2. CONCEPTUAL DESIGN AND FRAMEWORK

As dynamic and developing entities, children are exposed to an array of chemical, physical, and psychosocial environmental factors beginning in utero that affect their health, growth, and development and predispose them to later disease. Parental factors beginning prior to conception, and exposures in utero that continue through childhood, interact with the inherent genetic potential of the child to determine ultimate health. Most health outcomes are not the result of a single environmental exposure, an inherent genetic predisposition, or the interaction of a single environmental exposure and genetic factors. Rather, health outcomes are a complex amalgam of multiple environmental exposures over time that affect the inherent genetic makeup of each person.

2.1 Hypothesis Formulation and Study Design

The NCS is designed to respond to the significant challenge of delineating associations between single and multiple exposures over a long period of time, genetic factors, and health outcomes in children. The previous sections provide brief examples of hypothesized relations between various environmental exposures and major diseases or conditions of children, each having its own empirical and/or theoretical basis. To identify and confirm any one of these relations would best be accomplished with a prospective longitudinal study that incorporates the following criteria for design and data collection:

- Assessments of multiple exposures and multiple outcomes.
- Prospective and high quality data collection to decrease bias.
- Measurement of relatively rare outcomes (e.g., autism spectrum disorders, type 2 diabetes).
- Coverage of a long enough portion of the lifespan to measure and link early exposures with later outcomes.
- Repeated measures to capture the relation between exposures during critical time periods and trajectories of development.
- Generalizability to the U.S. population.

The NCS seeks to study the exposure-outcome relations for multiple exposures at earlier life stages and multiple outcomes related to the same or different exposures at later stages. This can only be accomplished through the use of a single large cohort with multiple exposure and outcome measures. This approach is more efficient and economical than multiple separate cohort studies and less biased than multiple retrospective case control studies.

The prospective cohort design is well suited to look at the multiple outcomes associated with a single exposure or set of exposures (Manolio, Bailey-Wilson, & Collins, 2006). For example, the NCS offers the opportunity to examine the potential impact of certain endocrine active compound exposures early in life on neurocognitive development and sexual maturation later in life (Cooper, Goldman, & Tyrey, 1998; Jahnke, Choksi, Moore, & Shelby, 2004; Landrigan, Kimmel, Correa, & Eskenazi, 2004;

Longnecker et al., 2003), thus linking these disparate outcomes to a single exposure or class of exposures and potentially identifying common underlying mechanisms.

Moreover, the breadth of measurements in the NCS will allow examination of the combined and independent contributions of multiple exposures on a single outcome. For example, exposure to certain pesticides, plasticizers, heavy metals, early life exposure to media, different parenting styles, and genetic predispositions are all hypothesized to have effects on cognitive development. Careful prospective collection of data across multiple exposure domains will allow assessments reflective of true exposure patterns of U.S. children. The general conceptual model for exposure-outcome relations in this large longitudinal study, including direct relations, mediated relations, and gene-environment and environment-environment interactions, is illustrated in Figure 2-1.

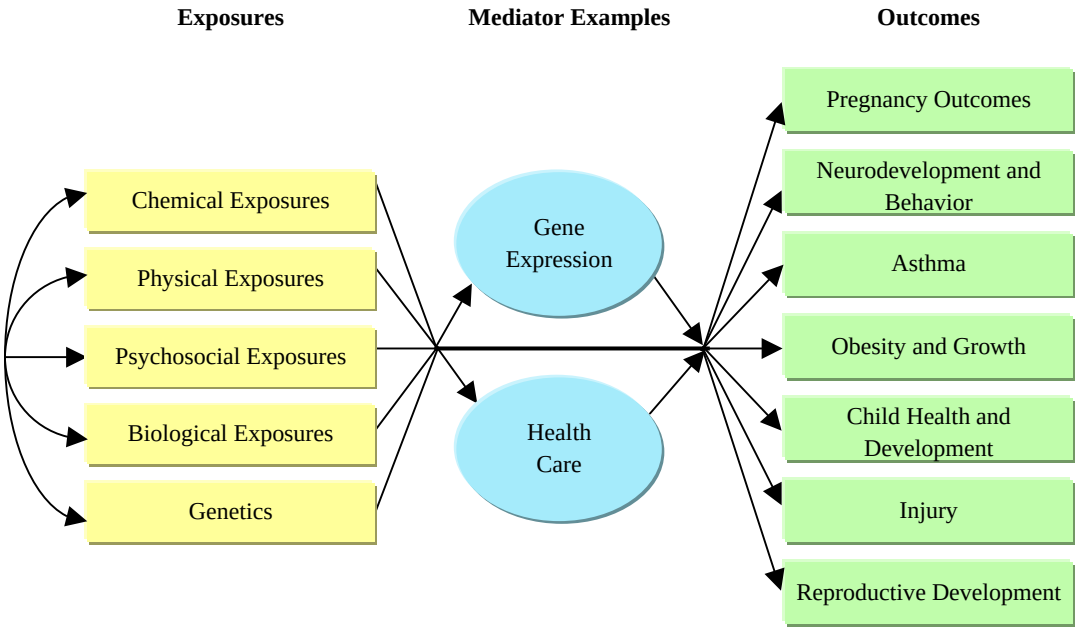

Figure 2-1. Conceptual Model of Exposures, Their Interactions, Examples of Mediators, and Outcomes

During the past several decades increasing numbers of reports have shown how different exposures interact with each other and/or with genetic factors to affect outcomes not seen with a single exposure or genetic variation alone. For example, the now classic study by Caspi (Caspi et al., 2002) replicated by Foley (Foley et al., 2004) demonstrated that maltreated children whose genotype conferred low levels of monoamine oxidase A (MAO A) expression more often developed conduct disorder, antisocial personality, and adult violent crime than children with a high-activity MAO A genotype. Similarly, Wang et al. (2002) reported that maternal CYP1A1 and GSTT1 genotypes modified the association between maternal smoking and infant birth weights. In another example, Berkowitz et al. (2004) found that mothers with both a low paraoxinase polymorphism and maternal elevation of metabolites of the pesticide chlorpyrophos had infants who had a small but significant reduction in head circumference. Other examples of environment-environment interactions include allergen-air pollutants (chemical-biological) with regard to induction of asthma (Diaz-Sanchez et al., 2006; Platts-Mills, Vaughan, Squillace, Woodfolk, & Sporik., 2001) and mercury exposure and socioeconomic

characteristics of child caregivers (chemical-psychosocial) in relation to cognitive development (Davidson, Myers, Shamlaye, Cox, & Wilding, 2004).

These findings from relatively small samples suggest causal relationships that require larger samples for replication and conclusive results. Analyses of gene-environment and environment-environment interactions require identifying subgroups and conducting subgroup analyses that are only possible with a large original sample. The sample sizes required for such analyses are described with specific hypotheses (see Appendix A-2), but, in general, samples approaching the proposed 100,000 are required. Unraveling the increasingly recognized complexities of environment-environment and gene-environment interactions offers great potential for targeted interventions but also requires increasingly complex and demanding subgroup analyses. Underlying this enormous potential is a requirement that these interacting factors must be measured in the same individuals who are followed prospectively for the outcomes of interest at later life stages. As long as the key measures for the various classes of exposure are obtained in the same individuals of the cohort, the opportunities exist to examine a variety of moderating processes, from how socioeconomic factors interact with chemical exposure (Agyeman, 2005; Anderton, Oakes, Fraser, & Anderson, 1994) to how parenting styles and maternal nurturing interact with specific neurotransmitter polymorphisms to affect major behavioral outcomes (Suomi, 2004; Champoux et al., 2002).

2.2 Significance of the Longitudinal Database as a Platform for Future Studies

As a longitudinal cohort study of considerable size and complexity, the NCS will provide answers to many current hypotheses. Because biologic and environmental samples and the extensive database will be stored and available in the future, however, it will also constitute a significant national database and resource to answer many questions not yet conceived.

2.2.1 Development of Future Hypotheses

A listing of the publications from the largest longitudinal study of child health, the Collaborative Perinatal Project (CPP) conducted during 1963 to 1989, reveals 611 separate publications. The topics and contributions of these publications vary widely from the intended original aim of the CPP, which was to identify the relation between neonatal asphyxia and cerebral palsy. With more extensive exposure data, a larger sample size, and a longer follow-up period than the CPP, the NCS provides even greater potential for opportunities to investigate questions regarding the health and development of children. This capacity for research and analyses beyond the stated aims of the Study constitutes a major contribution and impetus for undertaking such a project. Because of this potential, it is important to collect the samples and measurements in ways that optimize opportunities for future testing and analysis.

2.2.2 Health Disparities

The NCS will be able to address many major health disparities that currently exist in the United States, and to collect sufficient data to address others not yet recognized. The Children's Health Act of 2000 specifically directed that the NCS "consider health disparities among children ... ." Several major features of the NCS will allow the Study to address health disparities, to advance the state of knowledge about disparities, and to provide information that can guide policy and practice to reduce and eliminate disparities.

The representative, probability-based sampling approach ensures the Study sample will reflect the broad racial and ethnic diversity of the United States. The full NCS sample will consist of approximately 78,000 white, 19,000 Hispanic, 15,000 Black, 5,000 Asian, and 2,000 American Indian participants. Also, approximately 20 percent of the cohort will be from rural areas. This makes the NCS the most comprehensive, long-term study of this size for Hispanic children, Black children, and children from rural settings.

The size and diversity of the cohort will allow the NCS to generate large amounts of data and better characterize the disparities among subpopulations, including uncovering more subtle disparities than previously recognized. Disparities of interest to the NCS cut across a variety of the Study exposures and outcomes, from differences in the prevalence of preterm birth to variations in exposure to pesticides across communities to differences in the types of injuries children suffer.

Future researchers may find even more topics that prove timely and important. The wide range of study hypotheses that can be addressed using the NCS data also translate into a wide number of disparities that can be examined using the large data set. For many possible study questions, adequate power exists to perform analyses at the subgroup level.

2.2.3 Case-Control Studies

Many of the Study hypotheses can be addressed effectively and most efficiently through nested case-control studies. The prospective collection and careful storage of all Study data (biospecimens, environmental samples, stored images, etc.) will allow investigators to limit many of the expensive analyses to smaller subsets of identified cases and their matched controls.

Perhaps most importantly, this approach will allow the investigation of hypotheses formulated in the future. Some of the greatest values of this Study are the establishment of a databank of longitudinal measures and a repository of both environmental and biologic specimens that will allow future investigators to address important questions of clinical and public health relevance.