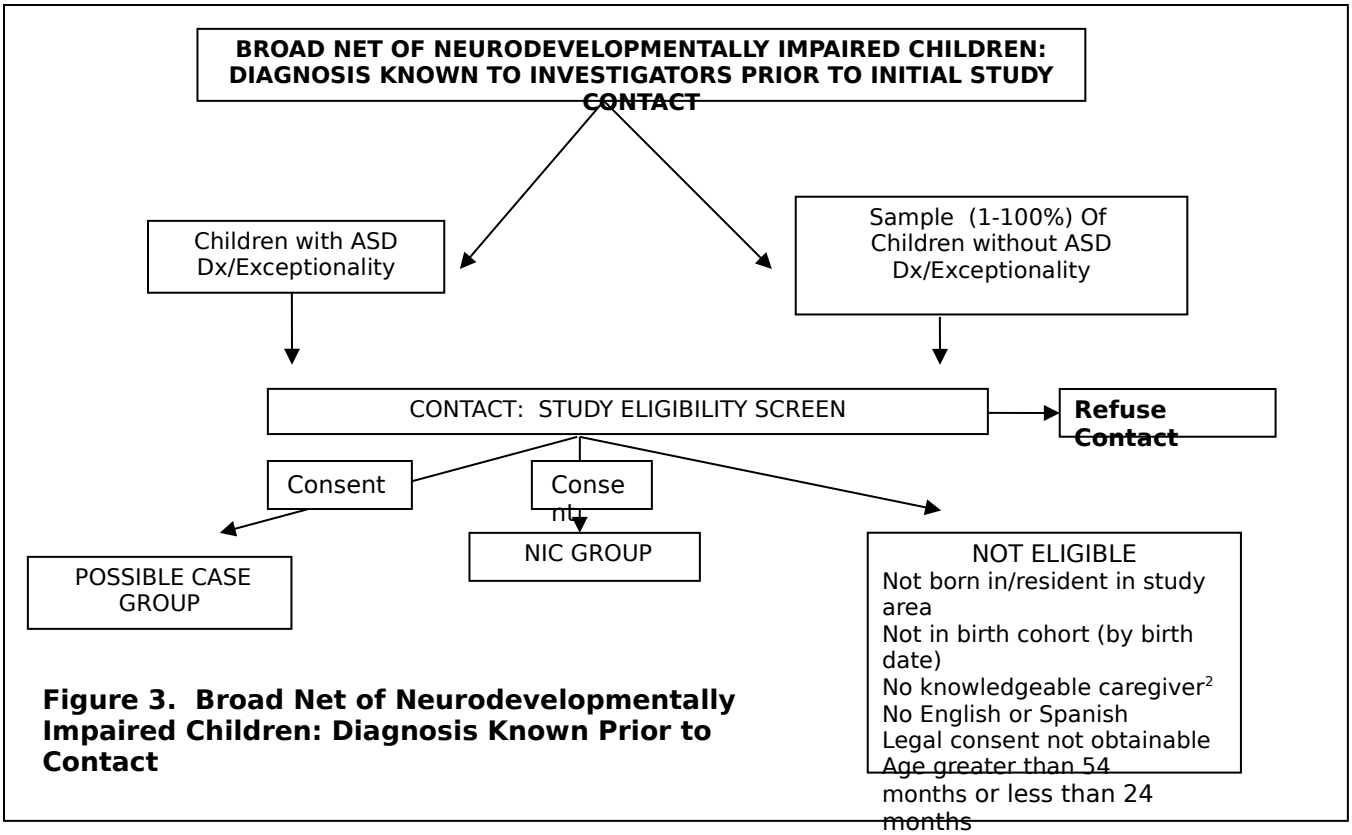


Figure 3. Broad Net of Neurodevelopmentally Impaired Children: Diagnosis Known Prior to Contact

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

Figure 4. (p. 63) describes the recruitment process for local sources that are unable to release the specific diagnosis type/exceptionality to the National CADDRE Study investigators without consent from the potential participant. These sources have a slightly different process for determining their potential case and NIC groups. Sites with this type of local sources will go through the additional step of obtaining consent to contact prior to ASD screening and determining the eligibility of the potential participant.

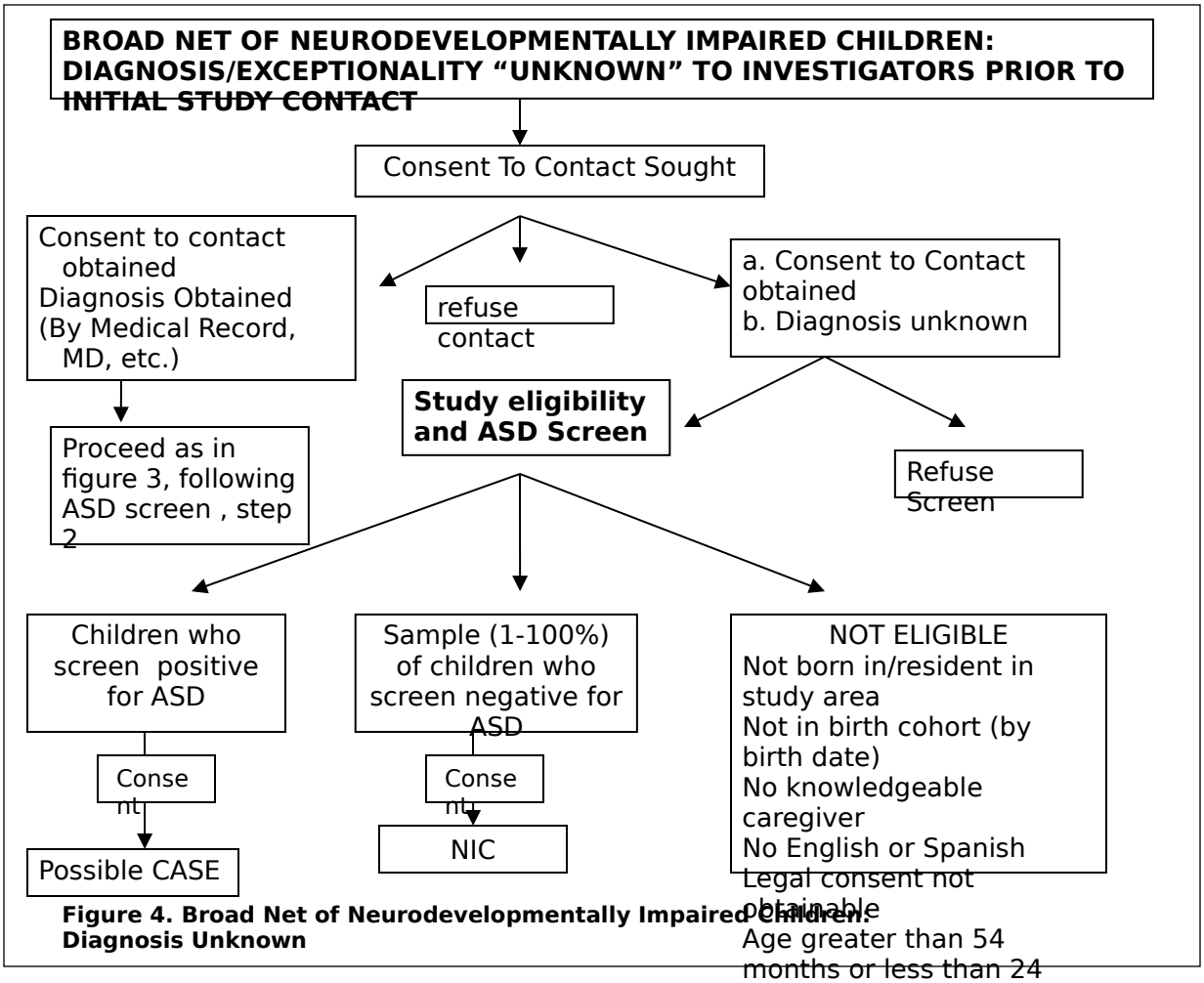


Figure 4. Broad Net of Neurodevelopmentally Impaired Children: Diagnosis Unknown

c. Self Referrals

Parents, or legal guardians, who contact the study site, have a child with a previous ASD or ASD-related diagnosis or concerns that their child might have an undiagnosed ASD or ASD-related disorder, and are interested in having their child participate, will be considered Self-Referrals. Children with neither a previous diagnosis nor parental concern about an ASD or ASD-related condition will be ineligible for Self-Referral. Self-Referral procedures are detailed in Appendix K. Children in this group who either, 1) meet specific ASD or ASD-related

criteria (Appendix L), 2) have documentation of ASD or ASD-related criteria (Appendix L), or 3) screen positive for ASD by medical record review or SCQ, and meet all study eligibility criteria, will be enrolled and proceed through the study as a member of the Broad Net, as shown in Figures 3 (p.62) or 4 (p.63) (depending on the site-specific recruitment process). Children who do not have a previously diagnosed condition listed in Appendix L and who screen negative (autism not likely) will be excluded. Regardless of the results of the screening, parents who express concern about their child's development will be given referral information for further developmental evaluation.

Identification of Potential Subcohort Children

Subcohort children will be identified from birth certificates on the basis of birth date range and residence in the catchment area at the time of birth. In addition to IRB approval, most sites will also require approval from the State Registrar or Vital Records Department to obtain files containing personal identifiers. Once approvals are obtained, potential subcohort member children will be randomly selected from among all cohort children. When possible, sites will link birth records to state death certificate files to remove any children from the contact list who are deceased (Site-specific information on subcohort ascertainment is provided in Appendix K). The current residence in the study catchment area will also need to be established.

Based on these steps, contact will take place with each child's identified legal guardian to obtain consent and to determine if the child meets the study eligibility criteria.

If the legal guardian consents to have the child participate in the study and the child meets the eligibility criteria, then the child will be enrolled as a subcohort member as shown in Figure 5 (p. 65). All children enrolled in the subcohort will receive the SCQ for the ASD screen. Children whose SCQ results indicate that the child is likely to have ASD will be invited for a clinic visit to undergo a developmental evaluation following the data collection of the Case group. This evaluation includes autism specific assessments.

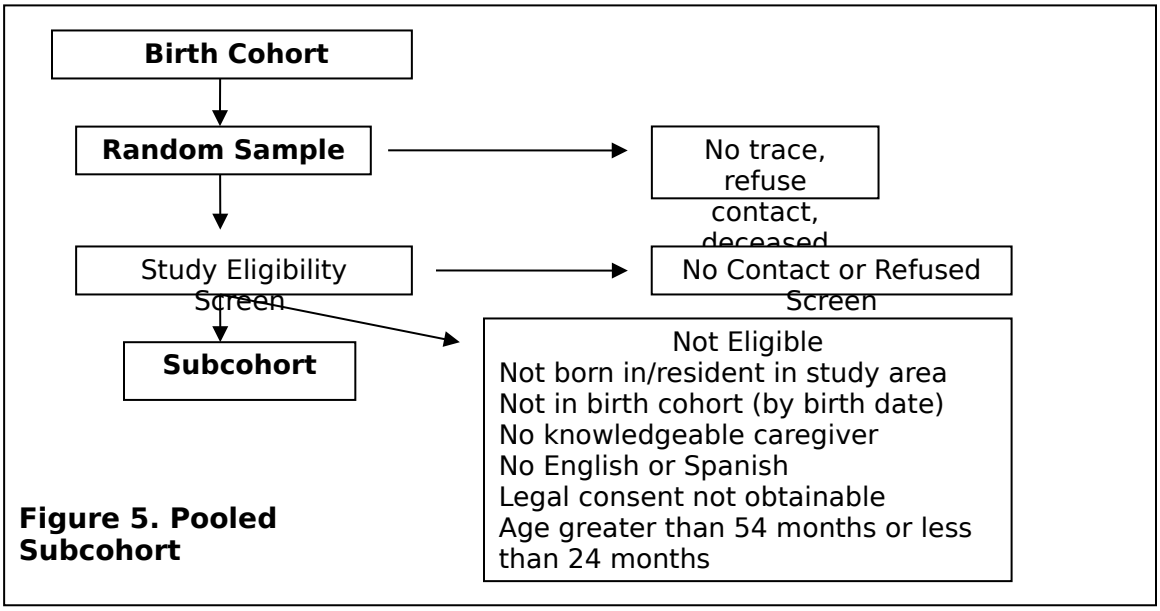


Figure 5. Pooled Subcohort

Final Subject Group Classification

1) ASD: ASD case confirmation will be established by the current state-of-the-art methods by administering the Autism Diagnostic Interview (ADI-R, Appendix F.1) to the primary caregiver, and the appropriate module of the Autism Diagnostic Observation Schedule (ADOS, Appendix G.1) to the child. In some instances, (e.g. refusal, loss to follow-up) these subjects may not complete the study's developmental evaluation and,

thus, will retain their initial assignment as Possible Case group member for analysis purposes. Consequently, analyses as outlined in A.16 may be performed separately on subjects classified on the basis of the developmental evaluation and subjects classified on the basis of the ASD screen.

2) NIC: A child is considered a member of the NIC when they have documentation of a condition listed in Appendix L indicative of a neurologically impaired developmental disability diagnosis and receive a score on the SCQ indicating that autism is unlikely.

Possible Case children who do not meet ASD criteria based on the complete developmental evaluation will also be designated a member of the NIC group.

3) Sub-cohort: A child that is sampled from the birth cohort, and meets study eligibility criteria, will remain a member in the sub-cohort comparison group. If a member of the sub-cohort group has a positive screen on the SCQ, they will also complete the same evaluation as that for a possible case. Analyses using traditional case-cohort methods recognize that these subjects are cases as well as sub-cohort members.

Sample Sizes and Response Rates

The proposed cohort size for the study (approximately 485,000 births combined across all 6 study sites, based on a range of 34,000-50,000 births per year in each study site catchment area) and therefore, the anticipated number of individuals in the case and comparison groups, was partly determined by time and resource considerations. Details

on expected case sample size and study power are provided in B2. In short, our target final sample size – i.e., the number of enrolled children with complete data collection - is about 650 children in each of the three subject groups (Case, NIC, sub-cohort), or about 1950 children in total. Based on the experience of the CHARGE Study regarding the percentages of cases and sub-cohort members who were eligible for the study, agreed to participate, and completed data collection (CHARGE investigators, personal communication; Hertz-Picciotto et al 2006), the following expected samples sizes (all sites combined) at different phases of the study are provided below (we expect the corresponding samples sizes for the NIC group to fall within the range of the ASD and sub-cohort estimates, with rates of participation assumed to be more similar to the sub-cohort group than ASD group):

Sub-cohort: 5000[1] x 50% traced = 2500[2] x 78% eligible = 1950[3] x 46% participation = 900[4] x 72% complete data = 650[5]

ASD: 1550[2] x 80% eligible = 1240[3] x 73% participation = 900[4] x 72% complete data = 650[5]

NIC: 2440[2] x 80% eligible = 1950[3] x 46% participation = 900[4] x 72% complete data = 650[5]

[1] random selection of potential sub-cohort children from birth cohort
[2] number of children (families) contacted; for ASD cases, this number corresponds to the number of potential cases in the cohort (with a previous diagnosis or true positives on the ASD screen) based on a conservative estimate of the ASD rate of 3.2 per 1000 among all cohort births – we expect the prevalence rate to be at least this value based on recent ASD prevalence estimates in older children; for NIC, this number corresponds to the minimum number of children with a broad net diagnosis who screen negative on the ASD screen.
[3] number of eligible children (families) invited to participate
[4] number of enrolled children (families)
[5] number of enrolled children (families) with complete data collection

B2.Procedures for the Collection of Information

Sample size and Study Power Estimation

The proposed cohort size for the study (approximately 485,000 births combined across all 6 study sites, based on a range of 34,000-74,000 births per year in each study site catchment area) and therefore, the anticipated number of cases and comparison groups, was partly determined by time and resource considerations. Further, few data are available to estimate anticipated effect sizes for particular exposure-outcome relationships of interest. Given these constraints, we calculated what the minimum effect sizes (relative risk estimates) would be with the anticipated sample size.

The anticipated sample size was based on CDC-provided prevalence estimates from the Metropolitan Atlanta Developmental Disabilities Surveillance Project (MADDSP) corresponding to 3.4 ASD cases per 1,000 children ages 3-10 years in the population in 1996 (published range in the literature: 2-6 per 1000). Because we are focusing on preschool children, we chose a more conservative, minimum prevalence estimate of between 3.0-3.4 per 1000 children. Thus, with 485,000 births in the birth cohort across all study sites, we anticipate identifying about 1550 ASD cases (based on a prevalence of 3.2 per 1000), including previously diagnosed children and CADDRE Study-identified ASD children. Based on the experience of the CHARGE Study regarding the percentages of cases who were eligible for the study, agreed to participate, and completed data collection (CHARGE investigators, personal communication; Hertz-Picciotto et al 2006), the following expected final sample size (all sites combined) is:

1550 ASD cases x 80% eligible = 1240 invited ASD cases x 73% participation = 900 enrolled ASD cases x 72% complete data =

650 enrolled ASD children (families) with complete data collection.

We will enroll subjects in each comparison group in a 1:1 ratio to cases. Although the study will have two comparison groups, only one group will be compared to the cases at a time.

Satisfactory, simple methods for estimation of sample-size (or minimum detectable relative risks (mdRR)) have yet to be developed for case-cohort designs. However, a number of simulation studies have compared the efficiency of like-sized case-control and case-cohort designs conducted on the same population and have found estimates of precision to be very similar (Wacholder, 1987). Confidence in the appropriateness of using case-control approaches for case-cohort designs increases if 1) analyses are going to be based on cumulative incidence rather than time-to-event and 2) if there is limited censoring in the data. National CADDRE Study will typically rely on cumulative incidence and censoring is not a major issue because we are defining the source population as born-in-plus-still-residing and because there are few competing risks. Therefore, estimation of minimum detectable Relative Risk (RR) estimates (or minimum detectable Odds Ratios (OR)) was based on simple case-control approaches. NCSS PASS 6.0 software was used for all calculations.

A number of estimations were made to gauge the effect of different assumptions (case

group size – all cases or stratified on specific ASD subgroups - and exposure prevalence rates) on min Odds Ratios (OR). For all estimates, conventional alpha and beta error tolerances were applied (0.05 and 0.20, respectively). The assumptions and range of min OR are tabulated below:

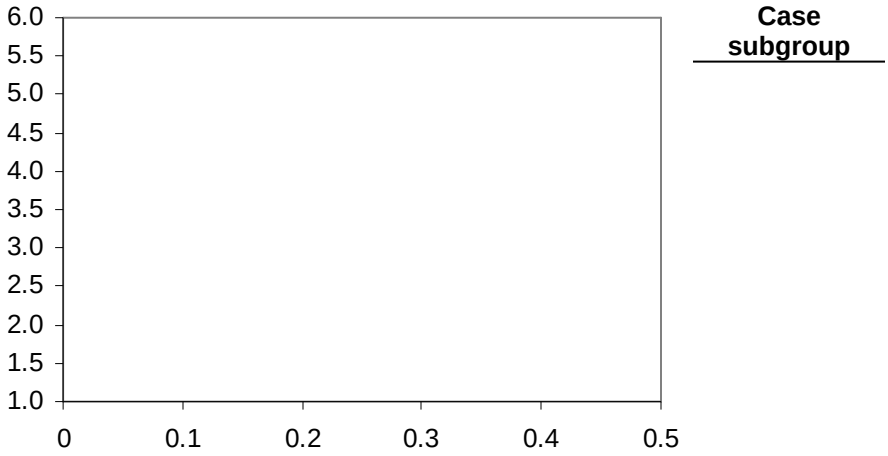
Size of ASD case subgroups of etiologic interest

Proportion of full case group	Corresponding Subgroup
20%	Complex autism
30%	Nonverbal; With Regression
40%	With MR
60%	No MR
70%	Verbal; Without Regression
80%	Essential autism

Select candidate exposure prevalence rates

Subcohort Exposure Prevalence Estimate	Corresponding Candidate Exposures
1%	ART, OCs
5%	Any infertility tx, maternal thyroid disorder hx, seizures, FraX
10%	Any maternal autoimmune hx, ADHD, GI dysfunction
20%	Pitocin, fever in pregnancy
50%	Any infection in pregnancy

CADDRE CASE-COHORT STUDY POWER ANALYSES



Minimum detectable OR (80% power, alpha error tolerance 5%) under alternate subcohort exposure prevalences and ASD case subgroup sizes

Subcohort Exposure	Size of case subgroup as a % of total case group (650)						
Prevalence	20%	30%	40%	60%	70%	80%	All cases
1%	5.28	4.55	4.15	3.71	3.57	3.46	3.31
5%	2.60	2.33	2.18	2.02	1.97	1.93	1.87
10%	2.12	1.93	1.82	1.71	1.67	1.65	1.61
20%	1.84	1.69	1.61	1.52	1.50	1.48	1.45
50%	1.72	1.59	1.52	1.43	1.41	1.39	1.37

Last, minimum detectable effects for interaction were calculated:

Interaction assessment:
Exposure 1 prevalence: 0.025, 0.05, 0.20, 0.35, 0.50; Exposure 2 prevalence: 0.05, 0.10, 0.25, 0.50

For the minimum detectable interaction odds ratio (assuming no stratification and a 1:1 case-control ratio), the values ranged from about 5.0 to 12.0 for combinations of the lowest prevalence rates for Exposure 1 with different Exposure 2 prevalence rates; they

declined to about 2.0 to 3.5 for combinations of the highest prevalence rates for Exposure 1 with different Exposure 2 prevalence rates.

Considering the factors used in the above calculations (e.g., exposure prevalence rates, ASD subgroup size), the factors that are possibly modifiable by design to enhance study power are the case: control ratio and case group size. However, compared to a 1:1 case-control ratio, a 1:2 ratio had relatively small effects on the mdRR. Thus, it was determined that the greatest benefit would be gained by enhancing case group size by taking steps to increase participation and/or enhancing case ascertainment (e.g., the "Broad Net of Neurodevelopmentally Impaired Children") from the proposed study cohort. An additional step, to be planned for the future (if funding levels permit), will be to expand the study cohort and continue enrollment into a second phase of field data collection.

With regard to power for gene-environment interactions, we acknowledge that 650 cases/ sub-cohort members is on the lower end of desired sample sizes for tests of interaction and, if feasible, we would certainly advocate enrolling more subjects. However, such an ideal world is rarely possible, and our task is to use our expertise to balance "perfect study" intentions with the practical limitations of the real world in a way that retains meaningful information. After such an exercise, we decided upon our current approach as the best way to achieve that balance.

Our reasons include consideration of phenotype precision versus sample size and the

current lack of agreement in the field about how to best address the multiplicity of testing when considering large-scale genotyping efforts. The first point has been a driving force. The power to detect a causal effect is driven not only by sample size, but also by the amount of error around each measurement used in an analysis. We believe that careful phenotyping will provide much less misclassification and much greater precision than less costly protocols that would allow larger sampling. Further, the main concern regarding power for gene-environment interactions appears to stem from an assumption of genome-wide genotyping, or at least very large numbers of genotypes to be generated, and the need to correct for multiple tests, which reduces power estimates by requiring a very strict threshold for significance. While this is worthy of consideration, the genetic epidemiology field as a whole has not yet agreed upon an approach. Some researchers have promoted the use of a Bonferroni correction for all markers genotyped, although most agree that this is too conservative. Several would argue in fact that no correction is needed, and that biological plausibility and replication will remove false positive findings. Others promote methods such as allowing a specified false discovery rate, or applying a Bayesian approach relying on priors for different sets of GxE tests. Each of these has merit, and would suggest a different "significance" threshold for calculations of power. Under the assumption that correction is irrelevant, we would have very high power estimates for our approach. Under the opposite extreme, using a Bonferroni correction for half a million tests, power estimates would be quite low. Most of the other approaches would yield results in between, but would suggest we have ample power to detect moderate to large GxE effects. Given the various aims of this project, we feel this is the appropriate compromise of size versus precision for this study.

Procedures for the Collection of Information

Identification of case, NIC, and subcohort subjects is described in Section B.1. Once the children are identified for the study, an introductory letter (Appendix M.1), study brochure (Appendix M.2), pre-paid response card (Appendix M.3), and incentive valued at one dollar will be mailed to the primary caregiver. This introductory packet will be mailed from the study site (e.g., CDC) or collaborating local sources (e.g., Montgomery County Department of Education) depending on the agreements each site has with their various sources. All introductory packets will be sent in English. For sites including Spanish speakers (California and Colorado), the materials will be sent in both English and Spanish and the response card included in this packet will ask the primary caregiver to indicate language preference, so that all future materials can be mailed in the appropriate language.

If no response is received within two weeks, a second introductory packet will be mailed to the household. If no response to the second packet is received from the subject, a third set of packets will be mailed at a later date (e.g., 6 months). This additional mailing will be identical to the first packet described above. Some sources may require an affirmative response before proceeding with other recruitment measures; in this case, if no response card is received no further contact will be attempted.

If the invitee returns the response card indicating interest, recruitment will proceed with the invitation telephone call. In addition, if a source allows study staff to contact the potential participants by telephone prior to receiving a positive response from the

response card, sites will proceed with the invitation phone call after the second introductory packet. CA CADDRE will not send an introductory packet and instead will initiate contact with the invitation phone call.

The invitation phone call will include verbal consent for the study, eligibility screen, and the SCQ autism screen (Appendix J). Scoring of the SCQ can occur during the call. If the SCQ score falls into the range indicating autism, the participant will be invited during this call to participate in the full study. Of participants whose SCQ scores do not indicate autism, only a sample will be invited into the study. Depending on the sampling process, individuals who are selected may need to be re-contacted with a second phone call to be invited into the study. The script used during the telephone call is in Appendix N.

If a participating primary caregiver is not the biological mother, the caregiver will be asked to provide the biological mother's contact information. The biological mother will be contacted by a similar Invitation Phone Call, omitting the eligibility and developmental screening. If she consents to participate, she will be contacted for only the data collection items that the primary caregiver is unable to complete. In addition, the participant will be asked if the biological father resides in the same household. If he does not reside in the household, the caregiver will also be asked to provide the biological father's contact information. If provided, the biological father will be contacted by phone and invited to complete the paternal medical history form, a buccal swab kit, and will be asked to come in for a blood draw. The biological father is only considered essential to the study if he is the primary caregiver (participants without biological father data will

still be considered as having a complete data set).

The Enrollment Packet (Appendix E) will be mailed to participants who verbally agree to be in the study. This packet will include a cover letter (Appendix E.1), study flow diagram (Appendix E.1), consent documents to review (Appendix E.3), prep guides for the remaining data collection components (Appendices E.4 and E.5), a checklist of questionnaires (E.23) HIPAA medical records release forms (Appendix E.18), and a cheek cell collection kit (Appendix E.19-E.21). Sites also will include a picture story (Appendices E.22) in the enrollment packet. This explains what the child will do during the clinic visit and will make them feel more comfortable with the process.

The general flow of the data collection will begin with the enrollment packet. Please refer to Appendix D for a study flow diagram and a data collection instruments summary table. Ideally, the Caregiver Interview (Appendix B) will occur prior to the clinic visit. The medical record abstraction portion (Appendix S) of the study will be completed by study staff independently of other study components.

Each check swab kit will be assembled based on the availability of targeted participants in the household (e.g., a household including the biologic mother and child, but not the biologic father, would receive a kit including only consent forms and brushes for mother and child). Buccal cell sampling kits will be mailed to multiple households, as needed, e.g. if biological parents live in separate households from the index child. Kits will include an instruction sheet (Appendix E.19) and a consent form (Appendix E.20).

Participants will be asked to return the kit with a labeled Federal Express mailer provided by staff. Buccal cell samples will be sent directly by the participant to the study Central Lab where they will be stored at -80ºC until DNA is extracted. Participants whose samples are not received by the time of the clinic visit will be given an opportunity to complete the sample at the time of the clinic visit. In addition, subjects providing samples that fail to amplify on a test PCR will be asked to supply another sample (this could be done at the clinic visit or may require another mailing).

Study staff will call the participant one week after the enrollment packet has been mailed to answer questions about the packet, explain and schedule the caregiver interview, and to schedule the next data collection step – the questionnaire packets.

Participants will complete questionnaire packets with assistance from study staff at a clinic or home visit or over the telephone. We will allow participants to choose the option that works best for them. In addition, we will allow participants to complete the packets as self-administered questionnaires; however, we will recommend that they do not choose this option. Assisting participants with the completion of the packets ensures that the forms are completed fully, allows staff to answer questions, and builds rapport with participants.

The First Questionnaire Packet includes the Paternal and Maternal Medical Histories (E.13-14), the Autoimmune Disease Survey (E.9), the GI Questionnaire (E.12), and the Paternal Occupational Questionnaire (E.15). Case participants will also complete the

Services and Treatments Questionnaire (F.3) and the Early Development Questionnaire (F.2). The Second Questionnaire Packet includes the Child Behavior Checklist (E.11), the Carey Temperament Scales (E.10), the Child Sleep Habits Questionnaire (E.16), and the Social Responsiveness Scales (E.17).

Staff will call the participant after the First Questionnaire Packet is complete to answer any questions and to schedule the clinic visits and the Second Questionnaire Packet. If the participant has opted to complete the first packet as self-administered questionnaires, the staff member will check on the participant's progress and answer any questions.

All primary caregivers of case and comparison children will be asked to complete a caregiver interview by telephone. If the primary caregiver is not the biological mother, they will only be asked questions related to sociodemographic factors and the child's postnatal medical and developmental history (Sections A, B, and H of the main caregiver interview. The caregiver interview can be found in Appendix B. Sites will either contract with an independent agency or use their own staff to conduct the CATI.

The clinic visit will include four main components that can be split into different clinic visits or combined for one longer visit. Ideally, subcohort and NIC members will be scheduled for one clinic visit. The Case group will likely require more than one face-to-face visit, as the data collection is lengthier for these groups.

The clinic visit component will consist of four elements: 1) Intake, 2) Child Clinical, 3)

Parent Biosample, and 4) Parent Interview. During Intake, study staff will answer questions related to the study, review the self-administered questionnaires and HIPAA release forms, and consent the participant. The child clinical visit will include three components: a developmental assessment battery (Appendix G), a physical examination of the child (Appendix P), and a child blood and hair sample. The developmental battery includes the Mullen Scales of Early Learning, which all children will complete, and the ADOS, which only case children will complete.

Biological mothers and fathers in all three study groups (Case, NIC, and the subcohort) will be asked to provide venous blood samples. Blood tubes will be retained at ambient temperature and shipped via Federal Express to the Central Lab the same day they are drawn. Filter paper cards will be shipped to the Central Lab within seven days of collection. Please see Appendix R.1 for a more detailed description of the handling, shipment, and storage of these blood samples

All primary caregivers of case children will be asked to complete an interview asking questions about their child's development. This interview includes 4 separate components: the ADI-R, the Vineland, the Early Development Questionnaire, and the Services and Treatments Questionnaire. Please see Appendix F for the parent interview.

The third packet of self-administered questionnaires (Appendix H), which includes the Diet and Stool Diaries, will be provided to the participant during the clinical visit. During the clinical visit, study staff will provide the caregiver with a brief explanation for

how to complete these forms and provide each participant with a packet to be returned to the study.

Medical record abstraction includes the maternal prenatal and, labor and delivery records as well as the records from other medical providers (i.e., gynecologist, allergists, psychiatrists) the biological mother might have visited three years prior to the child's date of birth; and the neonatal and pediatric records of the child including any specialty providers (i.e, developmental pediatrician, allergist, gastroenterologist, psychiatrist) the child might have visited since birth. After the HIPAA medical records release authorization forms (Appendix E.18) have been received, medical record abstractors will contact relevant sources (clinics, hospitals) in order to obtain access to the medical records of participants. The Medical Records Request Script and Fax letter are Appendix S.1. All data will be collected electronically using the CIS. The CIS screens will be based on hard copy forms developed by the Data Collection Instruments Working Group (hard copies can be found in Appendix S).

B.3.Methods to Maximize Response Rates and Deal with Nonresponse

We expect to achieve a response rate of at least 75%. We expect to have several follow-up calls and face-to-face visits to build rapport with the participant and to keep abreast of the participant's progress and needs. This strategy will facilitate the participant's completion of the data collection protocol, and it will also enhance the quality of data obtained from each participant. In the first and all subsequent contacts with the

participant, we will emphasize that we will work with the participant to help them complete the data collection process – in essence, a tailored approach. If the proposed data collection plan needs adjustment to fit the individual participant's needs, it will be done.

As described previously, all letters of invitation will include a description of the study and other information to facilitate the respondent's understanding of the project. If the respondent has additional questions about the study they are encouraged to call the Principal Investigator. The respondent is asked to return a response card indicating whether or not they are interested in participating in the study.

If there is no response from the invitee within 2 weeks, follow-up phone calls may be made by CADDRE Study staff in order to contact the respondent and explain the purpose of the study, answer questions and encourage participation. At least 9 attempts at phone contact will be made; preferably on different days and different times of the day. For the first six calls, if the caregiver is not there, we will not leave a message. If after the 7th call, we have not been able to reach the caregiver, we will leave a generic message asking the caregiver to call back. If we have not heard back from the caregiver within two weeks, we will call again and leave another message if necessary. If after 9 attempts, the potential participant is not able to be contacted, no further attempts to contact will be made.

Verification of the telephone number and address information will be made using

telephone and crisscross directories. If an individual can not be reached by telephone, although their address is known, attempts may be made to visit the subject, obtain informed consent, and schedule the next study component.

All families will be compensated for the effort it takes to complete the study and costs incurred by the participant. Please see Section 9 (*Payments to Respondents*) for a detailed description of the compensation schedule.

Families may refuse participation in one component of the study and still be considered a study participant. We hope to enhance overall participation by not insisting that the family consent to all components of the study; although all participants will be encouraged to complete the entire study protocol. Also, study staff will be trained in approaches to alleviate concerns about participation in order to increase the likelihood of obtaining consent.

The National CADDRE Study will also implement extensive public relation campaigns in each of the study catchment areas. These campaigns may consist of advertisements via television, radio, magazines, physician referrals, advocacy and support group meetings, and other media that are appropriate to the potential study population.

In the CHARGE Study, recruitment rates for control families with other developmental disabilities were significantly lower than those for the case families and the general population control families. In response, CHARGE created a second recruitment

brochure that was less autism-centric and emphasized the study as looking at child development. These changes increased recruitment rates among non-case families (Culp, Personal Communication, April 6, 2007). CADDRE has used the CHARGE experience to develop a recruitment brochure that will be attractive to non-case families.

As described above, our population-based approach to subject ascertainment includes multiple clinical and education sources in each study area to identify potential CASE and NIC participants and birth certificates to identify potential subcohort participants. Some sites will have access to demographic or address information of individuals in the pool of potential participants (e.g., from clinic or school lists or birth certificate information) prior to sending the invitation letter. Thus, these sites will be able to assess differences in respondents and non-respondents to the invitation letter. In most cases, however, the sites will not have access to this information since they are not permitted access to personal information prior to positive response. In such instances, the sites are negotiating with the sources (clinic, school, or vital records office) to provide aggregate data on the pool of potential participants from that source so that the site can assess possible differences between the positive responses and the total pool of potential participants. The sites' ability to get these aggregate data, however, will vary from source to source. Consequently, we do not expect that we will be able to systematically adjust for non-response to the invitation letter.

B.4. Tests of Procedures or Methods to be Undertaken

The maternal interview (Appendix B) was adapted from a questionnaire used for the

National Birth Defects Prevention Study (OMB # 0920-0010, expiration 5/31/2009). In the process of modifying the questionnaire for the National CADDRE Study, some individual questions were adopted without any changes, but all sections were modified in some way to make the questionnaire more appropriate for this particular protocol. Additionally, investigators pilot tested the study instruments and materials, including recruitment methodology and materials, medical record ascertainment and abstraction, laboratory methods, self-administered questionnaires, and the caregiver interview.

B.5. Individuals Consulted on Statistical Aspects and Individuals Collection and/or Analyzing Data

The following individuals will act as statistical consultants on the project.

Principal investigators at each study site:

- Diana Schendel, Ph.D./Centers for Disease Control and Prevention, Atlanta, GA 404.498.3845, dcs6@cdc.gov
- Lisa Miller, M.D./Colorado Department of Public Health and Environment 303.692.2663, lisa.miller@state.co.us
- Lisa Croen, Ph.D., Kaiser Permanente Division of Research, 510.891.3463, lisa.a.croen@kp.org
- Dani Fallin, Ph.D., Johns Hopkins University, 410-955-3463, dfallin@jhsph.eduJennifer Martin, Ph.D./University of Pennsylvania 215.898.4726, pinto@nursing.upenn.edu

- Julie Daniels, Ph.D./University of North Carolina at Chapel Hill 919.966.7096, Julie_daniels@unc.edu

Co-investigators at each study site:

- Judy Grether, Ph.D./California Department of Health Services, Environmental Health Investigations Branch 510.620.3626, jgrether@dhs.ca.gov
- Craig Newshaffer, Ph.D./Johns Hopkins University 410. 955.1268, cnewscha@jhsph.edu
- Rebecca Landa, Ph.D., Kennedy Krieger Institute, 443.552.1000, landa@kennedykrieger.org
- Susan Levy, M.D., Children's Hospital of Philadelphia, 215.590.7528, levys@email.chop.edu
- Cordelia Robinson, Ph.D., JFK Partners/UCHSC, 303.864.5261, robinson.cordelia@tchden.org
- Laura Schieve, Ph.D./Centers for Disease Control and Prevention, Atlanta, GA 404-498-3888, ljs9@cdc.gov

Other CDC personnel that act as consultants for the GA CADDRE Study include:

- Marshalyn Yeargin-Allsopp, M.D./Centers for Disease Control and Prevention, Atlanta, GA 404-498-3842, mxy1@cdc.gov
- Catherine Rice, Ph.D./Centers for Disease Control and Prevention, Atlanta, GA 404-498-3847, cqr8@cdc.gov

Each of the principal investigators and co-investigators helped design the overall study and data collection plan. These investigators have extensive experience in epidemiologic research concerning adverse reproductive and developmental outcomes, especially developmental disabilities, and have authored or co-authored numerous peer-reviewed publications in the epidemiology of developmental disabilities.

DCC personnel will also act as statistical consultants. It is expected that these personnel will provide the necessary expertise in order to accomplish the DCC goals outlined in Section 3 (*Use of Improved Information Technology*).

Note Regarding Title of ICR:

The title of this ICR has changed from the National CADDRE Study: Child Development and Autism to the Study to Explore Early Development (SEED). Pending OMB approval, CDC will submit a change request to implement this change.

References

Achenbach TM, Ruffle TM (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. Pediatric Review Aug;21(8):265-71.

Achenbach, T. M., & Rescorla, L. A. (2000). Manual for ASEBA Preschool Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Ackerman, T. F. (1989). An ethical framework for the practice of paying research subjects. IRB, 11(4), 1-4.

Bacchelli E, Maestrini E. 2006. Autism spectrum disorders: Molecular genetic advances. Am. J. Med. Genet. C. Semin. Med. Genet. 142 : 13-23

Barlow, W. E., Ichikawa, L., Rosner, D., & Izumi, S. (1999). Analysis of case-cohort designs. Journal of Clinical Epidemiology, 52(12), 1165-1172.

Burd, L, Severud, R, Kerbeshian, J, Klug, MG. Prenatal and perinatal risk factors for autism. Journal of Perinatal Medicine 1999; 27: 441-450.

Croen, L., Grether, J., Hoogstrate, J., Selvin, S. (2002). The changing prevalence of autism in California. Journal of Autism and Developmental Disorders, 32 (3), 207-215.

Culp, L. & Croen, L. (Personal Communication, November 27, 2006).

Culp, L. & Croen, L. (Personal Communication, April 6, 2007).

Dickert, N. & Grady, C. (1999). What's the price of a research subject? Approaches to payment for research participation. New England Journal of Medicine, 341(3), 198-203.

Dunn, LB., Gordon, NE. (2005). Improving Informed Consent and Enhancing Recruitment for Research by Understanding Economic Behavior. JAMA, 293: 609-612.

Garcia-Closas, M., Egan, K. M., Abruzzo, J., Newcomb, P. A., Titus-Ernstoff, L., Franklin, T., Bender, P. K., Beck, J. C., Le Marchand, L., Lum, A., Alavanja, M., Hayes, R. B., Rutter, J., Buetow, K., Brinton, L. A., & Rothman, N. (2001). Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiology, Biomarkers and Prevention, 10(6), 687-696.

Glasson EJ, Bower C, Petterson B, de Klerk N, Chaney G, Hallmayer JF. Perinatal factors and the development of autism. Arch Gen Psychiatry 2004; 61:618-627.

Groves, RM. (2006). Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70: 646-675.

Hansen, E., Sullivan, N, Thurm, A. Ware, J., and Lord C. (2002) Social Communication Questionnaire (SCQ). International Meeting for Autism Research. Orlando, FL.

Harlow, S. & Linet, M. (1989) Agreement between questionnaire data and medical records: The evidence for accuracy of recall. American Journal of Epidemiology, 129(2): 233-248.

Hultman CM, Sparen P, Cnattingius S. Perinatal risk factors for infantile autism. Epidemiology 2002; 13: 417-422.

Juul-Dam N, Townsend J, Courchesne E. Prenatal, perinatal, and neonatal factors in autism, pervasive developmental disorder-not otherwise specified, and the general population. Pediatrics 2001; 107.

Kaaks, R., van der Tweel, I., van Noord, P. A., & Riboli, E. (1994). Efficient use of biological banks for biochemical epidemiology: Exploratory hypothesis testing by means of a sequential t-test. Epidemiology, 5(4), 429-438

Klauck SM. 2006. Genetics of autism spectrum disorder. Eur. J. Hum. Genet. 14 : 714-20

Larsson HJ, Eaton WW, Madsen KM, Vestergaard M, Olesen AV, et al. 2005. Risk factors for autism: Perinatal factors, parental psychiatric history, and socioeconomic status. Am. J. Epidemiol. 161 : 916-25

Lord C.,& Risi S. (unpublished). Unpublished draft of the Diagnostic Criteria for CPEA.

Lord, C., Risi, S., Lambrecht, L., Cook, E., Leventhal, B., DiLavore, P., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders, 30, 205-223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders, 24, 659-685.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. Journal of Autism and Developmental Disorders, 19, 185.

Macklin, R. (1981). “Due” and “undue” inducements: On paying money to research subjects. IRB, 3, 1-6.

McGee, G. (1997). Subject to payment? Journal of the American Medical Association, 278, 199-200.

Newschaffer, Craig & Lee, Nora. Personal Communication, December, 2003.

Newschaffer CJ, for the CADDRE Investigators. The epidemiology of autism spectrum disorders. Annual Review of Public Health (in press)

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika, 73, 1-11.

Rice, C., Schendel, D., Cunniff, C., & Doernberg N. (2003). Public health monitoring of developmental disabilities with a focus on the autism spectrum disorders. American Journal of Medical Genetics (Sem in Med Gen 00:1-6).

Rockenbauer, M., Olsen, J., Czeizel, A.E., Pedersen, L., Sorensen, H.T., & EuroMAP Group. (2002) Recall bias in a case-control surveillance system on the use of medicine during pregnancy. Epidemiology,12: 461-466.

Rutter M. Aetiology of autism: findings and questions. Journal of Intellectual Disability Research 2005; 49:231-238.

Sato, T. (1992). Estimation of a common risk ratio in stratified case-cohort studies. Statistics in Medicine, 11, 1599-1605.

Singer, E., Bossarte, RM. (2006). Incentives for survey participation: are they coercive?. Am J Prev Med, 31: 411-418.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. Journal of Official Statistics, 15 (2): 217-230.

Tilley, B., Barnes, A., Bergstrahl, E., Labarthe, D., Noller, K., Colton, T., Adam, E. (1985) A comparison of pregnancy history recall and medical records: Implications for retrospective studies. American Journal of Epidemiology, 121 (2); 269-281

Tomeo, C., Rich-Edwards, J.W., Michels, K., Berkey, C., Hunter, D.J., Frazier, A., Willet, W., Buka, S. (1999) Reproducibility and validity of maternal recall of pregnancy-related events. Epidemiology, 10 (6): 774-777.

Wacholder, S., & Boivin, J. F. (1987). External comparisons with the case-cohort design. American Journal of Epidemiology, 126, 1198-1209.

Weinberg, C. R., Wilcox, A. J., Lie, R. T. (1998). A log-linear approach to case-parent triad data: Assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. American Journal of Human Genetics, 62, 969-978.

Yu J, et al. (1983). A quantitative review of research design effects on response rates

to questionnaires. Journal of Marketing Research, 20:36-44.