March 19, 2004

White Paper

On

Evaluation of Sampling Design Options for the National Children's Study

by

Warren Strauss, Jeff Lehman, John Menkedick, Louise Ryan[1],
Tim Pivetz, Nancy McMillan, Ben Pierce, and Steve Rust

BATTELLE
505 King Avenue
Columbus, OH 43201-2693

With Critical Review and Input from
Colm O'Muircheartaigh[2] and Alan Zaslavsky[3]

Contract No. 282-98-0019

Dr. Peter Scheidt
Project Officer

Dr. Jim Quackenboss
Work Assignment Officer

National Children's Study Program Office
National Institute for Child Health and Human Development
6100 Executive Blvd - 5C01
Rockville, MD 20892

[1] Louise Ryan is a Professor of Biostatistics at the Harvard University School of Public Health and Dana Farber Cancer Institute.
[2] Colm O'Muircheartaigh is a Professor at the Harris Graduate School of Public Policy Studies at the University of Chicago and Vice President for Statistics and Methodology at NORC.
[3] Alan Zaslavsky is a Professor of Statistics in the Department of Health Care Policy at Harvard Medical School.

DISCLAIMER

ACKNOWLEDGEMENTS

The authors of this report wish to acknowledge the contributions of various people from different organizations who contributed to the design effort. From the NICHD Program Office, we wish to thank Peter Scheidt and Jim Quackenboss for their technical guidance and assistance throughout the project. We also wish to thank Mark Klebanoff (NIH), Pauline Mendola (US EPA), and Frank Speizer (Harvard) for their input and guidance. From Battelle, we wish to thank Vince Brown, Cody Hamilton, Pamela Hartford, Nicole Iroz-Elardo, Jyothi Nagaraja, Nancy Reiches, Rachel Sell, Lorraine Sinnott, Shannon Stetzer, and Amy Thomas for their technical contributions, and Allison Marshall for her secretarial support. Finally, we wish to acknowledge the input that the NCS Sampling Design Workshop Committee had on this effort, whose members represent the NCS Advisory Committee (Robert Michael, Univ. of Chicago; Barry Ryan, Emory), Study Design Working Group (Jonas Ellenberg, Westat; Frank Speizer, Harvard), Interagency Coordinating Committee (Jim Quackenboss and Sherry Selevan, US EPA), and the NCS Program Office (Peter Scheidt, NICHD; Lester Curtin, NCHS).

Table of Contents

Glossary of Terms

Analytic Study: a study in which action will be taken on a process or cause-and-effect system with the aim of improving future conditions.

Attrition: typically refers to the case where a member of a longitudinal study drops out of the study.

CDC: the U.S. Centers for Disease Control and Prevention.

Centers: a purposively selected medical center capable of performing data collection activities for the NCS – most likely selected through a competitive Federal procurement process.

Certainty Strata: any subset of the study population that can be enumerated which is selected with certainty (weight=1) in a multistage probability-based sampling approach.

Cohort: a group of subjects that are studied over a period of time as part of a scientific investigation.

Confounding: occurs when two factors are associated with each other or "travel together" and the effect of one is confused with or distorted by the effect of the other.

Core Hypotheses: a series of specific research hypotheses deemed by the ICC as sufficient to support the determination of sample size and design for the NCS and essential to assure that specific research questions can be addressed by the study.

Contract research organization: any organization that may be hired through competitive bids (e.g., universities, nonprofit organizations, hospitals, commercial research corporations, etc.) to perform a scope of work for the NCS.

Convenience sampling: a nonprobability sampling approach that selects members based on convenience.

Covariate: a variable that is related to, or has influence on, an outcome of interest.

Cluster Sampling: a method of sampling in which, at some stage, elements (e.g., children) are selected from the population in groups or clusters. In multistage cluster sampling, a sample of elements within a selected cluster may be taken during a subsequent stage of sampling.

Design Effect: a measure of the information loss due to the selected design. Typically defined as the ratio of the parameter estimate variance under a specified sampling design to the parameter estimate variance under a simple random sample.

Design Variables: the set of variables required to implement a probability-based sampling process, including stratification variables and any variables used to calculate probabilities of inclusion.

Effect Modifier: a variable that interacts with a risk factor so that a different association between the risk factor and the outcome of interest is apparent for different values of the effect modifier.

Enumerative Study: a study in which action will be taken on the elements in the frame studied where the term frame is used to refer to an aggregation of identifiable units, any of which may be studied.

External Validity: relationships identified in a study are considered to be externally valid if they are valid for the reference population associated with the study.

EPA: the U.S. Environmental Protection Agency.

Exposure: in this work, exposure is broadly defined as physical, chemical, biological, and/or psychosocial influences that may be related to adverse health outcomes.

GEE: Generalized Estimating Equations – a statistical modeling approach that allows for analysis of correlated data under the conceptual framework of generalized linear models (such as logistic regression models).

Generalize: refers to the ability to draw general conclusions relevant to some population (e.g., apply conclusions to the reference population).

ICC: the interagency coordinating committee – Investigators from each of the four lead agencies (NICHD, CDC, EPA and NIEHS) serve on an Interagency Coordinating Committee (ICC) that is charged with leading the planning and implementation of the NCS.

Inference: a conclusion drawn from evidence.

Internal Validity: relationships are considered to be internally valid if they are statistically significant for the study sample, if the effects of extraneous variables, plausible confounders, and plausible effect modifiers have been properly accounted for, and if hypothesized causal factors precede the effect.

Logistic Regression Model: a statistical analysis method used to model binary or binomial response variables. Parameter estimates from logistic regression models carry log-odds-ratio interpretation.

Model-based Analysis: refers to an inference procedure that implicitly assumes the sampling mechanism does not depend on the survey outcomes.

MSA: Metropolitan Statistical Area.

Multistage Sampling: multistage sampling methods allow selection of groups of elements from the sampling frame at one stage and then subsequent sampling from the selected groups of elements at a subsequent stage.

NCS Cohort: the *study sample* for the National Children's Study.

NCSAC: National Children's Study Advisory Committee (NCSAC), chartered under the Federal Advisory Committee Act, serves as the formal mechanism for providing advice and recommendations to the ICC.

NIEHS: the National Institute for Environmental Health Sciences.

NICHD: the National Institute for Child Health and Human Development.

Non-coverage: refers to the inability to completely identify or enumerate the reference population.

Nonprobability sampling: sampling from the population in some nonrandom manner (i.e., not all members of the population have a known non-zero probability of selection).

Nonresponse: occurs when a member of the population is selected as part of the sample, but, for whatever reason, does not become a participating member of the sample (e.g., a selected person refuses to participate in the study).

NPBS: National Probability-Based Sample.

Odds Ratio: a statistical measure of association. In the context of the design work presented in this report, it is a measure of the relationship between an adverse health effect and a binary measure of exposure. Specifically, it assesses the odds of disease among exposed individuals divided by the odds of disease among unexposed individuals.

Population of interest: could also be called the reference population or the target population (i.e., the population of subjects or units that are the target of the investigation). Typically, inference and/or conclusions are targeted at the population of interest.

Power: probability of correctly concluding that there is an effect when an effect of specified size is present.

Power Studies: studies involving calculation of power under different scenarios.

Probability-Based Random Sampling: a probability-based sampling method for which each element has a probability of being included in the target sample that is strictly greater than zero and strictly less than one, and that uses a random procedure to select elements into the *target sample* according to these probabilities.

Probability-Based Sampling: a method for selecting a *target sample* from a *sampling frame* in which the probability of occurrence for each and every possible *study sample* is a function of a set of *design variables*; an important property of a probability-based sampling process is that the probability of inclusion in the *target sample* is known for each and every element (e.g., child) in the *sampling frame*.

Proportional to Size Sampling: sampling of units with probabilities proportional to the unit size.

PSU: Primary Sampling Unit.

Purposive sampling: nonprobability sampling with some purpose in mind (e.g., purposely sampling a portion of the population that has previously been representative of the population).

Quota Sampling: a method of sampling in which certain characteristics of potential study participants are measured and participants are included in the *study sample* in such a manner as to obtain pre-determined numbers of participants in specified classes defined by values of the measured characteristics.

Recruitment Rate: the ratio of the number of subjects initially enrolled in the NCS cohort divided by the number of subjects for which a recruitment attempt is made.

Reference Population: the population about which valid inferences are desired and to which study inferences will be extrapolated in one form or another.

Representative: used in the context of a representative sample and generally meaning that the sample is "similar to" the population from which it is selected.

Response Rate: the ratio of the number of cohort members providing sufficient data for a particular line of inquiry divided by the number of cohort members for which an attempt is made to collect such data.

Retention Rate: the ratio of the number of actively enrolled cohort members at a given point during the data collection phase of a study divided by the number of cohort members initially enrolled.

Sampling Frame: that portion of the *study population* that has a positive probability of being included in the *target sample*; in practice, the sampling frame is constructed to be as close to the *study population* as possible subject to the requirements that (1) the sampling frame can be fully enumerated and (2) *design variable* values are available for each element of the sampling frame.

Sampling unit: refers to the elements or units that are to be sampled.

Sample Weights: refers to the number of elements/units that are represented by the observation and is typically defined as the inverse of the sampling probability.

Selection bias: a systematic tendency on the part of the sampling procedure to exclude or include one (or more) type(s) of study subjects from the sample.

Simple Random Sampling: simple random sampling methods select the target sample from the sampling frame in a totally random fashion without replacement.

Stratified Sampling: stratified random sampling methods control the subsample sizes for subsets (strata) of the sampling frame defined by one or more design variables.

Study Population: the population of elements that would be included in the *sampling frame* if full enumeration of the sampling frame and values for the design variables were not required.

Study Sample: all elements of the study population that are successfully recruited into the study, are successfully retained as study participants, and produce the required study data.

Target Sample: those elements of the study population for which a recruitment attempt is made; the target sample is the union of the study sample, the set of recruitment failures, the set of retention failures, and the set of retained study participants that fail to produce the required data.

Validation sample: a small sample that is designed to provide information related to the bias or error introduced into the main cohort by nature of the design. The information gathered from the validation sample is designed to allow for appropriate statistical adjustments to the data collected in the larger cohort to address bias and error.

Weighted Analysis: an analysis procedure that appropriately accounts for the sampling weights assigned to each observation.

1 INTRODUCTION

The National Children's Study (NCS) will study the complex relationship between health and the environment for approximately 100,000 U.S. children and their families. Enrollment will begin before birth and follow-up will continue for at least 21 years. Planning for the NCS was initiated by the President's Task Force on Environmental Health Risks and Safety Risks to Children, which was established in 1997. The Task Force was charged with developing strategies to reduce or eliminate adverse effects on children caused by environmental exposures. However, the Task Force soon recognized that such strategies required a much clearer understanding of risk factors, and therefore proposed a longitudinal cohort study of the effects of environmental exposure on the health and development of children (Branum et al., 2002). Title X of the Children's Health Act of 2000 subsequently authorized the National Institute of Child Health and Human Development (NICHD), in collaboration with the Centers for Disease Control and Prevention (CDC), the U.S. Environmental Protection Agency (EPA), and other appropriate Federal agencies, to plan, develop and implement the study.

1.1 NCS SCOPE, OBJECTIVES, GUIDING PRINCIPLES, AND GIVENS

The language in the Children's Health Act of 2000 (Title X, Section 1004) calls for "a national longitudinal study of environmental influences (including physical, chemical, biological, and psychosocial) on children's health and development." The additional direction in the legislation is sparse but critically important. It calls upon the Director of NICHD to "establish a consortium of representatives from appropriate Federal agencies (including the Centers for Disease Control and Prevention, and the Environmental Protection Agency) to (as quoted in subsection (b) of Section 1004):

(1) plan, develop and implement a prospective cohort study, from birth to adulthood, to evaluate the effects of both chronic and intermittent exposures on child health and human development; and

(2) investigate basic mechanisms of developmental disorders and environmental factors, both risk and protective, that influence health and developmental processes.

Finally, the legislation requires that the study shall (as quoted in subsection (c) of Section 1004):

(1) incorporate behavioral, emotional, educational, and contextual consequences to enable a complete assessment of the physical, chemical, biological and psychosocial environmental influences on children's well-being;

(2) gather data on environmental influences and outcomes on diverse populations of children, which may include the consideration of prenatal exposures; and

(3) consider health disparities among children which may include the consideration of prenatal exposures."

The five legislative statements quoted above provide the overall objectives for the NCS.

The legislation and its requirements and their interpretation by the responsible government agencies lead to a set of basic requirements or assumptions for the NCS, which have been referred to as "givens" for NCS sampling designs in the past by government study leaders. These include:

(a) The study will be observational in nature and will address multiple environmental influences.

(b) The study will be national in scope, but not necessarily nationally representative. The sample should be broad-based, inclusive of a wide range of populations and geographic diversity, and as representative as possible given tradeoffs with other features of scientific value to the study objectives. The primary purpose of the study is to investigate exposure-response relationships, not to provide estimates of disease and exposure incidence and prevalence.

(c) The study will include a large sample (approximately 100,000) – to allow for evaluation of rare exposures and outcomes; and of interaction of environmental factors and genetics.

(d) The study will include prenatal recruitment, as early in pregnancy as possible.

(e) The study will include clustering of samples to allow for efficient collection of exposure and outcome measures, and measurement of context (physical and social).

(f) The study will consider stratification to obtain a) an adequate range of exposures (including social), b) socioeconomic, racial/ethnic/geographic diversity, and c) population subgroups of interest.

(g) The study will have locality-based aspects to encourage community engagement.

(h) The study will include infrastructure to support specialized measures (e.g., medical facilities with technologies such as 3D ultrasound).

(i) The study will provide access/collection of appropriate specialized measures or biological samples during pregnancy and birth, for example, placenta or cord blood samples from the delivery room.

(j) The study will provide flexibility to conduct special studies (e.g., special population groups, preconception recruitment, or topics of community interest).

The distinguishing features of the NCS – what makes the study an unusual if not unique research opportunity – are its size (100,000 children), its duration (prenatal, and most likely for a subgroup, peri-conceptional, to adulthood) and its comprehensive charge to assess multiple effects on diverse populations. The legislative requirements that translate to study objectives, and the "givens" stated in terms (a) – (j) above provide the overall boundaries and the guiding principles for the study design. Within these boundaries the overarching goals of the NCS articulated by the Interagency Coordinating Committee (ICC) are to:

- Identify the presence or absence of adverse effects from environmental exposures of concern to development
- Identify possible causal environmental factors for various conditions and developmental and health problems in children and adults
- Provide valuable resources for additional, future studies of health and environment.

1.2 ORGANIZATION AND GOVERNANCE

The Children's Health Act stipulated that the study be carried out with participation of the multiple Federal agencies concerned with children's environmental exposures and possible outcomes. Since fiscal year 2000, a number of interagency agreements have been put into place to carry out methods development studies, provide support services, and establish collaborations among the agencies (NCS Business Plan, 2002). In an effort to solidify this partnership, the four lead institutes and agencies (NICHD, NIEHS, CDC and EPA) signed a Memorandum of Understanding in February 2002. Investigators from each of these four lead entities serve on an Interagency Coordinating Committee (ICC) that is charged with leading the planning and implementation of the NCS, which is coordinated through an NCS program office established at NICHD. By legislative directive, the director of NICHD has overall responsibility and accountability for conduct of the study.

An NCS Advisory Committee (NCSAC), chartered under the Federal Advisory Committee Act, serves as the formal mechanism for providing advice and recommendations to the ICC. The NCSAC is supported by more than 20 Working Groups representing both Federal and private-sector scientists and other specialists focused on providing input on specific scientific questions and issues encountered in study design. In addition, any interested parties receive information on the study and provide input through regularly scheduled Assembly meetings. The overall structure of the NCS leadership is illustrated in Figure 1-1.

Additional information on the history of the evolution of the NCS is available in the following references: Branum et al. (2002), Children's Health Act (2002), NCS Business Plan (2002).

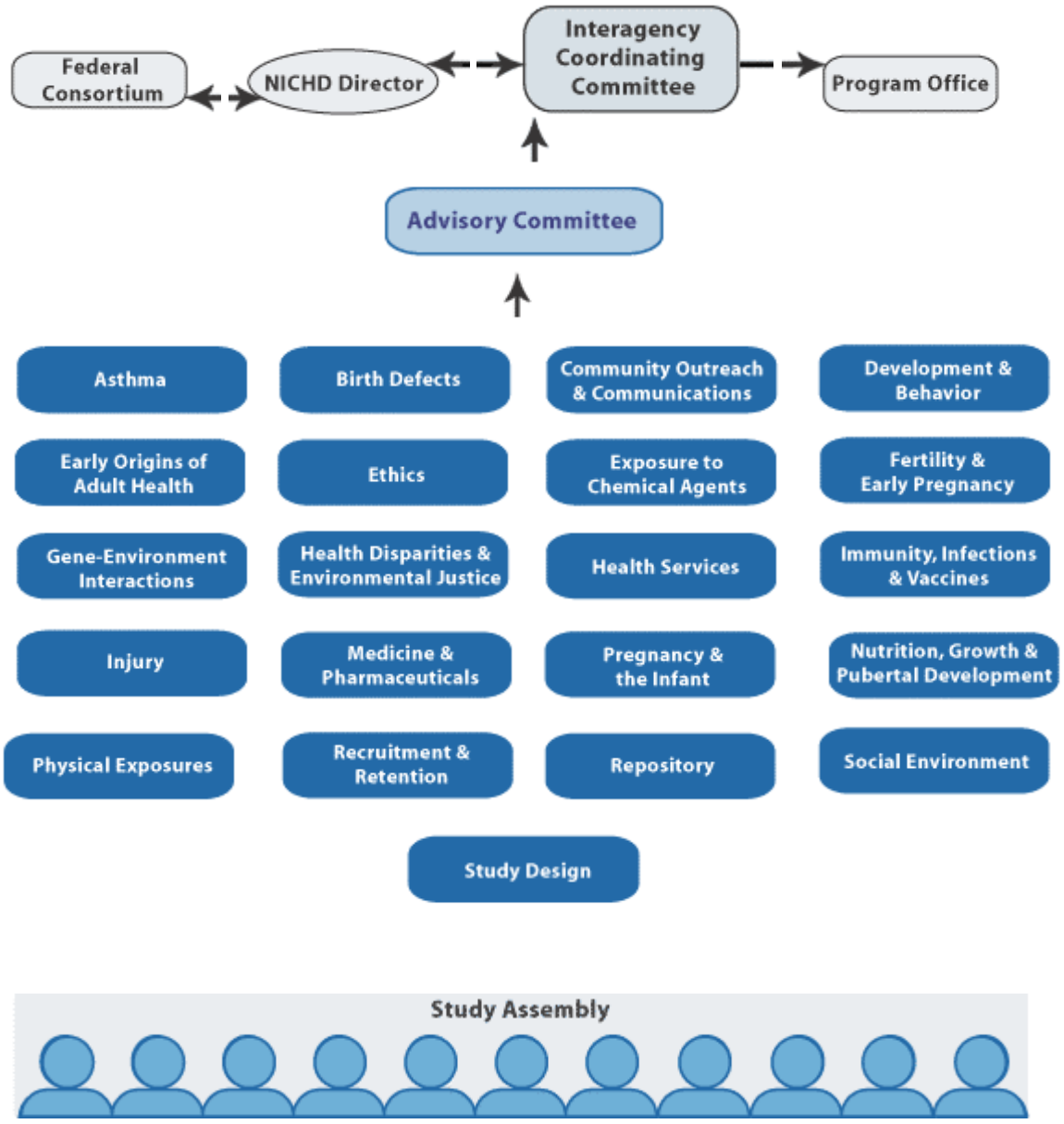


Figure 1-1. NCS Organizational Structure

1.3 SELECTION OF CORE HYPOTHESES

The italicized text that follows on selection of core hypotheses was taken from a November 25, 2003, document prepared by the ICC hypotheses subcommittee, and presented to the NCS Advisory Committee on December 15, 2003.

> *Hundreds of scientists and representatives from community groups and professional organizations have contributed to the identification of key children's environmental health questions. No single research question is of sufficient breadth or import to fulfill*

the entire mission of the NCS. The Study Design Working Group of the NCS Advisory Committee (NCSAC) proposed the development of core hypotheses encompassing exposures and child health outcomes of great public health significance requiring long-term follow-up and which cannot be reasonably studied with fewer children or a different study design. The set of research questions forming the foundation of the NCS must together provide: a rationale for a long-term, prospective study of approximately 100,000 children; the scientific framework to define the NCS, including sample design, data collection, etc.; as well as a "public identity" for the NCS.

The Interagency Coordinating Committee (ICC) has used the findings from 20 NCS working groups reported via the NCSAC, independent reviews of the children's environmental health literature, and comments from a broad-based Study Assembly to develop an initial set of these foundational, core hypotheses. These hypotheses are sufficient to support the determination of sample size and design for the NCS and are essential to assure that specific research questions can be addressed by the study.

However, a manageable set of core hypotheses cannot alone convey the true breadth of the NCS, nor do they, alone, assure the collection of data necessary to address the full range of topics to be covered by the NCS. The priority outcomes and exposures outlined below go further to convey the full scope of the NCS. Additional work is necessary to complete a study protocol that balances participant and family burden with data collection activities needed to address these important areas of children's environmental health.

Priority Outcomes. Based on the above criteria, the following child health areas have been identified as priorities for the NCS.

Pregnancy outcomes: Many pregnancy outcomes, including preterm delivery and birth defects, are plausibly related to environmental conditions and are understudied. These early life events can have profound impact on child health and development throughout life. These outcomes also provide a first set of results from the NCS right from the start.

Neurodevelopment and behavior: Assessment of child development and behavior is key to the mandate of the NCS. The NCS can address multiple environmental factors that are potentially associated with severe health concerns such as autism and schizophrenia, as well as more commonly occurring childhood disorders such as depression and learning disabilities. The NCS can also provide substantial data on variations in the course of normal child development and may provide insights into environmental factors related to aspects of development such as aggression, adjustment, achievement and resilience.

Injury: A focus area of the President's Task Force, injury is a major cause of childhood morbidity and mortality. The NCS expects to measure childhood injuries, particularly those that require hospitalization or other medical attention,

and to evaluate a variety of environmental factors including aspects of the social and physical environment that may be associated with injury.

Asthma: While there is a substantial body of research into environmental factors that can trigger asthma attacks or exacerbate existing asthma, there is a need to understand more about contributions the environment and gene-environment interactions have on the development of asthma. Because asthma is relatively common among U.S. children, the NCS will have enough statistical power to be able to examine various constellations of environmental and genetic factors that may be related to asthma incidence and exacerbation.

Obesity and physical development: The NCS will likely have sufficient statistical power to examine disorders of physical development related to diabetes, obesity and altered puberty. The longitudinal nature of the data and the ability to examine the interaction of multiple environmental factors with an individual's genetic composition is expected to provide insight not only into growth-related disorders, but also to provide a strong study of variations in growth, physical and reproductive development that may be affected by the environment.

Priority Exposures. The priority exposures listed below are outlined by influence (either beneficial or deleterious) on child health and development:

Physical environment: The NCS will measure aspects of the physical environment, including housing quality and neighborhood and community conditions that may relate to child health and development. In addition, the influence of physical factors such as radiation (electromagnetic, ultrasound, microwave, x-irradiation), light, and noise may be studied.

Chemical exposures: Exposure to chemical environmental contaminants generally occurs through human contact with air, water, soil, dust, food or industrial products. Pollutant exposures currently of concern in the NCS include metals, PCBs and dioxins, phthalates, organic and inorganic pesticides and herbicides. Exposure to many of these compounds, and their mixtures, at low background levels is ubiquitous. The NCS can investigate the potential health effects associated with these complex low-level exposures. Additionally, the NCS may select specific populations with unique exposure scenarios for special sub-studies of related health effects.

Biologic environment: The biologic environment includes exogenous factors (e.g., infectious agents, endotoxin, diet) and individual response to those factors (e.g., inflammatory response, glucose metabolism). In utero and early life exposures have potential implications for a wide range of health conditions including birth outcome, developmental outcomes, asthma, obesity, and cardiovascular disease. The NCS will allow for elucidation of those associations as well as physiologic mechanisms underlying those relationships, including the influence of genetic composition on those interactions.

Genetics: The NCS offers a unique opportunity to investigate the genetic component of many health outcomes. Although it is recognized that genetic factors play a role in many conditions, the mechanism behind the genetic contribution to specific diseases, such as autism, remains unknown. In addition, the quantitative contribution of genetics to more general conditions, such as obesity, is also unknown. A complete understanding of the effects of the environmental factors listed above requires elucidation of the interactions between these factors and genes, including the roles played by various polymorphisms in environmentally responsive genes and the effects of exposures on gene expression. The large sample size will allow for examination of the interaction between genetic make-up and chemical, biologic, and social exposures on many outcomes. The longitudinal and prospective nature of the NCS offers the possibility of examining the potential development of somatic mutations in relation to specific exposures. The current state of the science likely does not allow for the genetic profiling of study participants but will, at least initially, require a focus on suspect candidate genes. This will change as the study matures.

Psychosocial milieu: The NCS expects to assess many potential aspects of the psychosocial environment including: families and households; socioeconomic status; social networks and social support; neighborhoods and communities; formal institutions; and public policy. These factors have the potential to influence a child's health either directly or indirectly, by affecting exposure to the chemical or physical environment. The NCS will be able to examine those associations as well as shed light on the physiologic mechanisms underlying, for example, potential relationships between psychosocial stress and asthma or preterm birth. In addition to the putative influence on the health of an individual, social environmental factors may be an important area of consideration for investigation of health disparities.

Integrating Priority Outcomes And Exposures. Based on input from hundreds of experts, the ICC has proposed a set of core hypotheses to define a framework for study design. Though the current list of core hypotheses [see Table 6.1 in Chapter 6] *is still under debate, it is largely accepted as being adequate to move forward with development of a sampling design. It is expected that, over the long course of the study, new questions will emerge and be added to the study and some of the core hypotheses here may become outdated (ICC Hypotheses 2004).*

1.4 GUIDANCE FOR THE STUDY DESIGN

The ultimate purpose of the study design is to define all study specifications – cohort selection, measurement specification, and implementation details – in a manner that will best meet the overall objectives, requirements, and goals of the study described in Section 1.2 above. In addition, more specific guidance has been articulated (by government study leaders, the Study Design Working Group, and the NCSAC) that calls for the study design to:

- Emphasize hypotheses and science needs that require the unique longitudinal nature or sample size of the NCS;
- Go beyond characterization of associations, to provide understanding of the causal relationship between exposure and disease;
- Cover a sufficient range of exposures and outcomes to understand significant interactions;
- Be as representative as possible of the U.S. population, with relationships between exposure and disease able to be generalized to a broader population;
- Provide sufficient power to detect target associations of interest for selected core hypotheses;
- Provide a resource to test hypotheses to be identified in the future;
- Allow assessment, as possible, of populations at higher risk of exposures or outcomes;
- Be transparent, with assumptions, tradeoffs, and decisions well-documented; and
- Address ethical considerations, including cohort burden.

The difficulty (or challenge) in meeting the goals for the study design lies primarily in two areas – first, the fact that design choices must be made in the face of scientific and implementation uncertainties, and second, the fact that even with a good understanding of what might be expected there are tradeoffs between conflicting objectives.

The most notable example of the difficulty in meeting multiple goals for the study design is the ongoing difference in opinions over the feasibility and desirability of certain aspects of probability sampling. On the one side, many epidemiologists believe that a strict probability approach will result in fewer measurements, more attrition, and negative impact on the ability to measure exposures and outcomes sufficiently well to understand the etiology of disease. On the other hand, sampling statisticians and social scientists are concerned that the lack of probability sampling may introduce unknown biases into study results and leave the study with results that cannot be generalized to a broader population. This controversy arises first because there is uncertainty over the degree to which a probability sample will result in more attrition and less measurement in comparison to a convenience sample; and second because there are tradeoffs involved between maximizing a probability component to the sample and many other desirable features such as efficiency and accessibility of measurements, local community involvement, and support for major research institutions. The white paper on the Advantages and Limitations of Probability-Based Sampling for the National Children's Study included in Appendix A and the two white papers on Criteria and Design Options included in Appendices B1 and B2 provide more discussion on the specific tradeoffs and uncertainties associated with study design choices in the NCS.

It is for this reason that study leaders have convened the sampling workshop to discuss tradeoffs, and identify a study design approach that maximizes advantages and minimizes disadvantages, in light of uncertainties and conflicting objectives.

1.5 HISTORY OF THE DESIGN EFFORT

Work on issues associated with the study design for the NCS began with the creation of the ICC. Members of the ICC, NICHD program staff, and members of the NCSAC have engaged in a rich discussion of options and possibilities. In addition to this ongoing dialogue, there were three directed efforts at preparing for a study design that deserve particular mention.

The first (and ongoing) effort is the contribution of the Study Design Working Group. Beginning in 2001, the Study Design Work Group has met, discussed study design needs, provided findings through the Advisory Committee to the ICC and Program Office, and requested pilot studies necessary to help inform design decisions. The Working Group originally focused on helping identify candidates for core hypotheses for the study and the criteria that might be used to judge candidate hypotheses. Later the Working Group focused on review of sampling designs proposed in the Westat report discussed below. This included comments and findings provided through the NCSAC to the ICC. The Sampling Workshop Planning Committee that planned the March 2004 NCS Sampling Workshop includes two members from the Study Design Working Group who continue to provide input from this working group.

The second effort is a report prepared by Westat, under contract to the National Center for Health Statistics, with guidance from members of NICHD and NCHS. The purpose of the Westat report was to develop and evaluate a number of candidate sample frames and sample designs for NCS enrollment. The report discusses three sampling models for initial consideration: a Household Model (door-to-door screening for fecund women), an Office Model (recruitment of pregnant women during ordinary prenatal care visits), and a Center Model (recruitment of pregnant women through a small number of formal centers that would be responsible for executing all aspects of the study protocol for their own recruits throughout the life of the project). Two variants of the Household Model with different degrees of clustering were examined, resulting in evaluation of four candidate designs. The report discussed the type and degree of clustering in the four evaluated designs, initial sample size determination, detailed costs for the sample recruitment, some aspects of the relative difficulty of various measurements of exposure and outcomes under the alternative designs, and statistical power for various tests (Westat, 2002). The Westat report significantly advanced the study design effort by providing detailed candidate options for consideration.

The third effort is a report prepared by Battelle, under contract to the U.S. Environmental Protection Agency, that examined optimal design considerations for measuring environmental exposures in the NCS, including methods for improving estimates of exposure through the use of detailed sub-studies that collect more precise exposure information on a small validation subsample of participants and use latent variable models to assess the relationship between health outcome and environmental exposure in the presence of measurement error. The methodology presented in the Battelle report is relevant to the overall sampling design in that it provides a tool that can reduce burden across the cohort and therefore potentially impact the feasibility of different sampling designs. A summary of the Battelle report to EPA is provided in Appendix C.

In September 2003, the NCS Program Office contracted with Battelle to prepare a white paper (Appendix B2) outlining a range of design options for selecting the longitudinal cohort into the study, building off the sampling design work described above. The Battelle paper first discussed options for three primary design elements which were seen as fundamental aspects of any proposed design. The design elements were: choice of the sampling frame for the population, method of selecting participants for the cohort, and organizational structure of the study.

For the first design element, the options paper presents three primary candidates for a sampling frame, largely synonymous with those presented in the Westat report. The first candidate was a household sampling frame that consists of a set of identifiable households in the U.S., and operationally would involve screening a sample of households to identify pregnant women, women of childbearing age, and/or couples attempting pregnancy. The second was a physician's office sampling frame which would allow for the selection of a sample of physicians and/or medical offices during a first stage of sampling, and the recruitment of a sample of pregnant women and/or women of childbearing age seen in their practices during a second stage of sampling. The third candidate was a community or university medical center sampling frame that involves selecting a sample of large health centers during the first stage of sampling that have previously demonstrated their ability and interest in conducting the NCS data collection protocol (e.g., through a competitive proposal process). These centers would recruit pregnant women and/or women of childbearing age either in proximity to or currently being served by their center or associated physician's offices.

The second design element discussed in the Battelle options paper addressed the methods for sampling the cohort of subjects from the sampling frame (i.e., selecting the subjects that will participate in the NCS). The range of options for selecting the cohort began with a set of fundamentally simple sampling design options that result from a choice of whether (1) the Primary Sampling Units (PSUs) are selected via probability-based sampling, quota methods, or some type of other non-probability method, and (2) participants within the PSUs are chosen via probability-based sampling, quota methods (to ensure some diversity and/or some similarilty with the larger population), or some other type of non-probability-based method. In addition to these fundamentally simple sampling designs, a class of hybrid design options was described. These hybrid design options combine probability-based sampling and non-probability-based sampling by selecting a portion of the sampling units on a probability basis and selecting all other sampling units on a quota or other non-probability basis, both for PSU selection, as well as for selection of participants within a PSU. The methods for specifying hybrid options for cohort selection introduced in the initial options paper (Appendix B2) represents a starting point, with subsequent development of a framework for the family of designs presented for consideration in Chapter 3 of this report.

Finally, the last design element covered in the options paper was the choice of an organizational structure for conducting the NCS and implementing the data collection protocols. The options for the organizational structure discussed included primarily University medical centers or large hospitals, contract research data organizations, health care providers, or some combination of the three.

After discussion of the design elements, the Battelle options paper discussed six general design categories or classes for recruiting and retaining the NCS cohort and the advantages and disadvantages of each. These included:

1. Complete probability-based design (all units at all levels are selected on a probability basis).
2. Convenience or quota sampling of PSUs and within PSU probability-based sampling.
3. Complete convenience or quota sampling.
4. A combination of convenience and probability-based sampling of PSUs, and complete probability-based sampling within PSUs.
5. A combination of convenience and probability-based sampling of PSUs and within PSUs.
6. A multiple cohort design with convenience selection of one (or more) cohort(s) and probability-based sampling of another (or other) cohort(s). The multiple cohorts could undergo varying levels of data collection (e.g., less burdensome environmental, behavioral, and health outcomes sampling for the probability sampled subjects), and could be followed for varying periods of time.

This Battelle options paper, along with a companion paper on criteria for evaluating the design options (see Appendix B1), served as the basis for discussions between NICHD Program Office representatives, Battelle staff members, and two consultants, Dr. Alan Zaslavsky of Harvard Medical School, and Dr. Colm O'Muircheartaigh of the University of Chicago's Harris School of Public Policy Studies. The complete summary of these meetings is included in Appendix B3. As discussed in the following Section, the final outcome was consensus agreement to explore a family of designs rather than pursue purely probabilistic or non-probabilistic designs.

1.6 THE FAMILY OF DESIGNS

At the Battelle meeting, all participants acknowledged that both the probability-based selection approach and the non-probability-based selection approach offer advantages and disadvantages, and both approaches have certain limitations in light of the objectives and constraints of the NCS. As discussions progressed, the meeting participants began to share the opinion that both of these sample selection methods offer important components to the NCS and may be able to be accommodated in the design. The group recognized that different categories of study users had legitimate scientific objectives that would favor probability sampling in some instances and restrictions on probability sampling to achieve other scientific objectives in other instances. For example, probability-based sampling offers the ability to generalize the results of the study with minimal assumptions; however other types of sampling approaches might offer more flexibility in obtaining previously collected medical history information from a more narrowly defined subset of potential respondents. Therefore, the group recognized a continuum of sampling methods in which a complete non-probability sample is at one extreme of the continuum and a complete probability-based sample is at the other extreme. Somewhere in the middle of these two extremes (i.e., a design that selects some portion based on probability and some portion non-probabilistically) may lie an optimal design that can satisfy *most* (ideally all)

of the objectives of the NCS. In keeping with the Battelle sampling design options paper, this might be called a "hybrid" design, but it may be better referred to as a *family of designs*. In other words, the NCS may not be composed of a single design, but rather a variety, or family, of designs that can be combined to address the multiple objectives of the NCS.

With respect to the concept of a family of designs, the meeting participants agreed that this type of design would be used to tackle multiple hypotheses and objectives. Thus, different parts of the design would be best suited to service different hypotheses and research demands. Some parts of the design would be essential for measures where data could be collected only in or by major medical centers; other parts of the design would protect against unforeseen circumstances and biases, protect against under-coverage of particular parts of the population that might undermine the validity of an inference, and allow statistical inferences to be extended to the whole population of the U.S. In terms of the application of evaluation criteria to the family of designs, the participants thought that it would be useful to check designs explicitly against criteria such as those proposed in the white paper included in Appendix B. By thinking of a family of designs, however, it is quite possible that a particular member of the family may fail a critical criterion, but may contribute enough on other criteria to make its inclusion not only worthwhile, but essential. Considering the array of designs and the array of criteria jointly, as well as the features and needs of "family members," is what will make the overall design a success.

The discussions also identified other proposed rationales for using a family of designs for the NCS. These rationales are generally related to the size of the study and the ability to propose a design that will meet the objectives of a variety of researchers (medical researchers, epidemiologists, social scientists, health researchers, clinicians, etc.), for whom the values of probability sampling, intensity of data collection, and exposure measures, etc., are of differing relative importance. First, since the sample size for the NCS is so large (100,000), the possibility of splitting the cohort into a portion selected non-probabilistically and a portion selected randomly could result in large sample sizes for both groups of individuals (whereas, in most studies that involve a small cohort of individuals, splitting of the cohort would not produce reasonable sample sizes). Second, since there are a variety of opinions as to the appropriateness and limitations of probability and non-probability-based selection for the NCS, incorporation of both types of sampling through a family of designs may provide a sampling design that can meet the objectives of a variety of NCS stakeholders. Finally, a family of designs might provide adequate coverage of populations that might not be served or included by more limited sampling frames.

1.7 ROADMAP TO THE REST OF THE REPORT

The purpose of the remainder of this report is to provide technical details on the options for a family of designs in sufficient detail to allow the NCS Sampling Workshop participants to make recommendations on the NCS Study Design. The hope is that the design framework presented in this report is sufficiently clear and reasonable to allow recommendations to be made for a study design.

Chapters 2 – 4 outline the options related to sampling frame, selection of the cohort, and organizational structure for implementation. Chapter 2 discusses target populations and the candidate sampling frames that might be chosen to represent those populations. Chapter 3 focuses on candidate methods for selecting participants, further introducing the family of designs concept and terminology. Chapter 4 introduces the candidate organizational structures for implementation.

Chapters 5 – 9 then discuss technical details critical to the evaluation of the options presented in Chapters 2 – 4. Chapter 5 discusses technical details on implementation of various sampling strategies incorporated in the family of designs, and the impact of the choice of a mixture of sampling strategies on the precision of estimates for the relationship between health effects and exposure (the design effects). Chapter 6 reviews core hypotheses and the measures that are critical to testing these hypotheses, providing a basis for the specific hypotheses chosen to be investigated in the power studies. Chapter 7 provides an overview of the assumptions used concerning recruitment and retention. Chapter 8 presents a model for estimating costs associated with the study as well as initial cost estimates and the assumptions on which they are based. Chapter 9 presents the results of analyses to characterize the power of different design options to detect significant associations for selected core hypotheses.

Finally, Chapter 10 summarizes the results, discusses caveats and limitations, and makes recommendations related to future work, including pilot studies that would help inform design decisions.

2 TARGET POPULATIONS AND CANDIDATE SAMPLING FRAMES

Broadly speaking, the main objective of the NCS is to study relationships between exposures (including chemical, physical, biological, and psychosocial exposures) and outcomes. Chapter 6 provides a description of the five priority outcome areas that have been proposed by the ICC for the NCS, and, for each outcome, identifies one or more hypotheses that focus on specific relationships between the outcome and some exposure of interest. In general, the aim of each hypothesis is to evaluate whether exposure is associated with the occurrence of a disease, or changes in the associated outcome measures, so that appropriate action can be taken for the affected populations. For example, in some cases regulatory action may be necessary to modify the chance of exposure to certain risk factors (e.g., more stringent controls on newly identified harmful chemicals), while in other cases, educational action may be necessary so that appropriate individuals (e.g., doctors, parents, guardians, etc.) have a working knowledge of the potentially dangerous exposures.

Since the NCS will necessarily study contemporary children (e.g., children born in the U.S. during the NCS recruitment period), by the time conclusions are drawn from the NCS data, it will in most cases be too late to take effective action for this contemporary population. Thus, in the terminology of Deming (1953) and Hahn and Meeker (1993), we consider the NCS to be primarily an "analytical" study rather than an "enumerative" (or "descriptive") study. The term "enumerative study" is used to refer to a study in which action will be taken on the elements in the frame studied where the term frame is used to refer to an aggregation of identifiable units, any of which may be studied. The term "analytical study" is used to refer to a study in which action will be taken on a process or cause-and-effect system with the aim of improving future conditions.

"Analytical" and "enumerative" studies engender somewhat different conceptions of the representativeness of the study. To illustrate this, consider a plausible descriptive study – summarizing the distribution of levels of prenatal exposure to environmental lead, overall and by the educational level of the parents – and a possible analytical study – characterizing the relationship between prenatal exposure to environmental lead and the incidence of severe learning disabilities. To describe the national distribution of exposure levels, a representative national sample is required, designed in such a way that a well-founded inference can be made to the distribution of levels nationally and in each of the groups of interest. In many cases, for such an inference to gain general acceptance, it must be based on only minimal assumptions, primarily those concerning the method by which the sample was selected. In particular, it may be desirable that the inference not depend upon model assumptions about the consistency of the distribution between parents' education levels and exposure levels. For example, it is possible that on average the children of lower-education parents tend to have higher prenatal exposures to lead, but this would not imply a universal conclusion since in some areas, more highly educated parents might happen to live in neighborhoods in which lead levels are more elevated than elsewhere in the area. The "descriptive" study described requires summarizing across enough such areas that the inference is representative of national rather than local problems. Furthermore, this type of analysis suggests taking into account the impact of survey design through adjustments for weighting (correcting for the varying probabilities that different children will be enrolled into the NCS sample) and clustering (reflecting the potential underestimation of

population variability introduced into survey results if the design leads to enrollment of groups of children that tend to be similar, e.g., because they are clustered in certain areas).

On the other hand, when we conduct an "analytical" study of relationships, such as that of associations of lead exposure with retardation, we often have a stronger causal relationship in mind and a larger population of interest (e.g., all U.S. children born over the next 50 years). For example, we may believe that the physiological mechanism by which lead affects brain development is consistent across the country and over time. Furthermore, we might study the effects of this underlying mechanism by collecting data on many other variables that may confound it in an observational setting. We would then apply statistical modeling techniques to isolate the effects of interest. For a study of this type, it might be less important that the sample from which the data are collected be fully representative of the national population (e.g., since the current national population is not fully representative of the larger population of interest). Indeed, we might be fairly confident that if our analytic methods are able to validly estimate the relationships of interest (by controlling for other variables such as nutrition, quality of prenatal care, and so forth), then those relationships should be fairly consistent across areas and subpopulations and over time. In fact, we may even want to deliberately seek out a study sample that is unrepresentative of national distributions but better supports estimation of the effects of interest (e.g., by including areas with unusually high levels of environmental lead).

Of course, rarely, if ever, are we so confident in our theories and analyses that we would be completely unconcerned with the representativeness of our sample. Including a variety of areas and populations in our study provides an important check on the robustness of our findings; indeed, investigations of this sort are one of strengths of the NCS, relative to a small, focused research project. Furthermore, statistical methods might be able to control for some but not all types of bias due to unrepresentativeness in the sample selection. For example, if the enrollment procedure had an uncontrolled and unmeasured tendency to miss developmentally disabled children from high-lead areas, our inferences could be biased. Thus, for this type of study the greatest concern might be about using well-defined sampling mechanisms *within* each area or subpopulation that has been selected for study. Nonetheless, representativeness of sampling is likely to be a lower priority in an "analytical" study relative to the ability to collect all of the desired variables in a consistent manner.

Additionally, while design and data collection planning of the NCS is driven by the study hypotheses, significant analyses will be conducted using the study data that are not closely related to these hypotheses but are serendipitous opportunities to investigate both future hypotheses and other relationships suggested by the data. (In fact, recall that this is one of the aims of the NCS identified in Chapter 1). Thus, it is not sufficient to ensure that the NCS sample is representative only for assessment of the study. Sound sample design and quality data collection in the execution of the study will allow these unexpected or unplanned opportunities to yield valuable findings. Good hypotheses provide the structure and the priorities for the survey and clinical data capture, but the payoff will include the ability to explore many additional relationships. Thus, it is all the more important that the sample be generalizable to some known and definable population, since we cannot know every particular group that will be our focus when we pursue one of those opportunities.

In this section, we attempt to take a broad view of sampling design presuming that valid statistical inferences of an analytical nature may be drawn from either a probability-based sample designed to be fully representative of a large target population or a less encompassing sample, which is representative of some portion of the target population. We consider how the NCS objectives and the desired target population combine to define a number of NCS candidate sampling frames from which the NCS cohort could be selected. In particular, Section 2.1 explains the notion of the ideal or true population of interest and how that differs from any of the populations available for study, Section 2.2 discusses several candidate sampling frames for the NCS, and Section 2.3 presents the idea of multi-frame sampling.

2.1 IDEAL AND POTENTIALLY REALIZED POPULATIONS

In order to have an impact on the health and well-being of children in the United States, the inferences drawn from the NCS cohort will have to be valid for a future population of children for whom some form of intervention is possible. Thus, the theoretical target population of the NCS is "all future children born in the United States." Figure 2-1 (duplicated from Appendix A) illustrates a context within which to consider various candidate sampling frames and design options for the NCS. Under any design scenario, data from the NCS cohort will be analyzed to test multiple hypotheses regarding environmental exposures in the "broadest sense" to help identify plausible cause-and-effect conclusions relating environmental exposure to health and developmental outcomes. The long-term objective of the NCS must be to influence public health policy and social behavior to bring about the application of effective environmental, behavioral, or clinical interventions that could prevent, detect, and treat diseases. Such interventions, when applied to a future national population of children, should lead to improved health and developmental well-being.

This future population of interest simply will not exist during the recruitment phase of the NCS and, therefore, cannot possibly be characterized statistically via an enumerative study. Furthermore, this future population may differ from the current population of children in numerous ways including:

- Magnitudes of environmental exposures may be different for reasons unrelated to the NCS,
- Magnitudes of environmental exposure may be different because of environmental interventions driven by cause-and-effect conclusions drawn from the NCS data, and
- Distributions of known and unknown confounders and effect modifiers that are correlated with environmental exposures and health/developmental outcomes may be different.

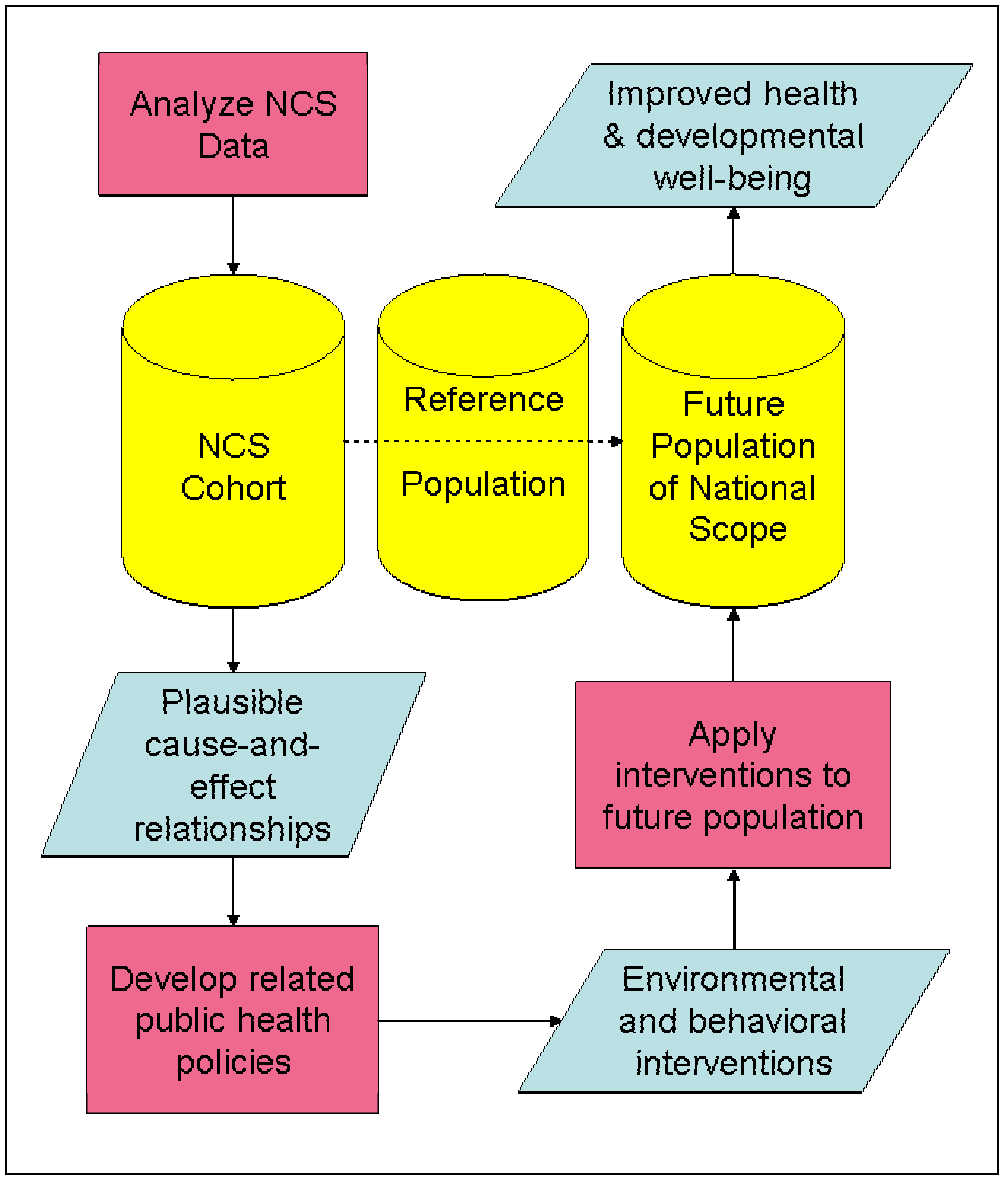


Figure 2-1. Context for NCS Sampling Design

In other words, since the future population of children cannot be studied in the NCS (i.e., since they will not be in existence during the recruitment phase of the study), it is logical, and necessary, to study an alternative population, such as children born during the period in which subjects are enrolled in the study (e.g., children born during the three to five year enrollment period). Thus, we must study analytical relationships in some alternative reference (or study) population, and implicitly assume that the relationships studied in this population do not change when applied to the larger future population, making the results irrelevant to the future population. In general, selecting an appropriate reference population should be done so that conclusions drawn from the study data (about the reference population) apply *to the maximum degree possible* to the future population of interest (see Appendix A).

Plausible choices for the NCS reference population include (a) all eligible children born in the U.S. during the NCS recruitment phase and (b) a subset of eligible children born in the U.S. during the NCS recruitment phase that has desirable properties with respect to recruitment, retention, and/or cost efficiency. (Note that eligible children is purposely ill defined here since specific eligibility criteria, such as pre-conception or prenatal recruitment, have yet to be determined.) The population of all eligible children is attractive because, lacking any knowledge of specific ways in which the future population of U.S. children will be different, the available population is arguably most likely to be similar to, or to include attributes of, the future population of children. Imperfect recruitment and retention erode the attractiveness of this study population. On the other hand, focusing on a subset of eligible children that have desirable properties with respect to recruitment, retention, and cost-efficiency is attractive from the point of view of maximizing the amount of information produced by limited study resources. However, as the study population is narrowed to achieve better cost efficiencies, a population bias relative to the current (and future) population may be introduced. The smaller that ultimate study (reference) population, the greater the potential for bias (see Appendix A) for further discussion of this issue).

Within the subsequent subsections, we consider several candidate sampling frames designed to enumerate various potential NCS reference populations. In general, we contend that there is no one right or wrong way to define the NCS study population. The practical realities associated with a study of such unprecedented magnitude are many and result in a very complex study design problem. In this section of the report, we consider only candidate sampling frames for which the NCS reference population is identifiable, as we are interested in populations (candidate sampling frames) that are representative of the future population of children and it is difficult to assess the representativeness of a frame that does not clearly define the units available for study. This restriction excludes volunteer samples since it is unknown who might volunteer for a study; however, volunteer samples are considered elsewhere within the design options and we would like to note that they may provide benefits to the study beyond representativeness (see Chapter 3). As suggested previously and illustrated in Figure 2-1, the NCS reference population serves as a stepping stone between the NCS cohort and the future population of interest. In the following sections, the potential NCS populations enumerated by the candidate sampling frames considered provide varying degrees of similarity to the future population of interest (again, see Appendix A for further discussion of this issue).

2.2 CANDIDATE SAMPLING FRAMES

The first element of any design option, the sampling frame, involves selection of the methodology for enumerating or identifying the pool of subjects from which the sample will be selected. As suggested above, perhaps the optimal NCS population is all children born in the United States during the three to five year NCS recruitment period (since it is arguably the available population that is most likely to be similar to the future population of interest); however, in order to accommodate the NCS objectives of obtaining prenatal health and exposure measurements (and in some cases pre-pregnancy health status and environmental exposures of the mother) it is necessary to sample either pregnant women early in their pregnancy, women of childbearing age, and/or couples considering or attempting pregnancy. The sampling frame must

reflect this necessity, and have the ability to identify these types of individuals. In addition, while practical considerations will likely limit the sampling frame to a subset of all potential participants, the sampling frame should provide broad coverage of the reference population(s) for the NCS. Several candidate-sampling frames and their strengths and weaknesses are discussed here.

Some of the sampling frames discussed in this section will identify the pool of subjects through a multistage description. This multistage description partitions population elements into groups within which they are enumerated. The term primary sampling unit (PSU) is used to describe the highest level of grouping in a multistage description of the population. Of course, several possibilities exist for defining the PSU. For example, counties, established Centers (e.g., university hospitals), metropolitan areas, or states could be considered the PSUs. Subsequent stages of sampling (e.g., secondary sampling units) will also depend on the sampling frame as well as the PSU definition. As an example, if the sampling frame calls for selection of households and counties are selected as PSUs, secondary sampling units could consist of census tracts within the county, and the final stage of sampling could select households within the census tracts. Alternatively, if established Centers are the PSUs and the sampling frame calls for selection of pregnant women attending these Centers, the second stage of sampling could simply involve sampling of pregnant women within the selected Centers.

In a pilot study conducted by CDC and Westat (2002), three basic approaches to the NCS sampling frame were proposed. Each approach has the potential for providing a sample of women in early stages of pregnancy (and in some cases prior to pregnancy). They are as follows:

1. **Household model**: This model involves screening a sample of households to identify pregnant women, women of childbearing age, and/or couples attempting pregnancy.

2. **Physician's office model**: This model involves selecting a sample of physician's and/or medical offices that would recruit pregnant women and/or women of childbearing age seen in their practices.

3. **Community or university medical center model**: This model involves selecting a sample of large health centers that would recruit pregnant women and/or women of childbearing age either in proximity to or currently being served by their center or associated physician's offices. (We will refer to this model as the Centers or "Centers of Excellence" model.)

In Table 2-1, we highlight several important considerations when comparing these three models in light of the objectives of the NCS. Note that this table focuses on the choice of a single sampling frame to highlight the differences between approaches. Multiple or layered sampling frames may combine various pieces of the different sampling frames in order to capitalize on the strengths of each possibility while attempting to minimize their weaknesses. These are discussed in Section 2.3.

Table 2-1. Broad Considerations of Household, Physician, and Center Models With Respect to NCS Objectives

Broad Consideration	Household Model	Physician Model	Center Model
Coverage	Likely offer most geographically representative sample depending on details related to stratification; does not cover homeless and institutionalized individuals, including college students.	May involve some form of geographic non-representativeness since presumably some regions will not be "covered" by a Center or a Physician's Office that is willing to participate in the study. Additionally, lacks coverage of those not served by a Center or physician, e.g., those without health insurance.	
Ability to sample pre-pregnancy	Offers the possibility of sampling women prior to pregnancy; may provide largest degree of pre-pregnancy measurements since it could select and follow women of childbearing age	Offers the possibility of sampling women prior to pregnancy.	
Screening requirements	Would involve a potentially sizable amount of resources in requiring tracking and measurements for women who never become pregnant; May include household screening efforts in order to eliminate households without age appropriate women.	The Physician's Office model and the Centers model would provide a less costly means of identifying and selecting women that are already pregnant.	
Community-based	Has the potential to be community-based	Not only has the potential to be community-based, but capitalizes on a community-based infrastructure that exists prior to the study.	
Primary Sampling Unit	PSUs are most likely geographic regions, such as counties.	PSUs may consist of geographic regions, with Physician's Offices being selected at a lower stage of the design.	PSUs could be the Centers themselves, since centers include a large number of possible study subjects, or PSU could be geographic regions around a Center.

There are also other options that do not include any of the above sampling frames. For example, if it were acceptable to collect prenatal health and exposure related information retrospectively (probably not the case for the NCS), a sample of infants could be selected from the births reported in the Birth Registration System (see Westat, 2002). Alternatively, there may be other existing sampling frames (e.g., the Census American Community Survey database, or the NHANES sampling frame) that may be possibilities for constructing the NCS sampling frame. Finally, although perhaps not as likely to produce a sample that conforms to the needs of the NCS, other possible sampling frames could involve selecting women of childbearing age based on motor vehicle records, based on participant records from cooperating HMOs, public

health care clinics, or based on random digit dialing. At the present time, and given the necessary components of the NCS, the household, physician, and centers models (and combinations of these three) appear to be the best options for creating a viable sampling frame for this study. In the following section the use of multiple sampling frames, such as a combination of the household, physician, and centers sampling frames, is discussed as a promising method for capitalizing on the strengths associated with different sampling frames.

2.3 DESCRIPTION OF MULTI-FRAME SAMPLES

Dual and multi-frame samples have a long history, going back to Hartley (1962), who noted that such designs can result in considerable cost savings over a single-frame design with similar precision. The general idea is that by drawing a study population using a combination of several different sampling frames, one can benefit from the advantages while minimizing disadvantages associated with each individual frame. Lohr and Rao (2000) provide some excellent examples of how dual-frame sampling might work. For example, a sample of individuals with Alzheimer's disease might be constructed by drawing some individuals from the general population (in order to ensure representativeness) and drawing others from senior care facilities (in order to reduce costs by sampling from a high prevalence population). They cite this Alzheimer's example as an illustration of the general principle of generating a sample of individuals with a rare disease by augmenting a population-based sample with one drawn from a high prevalence (or high risk), yet incomplete, population. Lohr and Rao also describe an example of particular relevance to the National Children's Study, namely Canada's National Longitudinal Survey of Children and Youth which was based on three different sampling frames. Two of the frames correspond to one used for the Labour Force Survey, before and after a redesign in 1995, while the third frame is the one used by the National Population Health Survey.

For the NCS, a dual- or multi-frame sampling strategy would combine a broad probability-based population-wide sample, such as a national household sample or a household sample restricted to MSAs of qualified Centers (call this frame A representing the Household Model) with a sample selected from patient lists of qualified Centers or physicians' offices (call this frame B representing recruitment through university-based medical centers or physicians' offices). By incorporating a sampling strategy based on frame A, the NCS will have a greater chance of being truly representative of the entire United States or of the selected areas. For instance, such a sample could ensure appropriate representation of low-income subjects or subjects from minority ethnicities. However, the downside to this is that some of the subjects sampled from frame A might be more likely to refuse to participate in the study, or might be more difficult to retain (i.e., being more likely to drop out before study completion). A careful choice of frame B can potentially identify a more compliant population (lower refusal rates, higher retention rates, easier tracking, greater cooperation with follow-up appointments, etc.). For example, study subjects recruited through a university-based medical center already have built-in alternative tracking and contact mechanisms, as well as incentives to maintain contact with study staff as part of receiving ongoing care for their child.

In other words, a possible multi-frame sampling strategy for the NCS would combine all three models (Household, Physician's Office, and Centers) into a single framework. While this use of multi-frame sampling is appealing from a heuristic perspective in terms of enhancing

study validity by overcoming weaknesses associated with each approach (e.g., weaknesses in coverage, anticipated retention rates, efficiency, varying degrees of willingness to undergo burden, etc.), usage of multiple frames does not come without cost as there are a number of challenges associated with how data from the separate cohorts should be combined. For example, statistical analysis of data collected in such a manner poses considerable challenges (e.g., determining an appropriate approach to assigning sample weights to all study participants), as does determining the appropriate "mix" of the multiple frames given the numerous, and at times competing, objectives of the study (e.g., national probability-based sampling may provide greatest generalizability of the results but may result in relatively small retention rates over the course of the study). In Section 3 of this report we describe a family of designs that attempts to combine the Household Model and the Centers Model in a dual-frame sampling approach. Further details of multi-frame sampling can be found in Section D-1 of Appendix D, including a discussion of a framework for statistical analysis of data from multi-frame studies.

3 DESCRIPTION OF A FAMILY OF DESIGNS

Much of the initial sampling design work for the NCS that preceded this current effort was aimed at developing relatively straightforward designs such as those detailed in the Westat report (2002). The design options and modes of recruiting women/children into the study that were characterized in this report, namely the Household and Centers of Excellence Models, each have strong support from different members of the scientific community that have been involved in the planning process for the NCS.

The Household Model, which employs a strict probability-based sampling approach to recruit women into the study, has strong support from scientists that want to ensure that the results of the NCS can be generalized to a broader population beyond the cohort of participants. Use of a probability-based approach to selecting study subjects will allow researchers to minimize unintentional bias in the analysis of data from this study, particularly when there are factors related to the outcome of interest that are not observed as part of the study protocol. The specific implementation option for the Household Model advocated in the Westat work involved sampling from approximately 800 counties (approximately 25% of counties in the U.S.) based on an analysis of design effects associated with the estimation of disease prevalence.

Critics of this particular implementation option of the Household Model cite that recruitment and retention of study subjects in such a large number of locations throughout the U.S. would be operationally infeasible for a study as complex as the NCS, that the high overhead associated with maintaining presence in a large number of counties would divert resources away from important data collection activities, and that the lack of community connection would adversely affect recruitment and long-term retention of participants.

Other concerns that have been raised about probability-based sampling approaches (not necessarily specific to the Household Model) for the NCS relate to anticipated low recruitment and retention rates, compared to other approaches that might capitalize on a sample of well motivated volunteers. For example, if the initial response rate during the recruitment phase of the NCS is very low (e.g., between 15 and 30 percent) under a probability-based selection approach as some would suspect, then the results of the study could not necessarily be extrapolated to the original sampling frame without additional strong assumptions. Rather the results of the study would only be generalizable to a much smaller subset of the original sampling frame that would agree to participate in a study like the NCS. Lower retention rates, if these were to result from a probability-based sampling approach, could also be devastating in the NCS, particularly when considering the ability of the study to address hypotheses related to rare adverse health outcomes that may be detected only later in life, such as schizophrenia and congenital heart defects.

The Centers of Excellence Model suggested in the Westat report assumes that a more limited sample of purposively selected qualified medical centers (e.g., 100 university hospitals) would be selected at the first stage, and that these centers would then conduct the NCS data collection activities from within their areas of service. An assumption was made in the Westat report that the Centers of Excellence Model would likely result in the centers recruiting study

participants through convenience sampling rather than a probability-based sampling approach. Assuming that the centers recruit women who are motivated to participate in the NCS, it is likely that the long-term retention rates under this model would be higher than probability-based sampling approaches, which would attempt to enroll some fraction of participants who were initially reluctant. In addition, it is likely that the prestigious nature of the University Centers of Excellence would enhance initial recruitment and long-term retention rates. Supporters of the Centers of Excellence approach cite other major studies such as the Women's Health Initiative, the Collaborative Perinatal Project, the Framingham Study, and the Nurses Study as having made tremendous contributions to the fields of medicine and public health without following a probability-based approach. The relationships between disease and important risk factors observed in these studies have internal validity, and the causal inferences that have been drawn from these studies have never been shown to be severely biased when applied to a more general population. These historic studies gained their internal validity based on their ability to assess important risk factors, covariates, effect modifiers and confounders across the study population. However, these studies were also much more limited in scope and complexity in comparison to the NCS. With the broad range of health outcomes, potential exposures (including psychosocial, chemical, biological, and physical environments), and critical stages of vulnerability that fall within the scope of the NCS, many members of the scientific community fear that it will be operationally infeasible to assess all of the important risk factors, covariates, effect modifiers, and confounders across the entire cohort over time. If important factors are not observed within this sample, the potential for misleading inferences due to a biased sampling approach becomes much more likely.

The above discussion provides a very quick overview of a complex and ongoing debate that has yet to be resolved for the NCS. A more detailed discussion of the advantages and limitations of probability-based and non-probability-based sampling approaches can be found in Appendix A. Both the probability-based sampling approach and the clinical/biomedical/ epidemiologic center-based approach have strong and compelling arguments that must be taken into consideration when planning this study. Initial review of the Westat work suggested that so-called Hybrid Options should be considered. These Hybrid Options would hopefully capitalize on the strengths of the various different sampling approaches discussed above, while minimizing their weaknesses.

While the sampling approaches that are discussed in this report can legitimately be considered as Hybrid approaches, we would like to introduce them as a Family of Designs for the NCS. Within the Family of Designs will be multiple methods of recruitment into the NCS, with some study participants recruited from one or more well-defined sampling frames that have broad coverage of the population of interest, other study participants recruited from more narrowly defined sampling frames that represent study subjects who are easier to access and retain, and still other study subjects who are recruited through an opportunity or convenience sample because they comprise a segment of the population that is of interest (e.g., children born outside of the hospital environment using a midwife) but difficult to access, or highly motivated volunteers who are likely to undergo higher-burden data collection activities without losing interest in participation. It is assumed here that a minimal core data collection protocol, that allows researchers to address some or all of the core hypotheses for the study, would be applied to all study subjects, regardless of their mode of recruitment into the study. That is not to

suggest that every study subject will undergo repeated waves of biological sampling during pregnancy or extensive environmental sampling over time – as study planners will need to identify methods to minimize the data collection protocol to the extent possible among the large NCS cohort while taking advantage of the fact that additional data collection activities will be performed as an add-on to the minimal protocol for certain segments of the study population. These additional data collection activities (more extensive environmental assessment, repeated waves of biological sampling during pregnancy, detailed neurological assessments, etc.) would likely be planned with higher frequency among study subjects that are easier to access or more willing to participate in additional or more burdensome sub-studies. Even here, matrix sampling may be used to assign participants into these sub-studies to minimize their total burden.

Conceptually, the Family of Designs provides a multiple-approach solution for planning the study, in which part of the study population will be recruited in a manner that maximizes the opportunity for detailed and rigorous data collection, while another part of the study preserves the ability to generalize important study results to the population of interest. The intent is to maximize the advantages of different approaches while minimizing their limitations, resulting in a study design that is more optimal overall than one that is limited to a single approach.

Figure 3-1 provides a conceptual layout of the Family of Designs considered throughout the rest of this report. The Family of Designs initiates with identifying a fraction (P_1) of the NCS cohort that will be recruited through a national probability-based sampling approach. The remaining fraction of study subjects (1- P_1), located within purposively selected Centers of Excellence, will be recruited through a variety of mechanisms. Among the participants located within the Centers of Excellence, we assume that a fraction (P_2) will be recruited from a probability-based sample from areas in proximity to the Centers (e.g. from the metropolitan statistical areas (MSAs) surrounding the Centers), another fraction (P_3) will be recruited from a probability-based sample of Center patients, and the remaining fraction (1- (P_2+P_3)) will be recruited from an opportunity or convenience sample. Specific details about the different components of the Family of Designs are provided in the bulleted list below:

- Operationally, we assume that the P_1 fraction of the NCS cohort that is recruited under a national probability-based design will follow a multistage clustered sampling approach similar to those discussed in the Westat report (e.g., the Household or Physician Office Models) in which counties (or other groupings/clusters) are selected as Primary Sampling Units (PSUs) at the highest level of hierarchy in the design using a probability proportional to size sampling approach. It should be noted that some PSUs may be selected with certainty, as discussed further in Chapter 5. Study subjects from within the selected PSUs would be recruited into the study at a subsequent sampling stage also using a probability-based approach. For PSUs located in urban or suburban areas, a Household Model for recruitment is likely to be the most efficient method for mainstream recruitment, possibly with supplemental recruiting from homeless shelters and other institutionalized populations (college dorms, prisons, military) that would likely be excluded from the household sampling frame. This supplemental recruiting would be attempted using a probability-based sampling approach, so that these study subjects can be integrated into the national probability-based sample. In rural PSUs selected into the sample, it may be more efficient to use a physician's office sampling frame for the initial

recruitment of women in early stages of pregnancy (or prior to conception). This recommendation comes under the assumption that there will be a limited number of obstetrics/gynecology providers in rural areas, and that households are likely to be more geographically sparse, making door-to-door recruitment much more difficult. Again, we would consider supplementing the physician's office sampling frame in rural areas with additional sampling to recruit the segment of those rural women/children who would be excluded from this sampling frame – namely those without access to medical care. It is assumed that the majority of study subjects that are recruited through the national probability-based sample component would undergo the minimal core data collection protocol. It is also assumed that within the Family of Designs, the national probability sample component will suffer from the lowest rates of recruitment and retention, as discussed in Chapter 7. However, this component of the Family of Designs is important to the extent that study planners would like to generalize to the population of all children born in the U.S. during the period of recruitment. It is presented first here, as an extension of the initial designs proposed in the Westat report.

- The second component of the Family of Designs, representing a P_2 fraction of participants that are associated with the Centers ($1-P_1$) for a ($P_2*(1-P_1)$) fraction of the total NCS cohort, consists of probability-based sampling within the geographic area surrounding each purposively selected Center. An initial review of the University Medical Centers that are most likely to qualify for participation in the NCS (see Section 5.2) suggests that most (if not all) will be located within geographic areas designated as Metropolitan Statistical Areas (MSAs) by the U.S. Census Bureau. MSAs generally represent groups of counties in close proximity to large urban areas. For the design work conducted in this report, we assume that a probability-based sampling approach will be conducted to recruit a portion of study participants from the MSA counties that surround the purposively selected Centers – most likely through a Household Model of recruitment. Study participants that are successfully recruited through this mode of sampling will be referred to the qualified Center for all data collection and related study activities. It is assumed that this mode of sampling will have slightly higher rates of recruitment and retention than the national probability-based sample component, mainly because of the prestigious reputation of the participating Centers and the heightened sense of community awareness of the NCS that is likely to be associated with a Center. For this reason, it is likely that a significant part of the study population recruited through this mode of sampling could undergo more detailed and rigorous data collection protocols.

    - Note that in some cases, a county that falls within the MSA of a qualified and purposively selected Center will have been identified in the national probability-based sample component of the Family of Designs. In these cases, study subjects that are identified in these counties would be referred to the qualified center for all data collection and related study activities. The second component of the design [probability-based sampling of the surrounding MSA] would also exclude recruitment of study subjects from within that particular county to ensure that these counties are not overrepresented in the study.

- The third component of the Family of Designs, representing a P_3 fraction of the $(1-P_1)$ participants that are associated with the Centers for a $(P_3*(1-P_1))$ fraction of the total NCS cohort, consists of a probability-based sample of patients already affiliated with the purposively selected Centers. Recruitment of these study subjects would not occur at the household level, rather a more convenient probability-based sampling approach would be used based on a sampling frame constructed of eligible patients already serviced by the purposively selected Centers. It is assumed that probability-based sampling from this sampling frame would have a relatively high rate of initial recruitment and retention (compared to the previously described two modes of sampling) because of the existing relationships that these potential study participants will have with the medical center that will be responsible for the data collection activities. It is also assumed that the sampling frame of patients already affiliated with the qualified centers is very small relative to the previously described sampling frames (National PBS and PBS of MSAs surrounding a Center) – and therefore, study participants recruited in this third component of the Family of Designs would not have high influence in any weighted analyses that are conducted with the purpose of generalizing from a statistical basis to the population of interest. On the other hand, these study subjects would likely be available for more rigorous and detailed data collection activities, making them invaluable for more careful investigation of the causal relationships between exposures and adverse health effects. Despite the low influence of these study subjects in weighted analyses of the NCS data, we advocate probability-based (or other systematic) sampling approaches from the pool of patients already affiliated with the purposively selected Centers – thus allowing researchers to extrapolate the results of the study from this fraction of the NCS cohort back to a population with known or measurable characteristics. In this regard, simple quota sampling among patients of the Center would not be acceptable, as it merely specifies the number of participants in a given category, but leaves the choice of whom to select in the hands of the individual recruiter. This is unacceptable due to the fact that even well intentioned recruiters, if allowed to select individuals, will select a biased sample of individuals. But more importantly, the study will have no information on how these study subjects were selected to allow scientists to assess the generalizability of the study results that pertain to the subjects recruited from the population of patients affiliated with the chosen Centers. Thus, a strict convenience sample (i.e., allowing the recruiters to select whomever they like including well motivated volunteers) is likely not ideal for this component of our Family of Designs. Rather, there must be some level of design for which it will be "easy" to incorporate random sampling (e.g., when a pregnant woman enters the medical center, flip a coin to determine whether or not to recruit her).

- The last remaining component of the Family of Designs, representing a $(1-(P_2+P_3))$ fraction of participants that are associated with the Centers and a $(1-(P_2+P_3))*(1-P_1)$ fraction of the total NCS cohort, consists of an opportunity or convenience sample of study subjects recruited by the Centers. These study subjects may be comprised of volunteers who are highly motivated to participate in the study but were not selected through one of the previously mentioned probability-based sampling approaches. These study subjects may alternatively include segments of the population of extreme interest to the study that would otherwise be excluded or underrepresented in the previous sampling frames (such as women who utilize the services of a midwife

during delivery, women who plan on putting their babies up for adoption, women who become institutionalized during pregnancy, etc.). Study subjects recruited through convenience or opportunity will carry virtually no weight in the models for external validity (weighted analyses of the NCS data) that generalize results based on the sample weights from a probability-based sample. However, these study subjects may provide very valuable information for model-based inferences and add value to the NCS.

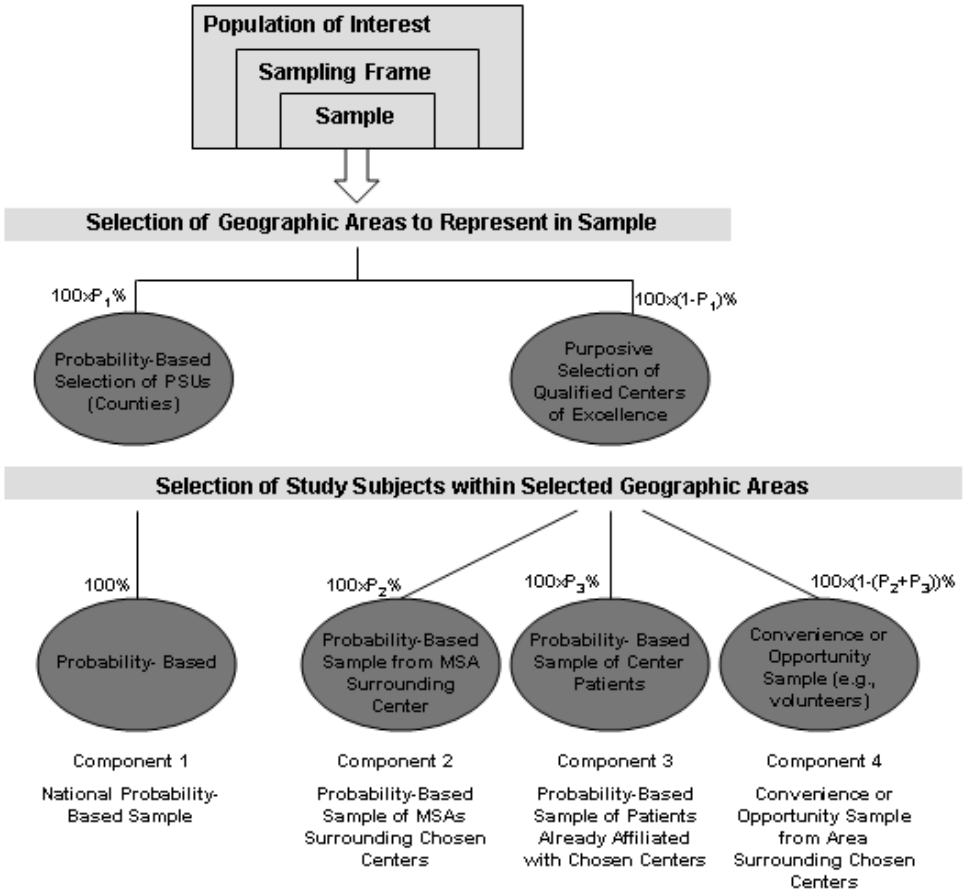


Figure 3-1. Conceptual Model for the Family of Designs.

Throughout the remainder of this document, a series of 23 different study designs for the NCS are explored by varying the fraction of study participants that are assumed to be recruited through the above four components of the Family of Designs. As shown in Table 3-1, the first two designs set the values of P_1 (the fraction of the NCS cohort recruited through a national PBS) and P_2 (the fraction of the center-based sample recruited through a PBS of the MSA surrounding the Center) to zero. These two designs limit the NCS sample to study participants already affiliated with the purposively selected Centers recruited via probability-based sampling and other participants recruited via non-probability selection methods. The next three designs set the values of P_1 (the fraction of the NCS cohort recruited through a national PBS) to zero, thereby limiting the study population to those that live in proximity to purposively selected

Centers at the time of initial recruitment. These designs allow for a mixture of household- and Center-based sampling frames and for probability and non-probability selection methods, and include a number of the "hybrid" models that were discussed following the Westat report. These first five designs that exclude a national probability-based sample are then followed by a series of 18 designs in which 25, 50, or 75 percent of the NCS cohort is recruited through a national probability-based sample. Among these 18 designs, the first nine involve the probability-based selection of 50 PSUs (counties) and the second nine involve the selection of 100 PSUs. It should be noted that the designs considered in this report exclude the option in which P_1 (the fraction of the NCS cohort recruited through a national PBS) is one. However, design options E14 and H23 in Table 3-1 come close to approximating what would occur if P_1=1, with 75% of the NCS cohort recruited through a national probability-based sample and 18% of the NCS cohort recruited through a probability-based sample of the geographic areas surrounding 13 purposively selected Centers (which could be viewed as 13 PSUs selected with certainty) resulting in 93% of the NCS cohort selected from a PBS of relatively unrestricted populations.

Table 3-1. Overview of Specific Designs Explored as a Function of the Fraction of the NCS Cohort Recruited through each Component in the Family of Designs

Design	P_1	P_2	P_3	Fraction of the National Children's Study Recruited From				Number of PSUs in National PBS	Number of Purposively Selected Centers
				National PBS	PBS of Center MSAs	PBS of Center Patients	Purposive/ Convenience Sample		
A1	0.00	0.00	0.97	0.00	0.00	0.97	0.03	0	50
A2	0.00	0.00	0.50	0.00	0.00	0.50	0.50	0	50
B3	0.00	0.24	0.72	0.00	0.24	0.72	0.04	0	50
B4	0.00	0.48	0.48	0.00	0.48	0.48	0.04	0	50
B5	0.00	0.72	0.24	0.00	0.72	0.24	0.04	0	50
C6	0.25	0.24	0.72	0.25	0.18	0.54	0.03	50	38
C7	0.25	0.48	0.48	0.25	0.36	0.36	0.03	50	38
C8	0.25	0.72	0.24	0.25	0.54	0.18	0.03	50	38
D9	0.50	0.24	0.72	0.50	0.12	0.36	0.02	50	25
D10	0.50	0.48	0.48	0.50	0.24	0.24	0.02	50	25
D11	0.50	0.72	0.24	0.50	0.36	0.12	0.02	50	25
E12	0.75	0.24	0.72	0.75	0.06	0.18	0.01	50	13
E13	0.75	0.48	0.48	0.75	0.12	0.12	0.01	50	13
E14	0.75	0.72	0.24	0.75	0.18	0.06	0.01	50	13
F15	0.25	0.24	0.72	0.25	0.18	0.54	0.03	100	38
F16	0.25	0.48	0.48	0.25	0.36	0.36	0.03	100	38
F17	0.25	0.72	0.24	0.25	0.54	0.18	0.03	100	38
G18	0.50	0.24	0.72	0.50	0.12	0.36	0.02	100	25
G19	0.50	0.48	0.48	0.50	0.24	0.24	0.02	100	25
G20	0.50	0.72	0.24	0.50	0.36	0.12	0.02	100	25
H21	0.75	0.24	0.72	0.75	0.06	0.18	0.01	100	13
H22	0.75	0.48	0.48	0.75	0.12	0.12	0.01	100	13
H23	0.75	0.72	0.24	0.75	0.18	0.06	0.01	100	13

For all designs considered besides the first two, P_2 (the fraction of Center-based participants recruited through a probability-based sample of the MSA surrounding the Center) and P_3 (the fraction of Center-based participants recruited through a probability-based sample of

patients already affiliated with the Center) take on paired values of (0.24, 0.72), (0.48, 0.48), and (0.72, 0.24) – thereby leaving a nominal proportion (between 1 and 4 percent) of the NCS cohort in each design to be recruited through convenience or opportunity sampling. It should be noted that among the last 18 design options presented in Table 3-1, the cost estimates provided in Chapter 8 and the power calculations provided in Chapter 9 are relatively robust to any assumption that a higher fraction of study subjects are recruited through convenience sampling, as long as the fraction of study subjects recruited through a probability-based sample of Center patients is reduced by a proportional amount. Thus, if study planners believe that the design options explored grossly underestimate the fraction of study participants that will be recruited through convenience or opportunity sampling, they could reduce the fraction of participants assumed to be recruited through a probability-based sample of Center patients and assume that those subjects are recruited through purposive sampling within any of the last 18 design options with confidence that the cost and power results are reasonably accurate. This robustness property is based on assumptions of similarities between the PBS of Center patients and purposive sampling with respect to (1) costs of recruitment, (2) retention rates, (3) relatively low sampling weights in an analysis of data from a study that includes some national probability-based sampling.

The choice of design will likely impact both model-based and weighted statistical analyses in the use of the NCS data. The previous work conducted by Westat included a detailed discussion related to design effects associated with the estimation of the prevalence of adverse health outcomes. The concept of design effects is important for the planning of this study, and provides useful insight into the possible reduction in effective sample size that results from both unequal weighting and clustering in a multistage sampling approach when compared to a similar study that is performed using a simple random sample of the population of interest. However, the focus of the NCS is evaluating relationships between adverse health effects and measures of exposure or other risk factors, so the focus here is on calculating design effects for these estimates. This concept is explored in more detail in Chapter 5. Regarding design effects, we can provide some intuitive results on the likely reduction in available sample size that is introduced at the highest level of design with the choices of the fractions P_1, P_2 and P_3 in the family of designs:

- For model-based inferences, which we assume for the moment to be unweighted analyses of the NCS data, the maximum effective sample size is represented by the number of study participants who remain in the study at the time the response of interest is observed. To the extent that a particular model-based inference requires measures from additional data collection procedures that are beyond the minimal sampling protocol, the effective sample size may be reduced to a smaller number (perhaps approximated by the number of participants remaining from the probability-based sample of Center patients and purposive sampling).

- For weighted analyses of the data, the maximum effective sample size is best approximated by the number of study participants recruited from the national probability-based sample and the probability-based sample of MSA counties surrounding the purposively selected Centers who remain in the study at the time the response of interest is observed. Of course, the true effective sample size for the weighted analyses are

further reduced by any unequal weighting and clustering that occurs as a result of the multistage sampling approach.

It should be noted that the reference population to which study results from weighted analyses may be generalized changes as a function of P_1 and P_2. For example, the first two designs in Table 3-1, in which P_1 and P_2 are both zero, only represent patients affiliated with purposively selected Centers and patients recruited using non probability-based methods. These first two designs may result in a biased NCS sample, based on the selection of study participants that have access to (and utilize) healthcare, as well as any geographic biases connected to the purposive selection of Centers. The next three designs in Table 3-1, in which P_1 is zero, will suffer from similar geographic biases as discussed above, but will not necessarily suffer from any strong selection bias with respect to the selection of study participants from those locations. Weighted study results from these three design options can be generalized to the areas surrounding the Centers, under assumptions of minimal non-response biases. Weighted analyses of the remaining 18 design options can be generalized to the more broad national population under similar assumptions.

Finally, it should be noted that each of the 23 designs described in Table 3-1 and explored in further detail throughout this report are consistent with the recommendation from the NCS Federal Advisory Committee that the NCS be based on a probabilistic design (NCSAC, 2003).

4 ORGANIZATIONAL STRUCTURES FOR IMPLEMENTING THE NCS

Successfully conducting and completing the many aspects of the NCS will involve many different government, private, and nonprofit institutions. Possible organizational structures are described here to identify assumptions which were made to provide cost estimates of alternative designs in Section 8. The government members of the team will include the multiple Federal agencies involved in the Interagency Coordinating Committee (i.e., NICHD, CDC, EPA, NIEHS) as well as state and local government health departments and, possibly, laboratories. Private companies and other types of organizations will likely participate as contractors helping to enroll participants, collect data, staff phone banks, analyze and store data, and perform software development and data management services. Participating nonprofit institutions may range from hospitals and universities operating as study centers to local neighborhood groups encouraging the participation of local residents. During the design stage, it is important to identify which types of institutions are best suited for certain activities. This section of the report will first discuss the characteristics of each type of institution and then review the seven major work areas involved in implementing the study and discuss the organizations required to complete each work area under the four sampling approaches (components of the Family of Designs) described in Chapter 3.

4.1 CANDIDATE ORGANIZATIONAL STRUCTURES

Both governmental and nongovernmental organizations will have a role in implementing the NCS. This section provides a brief introduction to these two main categories of organizations and eight specific types of organizations within these categories, discussing the potential role of each organization in the study.

4.1.1 Government Organizations

Federal Government

The various agencies of the Federal government involved in planning the NCS will oversee the implementation of the study. NICHD will lead the effort, but, as in the planning stages, it will be supported by EPA, CDC, and NIEHS and the Interagency Coordinating Committee. These agencies have responsibility for deciding the study protocol and organization with final decision-making power resting with Dr. Duane Alexander, Director of NICHD. As with other major government health and environmental studies, rather than hire additional Federal employees or utilize existing staff to implement the study, the government likely will issue grants or contracts to fund the study implementation activities while maintaining its oversight and overall management role.

State and Local Government Organizations

The Federal government may involve state and local health departments in the planning and execution of the study. For the most part, these organizations might best serve as advisors to the centers and contracted organizations (COs) operating in their area. State and local government agencies can ensure that study designers and data collectors are aware of any local health and environmental issues that should be accounted for in the data collection protocol. This advisory role assumes that funding for the study originates from the Federal government. If

state and local governments contribute to funding or provide staff for the study, they may require a larger role in both the planning and implementation of the study in their areas. In this situation, the Federal government may choose to establish a state and local government committee that would meet regularly to provide input to decision-making on study implementation.

An additional role that local government employees could play is performing environmental assessments and sample collection. Many local housing or health departments have equipment, processes, and personnel necessary to collect paint, dust, soil, and other samples in and around participants' residences. Similarly, local health departments may have staff that can also assist with collecting biological samples, conducting some interviews, accessing information on local environmental conditions, and performing some medical tests. Utilizing these local government resources may be cost-effective and may also build goodwill with the local community.

4.1.2 Nongovernmental Organizations

The government agencies directing the study will likely enlist many other types of organizations to assist them with the study's implementation. Although the government has various mechanisms for utilizing nongovernment companies, personnel, or other organizations, it is likely that the government will want to enter into contracts in most cases to carry out the NCS. Contracts, as opposed to grants, will allow the government to control, when necessary, how the study is being conducted across the country. Under a contract, the government can ensure that each piece of the study is accomplished according to the overall plan—that an overall sampling plan is implemented consistently across multiple areas and organizations, that all data collection organizations follow specified protocols, and that specified timelines and deliverables are met. For this reason, we assume that most of the organizations below would operate under contracts with the Federal government or under subcontracts to another of these organizations under contract to the Federal government. We discuss five specific types of organizations that may have a role in implementing the study – medical centers, a central coordinating center, physicians' offices, local neighborhood groups, and laboratories. Although each of these five types of organization may operate under contract to the Federal government, the sixth type of organization – contracted organizations – is meant to be a catch-all category including all other organizations working under contract to implement the study.

University and Hospital Medical Centers

It is likely that the Federal government will select a number of large university and hospital medical centers to serve as data collection organizations serving particular geographic areas. Using these established research centers (Centers) will capitalize on the talented researchers and medical professionals currently working at these centers as well as their vast infrastructure – ranging from buildings to medical equipment to supporting organizations such as IRBs, accounting centers, and law departments. The scope of activities conducted by each Center may vary from Center to Center. For example, one Center may choose to use staff employed by the Center to assist participants with completing questionnaires and surveys, whereas another may choose to contract this work out to another organization.

Central Coordinating Center(s)

If the study's organizational structure includes a large number of medical centers or other data collection organizations, the government may consider awarding a grant or contract to one or more medical centers or other organizations to coordinate all aspects of the study. The central coordinating center (CCC) could be a single organization or could be multiple organizations, e.g., a clinical coordination center and a data coordination center. This central group would serve as the primary point of contact for several other medical centers and participating organizations. It would track all study activities and provide regular status reports to the government. This group might also coordinate and organize training and participant tracking activities.

Physicians' Offices

Within the national probability-based sample, local physicians' offices could potentially be enlisted to serve as mini-data collection centers responsible for performing health checks and collecting biological samples. Within the Center-based frames, a local physician's office could assist the study in a number of ways including 1) distributing or posting information on the NCS to encourage participation for patients who are selected for the study, 2) identifying and/or enrolling preconception women who are attempting or will soon attempt to become pregnant, and 3) serving as off-site mini-data collection centers for study participants who have problems with traveling to the Center's primary location. To conduct the study in rural areas of the country, it may make sense to utilize local physicians' offices to identify potential participants for the study. In these relatively sparsely populated areas, a small number of physicians may serve a large geographic area.

Nonprofit Community/Neighborhood Groups

Although they may not play as integral a role in the data collection process as the other organizations, nonprofit community and neighborhood groups could likely contribute to the study by helping with recruitment and retention efforts. These groups, with close ties to the communities targeted for NCS participation, could be enlisted to provide information to members of the community and encourage participation and continuation from those in the community selected for the study, especially in subgroups of the population who may be more hesitant to participate. In some situations, there may be nonprofit community health centers that could assist with the data collection protocol in various ways, for example by assisting non-English speaking participants with completing study questionnaires and enrollment forms.

Private and Government Laboratories

There will be a large amount of laboratory chemical analysis work involved in measuring various analytes in environmental and biological samples collected from participants. A decision that will have to be made early in the implementation process will be whether it is necessary to limit the number of laboratories used to analyze study samples. Benefits associated with use of one or relatively few laboratories include more likely use of consistent processes and methods and more efficient quality control and quality assurance (QA/QC) oversight. On the other hand, using a larger number of laboratories may lead to faster turnaround times and potential buildup of goodwill by using local laboratories to analyze samples collected in a given geographic area. Note that selected University Medical Centers may have associated laboratories that are logical choices to analyze samples collected from their Centers' participants.

Contracted Organizations

As noted above, the government and the participating medical centers may find it most efficient to hire outside organizations to implement and complete many study activities. These organizations could range from local government agencies to parts of the university system other than medical centers to private or nonprofit companies. As noted above, to maintain close control and coordination of the study, the government may hire these organizations under contract so that they would be required to follow specified procedures and meet specified deliverables. Because of the likely contractual nature of the relationship, we refer to these organizations collectively as COs. For some tasks, such as data management and participant tracking, the government may consider using a relatively large CO that could assume responsibility for the entire task. On the other hand, individual COs may be hired to support specific centers or geographic areas with the full range of study activities. When an organization is awarded a contract to perform a function for the study, that organization should have the flexibility to bring in other organizations and/or consultants at the government's request or if that option is more efficient to conduct the work.

Although it is likely that each participating center will have staff with QC/QA experience, it may make sense for an independent CO to handle all QA/QC documentation and assessment activities for the study. Similarly, training is another QA function that could be centralized to ensure that all study workers are trained to perform study protocols in a consistent fashion. (These functions might also be performed under the Central Coordinating Centers.)

4.2 TARGETING ORGANIZATIONAL EXPERTISE TO STUDY ACTIVITIES

Each of the eight organization types discussed in Section 4.1 has certain attributes that are conducive to their contributing to the study in a certain way. For example, large medical centers have expertise in performing medical tests and examinations on children and pregnant women, obtaining biological samples, and following specific protocols to perform those activities. Thus, they may surface as the logical organizations to perform the biological and medical data collection activities. State and local health departments have insight into local environmental and health issues that study designers may need to be aware of. Also, there are some COs with significant experience in conducting national probability samples. If probability-based samples are a component of the study, these firms may offer advantages in implementing the recruitment of participants in both the national probability-based sample and the probability-based sample of MSAs surrounding purposively-selected Centers. During the study design process, the government will need to consider organizational capabilities relative to specific implementation plans chosen for the study.

4.3 RELATIONSHIP BETWEEN ORGANIZATIONAL STRUCTURES AND SAMPLING FRAME

In this section, we will review how each of the organizations discussed above may best work together within each of the four sampling approaches under consideration (as described in Chapter 3). Within each of the options, we review the likely organizations responsible for the major study activities that could shift depending on the frame. The seven major study activities we have identified are:

- Study design and start-up: including survey tool and protocol development, OMB/IRB approval, contract/grant awards, and training;
- Recruitment: including initial enrollment of participants;
- Data collection: including biological/environmental sample collection, questionnaires, and QA/QC;
- Sample retention and tracking: includes providing incentives, managing a help desk, retaining children (when they become older), and tracking participants that have moved;
- Sample analysis and storage;
- Data management and software development: includes database/software development, data entry, and website design and management; and
- Project management: involving overall project coordination.

4.3.1 Study Design and Start-Up

The Federal government itself is and will be heavily involved in study design and start-up. The government may find it efficient to obtain assistance with start-up activities from selected medical centers or COs – whether they are universities, private companies, or other government organizations working under contract to the Federal government. Federal participants may focus primarily on setting policy and guidelines for the study while the COs and Centers interpret that guidance into products such as statistical sampling designs with the appropriate tools for data collection; data collection instruments such as informed consent forms, questionnaires, environmental assessment forms, chain-of-custody forms, etc; training material; and other study implementation documents such as Quality Assurance Project Plans (QAPPs), Office of Management and Budget (OMB) Supporting Statements, and Human Subjects Review packages. The Federal government consortium and supporting committees will, at a minimum, review and provide feedback on all draft tools.

It also seems most efficient to have Federal government staff members or a supporting CO prepare the supporting documentation for OMB review of the study and all the supporting data collection instruments. Federal government staff members also will lead 1) the preparation and distribution of Requests for Proposal (RFPs) for all necessary grants and contracts required to support the study, 2) review of submitted proposals, and 3) awarding of all funds. A supporting CO potentially could assist with the process of issuing and reviewing RFPs to select study research centers. Similarly, a CO (perhaps a single organization with QA/QC expertise, as

mentioned earlier) could prepare appropriate training material to instruct data collectors on proper use of all data collection instruments and study protocols.

One facet of the study design that may go beyond the Federal government and its supporting COs is the process of obtaining Institutional Review Board (IRB) approval for conducting the study using the network of selected research centers. The Federal government can likely lead the preparation of material for an initial Federal-level IRB meeting and gaining of approval; however, many of the selected Centers will have their own IRBs that may need to provide approval before the study can begin at that Center. Thus, the Federal government can provide each Center with the material developed for the Federal IRB process and can participate in individual hearings with local IRBs, but Center staff will need to coordinate the IRB process within their own Center.

Figure 4-1 illustrates the study design and start-up activities.

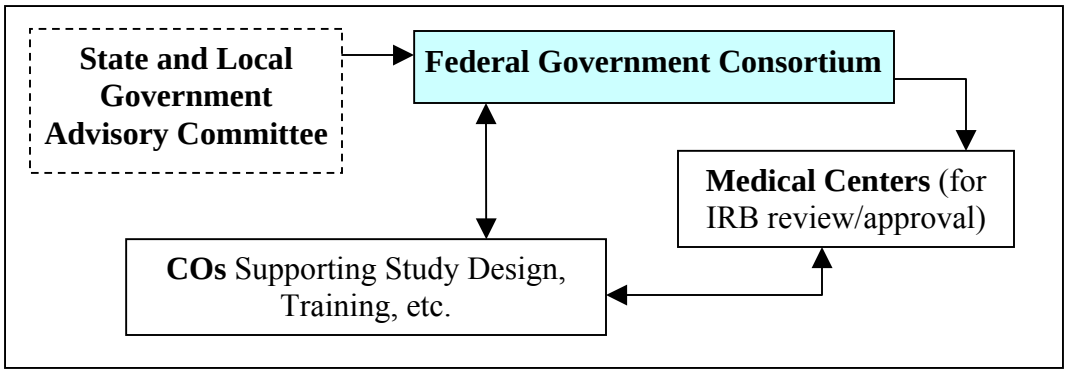


Figure 4-1. Primary Organizations Involved in Study Design and Start-Up

4.3.2 Recruitment of Study Participants

In contrast to the study design activities, the organizations involved in recruitment activities will likely differ depending on the sampling approach. For the probability sample of the geographic area around a Center (area PBS), the Federal government may want to encourage or require that Centers work with COs that specialize in performing recruitment activities for probability-based samples in their area. As some of the purposively-selected Centers will likely be located in cities with large populations that have already been selected in the national PBS, the CO may have recruiting infrastructure in place (or be able to establish the infrastructure quickly and efficiently) – trained interviewers, patient tracking capability, etc. – in the areas covered by the Centers. Thus, in the area PBS frame, a recruitment CO could perform the recruitment activities and pass the participant to the Center for data collection, in some cases. If this arrangement is not practical or feasible (or if the Center has established capability in participant recruitment), the Center would assume full responsibility for the recruitment effort.

Within a national probability-based sample (PBS), it may be most efficient for the government to hire a CO that specializes in survey sampling and recruitment of participants for large national surveys. Enrolling and retaining a sample of women/children through a national PBS will require data collection activities in a number of geographic areas (PSUs) around the country that are not necessarily associated with medical centers. As noted earlier in the report, the national PBS will involve random selection of households or physicians' offices in selected geographic areas. Regardless of the sampling approach utilized, the optimal organizational structure to conduct this type of sample would likely involve single or multiple COs that would assume responsibility for recruitment, data collection, and retention/tracking activities for the national PBS. Note that the CO for a particular area could be an organization such as a university that could lead the data collection effort, although not serve as a Center. The number of COs required may depend on the number and type of PSUs selected. An attractive quality of a relatively large, national CO specializing in survey operations would be its ability to efficiently mobilize staff in many of the geographic areas selected as PSUs for the national PBS. This primary CO could also subcontract to other COs in these areas where they do not have a current presence. The use of a centralized organization has the disadvantage of not having local or community connections that may be needed to maintain long-term retention and allow for ongoing scheduling of follow-up appointments.

To recruit pregnant mothers in the PBS of Center patients, it may be most efficient to allow individual physicians or their staff to handle the recruitment of their selected patients. If recruitment is done outside of normal visits to the office, nurses or other staff from within the office might be able to handle making telephone calls to selected patients to introduce the study and ask for their participation. Since physicians and their staff are generally not trained to recruit people into a study and may not have the time available to do so, they may not be the best people to explain the NCS and convince patients to participate. Physicians' offices performing this function may also experience problems in implementing a probability-based sampling approach, and in tracking the status of those who have been asked to participate over the course of the study. In place of physicians and their staff, a recruitment CO could also contact selected patients and enroll them in the study. The script or recruitment procedure might differ based on whether a participant's physician or an independent organization handled the recruitment, but the skills involved would be the same.

For recruiting pregnant women into an opportunity sample, the study coordinator within the Center will likely handle the recruitment efforts. When study coordinators encounter a patient with the characteristics needed for the opportunity sample (e.g., homeless women, women planning on using a midwife, institutionalized women), they can recruit them on the spot. Thus, they will need to have enrollment materials available in the office. Other physicians' offices in the same geographic area as a Center may assist in identifying patients for an opportunity sample. Center physicians can decide whether they have adequate numbers of patients within the Center to meet any associated quotas or whether they have to enlist the assistance of other physicians in the area to enroll the necessary number of participants.

Figure 4-2 illustrates recruitment activities.

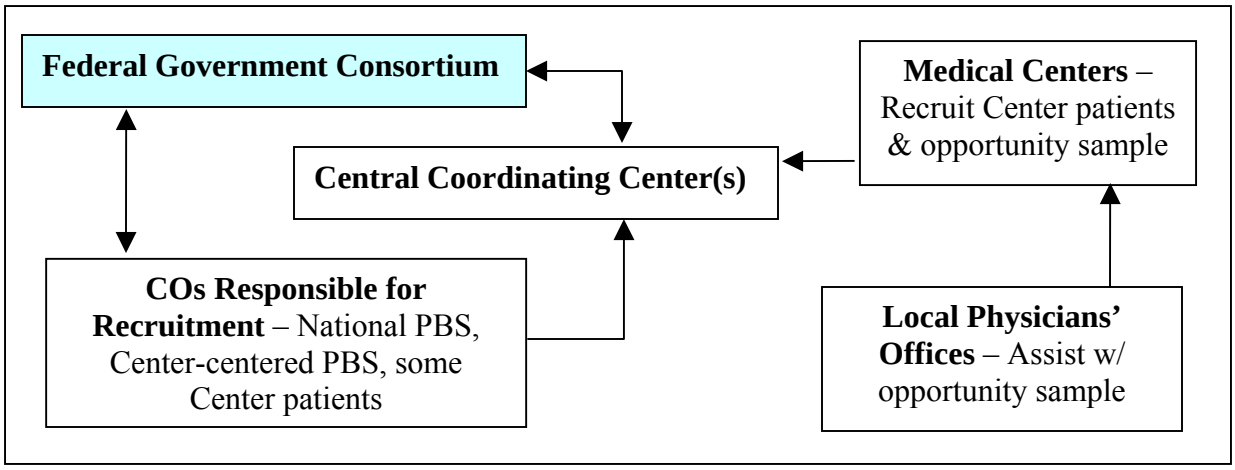


Figure 4-2. Primary Organizations Involved in Recruitment Activities

4.3.3 Data Collection

Many different data collection activities will be involved with the NCS data collection protocol – questionnaires, environmental sampling, physical exams, advanced medical tests, biological sampling, and psychological testing. Additionally, a significant effort will be required to coordinate the data collection, e.g., planning and scheduling appointments. The primary determinant of organizational responsibility for the data collection effort may be whether a participant is in a geographic area covered by a participating Center or not. In geographic areas not covered by a Center, it is likely that some type of CO will assume the responsibility of ensuring that all data are collected from each participant in a timely fashion. In geographic areas covered by Centers, it may make sense for the Centers to assume that role.

In areas recruited into a national PBS not covered by a Center, the data collection CO(s) (which may be the same as the recruitment CO) will coordinate all aspects of data collection. This may involve hiring subcontractors to perform environmental assessments and data collection or administer questionnaires; however, it may also involve working with local physicians and hospitals to obtain necessary medical tests and biological samples. For some portion of the medical testing, the CO may choose to set up its own NCS medical testing centers located within each selected PSU. In early stages of the NCS, when many measures are necessary in the early stages of life (e.g., preconception through early childhood), these testing centers would likely need to be staffed with full-time data collection personnel. In later stages of the study, when measures are less frequent and less time-sensitive, it is possible that these medical testing centers could be replaced by mobile testing centers such as those utilized in the NHANES study by CDC's National Center for Health Statistics (NCHS) and Westat. These mobile data collection centers could then efficiently travel to multiple PSUs throughout the year to perform data collection activities in a more cost-efficient manner.

For the other three sampling approaches that all involve the selected medical Centers (area PBS, PBS of Center patients, and a purposive sample of Center patients), the Centers would likely want to coordinate the data collection process. Center staff would handle the scheduling of appointments, and participants would physically travel to the Center to undergo the full range of data collection. The Center may have experience in environmental sampling, or could subcontract to an environmental assessment CO or to a local government agency to conduct the environmental testing at participant's homes. Likewise, a Center could choose to subcontract additional data collection work, such as assisting with completion of questionnaires/surveys. Local nonprofit community groups may be able to assist with ensuring that participants keep appointments and understand and provide all necessary information. Figure 4-3 illustrates the data collection activities.

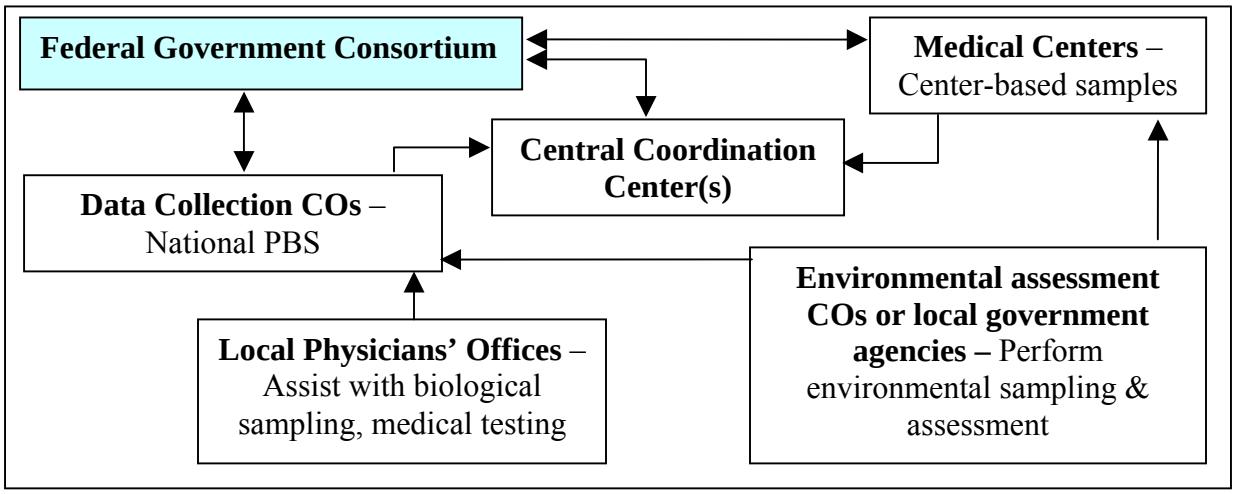


Figure 4-3. Primary Organizations Involved in Data Collection Activities

4.3.4 Retention/Tracking

Ensuring that enrolled participants remain in the study is vital to long-term success. Short-term measures for retaining participants include providing incentives and reimbursing for transportation to Centers or physicians' offices. Provision of these short-term retention measures will be led by the organizations coordinating the data collection activities, which were discussed in the section above. These organizations would likely be Centers and other data collection COs.

Long-term tracking of participants over the course of the study is a different type of activity which may involve tasks such as 1) annual (or more frequent) verification of addresses and contact information via telephone or postcard mailings, 2) searching for individuals/families via public databases when they are no longer at their previous address and have left no forwarding contact information, 3) notifying the data collection coordinating organizations of participants moving into and out of their areas, and 4) developing and providing regular newsletters and/or other mailings (e.g., birthday cards) to participants. These tracking and long-term retention tasks might best be conducted by the Central Coordinating Center (CCC) or a Participant Tracking CO hired by the CCC. The CCC or other responsible organization could

maintain a master database of all participants, develop a system to update the database with all contact information changes, notify appropriate organizations of changes and moves, and coordinate continued follow-up at new locations. Figure 4-4 illustrates the relationships of the organizations involved in tracking activities.

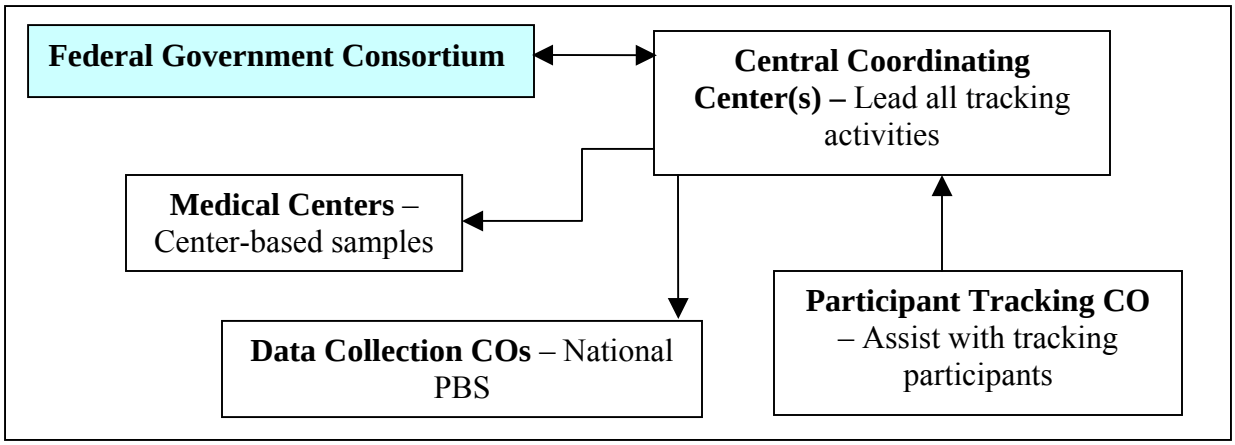


Figure 4-4. Primary Organizations Involved in Retention/Tracking Activities

4.3.5 Sample Analysis and Storage (Repository)

Given that specific technical skills are required for laboratory analysis of samples collected during the study, the pool of potential organizations conducting the work is limited. Three different types of laboratories might be considered for this work – private, government, or Center. As mentioned earlier in this section, the government will need to decide during the study design phase whether it is essential to use a single, central laboratory for archiving and analyzing all study samples (because of potentially better standardization and more efficient QA/QC) or whether it is preferable to use a distributed laboratory network. If a distributed network is used, the type of laboratory actually used by each PSU and Center may differ based on local resources. The Federal government may want to make another policy at the start of the study regarding whether a Center's own laboratory can analyze samples collected by that Center.

If a single laboratory is utilized, that laboratory would be directed by and provide results to the Central Coordinating Center. The CCC could decide whether the lab should also provide results directly to the originating Center or CO or whether the CCC should take responsibility for distributing results. In a distributed network of laboratories, it might make more sense to have each data collection organization coordinate the gathering of results from analysis of samples it collected. Regardless of the number of laboratories used, the government way want to hire a separate, independent laboratory to prepare QC samples that would be shipped to the various data collection organizations for blind insertion into the sample stream, and to conduct audits of analytical and data handling/processing operations. Figure 4-5 illustrates the sample analysis and storage activities.

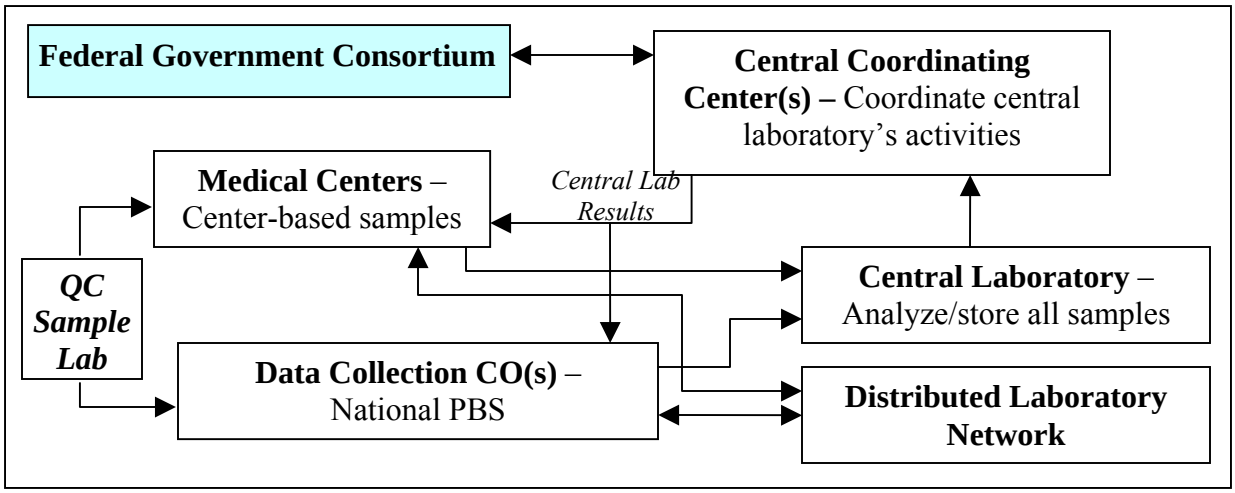


Figure 4-5. Primary Organizations Involved in Sample Analysis and Storage

4.3.6 Data Management and Software Development

The NCS will involve complex software development and data management in order to develop and maintain support systems, including a study website, and study databases. Regardless of the sampling frame, it may make the most sense to have a single organization coordinating all these tasks. This organization likely would be a single CO, potentially supported by other COs. The lead data management CO would interact with the Federal government, the Central Coordinating Center, the data collection organizations, and any laboratories providing electronic results. Utilizing a single data management CO would ensure that data are recorded and stored in a consistent manner across the study, whereas if each Center handles its own data management, problems would arise in trying to integrate data from across the study. Figure 4-6 illustrates the data management and software development activities.

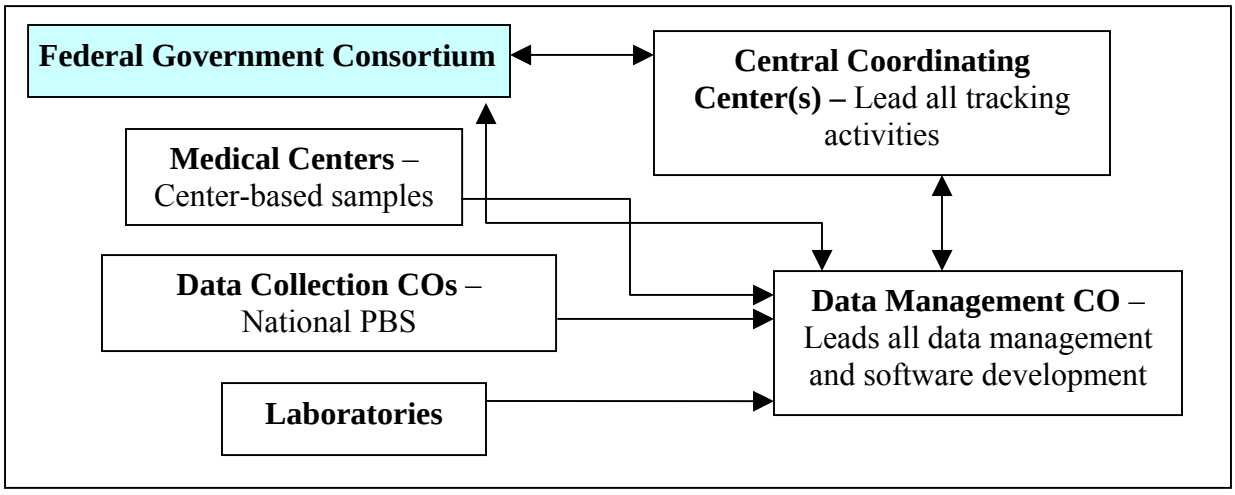


Figure 4-6. Primary Organizations Involved in Data Management Activities

4.3.7 Project Management

Overall project management duties will largely be handled by the government and by its Central Coordinating Center(s). The government needs to ensure that the study is proceeding as planned and monitor implementation costs. The lead agency may wish to provide regular status reports to the other involved agencies and to Congress, as required. The Central Coordinating Center will monitor all selected Centers and COs to ensure that they are 1) meeting enrollment and participation goals, 2) following standard methods and protocols, and 3) spending appropriately to complete their required tasks. The CCC will maintain frequent communication with both the government and all the data collection organizations. The only situation that might not require a CCC to oversee study implementation would be if a single organization were responsible for all data collection activities, e.g., in a complete national probability sample design. We assume, however, that the full study design will include Centers responsible for some portion of the data collection. Figure 4-7 illustrates project management activities.

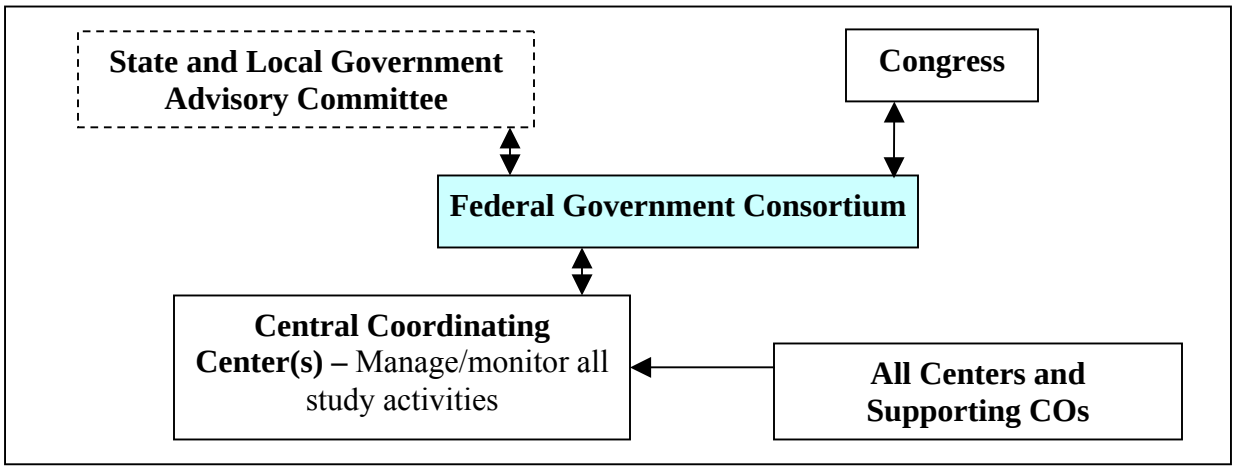


Figure 4-7. Primary Organizations Involved in Project Management Activities

5 IMPLEMENTATION DETAILS FOR DIFFERENT SAMPLING APPROACHES, OVERSAMPLING, AND DESIGN EFFECTS

In many cases, large-scale national surveys of health, environmental, and even housing and economic conditions employ probability sample designs based on sampling frames that include a large percentage of the population of interest. Probability designs are chosen so that estimates of population characteristics (e.g., prevalence of a disease, exposure, or condition) will not depend on untestable assumptions (see CDC 1992, USEPA 1995a, USEPA 1995b). On the other hand, epidemiological studies of the effect of a treatment or exposure on a disease or health outcome do NOT always employ probability sample designs. When they do use probability sample designs, the sampling frame considered is often restricted to a small, manageable proportion of the population. The NCS is an epidemiological study in that estimation of the relationship between environmental exposures (with environmental being broadly defined as chemical, physical, biological, and psychosocial) and biological, emotional, social, and behavioral outcomes is the goal, rather than estimation of population characteristics, such as the distribution of an exposure or prevalence of a disease (Branum et al., 2002). There is, however, a strong sentiment that probability sampling methods should be employed if feasible within the NCS to ensure that the exposure-outcome relationships are not biased from known or unknown factors that may be introduced by limiting the selection to very small segments of the population or by relying on volunteerism.

The purpose of this chapter is to discuss methods for implementing probability sampling strategies within the NCS, and to discuss the effect that these strategies may have on estimation of exposure/outcome relationships (e.g., through calculation of design effects). In particular, since the family of designs outlined in Chapter 3 includes the possibility of national sampling, sampling through purposively selected qualified centers, and combinations of the two, we discuss methods for each of these approaches here. Section 5.1 discusses probability sampling strategies based on the selection of qualified centers, and Section 5.2 discusses national probability sampling strategies and their implementation. As discussed in Chapters 2 and 3, combining these two sampling frames (i.e., combining multiple probability sampling approaches) within the NCS may offer important advantages by capitalizing on the strengths of the various frames while minimizing the disadvantages associated with each individual frame. Thus, Section 5.3 goes on to discuss the methods used in combining the multiple frames (e.g., calculation of the weights for each individual). Finally, Section 5.4 provides a brief discussion of the issue of oversampling certain population groups, such as highly exposed populations or certain spatial regions of interest, and Section 5.5 provides a discussion of the calculation of design effects for estimation of relationships between an exposure and a health outcome.

5.1 REVIEW OF SAMPLING STRATEGIES CONSIDERED FOR SAMPLING WITHIN PURPOSIVELY SELECTED QUALIFIED CENTERS

In this section we discuss sampling strategies based on the selection and funding of a small number of health care institutions, termed "Centers." The Centers would be responsible for the selection of study participants from within their geographical areas. Each Center can be

thought of as a primary sampling unit (PSU) and is responsible for recruiting a number of children, prenatally or even prior to conception, for the NCS. The Center's geographical area could be defined as the county or the MSA of the Center, depending on the desired size of the population for each Center.

This sample design represents a two-stage probability sample, with Centers (or MSAs corresponding to Centers) selected at the first stage and households within the Center's geographical area or Center patients selected at the second stage. For this type of design, the sampling frame could be narrowed for both the first- and second-stage sampling steps, and it is important to consider whether this narrowing is acceptable. For example, the Centers (or PSU) sampling frame would presumably be limited to a small set of qualified Centers. Assuming sampling of the Centers geographical area is conducted at the second stage, only children born within the geographical areas of these qualified Centers would be in the sampling frame due to this purposive narrowing of the first stage sampling. On the other hand, if Centers identify study participants by only sampling from their existing patient lists, then only children of current patients of the Centers are in the sampling frame population.

The main advantage of the Center model is that these medical centers likely have experience in coordinating large, multi-site, longitudinal research projects as well as small specialized studies requiring access to certain populations of interest (e.g., mothers experiencing a variety of exposures). Additionally, they have likely demonstrated the capability to successfully recruit and retain a diverse population of study subjects, and will presumably have the infrastructure needed to ensure efficient data collection and adequate quality control practices. The remainder of this section discusses the methods used for selecting qualified Centers (Section 5.1.1), and the methods used for sampling individuals within those Centers (Section 5.1.2).

5.1.1 Selection of Centers

If a Centers approach is used for the NCS, then selection of the qualified Centers will likely be performed through a competitive procurement process, in which the Centers would demonstrate their ability and capacity to perform appropriate data collection activities. Since a formal process is not available at this stage, as a surrogate for this type of merit-based selection a list that includes 105 medical research institutions with affiliated hospitals and their total dollar amount of National Institutes of Health (NIH) research grants (averaged over 2001 and 2002) was obtained from *U.S. News and World Report* (Table D-1, Section D-2 of Appendix D displays these 105 Centers). Figure 5-1 illustrates the distribution of these Centers across the U.S. (note that one of the institutions is in Puerto Rico and is not displayed in Figure 5-1). Note that the list includes only university-based medical research centers that have had National Institutes of Health (NIH) grant funding during FY2002 so that some large medical centers not affiliated with a university, like the Cleveland Clinic, and some research universities not affiliated with particular hospitals, like the University of California, Berkeley are excluded from the list. (Note that we do not assume that these respected institutions would be excluded from participating in the NCS. However, as a starting point for understanding the portion of the population that might be within the geographical area of a Center, the above list, although not a

comprehensive enumeration of every institution that might compete to participate in the NCS, was utilized.)

As mentioned above, for each of these 105 Centers the total dollar amount of NIH research grants awarded to the medical school and its affiliated hospitals (averaged over 2001 and 2002) was available. Additionally, merging Census data with the list of Centers provided counts of the total population, the population of children 0 to 3 years of age, the population of females of child-bearing age (15 – 44), and the total number of households in the respective counties and MSAs where the Centers are located. Finally, data on the annual number of births at each Center were obtained directly from each Center for all but 34 of the Centers.

Most of the Centers on the list are teaching hospitals located in the inner city, and, thus, sampling from only these Centers may overrepresent certain populations. Additionally, some of the Centers are specialized for high-risk patients, again possibly overrepresenting certain populations. It is likely the case that including more Centers on the list (e.g., all medical Centers of a certain size) and stratifying these centers based on geographic areas (north, south, east, west) and possibly by socio-demographic or socioeconomic groups would remove some of these biases, and is likely an important consideration if a Centers type model is utilized for the NCS. Nevertheless, even with the limited list utilized here, 55% of the U.S. population lives within the MSA of one of the 105 University Centers and 26% lives within the county of one of the Centers. Thus, a Centers approach to the NCS has the potential to include a relatively large percentage of the U.S. population.

In terms of the selection of Centers for the family of designs considered in this report, several methods, some of them purposive, have been considered. Probabilistic approaches included PPS sampling from the list of Centers, with size defined by 1) total amount of NIH funding, 2) number of households or children aged 0 to 3 years in the geographical area, and 3) number of births annually at the Center (see Section D-2 of Appendix D for an example). Each of these criteria could also be applied purposively to pick the top ranking Centers. As mentioned above, a procurement process would be the likely means of selecting Centers for the NCS, in which case the selection process would not produce a probability sample. [By not selecting the Centers probabilistically, the sampling frame of the Center approach is inherently reduced to only the portion of the U.S. population that is in the MSA (county) of one of the *selected* Centers.] However, it was decided to focus Center selection on purposive methods since this may be a more likely scenario. In particular, for the sampling plans outlined in this report, Centers with the most annual births were selected from the list of Centers. It is assumed that each Center can recruit 2000 study participants over the course of the four-year recruitment period, and follow those participants over the course of the study. For example, if the sample design called for 50 percent of the cohort to be selected in the Centers frame (i.e., 50,000 individuals), then a total of 25 Centers would be assumed.

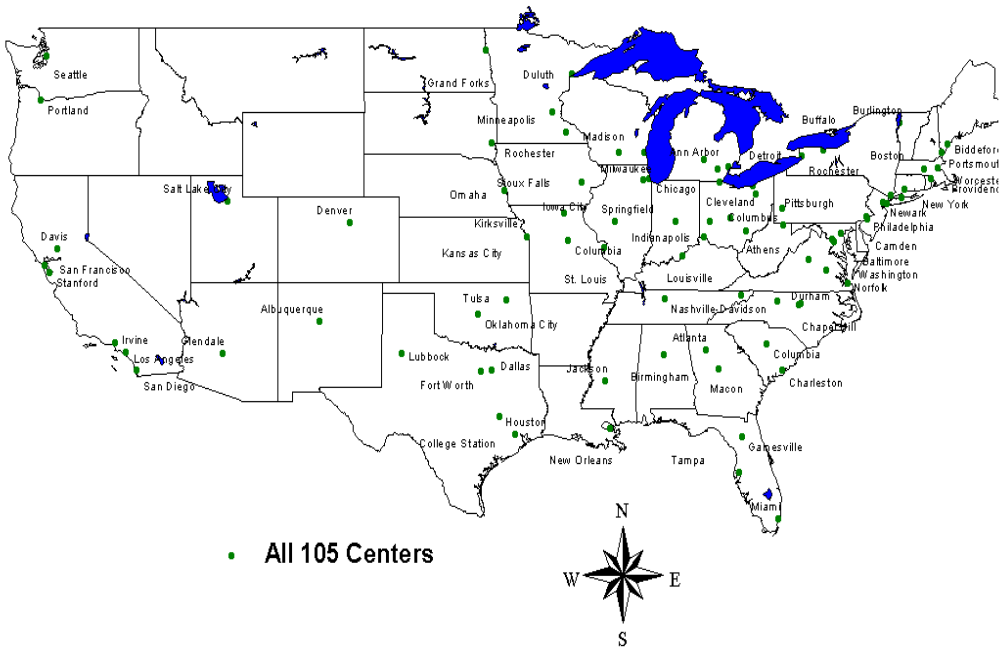


Figure 5-1. Map of U.S. Indicating Locations of All 105 Centers.

5.1.2 Selection Within Centers

To select NCS study participants within a selected Center, two probability sampling approaches have been considered: probability sampling within the geographical area corresponding to the Centers and probability sampling of Center patient lists. The area sampling approach is appealing due to the large percentage of the U.S. population that would be included in the sample frame (under the assumption that each of the 105 university centers has a non-zero probability of selection, this sampling frame includes 55% of the population based on MSA geographical areas and 26% of the population based on county geographical areas). On the other hand, sampling of patient lists may be appealing as it is easier to do and may result in higher recruitment and retention rates due to pre-established doctor-patient relationships. As mentioned above, the annual birth rates for each of the Centers (note that for the 34 Centers with missing data, we imputed the median number of births at the other Centers to estimate their annual birth rates) were obtained and suggest that approximately 573,438 births occur per year at the105 Centers. The CDC estimates that approximately 4 million births occurred in the U.S. in 1999 (CDC 1999), and thus, approximately 14% of the births in the U.S. are covered within the

sampling frame if probability sampling of the patients from the 105 Centers is used to identify NCS study participants.

We assume that a portion of the patients sampled at each Center are sampled using area sampling of the Center's MSA with all households having an equal probability of selection (i.e., we assume random sampling of the Center's MSA; however, a multi-stage design could select area segments of the MSA, such as census tracts, and select households within these area segments). Thus, the chance of inclusion of a household is equal to the number of households selected by this method divided by the estimated number of households. The number of pregnant women over the recruitment period is estimated as the number of households in the geographical area divided by 12 using the reasoning of Section 5.2.1 to determine that one live birth is expected to result from 12 households. The remaining subjects are sampled with equal probability from the Center's patient list or through convenience sampling as described in Chapter 3 of this report. Thus, the chance of inclusion assigned to these subjects is equal to the number of subjects selected by this method divided by the size of the patient list. Figure 5-2 provides histograms of sample weights for three Center designs where we allow the portion of the sample selected using probability sampling from the Center's MSA to be 25, 50, and 75 percent. Note that the sample weights for the Center designs are quite variable, because Centers are each assumed to sample 2,000 patients, regardless of the size of their geographical area and because, within each Center, list-based patients have much smaller weights than area sampled patients. Of course, alternatives to this approach, such as selecting the number of subjects proportional to the number of births in the Centers geographic area (i.e., considering the geographic reach of each Center), are also plausible, and may offer a means of reducing the variability in the sample weights associated with the Center designs.

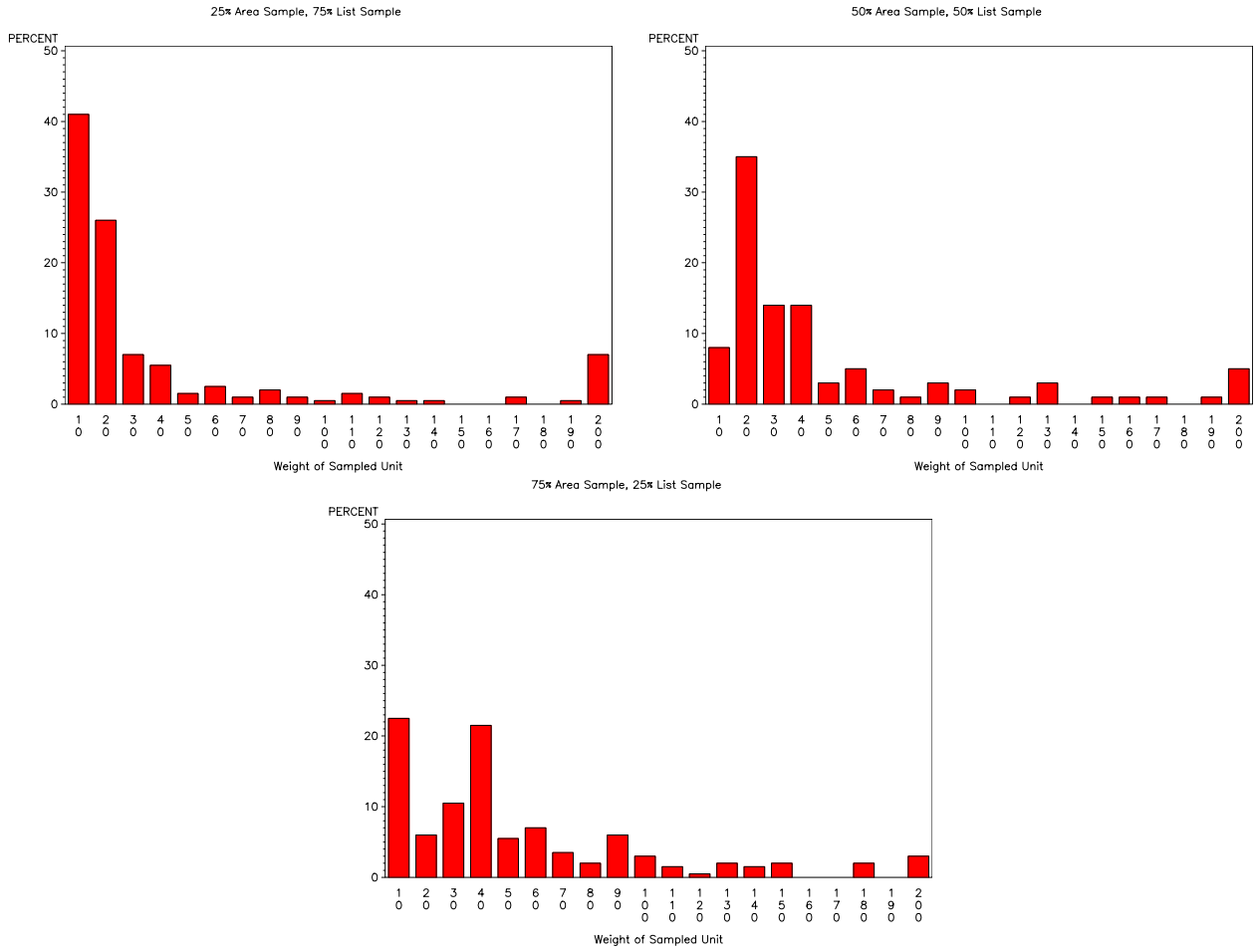


Figure 5-2. Histograms of Sample Weights for Three Center Designs

5.2 REVIEW OF SAMPLING STRATEGIES CONSISTENT WITH A NATIONAL PROBABILITY-BASED SAMPLE

The starting point for a probability sample is identification of the reference population and the sampling frame needed to capture (as much as possible) this reference population. Given the sampling frame, the next step is identification of an efficient sample design for selecting elements from that frame. The sample design assigns a known probability of sample inclusion to each element in the sample frame, and typically represents a compromise between selecting the sample in a manner producing the most information from the resultant data and selecting the sample in a manner that aids the data collection process (e.g., through reducing the cost of collection). In this section, we generally consider designs in which the sampling frame (theoretically) includes all infants born in the United States, or in selected areas, during the three to five year NCS recruitment period.

In general, multistage designs based on area sampling frames are chosen for national surveys that require in-home or medical center/clinic data collection from each participant. The

reason for this typical approach is that it aids in the data collection process by localizing significant numbers of participants to small areas. It should be noted, however, that there is a "cost" associated with this type of sample design. The cost is the result of clustering the participants in small areas which often produces a loss of information when compared to a simple random sample of the same size (assuming there is some cluster effect). This loss of information can be characterized by the calculation of a design effect, which essentially compares the loss of information due to the sample design employed to that resulting from the use of a simple random sample of the same size. In other words, the design effect can be thought of as the loss incurred for using the selected design instead of a design that selects a simple random sample. Thus, large design effects mean that a larger sample will be required to obtain information comparable to that obtained using simple random sampling. Section 5.5 includes further discussion of design effects and their calculation.

Currently two multistage probability-based sampling strategies for identifying potential participants in the NCS have been proposed (Westat, 2002). One is a multistage area sample, which identifies and screens households to identify pregnant women for study participation. We refer to this strategy as the Household Model. The other strategy is a multistage list-based sample that identifies pregnant women for study participation through patient lists of registered physicians that practice obstetrics. We refer to this strategy as the Physician Model. In this report, the Household Model for recruiting study participants within selected PSUs of the national probability-based sample and the probability-based sample of MSAs surrounding purposively selected Centers was assumed for the cost and power analyses, out of convenience. However, it should be noted that there are several other options available for the probability-based selection of study participants within selected geographic areas that may be more optimal with respect to implementation of the NCS. For example, the Physician's Office Model may be a more resource efficient method for identifying women planning pregnancy or in early stages of pregnancy. In addition, use of alternative methods for the probability-based selection of study subjects within selected locations will result in similar power for addressing study hypotheses, assuming no substantive difference in rates of attrition and similar coverage of the target population.

The remainder of this section is concerned with reviewing the possible methods for implementation of the Household Model and any corresponding implications. Specifically, Section 5.2.1 is concerned with the selection of primary sampling units (PSUs) and of households within PSUs (i.e., methods for sample selection). Section 5.2.2 continues by discussing the determination of sample weights (i.e., inverse of the probabilities of selection) for each unit sampled.

5.2.1 Sample Selection

In the typical Household Model, counties are selected as PSUs at the first stage; segments, such as combinations of Census blocks, are selected at the second stage; households are selected within segments at the third stage; and all age-eligible women in the household are selected at the fourth stage. Infants of age-eligible women who are pregnant at the beginning of the study or become pregnant within four years of follow-up become part of the sample of participants (assuming a four-year NCS recruitment period). As discussed in Chapter 3, the

family of designs calls for selection of a portion (e.g., 25,000 subjects) of the cohort in a national probability sample. For the 25% national PBS (NPBS) example, in order to obtain a sample of 25,000 live births in the household model, we assume that 90% of pregnancies result in live births, 94% of infants participate in the study, the fertility rate of women aged 15-44 years is 65 pregnancies per 1000 women (per year) (http://www.cdc.gov/nchs/fastats/births.htm), women have the same chance of getting pregnant each year, 80% of women agree to participate in the study, there are 568 women aged 15-44 per 1000 occupied households, 95% of occupied households can be screened, and 88% of households are occupied (see Westat, 2002 report). Based on these assumptions, approximately 300,000 households must be selected in order to obtain 25,000 participants [25,000/(0.90 * 0.94 * 0.065 * 4 * 0.80 * 0.568 * 0.95 * 0.88) ≈ 300,000]. In other words, we assume that there is approximately 1 live birth for every 12 households contacted. (Note that this ratio could be increased if follow-up was focused on households with women that intended to become pregnant over the next several years. Additionally, note that the above estimates should be further verified as suggested in Chapter 10)

It should be noted that there are alternatives to selecting counties as the PSU. One possible alternative is to use metropolitan statistical areas (MSAs), or groups of counties, as the PSU. However, since these areas may be deemed too large for the NCS (e.g., since travel times across an MSA may be significant and this could adversely affect study response rates), the more compact counties were selected as the PSU to reduce the possible travel time for sampled persons. Additionally, as noted above, it may be desired to limit selection of PSUs (counties) to a subset that have desirable properties relevant to the NCS (e.g., have a major medical institution). Further evaluation of the PSU issue may lead to possible alternative definitions for the PSU; however, at this stage we assume counties are the appropriate PSU level elements for the national probability-based sample component in our family of designs.

To further investigate the Household Model approach to sampling, and to identify the sample frame for selection of PSUs, data on each of the 3,140 counties in the United States, including Hawaii and Alaska but excluding Puerto Rico, were obtained from census data files. These data include number of households, region, and an indicator of whether or not the county is within a MSA. A stratified sample design is proposed for PSU sampling in order to ensure the desired representation of different kinds of PSUs. We consider two plausible stratification variables that may be appropriate: region of the country and county urbanization (see Table 5-1). Admittedly, there are likely other stratification variables that will ultimately be deemed important for the NCS, and will need to be incorporated in the sampling framework; however, since these variables have not yet been identified, the examples presented here consider just these two stratification variables. Thus, we begin the design process by stratifying the selection of PSUs into eight strata: four levels of region (East, Midwest, South, and West) and two levels of urbanization (MSA versus non-MSA, representing urban and rural areas, respectively).

Table 5-1. Example Stratification Factors for Selection of PSUs

Factor	Levels
Region	East, Midwest, South, West
Urban	MSA, non-MSA

In addition to stratifying the sample to ensure coverage of the entire nation and coverage of both rural (non-MSA) and urban (MSA) areas, another design objective that we pursue is an approximately self-weighting sample (i.e., ultimately we would like sample weights of study participants to be nearly equal). To support this objective, the total number of PSUs available to the design will be allocated to the eight strata proportional to each stratum's size (with size defined as the number of households in a stratum).

One approach to determining the sample size within each selected PSU is to attempt to select approximately the same number of study participants within each PSU. This approach may offer certain efficiencies since following the same number of study participants would likely require similar infrastructure in all the PSUs, making study planning somewhat uniform across the PSUs (although note that population density could also play a significant role in the necessary infrastructure as some PSUs may be very spread out and may require mobile examination centers). Additionally, this approach attempts to avoid setting up infrastructure to support data collection for only a small number of participants (e.g., if sample size proportional to population size were utilized instead). To allow selection of equal numbers of study participants in each PSU, while maintaining the objective of a self-weighting sample, within strata probability proportional to size sampling (PPS) of PSUs is utilized. In other words, by employing PPS sampling of counties and selecting equal numbers of study participants within each PSU, an approximately self-weighting sample is obtained.

We offer the following simplified example of how participants might be enrolled within a single PSU to illustrate the complexities that will be involved in planning the required infrastructure to support the data collection. In a longitudinal study, sample size accumulates quickly and the data collection issues compound even more quickly. Assuming for the moment that the period of enrollment for the NCS is four years, by the end of the period there may be an operational problem of keeping track of the sample, the age of the sample, and who is getting what data collected at any given point in time. For any specific design model, this has implications for type and level of staffing required for the data collectors, and, thus, implications on the costs. There are also possible implications for the feasibility and quality of the various types of data collection required in this time period. These complexities are one reason that developing a single consistent plan for study infrastructure that can be used as uniformly as possible across PSUs is advantageous; however, there remains the need to provide some flexibility in order to respond to local conditions and situations.

Assuming 25 percent of the cohort would be selected in the national probability-based sample, and assuming 100 PSUs will be utilized, the number of households per PSU would be on the order of 3000 (since approximately 300,000 households would lead to 25,000 live births) while the number of pregnant women per PSU would be on the order of 280 (producing approximately 250 live births per PSU). Additionally, assuming uniform allocation of gestational month, approximately six pregnancies would occur each month (i.e., six women would be enrolled each month per PSU). By December of the first year, approximately 65 women/participants would be enrolled, and an additional six women would be expected to be pregnant during the month of December. Here, the study operations would require age-specific data collection for the children born prior to December (the longitudinal follow-up), data collection for the currently pregnant women, data collection for the births that occur during

December, and continued tracking of the women in the remaining households to determine when/if they become pregnant. At the end of four years, approximately 280 pregnant women will have been enrolled in each PSU, and approximately 250 live births will have resulted. At this point, any remaining pregnant women and any remaining households would become "out-of-scope" and would be dropped from the birth cohort study. Thus, the data collection effort, as well as the type and number of data collection staff, could be estimated, and could provide a means of identifying the type and number of staff required to implement any particular design option. Additionally, depending on the feasibility of the presumed data collection effort, certain design options may become impractical.

Thus, in general, PPS sampling of PSUs and selection of an equal number of study participants within each PSU will be conducted for these examples. However, there are two conditions where this approach falters. First, there is the case where one of the selected PSUs is too "small" to provide the desired number of participants in each PSU. An approach to dealing with these "small" PSUs is outlined in Section 5.2.2. Second, there is the case where the use of PPS sampling results in some large PSUs being selected with certainty (i.e., their probability of selection is one), which we then refer to as certainty strata. This occurs more and more frequently as the number of PSUs increases (i.e., there are more certainty strata as the number of PSUs increases). Fortunately, for the 50 and 100 PSU designs considered here, there a very few certainty counties. Table 5-2 displays the certainty counties for the 50 and 100 PSU samples. Note that there is only one certainty county when 50 PSUs are sampled (Los Angeles County), and there are four certainty counties when 100 PSUs are sampled (Los Angeles, Cook, Harris, and Maricopa counties).

Table 5-2. Certainty Strata for 50 and 100 PSU Designs

Design	Region	Urbanicity	County Name	Number of Households
50 PSU	4	Urban	Los Angeles County	3,270,909
100 PSU	4	Urban	Los Angeles County	3,270,909
	2	Urban	Cook County	2,096,121
	3	Urban	Harris County	1,298,130
	4	Urban	Maricopa County	1,250,231

Within PSUs, segments are usually selected at the next stage of the design. In many cases, census blocks, or combinations of census blocks, are the segments selected within the PSUs. Generally, the segment would be defined to consist of a large enough area to yield the desired segment sample size without too great a concentration of sampled elements within the segment (note that this may have a number of ramifications for assessment of common environmental conditions and having a sufficient number of participants in an area for logistical reasons and to engage the local communities). Of course, sampling of segments could also be stratified (in addition to the stratification of the PSU sampling) to ensure appropriate representation of different types of segments within the selected PSUs. However, since the basis of the segment stratification has not been defined at this point, we assume only equal probability of selection for each household within a PSU (i.e., within a county a PBS is used), and, at this

stage of design development, we do not attempt to explore the effects of clustering within PSUs due to possible segment sampling.

5.2.2 Sample Weighting

As mentioned above, the sampling design has two basic objectives:

1. Achieve a self-weighting sample and
2. Sample the same number of study participants within each PSU

The goal of a self-weighting sample must be slightly compromised to deal with integer allocation of PSUs across strata (i.e., only an integer number of PSUs can be selected in each strata), integer allocation of households within counties, and the certainty strata. On the other hand, the goal of sampling the same number of households within each county had to be compromised to deal with counties that are too small to support the sampling requirements of the design (i.e., counties with fewer households than the number that must be sampled). One approach to dealing with these very small counties is to combine them with contiguous counties to form a larger pseudo-county that would have a large enough population to support the desired sample size. However, this approach was not pursued here since it may be necessary, especially in rural strata, to combine large land areas in order to provide sufficient numbers of households to support the desired sampling (this is particularly problematic when the total number of PSUs available to the design is small, e.g., 50). Instead, when a small county is selected as a PSU, the number of units selected in that county is made as large as possible given the population limitations for the county, and the sample size of the other PSUs is increased by an appropriate amount to still obtain the desired overall sample size.

Table 5-3 provides a description of the distribution of counties and households across the eight strata (see Table 5-1). Supposing that 25 percent of the cohort is selected in the national PBS (i.e., 25,000 individuals), if the sample is selected from 100 PSUs, approximately 3,000 households must be screened in each selected PSU to identify 250 pregnant women per PSU. Clearly in the rural Midwest, South, and West regions these sample sizes may not be possible in some sampled PSUs, due to their small population/household counts. Additionally, it should be noted that as the proportion of the cohort selected in the national PBS increases, the number of counties that have population/household counts that are too small also increases. This indicates that it may be necessary to define slightly larger PSUs (e.g., combinations of several contiguous counties) in areas with small population/household counts (e.g., rural areas), or increase the number of PSUs planned for the rural portion of the national probability-based survey (having obvious cost implications).

Table 5-3. Distribution of Counties and Households Across Strata

Region	Urban	Number of Counties in Stratum	Number of Households		
			Mean	Min	Max
Northeast	Rural	120	52,991	4,592	353,022
	Urban	97	163,109	12,713	930,866
Midwest	Rural	834	9,139	273	46,438
	Urban	221	87,518	2,424	2,096,121
South	Rural	1016	10,822	174	93,070
	Urban	407	77,118	2,702	1,298,130
West	Rural	348	10,702	172	81,730
	Urban	97	212,927	3,767	3,270,909

There exists a rich literature for dealing with issues related to nonresponse during initial recruitment, summarized briefly in Appendix E. For the purposes of discussion here on sample weighting, we ignore the issue of nonresponse. That is not to suggest that recruitment nonresponse will be a trivial issue for the NCS – in fact we anticipate this to be an important problem that will require careful consideration. However, the discussion in this section focuses on assignment of sample weights and components of sampling design that attempt to minimize any design effects due to unequal selection probabilities.

Ultimately, the weight assigned to each sample unit is the inverse of the product of the PSU probability of selection and the household probability of selection within the PSU. Although the objectives outlined above specify that an attempt is made to obtain a sample with equal weights for all individuals, the resulting quantities are not all equal due to the considerations outlined above. Figure 5-1 provides two histograms of the sample weights for example designs including 50 and 100 PSUs. In the sample of 50 counties, most sample units had weights of around 85, while the units selected in smaller counties had weights up to approximately 780 (note that the histograms are truncated at a weight of 300). In the sample of 100 counties, most units had weights of around 90 while the units selected in smaller counties had weights up to approximately 500. The degree of unequal weighting displayed in Figure 5-3 will likely have an effect of reducing power in the NCS (at least for the national PBS portion of the design). Normally, additional work would be conducted to minimize variability in survey sampling weights to the extent possible in the national PBS portion of the design (i.e., there is likely little need to have a wide range of weights when selecting a national probability sample). However, under the family of designs concept, the national PBS is just one of several sampling approaches that will be conducted, and we expect there to be large variations in weights across the cohort. This feature of the family of designs will likely allow us to relax the normal survey design preference for equal weighting across the design.

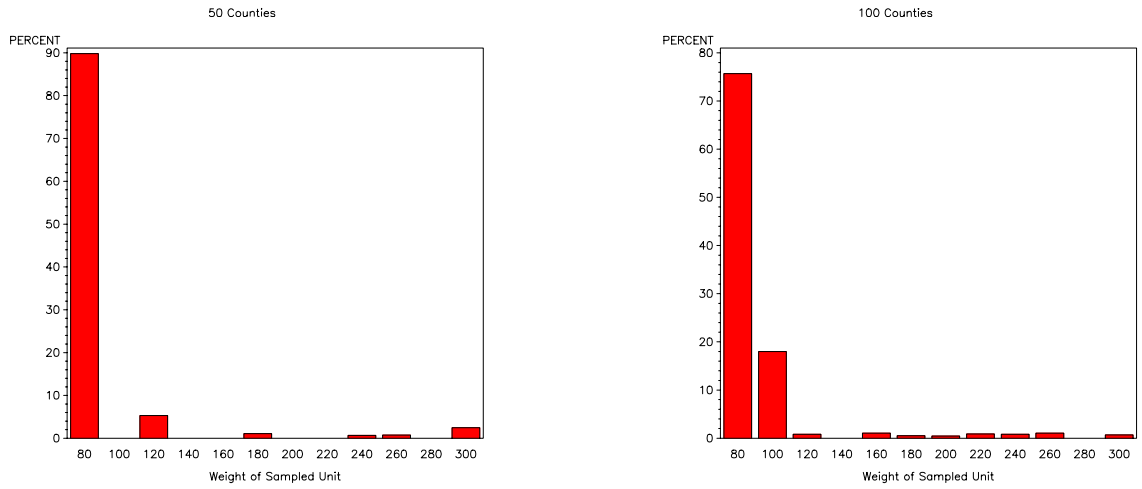


Figure 5-3. Histograms of Sample Weights for Two Example Designs

5.3 COMBINING INFORMATION ACROSS MULTIPLE PROBABILITY-BASED SAMPLES

Instead of simply selecting NCS participants using just one of the above approaches, the family of designs described in this report calls for selection of NCS participants using a mixture of national probability sampling and purposively selected qualified Center sampling (i.e., a multi-frame sample). Here, we provide a brief review of this family of designs, and identify the designs that will be considered in the power analyses of Chapter 9 of this report (Chapter 3 provided a more detailed description). Figure 3-1 in Chapter 3 illustrates the selection of a cohort using both of these approaches to sampling. The family of designs begins by splitting the cohort into two recruitment arms by assuming that a portion of the cohort will be selected from a national probability-based sample, and the remaining portion of the cohort will be selected from a set of purposively selected Centers. Thus, the first design parameter in Figure 3-1 is the fraction of the cohort, denoted as P_1, that will be selected using the national probability-based sampling (national PBS) approach described in Section 5.2. On the other hand, the individuals (the 1-P_1 fraction of the cohort) selected from a set of purposively selected centers (note that we assume a total of 2000 individuals will be selected from each Center) are further split into two recruitment arms with a fraction (P_2) of the individuals selected using a probability-based sample of the Center's metropolitan statistical area (MSA), and the remaining portion selected using probability-based sampling of the Center's patients. Note that Figure 3-1 also allows a fraction of the cohort to be selected using convenience/opportunity sampling; however, in all but one of the designs considered here we focus more directly on the probability sampling aspects of the design by assuming all participants are selected probabilistically (see discussion below). Thus, by defining the fraction of the cohort selected in the national PBS (i.e., specifying P_1), and defining the fraction of the Center sample selected from the Center's MSA (i.e., specifying P_2), the number of individuals selected from each of these sampling frames is specified. In the designs considered here we allow P_1 to take values of 75, 50, 25, and 0 percent, and we allow P_2 to take values of 75, 50, 25, and 0.

To complete the specification of the designs, the number of PSUs utilized in the NPBS (recall that we consider counties the PSU in the NPBS) and the number of purposively selected

Centers utilized must be specified. We consider two levels for the number of PSUs in the NPBS sample, 50 and 100 PSUs. Additionally, it should be noted that the example NPBS included in this report used proportional to size sampling from eight strata which represented four regions of the country and rural or urbanized counties within those regions. Currently, the NPBS selects a larger number of counties in urban areas due to this proportional to size criteria for selecting counties as PSUs. However, in subsequent refinement of the NCS design, we may wish to oversample rural areas due to the fact that urban areas will most likely be over-represented in the fraction of the cohort that is center-based.

Finally, under the assumption that each Center can recruit and follow 2000 individuals, the number of Centers utilized is defined by the portion of the cohort selected from the Centers (P_1). For example, if 75 percent of the cohort is selected from Centers (i.e., 75,000 subjects), then the Centers sample would be spread over a total of 38 Centers. These design specifications identify a total of 22 designs that will be considered in the power analyses of Chapter 9. Admittedly, these specific designs (i.e., the specific settings of the portion of the cohort selected in the NPBS, the portion selected in the Centers, the number of PSUs, etc.) do not cover all of the possibilities in the family of designs. Rather, they are meant to provide a range of possible designs so that an indication of the effect of changing the various design parameters can be obtained.

As mentioned above, the family of designs outlined in Chapter 3 also allows a portion of the cohort to be selected as a volunteer sample; however, the 22 designs described above concentrate more directly on the probability-based sampling aspects of the family of designs by assuming that the entire cohort will be selected in some probabilistic manner (i.e., national probability-based selection, Center MSA probability-based selection, or Center patients probability-based selection). In other words, these designs ignore the small portion of the cohort that is obtained through volunteerism. Assuming that the volunteer portion of the cohort is relatively small (e.g., 1 or 2 percent), this lack of consideration of volunteer subjects is justified since the cost and power analysis results of Chapters 8 and 9 are highly dependent on the choice of the P_1 and P_2 parameters, and relatively insensitive to the specific choice of the P_3 parameter in the Family of Designs. (Recall that the choice of P_3 helps to differentiate the number of participants recruited via a convenience or opportunity sample within the $(1-P_2)$ remaining fraction of center-based participants.) This is not to say that we consider volunteerism an unimportant aspect for the NCS, since there are a number of possible advantages in dealing with volunteer subjects (e.g., higher retention rates, increased motivation to participate fully in the study, etc.). It may be the case that NCS planners choose to include some significant portion of the cohort as a volunteer sample. Thus, we include a 23rd design that calls for selection of 50% of the cohort as a volunteer sample. Additionally, Chapters 3 and 9 provide some discussion of selecting a portion of the cohort as a volunteer sample in the other 22 designs outlined here.

It should be noted that the case of 0 percent of the cohort selected from the NPBS (i.e., the entire cohort selected from the patient lists or the geographic area of a set of purposively selected Centers) corresponds to a design that limits the sampling frame population to only the population associated with those Centers. In other words, the sampling frame population for these designs may be significantly smaller than those designs that include some portion of the cohort selected in a national probability-based sample (e.g., it may be necessary to require

Centers to recruit individuals in rural areas as suggested in Section 5.3.2 and implemented in two of the designs considered in Chapter 9). This set of design options is very similar to the "hybrid" designs that were initially discussed by study planners following the Westat report (Westat, 2002), and offer a promising avenue for focusing the NCS on a subset of eligible children that may have desirable properties with respect to recruitment, retention, and cost-efficiency. We refer the reader to Chapter 2 and Appendix A for more detailed discussion of this issue.

For each of the 23 designs there are admittedly steps that could be taken to further reduce the variability of the weights. For example, additional work could be conducted to more optimally obtain a self-weighting national PBS sample (e.g., by adjusting the probability of selection of small PSUs and/or by selecting a proportionally larger number of individuals in certainty counties), or to obtain improved self-weighting Centers samples (e.g., by selecting the number of individuals proportional to the size of the Center population). However, by combining data across all study participants (i.e. those recruited through a national probability-based sample, and those recruited from sampling frames associated with purposively selected Centers), there will be large discrepancies between the weights that cannot be resolved. Therefore, obtaining a self-weighting sample in any single sampling frame will only marginally improve the amount of unequal weighting that detracts from power in weighted analyses that address external validity. This is not a unique problem to the family of designs, as there is a rich history in sampling theory for introducing large discrepancies in survey weights when using stratified sampling approaches to over-represent small and important segments of the population, such as minority subpopulations.

In the remainder of this section we provide further discussion of the implementation details for these designs. In particular, Section 5.3.1 describes the methods used in calculating the weights associated with each participant (i.e., the inverse of their probability of selection). (Note that these weights are only relevant if a weighted analysis is conducted, and there is some debate over whether sample weights should be used in analysis of a relationship.) Section 5.3.2, on the other hand, provides a table that illustrates the demographic characteristics of the 23 designs.

5.3.1 Calculation of Weights for Weighted Analyses

The multi-frame designs combine sampled individuals from three sampling frames, a 'complete' national frame, the Center area (MSA) frame, and the Center list frame, each with its own sampling mechanism. Appropriate assignment of sample weights to participants requires consideration of the frame or frames from which the participant was sampled. To begin the discussion, we consider the simple case in which a participant belongs to two frames and can enter the sample based on the sampling mechanism of either frame. We label the first event that the participant is selected from the "frame A," and the second event that the participant is selected from the "frame B." The probability that this participant is selected for the study is the probability of A or B, which is calculated as the sum of the probability of A and the probability of B, minus the probability of both A and B.

$$P(A \text{ or } B)=P(A) + P(B) - P(A \text{ and } B) \qquad (5\text{-}1)$$

The formula can be used to adjust sampling probabilities, and thus weights, of participants in multiple frames.

Since the NCS family of designs samples can be built in a sequential manner, we first select the national probability sample, and then the Center sample is selected. By proceeding in this manner, Center MSAs can be redefined to exclude counties selected as PSUs in the national probability sample. If this is done, national probability sample patients do not need their probabilities of selection modified by equation (4) since they have no chance of selection in the Center probability sample. Additionally, since the sampling probabilities of the subjects selected from the Center patient lists are generally much larger than national probability subjects, their selection probabilities would only be slightly increased by use of equation (4), so the correction is ignored. Finally, the sampling probability of Center area subjects is equal to the sum of the Center area sampling probability and the probability of being sampled in the national probability sample (i.e., the third term in Equation 5-1 is zero because the probability that the patient could be selected in the Center area sample and in the national probability sample is zero). [It should be noted that some Center MSA geographical areas could not be redefined to exclude national probability sample PSUs because this would make them too small to sample 2,000 patients. For participants in these MSAs, Equation (5-1) was used to adjust their sampling probabilities.] Sample weights were then determined for all NCS participants as the inverse of their sampling probabilities. Further details on the calculation of sampling weights associated with NCS participants recruited under the Family of Designs hierarchy are provided in Section D-3 of Appendix D.

Note that it is also possible to begin with the selection of Centers and remove these MSAs from the national PBS frame. In this case, the national PBS could be used to supplement the Centers sampling in order to "fill in the gaps" that might be missed by the Centers frame. However, in this case, the "national" PBS may no longer be a national sample, but is rather a sample from the part of the nation not covered by the selected Centers.

5.3.2 Demographic Characteristics of the 23 Designs

Table 5-4 displays the set of 23 designs that we consider along with a set of demographic characteristics for each design. To construct this table, Census data that included various demographic variables were gathered at the county level. For each of the 23 designs, 50 sample realizations were obtained, and their corresponding characteristics were evaluated using the census data (e.g., for the NPBS sample the county characteristics were utilized and for the Centers sample MSA characteristics were obtained by aggregating the county-level information). Then, averaging over the 50 sample realizations, the average demographic characteristics for each design could be computed. Note that this calculation implicitly assumes that random sampling within the counties and within the MSAs is possible, so that the characteristics of the county (or MSA) can be applied to the sample. Additionally, to construct this table we will assume that Center patients and Center volunteers have the same demographic characteristics as the Center MSA demographic characteristics (note that this assumption makes the demographic characteristics for those designs with P_1=0 identical). If this is not the case (e.g., if certain racial groups are more likely to refuse participation in the study), then admittedly these numbers may not be accurate. However, they provide an example, albeit simplistic, of the possible

demographic characteristics that could be obtained in each of the designs. The demographics included in the table are:

- Proportion of the sample that is urban
- Average median income of the sample
- Proportion of the sample from white, black, Hispanic, and other groups
- Proportion of the sample from single parent households
- Proportion of the sample for which females have a HS diploma
- Proportion of the sample for which females have a college degree
- Proportion of the sample in poverty
- Proportion of vacant households in the sampled counties.

Note that for most of the characteristics there are very few differences between the 23 designs and the national estimates, indicating that each of these designs has the potential to provide a sample that is similar to the nation, at least in terms of these demographic characteristics (again, this does not account for the issue of nonresponse by certain population groups and other possible recruitment issues that may make it difficult to obtain a within-county, or within-MSA, sample with demographics that are similar to the county/MSA). However, one glaring difference is in the proportion of the sample from urban settings. For all 23 designs, this proportion is higher than the national average. This is likely due to the fact that the 105 Centers utilized in these designs were all in urban areas, resulting in an overrepresentation of urban regions. Note that as the proportion of the cohort selected in the NPBS goes up, the urban proportion begins to converge to the national estimate, as expected. This discrepancy between the characteristics of the design and the national characteristics could perhaps be remedied by (1) establishing Centers in more rural areas (e.g., such as the Children's Health Center operated by UC Berkeley in the Salinas Valley Farm Community), (2) including these

Table 5-4. Table of demographic characteristics for 23 designs.

Design	Number PSUs	P_1	P_2	P_3	Proportion Urban	Average Median Income	Proportion White	Proportion Black	Proportion Other	Proportion Hispanic	Proportion Single Parent Households	Proportion Females with HS Diploma	Proportion Females with College Degree	Proportion in Poverty	Proportion of Vacancies	Proportion of Rentals
National Estimates					0.81	43528	0.75	0.12	0.13	0.13	0.09	0.3	0.3	0.12	0.09	0.31
A1*		0	0	1.00	1	44847	0.78	0.13	0.1	0.08	0.09	0.28	0.34	0.11	0.07	0.33
A2*		0	0	0.50	1	44847	0.78	0.13	0.1	0.08	0.09	0.28	0.34	0.11	0.07	0.33
B3		0	0.25	0	1	44847	0.78	0.13	0.1	0.08	0.09	0.28	0.34	0.11	0.07	0.33
B4		0	0.5	0	1	44847	0.78	0.13	0.1	0.08	0.09	0.28	0.34	0.11	0.07	0.33
B5		0	0.75	0	1	44847	0.78	0.13	0.1	0.08	0.09	0.28	0.34	0.11	0.07	0.33
C6	50	0.25	0.25	0	0.95	45611	0.77	0.13	0.11	0.1	0.09	0.28	0.33	0.11	0.07	0.32
C7	50	0.25	0.5	0	0.95	45794	0.77	0.12	0.11	0.1	0.09	0.28	0.33	0.11	0.08	0.32
C8	50	0.25	0.75	0	0.95	45939	0.77	0.12	0.1	0.1	0.09	0.29	0.33	0.11	0.07	0.31
D9	50	0.5	0.25	0	0.9	44928	0.75	0.14	0.11	0.11	0.09	0.29	0.32	0.12	0.08	0.32
D10	50	0.5	0.5	0	0.91	45262	0.75	0.14	0.11	0.11	0.09	0.29	0.32	0.12	0.08	0.32
D11	50	0.5	0.75	0	0.9	45264	0.76	0.13	0.11	0.11	0.09	0.29	0.32	0.11	0.08	0.32
E12	50	0.75	0.25	0	0.87	44177	0.75	0.13	0.12	0.13	0.09	0.29	0.31	0.12	0.08	0.32
E13	50	0.75	0.5	0	0.87	44559	0.75	0.13	0.13	0.13	0.09	0.29	0.31	0.12	0.08	0.32
E14	50	0.75	0.75	0	0.86	44570	0.76	0.12	0.12	0.13	0.09	0.29	0.31	0.12	0.08	0.32
F15	100	0.25	0.25	0	0.95	45729	0.77	0.12	0.1	0.1	0.09	0.29	0.33	0.11	0.07	0.32
F16	100	0.25	0.5	0	0.94	45974	0.78	0.12	0.1	0.09	0.09	0.29	0.33	0.11	0.08	0.31
F17	100	0.25	0.75	0	0.95	46219	0.78	0.12	0.1	0.09	0.09	0.29	0.33	0.11	0.07	0.31
G18	100	0.5	0.25	0	0.89	44993	0.76	0.14	0.11	0.1	0.09	0.29	0.32	0.12	0.08	0.32
G19	100	0.5	0.5	0	0.89	45195	0.76	0.13	0.11	0.1	0.09	0.29	0.31	0.11	0.08	0.31
G20	100	0.5	0.75	0	0.89	45318	0.77	0.13	0.1	0.1	0.09	0.29	0.31	0.11	0.08	0.31
H21	100	0.75	0.25	0	0.85	44208	0.76	0.13	0.12	0.12	0.09	0.29	0.31	0.12	0.08	0.32
H22	100	0.75	0.5	0	0.85	44501	0.76	0.12	0.12	0.12	0.09	0.29	0.3	0.12	0.08	0.31
H23	100	0.75	0.75	0	0.84	44293	0.77	0.12	0.11	0.12	0.09	0.29	0.3	0.12	0.08	0.31

* Note that when implementing these designs in the power calculations of Chapter 9 we will assume that each Center is required to sample 20% of their subjects from a rural area in close proximity. In this case, their urban proportions would be on the order of 80%, and the demographic characteristics of these designs would be different then those displayed above.

types of Centers in the set of purposively selected Centers, (3) recruiting some number of Centers that demonstrate their ability to provide coverage of rural areas, or (4) requiring all Centers to recruit a portion (e.g., 20 percent) of their cohort from a rural area in close proximity. (In the designs of Chapter 9 we will nominally assume that designs A1 and A2 will require Centers to recruit 20 percent of their participants from a rural area.)

Note that although these designs may have the potential to provide a sample that is similar to the nation in terms of these demographic characteristics, it may be the case that the NCS would like to oversample certain population groups, such as groups that are thought to be highly exposed or groups for which group-specific analyses are desired. Thus, designs that are similar to the nation in terms of certain demographic characteristics might not be the goal. While these designs have overlooked the possibility of oversampling certain population groups, in the following section we provide a more detailed discussion of this issue and its likely importance for the NCS.

5.4 OVERSAMPLING OF SUBPOPULATIONS

The issue of oversampling certain population groups, such as highly exposed populations or certain spatial regions of interest, has been generally overlooked in many of the designs described in this report. For example, the sampling schemes described above and the sampling designs utilized in the power analyses of Chapter 9 generally do not suggest strategies for oversampling population groups that are considered important. This lack of consideration is not meant to imply that oversampling is thought to be an unimportant element in designing the NCS; in fact, oversampling is likely to be very important. For example, oversampling of highly exposed populations could provide better power to assess certain hypotheses, and oversampling of certain racial/ethnic groups may be important in providing sufficient information for race-specific analyses. Thus, we do not wish to suggest that oversampling is an unimportant design element; rather, the lack of consideration of oversampling reflects the fact that it is a difficult issue that requires a number of further design specifications, and, since oversampling can occur within any design, may not be a factor in selecting which designs are more appropriate for the NCS. In the following we provide a brief discussion of several elements that may impact design choices related to oversampling (the 2002 Westat report also provides a discussion of this issue).

One of the first difficulties in considering the issue of oversampling lies in the identification of the appropriate populations to oversample. For studies involving a single primary hypothesis, this may be a relatively straightforward decision that can be made based on the primary hypothesis. For example, if the primary hypothesis calls for estimation of characteristics for certain populations of interest, then sample sizes for those populations could be identified based on simple power analyses (e.g., the SRS analytical power formulas provided in Chapter 9). Alternatively, if primary interest is in studying the relationship between a single risk factor and some outcome, then one plausible strategy may be to oversample populations experiencing low levels of the risk factor and populations experiencing high levels of the risk factor (e.g., if a monotone relationship is assumed). Of course, this would require identification of populations that have high levels (and low levels) of the risk factor, which may involve cost implications and entails some difficulty in accurately identifying these populations.

For a study like the NCS, with a large number of primary hypotheses, it becomes much more difficult to identify populations that are appropriate for oversampling, since different hypotheses will imply oversampling of different populations. For example, oversampling of populations thought to experience elevated exposures to nonpersistent pesticides may be appropriate in addressing the hypothesis concerned with the effect of this exposure on neurobehavioral development, whereas oversampling of populations thought to experience elevated exposures to outdoor air pollution may be appropriate in addressing the hypothesis concerned with the effect of outdoor air pollution on the development of asthma. If all of the hypotheses were to suggest the same important populations (i.e., if the same population experiences elevated exposures for all of the exposures of interest), then there is certainly little difficulty in selecting the appropriate groups to oversample; however, since this is generally not the case, identification of populations to oversample for the NCS becomes a much more challenging problem that must involve some tradeoffs between the many NCS core hypotheses.

Another important consideration with regards to oversampling may be the anticipated degree of difficulty in identifying, recruiting, and following a selected population group of interest. For example, certain population groups may be difficult to identify (locate), and, therefore, may be difficult to sample. Other populations may be difficult to recruit, thereby lowering the recruitment rates and increasing the costs. Still other populations may be difficult to retain in the study for the desired duration, again introducing cost implications or sample size (i.e., power) implications. As an example, suppose interest was in oversampling migrant worker populations, since they may be highly exposed to pesticides and/or herbicides. Identifying a sampling frame that could access migrant worker populations, convincing a migrant worker to participate in a study as burdensome as the NCS, and tracking the migrant worker through all 20 years of the study could all be difficult and costly (both financially and scientifically).

Other important considerations regarding oversampling include:

- The effect of oversampling on power to detect the relationships of interest. It may be the case that while the power to detect the relationship for the oversampled population would increase, the power to detect the relationship for the entire cohort would decrease.
- The effect of oversampling on the characteristics of the design (e.g., design effects, national representativeness, etc.)
- The effect of oversampling certain populations on anticipated recruitment and retention rates.
- The time-varying nature of some population characteristics in a longitudinal study. There is no guarantee that oversampling of populations that currently undergo elevated exposures to a substance of interest will result in oversampling the population that undergoes elevated exposure to that substance five years from now.

In other words, there are a number of issues that must be considered and evaluated when (and if) oversampling of important populations is utilized in the NCS. Some of these issues are hypothesis-specific, some pertain to financial implications, and some are relevant to the scientific value of the study. Further consideration of appropriate strategies for oversampling, appropriate

populations for oversampling, and the effect of oversampling is an important area for further research in the design of the NCS.

5.5 DESIGN EFFECTS FOR RELATIONSHIPS

The purpose of this section is to explore the use of design effects for estimating relationships of interest in the presence of designs that involve both clustering and/or unequal weighting. In general, the clustering in a design will effect estimation precision in both weighted and model-based (i.e., unweighted) analyses, and unequal weighting will have an additional effect on estimation precision under a weighted analysis. While quite a lot has been written in the sample survey literature, much of this has been in the relatively simple context where the goal is to assess the precision that a planned study might have to estimate a summary quantity such as a mean (see Section D-4 of Appendix D). In the context of the NCS, however, the situation is substantially more complicated, since estimation of relationships between adverse health effects and exposures is of primary interest. To address this, we begin from first principles.

Suppose we are interested in exploring the relationship between an exposure and an outcome, based on data from clusters of individuals, each of whom has a binary response. Let X_{ij} be the binary exposure indicator for individual j in cluster i and let Y_{ij} be this individual's corresponding response. Suppose also that we are interested in fitting the following marginal logistic model:

$$\text{Logit}[\Pr(Y_i=1|X_i=1)] = \text{Logit}(\mu_{ij}) = \beta_0 + \beta_1 X_{ij}. \qquad (5\text{-}2)$$

In practice, of course, there will also be interest in including additional covariates and risk factors. For the purpose of power and sample size considerations, however, it is enough to consider just the main effect of interest. As discussed elsewhere in the report (see Chapter 9), generalized estimating equations (GEEs) provide an appropriate basis for analysis that accounts for both non-constant sampling probabilities, as well as for clustering of individuals (see Diggle et al., 2002). The introduction of sampling weights complicates the estimation of the variance of the parameter estimates; however there are a number of standard statistical packages (SAS Proc GENMOD, SUDAAN) that can appropriately solve the GEEs in the presence of sampling weights.

In certain cases, the expression for the variance of the parameter estimates simplifies. An example of this is provided in Section D-5 of Appendix D in which the exposure of interest, X, is cluster-specific so that x_{ij} is the same for all members of the same cluster (note that this is likely unrealistic for the NCS and may represent a worst-case situation for intracluster correlation and design effects; however, it provides a starting point for this discussion). Suppose furthermore that there is no within-cluster correlation with respect to the outcome, and also that the weights are independent of cluster membership and exposure. It follows in this special case that the variance of the parameter estimates in this setting is equal to the standard variance estimate based on a logistic regression, multiplied by a factor that involves the weights. The factor can be re-expressed as:

$$\frac{E(w^2)}{\left(E(w)\right)^2} = \frac{Var(w) + \left(E(w)\right)^2}{\left(E(w)\right)^2} = 1 + CV^2, \qquad (5\text{-}3)$$

or 1 plus the squared coefficient of variation of the weights. When the weights are constant, this factor equals 1 and the standard logistic regression variance formula applies. When the weights vary, then this factor will always exceed 1; hence the variance of parameters estimated using weighted estimating equations will always exceed those based on a simple logistic regression, a well known result among sample survey statisticians. This term is often referred to as a *design effect*, and represents the ratio of the parameter estimate variance under the selected design to the parameter estimate variance under a simple random sample. These design effects provide a very useful tool when it comes to study planning and design, since one can think in terms of the impact of various different weighting schemes on the estimated variances of parameters of interest, and adjust accordingly.

In a slightly more complex setting, the outcome intra-cluster correlation, ρ, is non-zero. Using a similar logic, it is relatively straightforward to show that the *design effect* (or the factor that multiplies the usual logistic regression variance) is:

$$1 + \rho(m-1) + CV^2 + \rho(m-1)\,\mathrm{cov}(w_{ij} w_{ij'}), \qquad (5\text{-}4)$$

where m is the average cluster size and the covariance term refers to the covariance between weights within the same cluster (note that as expected there is an effect of both unequal weighting and clustering in equation 5-4). In general, we would expect this covariance term to be zero. In the special case where the weights are all equal (variance and covariance of the weights equal zero), the design effect reduces to $(1+\rho(m-1))$, which is the usual inflation factor for a variance based on cluster data (see Diggle et al., 2002). The form of the design effect suggests that clustering will tend to inflate the variances of estimated parameters of interest. It also suggests that the best design strategy, from the perspective of the design effect, would be to have a large number of PSUs, with relatively little clustering. However, as indicated above, this design effect calculation has been obtained under the setting where there is perfect within-cluster correlation with respect to exposure, in other words, for the setting where all members of the cluster have the same level of exposure. In the context of the NCS, this is unlikely to be the case, and we expect exposure levels to vary substantially within-cluster. For this reason, it is not appropriate to use a standard design factor argument to guide design considerations for the NCS.

When the exposure of interest, X, is allowed to vary within-cluster (as is expected to occur in the NCS), all these calculations become considerably more complicated, as described in detail in Section D-5 of Appendix D. The design effects, however, can be calculated by using a computer package, such as R or S-plus, to compute the variance of the estimated exposure effects under varying assumptions of weighting and clustering. Figure 5-4 displays how design effects vary as a function of the response probability for an unexposed individual (denoted as μ_0 in the figure), the intraclass correlation in Y (the x-axis in the figure), and whether or not the exposure variable is cluster-specific (right-hand panel) or varies within-cluster (left-hand panel). Note that this plot assumes equal weights so that we can focus on just the effect of clustering. Additionally, note that the ratio displayed in Figure 5-4 is the ratio of the variance under simple

random sampling to the variance under the clustered design, which can be thought of as the inverse of the design effect.

The left-hand panel of Figure 5-4 corresponds to the case of a non-zero within-cluster correlation with respect to exposure. This means that each cluster is likely to have a mix of exposed and unexposed individuals, as is anticipated in the NCS for most hypotheses. The right-hand panel corresponds to the case where there is perfect within-cluster correlation with respect to exposure – that is, either all individuals in a cluster are exposed, or all the individuals in a cluster are unexposed. Note that the inverse of the "design effects" are much closer to 1 in the left-hand panel, suggesting that the effect of clustering is not nearly as severe when we have a within-cluster varying covariate. In other words, the impact of clustering on the estimated variances of parameter estimates is moderate compared to the more familiar case where covariates are constant within-cluster. This figure suggests that use of standard "design effects" arguments can lead to misleading results when designing a cohort study such as the NCS. In fact, preliminary explorations suggest that there are cases where the inverse of the design effect is greater than 1 (i.e., the clustering actually allows more accurate estimates of the relationships of interest). From a heuristic perspective, such a phenomenon makes sense and is an analogy to the well known argument that supports the use of a paired rather than an unpaired t-test when it is feasible to apply two different treatment conditions to the same experimental unit. Further work is needed to explore this issue and to better lay down the framework and assumptions that are inherent in calculation of design effects when estimating relationships.

For these reasons, the power results in Chapter 9 of this report estimate power via simulation under a number of assumptions regarding the specific regression relationship between response and explanatory variables. In other words, since design effects for relationships are not easily calculated, the power calculations presented in this report are done via simulation, rather than through the use of design effects.

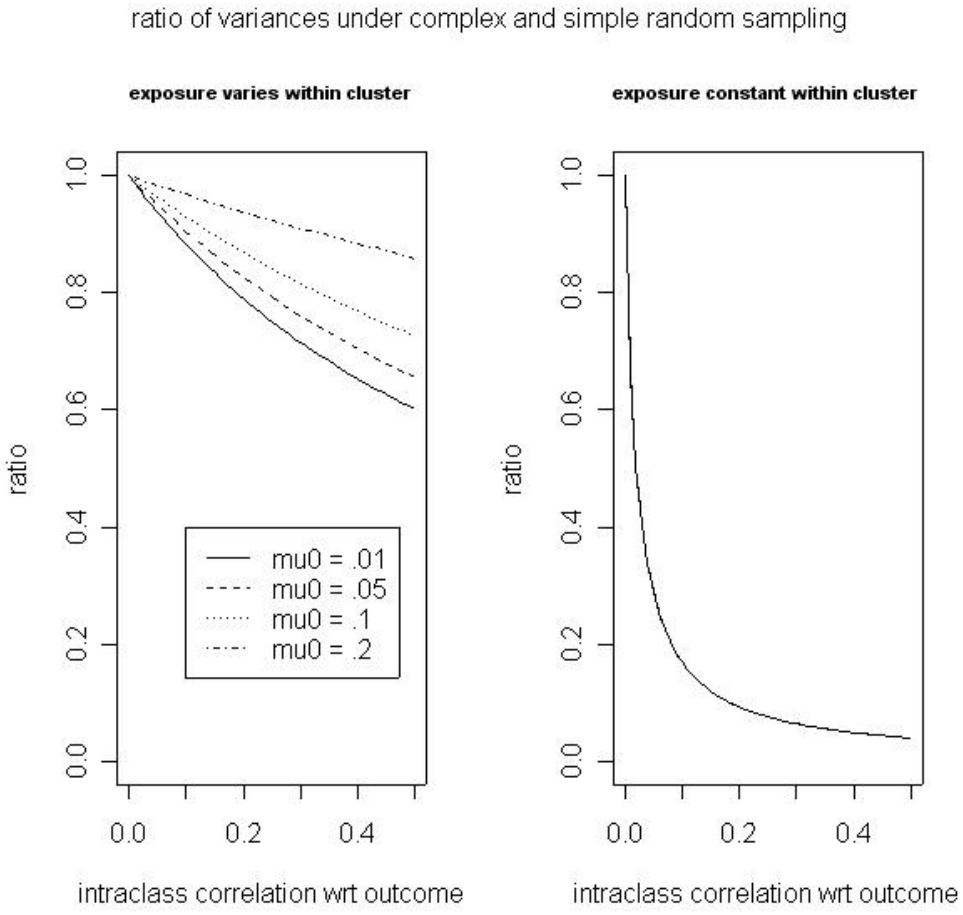


Figure 5-4. Ratio of the variance of the parameter (relationship between *Y* and *X*) estimate for a simple random sample to that for the clustered design (assuming equal weights).

6 REVIEW OF NCS CORE HYPOTHESES AND CRITICAL MEASURES

The National Children's Study has identified five priority outcome areas for the study to address. The study hypotheses relate to 1) pregnancy outcomes, 2) neurodevelopment and behavior, 3) injury, 4) asthma, and 5) obesity and physical development. Chapter 1 of this report discusses the rationale for selection of the priority outcomes and hypotheses. Related to each priority outcome are one or more hypotheses that focus on specific relationships between an adverse health outcome and either (1) measures of exposure or other risk factors, or (2) protective factors that help to lower risk of the adverse health outcome. Table 6-1 provides a list of the five priority outcomes and the 21 hypotheses that have been proposed by the ICC for the NCS. These hypotheses, along with the study goal that the NCS serve as a resource for future assessments, provide the basis for specifying a study design, and for assessing whether that design has sufficient power to adequately address priority research questions. The purpose of this Chapter is to:

1. Introduce the primary variables of interest as well as the methods, measures, and life stages of data collection that are necessary to address each specific hypothesis,

2. Briefly discuss their relevance to outstanding study design questions, and identify measurements that have implications/requirements for the design options, and

3. Document how a subset of the 21 specific NCS hypotheses were chosen for the power studies described in Chapter 9.

Battelle prepared a white paper on Measures for NCS Priority Outcomes, which is included in Appendix F. This Appendix provides a discussion of the different data collection activities that will be necessary to support the hypotheses of the NCS. The main body of Appendix F provides an overview of the key measures that must be obtained to support each specific hypothesis listed in Table 6-1 with many details relevant to sampling design decisions (e.g., pre-conception measurements, detailed environmental samples, delivery room biologics). In addition, Appendix F provides aggregated tables that may offer insight into requirements for a common data collection protocol that provides coverage across the current hypotheses of the NCS. (It must be noted that by no means does the paper serve as a consensus data collection protocol – rather, it offers a useful starting point for consideration and provides a basis for the costs associated with data collection that are assumed in Chapter 8). Finally, a series of detailed tables provide initial estimates of the primary, secondary, and perhaps tertiary measurements that should be considered for each specific hypothesis, as a function of method of data collection (blood sample, urine sample, other physical sample, medical record review, questionnaire, direct observation by medical professional) and life stage (preconception through early adulthood).

Table 6-2 provides a high-level summary, for each hypothesis, of the primary explanatory and outcome variables, the methods and measures that could be used to collect the data for each variable, and the life stages at which data might be collected. Where available, the expected prevalence of the outcome and explanatory (risk factor) variables are presented, which are integrated into the power analyses presented in Chapter 9. In the final column, a few examples of variables that may serve as covariates, confounders, or effect modifiers are provided.

Details on the measurement requirements for the NCS, such as those provided in the white paper in Appendix F, are critical to evaluating the real impact of many of the design issues discussed so far in this document. These include:

- The description of a Family of Designs provided in Chapter 3 of this report alludes to the fact that to allow multiple sampling approaches as part of the NCS, a standardized core (minimal) data collection protocol will need to be developed that can be used across all study participants regardless of their specific mode of recruitment into the study, with additional data collection activities planned for certain identifiable subpopulations of the study cohort that would be easier or more convenient to study. The aggregate tables in Appendix F provide insight into what a standardized minimal data collection might need to include to address all current hypotheses. Initial assessment indicates that the study may need to consider the use of matrix sampling to spread more burdensome measures out and minimize the burden for any one participant. For example, the amount of blood required for different analyses may be a significant limiting factor in and of itself.

- Estimates of initial response rate and retention rate are difficult because there is no good precedent of a study with the burden and duration requirements that the NCS will impose. Examination of the tables and text and in Appendix F provides insight into the detail of approximately what that burden might be to allow experienced researchers to apply professional judgment in estimating realistic response and retention rates. The fact that the burden is most intensive in the initial period (1-3 years) of involvement offers opportunities for a pilot study to better understand what retention rates can be expected and what factors most affect the rates. The details in Appendix F also provide a basis for estimating potential biases in a cohort agreeing to this protocol that might be expected or avoided, such as biases that may be introduced by inclusion of genetic information and archived biological samples in the study.

- Choice of organizational structure, sampling frame, and sample selection methods may all be affected by the need for specialized measures, or proximity to major medical care facilities. Appendix F provides an initial indication of the extent and importance of such measures, with a conclusion that in general, with the exception of delivery room samples, the measures required to address the NCS core hypotheses can be supported by data collection facilities that are widely available across all sampling frames considered in this report. Specialized measures that are not as widely available, such as 3-D ultrasound, may be more important for meeting the NCS goal to serve as a resource for future studies of health and environment.

- An unavoidable constraint of the NCS is that there is some maximum level of burden that can be imposed on the cohort for many reasons - ethical, practical, and financial. Appendix F provides a basis for study planners to determine whether that maximum level may be exceeded before all measurements for all hypotheses can be collected, thereby requiring designs that allow for designed missing measures or less precise and burdensome measures on some participants. Our contention is that the NCS will need to strongly consider integrating these types of design efficiencies, especially during the

earlier phases of the study when there are many sampling events planned to cover the pre-natal through infant stages of development.

Based on the information in Appendix F, it does appear that there is reasonable flexibility in the timing of measures, and that there are multiple ways to reduce the burden and cost of data collection and increase efficiency when addressing the hypotheses. These include consolidation of data collection across hypotheses collecting certain data only from a designed sample of participants rather than all participants, postponing the chemical analysis of archived biological and environmental samples until a later date, and collecting more precise and burdensome measures on a small subset of participants while collecting less precise measures on the full cohort. While all of these options require difficult tradeoffs between cost, burden, and quantity or quality of information, they do provide the flexibility to conduct a study under the kind of cost and cohort burden constraints that will inevitably be imposed upon it. Further discussion of the opportunities for reducing burden and cost is included in Appendix F.

6.1 CHOICE OF SUB-HYPOTHESES FOR ASSESSING POWER

The basis for selecting hypotheses to address in the power studies includes the following factors:

1. We wanted to select at least one hypothesis within each of the five priority outcome areas.
2. When reviewing multiple hypotheses within a priority outcome area (e.g., the hypotheses related to altered neurobehavioral development), we tried to select the specific hypothesis that we thought would be most challenging for the NCS to address with sufficient power. This generally led to the selection of hypotheses that focused on rare health outcomes and/or exposures, as seen in Table 6-2. For example, Hypothesis 2.2, which addresses cerebral palsy and autism as primary outcomes, may be difficult to assess due to the low incidence of cerebral palsy (0.20%) and autism (0.30% by age 3) occurring in the NCS population (CDC, 2003). The relatively small number of subjects expected to experience these adverse health outcomes may result in limited study data in the NCS with which to support this particular hypothesis.
3. When reviewing the list of hypotheses selected from the above two steps, we wanted to ensure that in total, they represented a range of life stages at which time the health outcome would be observed. This is an important factor because it will allow differences in retention rates among the various design options being considered to be taken into consideration as part of the power studies.

Based on the above factors, the following nine hypotheses were selected to examine in the power studies: Hypotheses 1.1, 2.2 and 2.3, 3.1, 3.2, 4.1, 4.2, 4.3, and 5.1 (highlighted in yellow in Tables 6-1 and 6-2). One feature that all of these hypotheses have in common is that they can be conceptually represented by a simple 2×2 contingency table focusing on both a single binary health outcome (presence or absence of the adverse health effect) and a single binary measure of exposure (exposed versus unexposed). Although some of the chosen hypotheses can certainly be represented using other measures of health outcome or exposure that are not binary (e.g., continuous, integer valued, ordered multinomial, etc.), we believe that in

most cases the choice of a binary representation leads to conservative estimates of power for the NCS. The choice of exploring power analyses using the binary representation also allowed the use of a consistent basis for presenting all of the power analysis results provided in Chapter 9.

An important limitation of assessing the power to address the NCS hypotheses using power studies that conceptually reduce to a simple 2×2 contingency table is that many of the hypotheses (e.g., hypothesis 5.6 on social, behavioral, and family factors that affect development of dietary preferences and physical activity patterns early in childhood and determine risk of childhood obesity and insulin resistance) focus on a number of potential exposures that may compete for explanatory power in describing the response of interest. In fact, when considering the impact of potential covariates, effect modifiers and confounders – none of the hypotheses truly reduce to a simplistic model of disease status versus a single exposure. However, earlier design work conducted by Battelle and Harvard under contract to EPA's National Exposure Research Laboratory on assessing more complex relationships between disease and multiple exposures suggests that these simple models will be reasonably accurate for the intended purpose of differentiating between different sampling design options based on their ability to detect important relationships (Appendix C).

Table 6-1. Hypotheses For The National Children's Study

1.0	**Pregnancy Outcome**
1.1	Among women without diabetes before pregnancy, impaired glucose metabolism during pregnancy is proportional to risk of major congenital malformations of the heart, central nervous system, musculoskeletal system, and all birth defects combined
1.2	Intrauterine exposure to mediators of inflammation due to infection of either vaginal, cervical, or uterine sites, or of more distal sites (e.g., periodontal disease) is associated with an increased risk of preterm birth
2.0	**Neurodevelopment and Behavior**
2.1	Repeated low-level exposure to nonpersistent pesticides in utero or postnatally increases risk of poor performance on neurobehavioral and cognitive examinations during infancy and later in childhood, especially, for certain agents, among those with genetically decreased paraoxonase activity
2.2	Prenatal infection and mediators of inflammation are risk factors for neurodevelopmental disabilities, such as cerebral palsy and autism
2.3	Infection and mediators of inflammation during pregnancy and the perinatal period are associated with increased risk of schizophrenia
3.0	**Injury**
3.1	Exposures early in life that lead to neurotoxic effects are associated with increased risk of injury***
3.2	Attributes of childcare and relationship with caregivers influence risk of injury***
3.3	Repeated head trauma has a cumulative adverse effect on neurocognitive development
4.0	**Asthma**
4.1	Exposure to indoor and outdoor air pollution and bioaerosols (including allergens, endotoxin, and mold) is associated with increased risk of asthma
4.2	Respiratory viral infection early in life is associated with increased risk of asthma
4.3	Maternal stress during pregnancy is associated with increased risk of asthma
4.4	Antioxidant constituents of diet decrease risk of asthma
4.5	Early exposure to bacterial and microbial products decreases risk of asthma (hygiene hypothesis)
4.6	Access to health care and management of asthma are strongly related to asthma hospitalization.
5.0	**Obesity and physical development**
5.1	Impaired maternal glucose metabolism during pregnancy is directly related to risk of obesity and insulin resistance in offspring
5.2	Intrauterine growth restriction as determined by serial ultrasound examination is associated with subsequent risk of central obesity and insulin resistance in offspring, independent of subsequent body mass index
5.3	Breast milk feeding, compared with infant formula feeding, and breastfeeding duration are associated with lower rates of obesity and lower risk of insulin resistance
5.4	Dietary predictors of obesity and insulin resistance include reduced intake of fiber and whole grains, and high glycemic index
5.5	Environmental factors such as distance to parks, availability of walking routes in the neighborhood, and neighborhood safety are associated with risk of obesity and insulin resistance
5.6	Social, behavioral, and family factors that affect development of dietary preferences and physical activity patterns early in childhood determine risk of childhood obesity and insulin resistance
5.7	In utero and subsequent exposure to environmental agents that affect the endocrine system (bisphenol A, atrazine, and lead) results in altered age at puberty

***These sub-hypotheses are under evaluation and will likely be changed or eliminated by the NCS ICC.

Table 6-2. Summary of Outcome and Explanatory Variables (Incidence or Prevalence where Available), Methods and Measures, Life stages, and Examples of Covariates for NCS Hypotheses

	Outcome Variables					Explanatory Variables					
Hypoth-esis	**Primary Outcome Variables**	**Incidence or Prevalence of Outcomes**	**Methods**	**Measures**	**Life Stages**	**Primary Explanatory Variables**	**Incidence or Prevalence of Risk Factors**	**Methods**	**Measures**	**Life Stages**	**Examples of Covariates, Confounders, Effect Modifiers**
1.1	Major congenital malforma-tions of the heart, central nervous system, and all birth defects combined	Congenital heart defects: 0.60%; CNS defects: 0.60%; All birth defects: about 3.00%	Direct Observation by Medical Professional	Any birth defects	Birth through Adolescence	Impaired glucose metabolism during pregnancy		Blood	Glucose Tolerance, Blood Glucose and Serum insulin levels	1st, 2nd, 3rd Trimes-ters	Family history, mother's medical history
1.2	Preterm birth	Approximately 2% preterm births	Direct Observation by Medical Professional	Gestation <37 weeks	Birth	Intrauterine exposure to mediators of inflammation due to infection	2% intra-uterine infection	Interview, Blood, Swabs	Cytokines, WBC. Antibodies	1st, 2nd, 3rd trimes-ters	Mother's medical history, recent infections
2.1	Neurobeha vioral and cognitive effects during infancy and childhood	Uncertain	Neuro & Psych Testing	Abnormal neuro and cognitive results	Infancy through Year 21	Repeated low level exposures to nonpersistent pesticides *in utero* or postnatal	Plasma of 1% pregnant women reveals OP exposures	Blood, Urine, Env Air and Dust Sampling	Mother's pesticide levels; environ-mental levels	1st, 2nd, 3rd trimes-ters through Year 7	Mother's medicine usage, occupational history, diet and nutrition; child's residential environment
2.2	Neurode-velopmen-tal disabilities, e.g., CP, autism	CP: 0.20%; Autism by age 3: 0.30%	Cord blood; Neuro & Physical Exams	Abnormal findings on autism screening test; umbilical cord pathology	Infancy through Year 7	Prenatal Infection and mediators of inflammation		Blood, Swabs, Obstetric Med Hx	Vaginal & cervical cultures, interleukins, infection serology	Preg-nancy	Mother's medical and obstetric history, family history

Table 6-2. Summary of Outcome and Explanatory Variables (Incidence or Prevalence where Available), Methods and Measures, Life stages, and Examples of Covariates for NCS Hypotheses

	Outcome Variables					Explanatory Variables					
Hypoth-esis	**Primary Outcome Variables**	**Incidence or Prevalence of Outcomes**	**Methods**	**Measures**	**Life Stages**	**Primary Explanatory Variables**	**Incidence or Prevalence of Risk Factors**	**Methods**	**Measures**	**Life Stages**	**Examples of Covariates, Confounders, Effect Modifiers**
2.3	Schizo-phrenia	Schizophrenia: 1.00% (older teens and adults)	Neuro & Psych Testing; Direct Observation	Neuro & Psych Testing Results	Infancy through Year 21	Infection and mediators of inflammation during pregnancy and perinatal period		Interview, Blood, Swabs	Maternal hormones, cytokines	1st, 2nd, 3rd trimes-ters, at birth	Family history, economic status, genetic polymorphisms, mother's medicine usage
3.1***	Increased risk of injury	Ave about 10% across age groups	Interview, Medical Record Review	Injury events	Every three months, Infancy through Year 21	Exposures to neurotoxins, e.g., PCB, mercury, Pb, pesticides, other metals		Blood, Interview, Env. Air and Dust	PCB, mercury, Pb, pesticides, other metals	Birth through Year 5	Occupational history, diet and nutrition; child's residential environment
3.2***	Increased risk of injury	Ave about 10% across age groups	Interview, Medical Record Review	Injury events	Every three months, Infancy through Year 21	Behavioral attributes of childcare; relationship with caregivers		Interviews	Social function measures	Birth through Adolescence	SES, residential environment
3.3	Neurocog-nitive develop-ment		Inter-views, school records, medical records	Behavioral, neuro, and develop-mental outcomes	Infancy through Year 21	Repeated head trauma	7/1000 children <10 years have ER visit for head trauma	Interview; Medical records	Traumatic brain injury	Every 3 months	SES, residential environment
4.1	Increased risk of asthma	Asthma 5-14 years: 6.00%	Physical Exam, Medical Record Review	Allergy, asthma in index child, airway reactivity	Year 1 through year 21	Indoor and outdoor air pollution, bioaerosols, inc allergens, endotoxin, mold		Env air and dust samples, interviews	Diesel exhaust, NO_2 , allergens, mold	Year 1 through year 21	Infections, inflammations, lymphocytes, urine cotinine, smoking, health care access

Table 6-2. Summary of Outcome and Explanatory Variables (Incidence or Prevalence where Available), Methods and Measures, Life stages, and Examples of Covariates for NCS Hypotheses

	Outcome Variables					Explanatory Variables					
Hypoth-esis	**Primary Outcome Variables**	**Incidence or Prevalence of Outcomes**	**Methods**	**Measures**	**Life Stages**	**Primary Explanatory Variables**	**Incidence or Prevalence of Risk Factors**	**Methods**	**Measures**	**Life Stages**	**Examples of Covariates, Confounders, Effect Modifiers**
4.2	Increased risk of asthma	Asthma 5-14 years: 6.00%	Physical Exam, Medical Record Review	Allergy, asthma in index child, airway reactivity	Year 1 through year 21	Respiratory viral infection		Medical histories, Physical Exams	lympho-cytes, cytokines markers	Birth through Year 5	Smoking, family lifestyle factors, health care access
4.3	Increased risk of asthma	Asthma 5-14 years: 6.00%	Physical Exam, Medical Record Review	Allergy, asthma in index child, airway reactivity	Year 1 through year 21	Maternal Stress during pregnancy		Interview, Blood	Mother's alcohol consump-tion, smoking, psychoso-cial stress, Cortisol	Preconce ption, 1st, 2nd, 3rd trimes-ters	Lifestyle factors, occupational history, mother's history of allergy and asthma
4.4	Decreased risk of asthma	Asthma 5-14 years: 6.00%	Physical Exam, Medical Record Review	Allergy, asthma in index child, airway reactivity	Year 1 through year 21	Antioxidant constituents of diet in mother, other adults, and index child		Diet and Nutrition Measures , exhaled breath conden-sate	Vitamin C, Vitamin E, , fatty-acid markers	Birth through year 21	Smoking, psychological history, history of infections in index child, allergic sensitization in index child

Table 6-2. Summary of Outcome and Explanatory Variables (Incidence or Prevalence where Available), Methods and Measures, Life stages, and Examples of Covariates for NCS Hypotheses

	Outcome Variables					Explanatory Variables					
Hypoth-esis	**Primary Outcome Variables**	**Incidence or Prevalence of Outcomes**	**Methods**	**Measures**	**Life Stages**	**Primary Explanatory Variables**	**Incidence or Prevalence of Risk Factors**	**Methods**	**Measures**	**Life Stages**	**Examples of Covariates, Confounders, Effect Modifiers**
4.5	Decreased risk of asthma	Asthma 5-14 years: 6.00%	Physical Exam, Medical Record Review	Allergy, asthma in index child, airway reactivity	Year 1 through year 21	Exposure to bacterial and microbial products		Medical history, blood, dietary measures	air survey, bacteria and other infection measures	Birth through Year 5	Smoking, psychological history, history of infections in index child, allergic sensitization in index child, medicine usage in index child
4.6	Asthma hospitaliza-tion	Asthma 5-14 years: 6.00%	Medical Record Review, Interview	Health Insurance claims, Hospital visits	Year 1 through year 21	Access to health care and management of asthma		Interview; Medical records	Neighborho od characteris-tics; health insurance; social function; SES; health care usage	Birth through year 21	Health-related knowledge; residential history; occupational history; content and quality of health care;
5.1	Risk of obesity and insulin resistance in offspring	Obesity: 15.30% ages 6-11; 15.50% ages 12-19; Insulin resistance may be as high as 25.00%	Medical Record Review, Physical Exam, Blood	Body size; serum insulin levels; blood pressure; growth hormones	Year 1 through year 21	Impaired glucose metabolism during pregnancy		Blood	Glucose Tolerance, Blood Glucose and Serum insulin levels	1st, 2nd, 3rd trimes-ters	Family history of obesity and diabetes; lifestyle factors

Table 6-2. Summary of Outcome and Explanatory Variables (Incidence or Prevalence where Available), Methods and Measures, Life stages, and Examples of Covariates for NCS Hypotheses

	Outcome Variables					Explanatory Variables					
Hypoth-esis	**Primary Outcome Variables**	**Incidence or Prevalence of Outcomes**	**Methods**	**Measures**	**Life Stages**	**Primary Explanatory Variables**	**Incidence or Prevalence of Risk Factors**	**Methods**	**Measures**	**Life Stages**	**Examples of Covariates, Confounders, Effect Modifiers**
5.2	Risk of central obesity and insulin resistance, indepen-dent of BMI	Obesity: 15.30% ages 6-11; 15.50% ages 12-19; Insulin resistance may be as high as 25.00%	Medical Record Review, Physical exam; Blood	Abdominal girth; serum insulin levels, blood pressure	Year 1 through year 21	Intrauterine growth restriction		Ultra-sound	Fetal ultrasound	1st, 2nd, 3rd trimes-ters	Diet and nutrition, physical activity, medical history of index child
5.3	Lower rates of obesity and lower risk of insulin resistance	Obesity: 15.30% ages 6-11; 15.50% ages 12-19; Insulin resistance may be as high as 25.00%	Medical Record Review, Physical Exam, Blood	Body size; serum insulin levels; blood pressure; growth hormones	Year 1 through year 21	Breast milk feeding and duration		Interview; Sample Breast Milk	Frequency and amount of feeding	Birth through Year 2	Physical activity, medical history of index child, family medical history
5.4	Obesity and insulin resistance	Obesity: 15.30% ages 6-11; 15.50% ages 12-19; Insulin resistance may be as high as 25.00%	Medical Record Review, Physical Exam, Blood	Body size; serum insulin levels; blood pressure; growth hormones	Year 1 through year 21	Reduced intake of fiber and whole grains, and high glycemic index		Interview	Diet and nutrition measures	Year 1 Through Year 21	Family history of obesity and diabetes; lifestyle factors; physical activity

Table 6-2. Summary of Outcome and Explanatory Variables (Incidence or Prevalence where Available), Methods and Measures, Life stages, and Examples of Covariates for NCS Hypotheses

	Outcome Variables					Explanatory Variables					
Hypoth-esis	**Primary Outcome Variables**	**Incidence or Prevalence of Outcomes**	**Methods**	**Measures**	**Life Stages**	**Primary Explanatory Variables**	**Incidence or Prevalence of Risk Factors**	**Methods**	**Measures**	**Life Stages**	**Examples of Covariates, Confounders, Effect Modifiers**
5.5	Risk of obesity and insulin resistance in offspring	Obesity: 15.30% ages 6-11; 15.50% ages 12-19; Insulin resistance may be as high as 25.00%	Medical Record Review, Physical Exam, Blood	Body size; serum insulin levels; blood pressure; growth hormones	Year 1 through year 21	Environmental factors such as distance to parks, availability of walking routes, neighborhood safety		Interview	Residential environ-ment; demograph-ic data, lifestyle factors, physical activity, cultural norms,	Year 1 Through Year 21	Cultural norms, residential environment, values wrt diet, social function
5.6	Risk of obesity and insulin resistance in offspring	Obesity: 15.30% ages 6-11; 15.50% ages 12-19; Insulin resistance may be as high as 25.00%	Medical Record Review, Physical Exam, Blood	Body size; serum insulin levels; blood pressure; growth hormones	Year 1 through year 21	Social, behavioral, family factors that affect dietary preferences and physical activity patterns		Interview	Health-related social, behavioral, factors	Year 1 Through Year 21	Smoking, SES, transportation methods, neighborhood characteristics
5.7	Altered age at puberty	Ave for girls: 8-13 years; Ave for boys: 9-14 years	Physical exam; urine	Tanner stages; age at menarche; presence of sperm in urine	Through Puberty	*In utero* and subsequent exposure to environmental agents that affect endocrine system		Blood, Urine, Interview	Metabolites levels of bisphenol A and atrazine	Prenatal Through Year 9	Lifestyle factors, smoking, medicine usage, exposure to environmental chemicals; reproductive history

** Table 6.2 was extracted from Appendix F, with all references for information contained in this table are provided there
*** These sub-hypotheses are under evaluation and will likely be changed or eliminated by the NCS ICC.

7 ASSUMPTIONS ON RECRUITMENT AND RETENTION

The recruitment and retention of subjects in a longitudinal study has been the topic of many research studies and much published literature. A recent review of the published literature on issues related to recruitment and retention in longitudinal studies did not reveal any single study that employed exactly the same scope, size, and design as that being considered for the NCS (see Appendix G). Differences in the identified studies include: much smaller sample sizes, lower burden on respondent, absence of the collection of environmental samples, geographically smaller areas, etc. However, many of the studies and papers in the current literature do provide an opportunity for gaining insight into the potential methods for recruitment and retention in the NCS, and can provide some foundation for the formulation of response and retention rates. Appendix G summarizes the salient issues with respect to recruitment and retention, as identified in the published literature, and suggests the implication to the NCS. The focus of this section, however, is to utilize the information from these studies to formulate estimates of the initial response rate and retention rates for the NCS.

The initial response rates and retention rates are drivers for the results in a number of other sections including power calculations and cost estimates. Section 7.1 discusses the formulation of estimates for the initial response rates that might be expected, while Section 7.2 discusses the approach used to estimate retention rates. Section 7.3 discusses some limitations and other considerations with estimating the initial response and retention rates including some alternative approaches. Because estimating response and retention rates is not an exact science, other methods and/or assumptions may be used to formulate estimates for these rates. All of the results included in this White Paper can be recalculated if desired to accommodate different assumptions/methods for calculating response and retention rates.

7.1 INITIAL RESPONSE RATES

Response in the NCS is a significant factor that needs to be considered in the overall design. The initial recruitment rate will be factored into the response rates for subsequent data collection stages because the response rate for a later stage in a longitudinal study is typically calculated as the initial recruitment rate multiplied by the retention rates for subsequent stages of the survey, unless additional participants are permitted to join the survey after the initial data collection effort. Although imperfect, response rates are viewed by many as a barometer for whether the survey will suffer from a sampling bias due to nonresponse. Certainly, it is possible to have a study with a low response rate and not have an issue with nonresponse bias, or be able to correct for nonresponse through weighting adjustments. It is equally possible to achieve a high response rate but still have a biased sample. However, generally speaking, higher response rates are desirable because they make easier the task of defending the study against criticisms of bias. They also reduce the need for weighting adjustments and they add credence to the study results. In short, nonresponse generates uncertainty in the data that should be avoided if possible.

Recruitment of women to participate in the NCS will be challenging. Because of the nature of pregnancy planning, this will be especially true for identifying and recruiting preconception women. A variety of different methods have been employed to identify women

for participation in longitudinal studies. In particular, women have been identified through many different sampling approaches including, probability-based methods such as household sampling, patient lists from hospitals, physicians, or centers, volunteers, etc. Within each of these approaches, there are several different methods that can be used to increase response rates. The literature would suggest that these methods have been successfully employed in all three approaches to obtain the participation of women/households for a longitudinal survey. The literature search yielded studies that reported initial response rates that ranged from 10% to 99%. No single approach seemed to stand out conclusively as more effective than another approach, though it is noteworthy that only two studies that employed a center-based approach were reviewed. Moreover, it does appear that a probability-based approach can be effective in recruiting participants, even in surveys with significant respondent burden such as the National Human Exposure Assessment Surveys (NHEXAS) (Whitemore et al. 1999; Callahan et al. 1995; Robertson et al. 1999) – though it is also important to note that these studies were not long-term studies, and did not involve health measures.

Historical studies can provide some insight into the initial response rate that could be expected for the NCS, though it is important to consider that every study employs different methods and these methods do have an impact on the initial response rates. In particular, the literature does indicate that response rates can be positively influenced by:

- Informative interviewers
- Well-communicated incentives
- Good communication of the study intent (i.e., good for humanity)
- Potential participants having good relationship with participating physician
- Face-to-face interviews
- Community involvement

while the response rates can be negatively impacted by

- Intrusive sampling over a period of time
- Long interviews
- Lack of incentives.

It will be very important to factor these general observations into the design of the NCS. For example, the use of incentives to offset participant burden or the perception of burden should be employed in the NCS. In-person recruitment appears to be more successful than other modes of recruitment, particularly when conducted by the potential participant's physician or by a very knowledgeable, informed interviewer who can form a bond with the potential participant. An inherent trust and bond between potential participants and their physician or a medical Center of Excellence is one primary reason why a center or physician-based approach is more appealing than a probability-based approach (where the relationship has to be developed essentially from scratch). If a probability-based approach is employed, it will be critical to provide recruiters with in-depth knowledge of the study, and to minimize turnover in the recruitment/data collection staff – facilitating the growth of a rapport between participant and data collector.

Initial response rates for 23 studies thought to be relevant to the NCS were obtained. Some of these studies are environmental exposure studies, others focus on policy or health issues, and most are longitudinal in nature. Table 7-1 summarizes the response rates observed in these studies (for additional details on the reviewed studies see Appendix G).

Table 7-1. Summary of Initial Response Rates for Reviewed Studies

Recruitment Approach	Number of Studies	Summary Statistics			
		Mean	Median	Minimum	Maximum
Probability[1]	13	77%	78%	55%	93%
Hospital/Physician[2]	7	89%	94%	60%	99%
Center-Based[3]	1	95%	95%	95%	95%

1. There were 14 Probability-based studies reviewed. There is some question on the accuracy of the response rate for one of the studies (NHEXAS-MD). This study was excluded from the calculation of these summary statistics.
2. One study conducted in Minnesota (The Diana Project) had an unusually low response rate due to the recruitment approach. This study was excluded from these summary statistics.
3. Initial response rates only could be identified for one of the two center-based studies.

A study by Buck et al. (2003) assessed recruitment rates of studies that were concerned with recruiting women prior to conception and that had at least three months of follow-up with the women, but no follow-up with the children. The study indicated contact rates ranging from 2% up to 67% and participation rates that ranged from 42% to 77%. Generally, letters were used as the method of recruitment, with some studies using media and some using physicians/fertility awareness clinicians. The two studies with the highest recruitment rates targeted women that had originally participated in a cohort study and targeted women at a specific company.

Certainly, there is a significant amount of variability in the initial response rates from study to study, which is a function of the various methods used by the study administrators for identifying and then recruiting participants for the study. Again, the method by which participants are contacted and recruited into the study will have a significant impact on the recruitment rate. Booth and Johnson (1985) have suggested that a probability approach using a telephone-based selection and recruitment method is a viable option for recruiting participants for a longitudinal survey. However, initial response rates for surveys conducted via telephone have been steadily declining over the past ten years due to increases in technology and the negative influence of marketing calls on the willingness of potential respondents to participate. Cox et al. (2000) examine the response rates of 40 surveys conducted in the 1990s through telephone methods and report interview response rates ranging from 35% to 84 percent, with a median of 61 percent. Although many of these studies were cross-sectional in nature, the initial response rates for longitudinal studies could be expected to be similar, if not somewhat lower. Booth and Johnson (1985) report an initial response rate of 65% and Brick et al. (1997) report an initial response rate of 77 percent. However, in terms of a longitudinal survey, initial response rates of 60% to 70% may not be sufficient to ensure that the response rates at subsequent stages are adequate, even if the response rates for each particular wave of the survey are very high. For example, Booth and Johnson (1985) report a follow-up response rate of 78 percent, which yields an overall response rate of 51% for the survey. Brick et al. (1997) report similar results (80% response rate in follow-up survey, yielding a 62% overall response rate), which led them to question the feasibility of employing a telephone recruitment method for longitudinal surveys. By contrast, initial response rates for personal interview surveys, where the initial contact is

made in a face-to-face contact, are generally higher than those found in surveys that employ telephone recruitment methods (Drew et al. 1998).

It is important to note that an exhaustive review of the literature was not undertaken. Thus, the statistics presented in Table 7-1 need to be carefully considered, since they are based upon a fairly small sample of studies. Moreover, it is noteworthy that the NCS is likely to be more burdensome than the reviewed studies because of the length of the study and the extent of data collection. Therefore, the NCS might experience lower response rates than the rates presented in Table 7-1 (or Table 1 of Appendix G). On the other hand, the NCS may have higher response rates than the observed studies because of its sheer size and notoriety. However, the published literature did not provide strong evidence that community-wide media, advertisement, and other information dissemination campaigns will have a substantial impact on the recruitment of women, based upon a review of studies conducted in the U.S.

In summary, it is difficult to estimate the initial response rates that will be obtained in the NCS. However, initial response rates between 70% and 90% would be reasonable guidelines under a probability-based approach with response rates nearer the higher end of this range if in-person recruiting is employed and closer to the lower end of the range if telephone methods are employed. Although only two studies that employed a Center-based patient list approach were reviewed, these studies may suggest that the initial response rates under a Center-based patient list approach may be in the range of 80% to 99%.

7.2 RETENTION RATES

In any longitudinal study there will be participants who discontinue participation for a variety of reasons, which can accumulate over time and yield significantly lower overall response rates for later waves of the survey. Therefore, it is important to understand potential factors that are related to loss of participation due to attrition, and take active measures to mitigate this effect.

Retaining participants in a longitudinal study is sometimes more difficult than initial efforts to recruit them into the study. There are generally three types of attrition that can occur in a longitudinal study: (a) attrition because the participant is no longer representative of the population of interest (such as when a person becomes institutionalized on a long-term basis), (b) refusals, and (c) failure to track and locate study participants from one survey stage to the next. The first of these is difficult, if not impossible, to control. However, there are efforts that can be undertaken to reduce refusals and to track study participants. Several authors suggest tracking study participants as the single largest action that can be taken to reduce attrition. Many of the other methods associated with retaining participants are targeted toward preventing refusals and are the same as those discussed for improving the initial response rates, such as the use of incentives and minimizing respondent burden. Retention of study participants can also be influenced by the following:

- Testing at schools, including medical exams, particularly among school-aged children;

- Day of the week sampling occurs (e.g., offering weekend appointments) for working parents;
- Flexibility in scheduling of visits, examinations, data collection efforts, etc.;
- Knowledgeable, well-trained, motivated, and persuasive study staff that establish good rapport with participants and are persistent in obtaining responses;
- Providing personalized attention to study participants;
- Providing feedback to participants, including quick responses to questions, and results of medical tests;
- Use of in-person visits;
- Imparting a sense of partnership/ "Good for Mankind" aspect of the study to motivate participants.

Many of these, and other methods, were employed in the reviewed studies with varying success. A few of the studies had retention rates in excess of 80% over a number of years. Generally, the longer studies saw declining retention rates over time. The final retention rates observed ranged from 31% to 57%. Possible explanations for higher retention rates in some studies, relative to others, may be excellent tracking of participants and/or the infrequent sampling/interviews required for a study. Additionally, there is some evidence that attrition by study participants is not uniform. The characteristics of participants that are lost-to-follow-up can vary significantly from study to study.

There were three steps to the formulation of estimates of retention rates that could be anticipated in the NCS. First, information on the retention rates observed in previous longitudinal studies was obtained. Second, models were fit to these data to establish estimates of retention rates over time that could be applied to the NCS. Finally, various assumptions were applied to the estimated model parameters to formulate retention rates for the different design options under consideration for the NCS. The following section provides additional details for each of these three steps.

7.2.1 Information on Retention Rates in Previous Studies

Only a limited number of studies are somewhat similar to the NCS where retention rates over time could be identified in the recent literature review summarized in Appendix G. For many of these studies, only the retention rates for a specific point in time or sampling round could be readily identified. Table 7-2 provides a brief summary of the initial response and retention rates for the studies that were reviewed. To maximize the amount of information available for estimating potential retention rates for the NCS under different scenarios, information on the long-term response rates presented in a recent report prepared by Strauss et al. (2003) were included. A copy of the long-term response rates presented in that paper are included in Table A-1 in Appendix G.

Table 7-2. Brief Summary of Studies Examined for Initial Response and Retention Rates

Study [Reference]	Identification Method	Types of Measurements[a]	Initial Response Rate	Length of Follow-up	Latest Retention Rate	Figure 7-1, 7-2 Plotting Symbol	Source of Retention Rate Cited in Figures
National Human Exposure Assessment Survey, EPA Region 5	Probability – Area Sampling	Q, F, E, B, Diary	72%	2 yr	28.4%	P	Review
National Human Exposure Assessment Survey, Maryland	Probability – Area Sampling	Q, F, E, B, Diary	35%	1 yr	86%	P	Review
National Human Exposure Assessment Survey, Arizona	Probability – Area Sampling	Q, F, E, B, Diary	79%				
National Longitudinal Survey of Children and Youth	Probability Area Sampling	Q, Skill assessment	86%	8 yrs	66%	P	Review
Danish National Birth Cohort	Physician	Q, B	60%	2-4 yrs	92%	D	Review
The Collaborative Perinatal Project	Center-based (Patient Lists and Volunteers)	Q,B, Medical Exam	95%	7 yrs	84%	9	Strauss et. al.
The Diana Project	Hospital/Physician – Patient Lists	Q, B, Medical Exam	7.2%	4 yrs	66%	E	Review
Avon Longitudinal Study of Parents and Children	Hospital/Physician – Volunteer	B, Observational, Medical Exam	85%	9 years	78% mothers, 81% children	B	Strauss et. al.
British Cohort Study, 1946 Cohort	Physician – Births in a 1-Week Period	Q, B, Medical Exam	90%	53 yrs	72% after 4 yrs 57% after 53 yrs	4	Strauss et. al. and Review
British Cohort Study, 1958 Cohort	Hospital – Births in a 1-Week Period	Q, Medical Exam	98%	33 yrs	66%	7	Review
British Cohort Study, 1970 Cohort	Hospital– Births in a 1-Week Period	Q,	93.8%	26 yrs	56%	F	Review
Bogalusa Newborn-Infant Cohort	Hospital/Physician	Q, B, Medical Exam	98.4%	7 yrs	30.5%	G	Review
The Boston Residential NO_2 Characterization Study	Probability	E	60%	1 yr	83%	P	Review
Early Childhood Longitudinal Study: Kindergarten Class	Probability – schools	Q, School records	66%	6-7 yrs			
Framingham Children's Study	Non-probability – family history w/study	B, Medical Exams, Diary	58%	3 yrs	94%	H	Review
Framingham Heart Study	Probability- list sample augmented with volunteers	B, Medical Exams, Medical history	68.7%	50 yrs			
National Cooperative Inner-City Asthma Study	Center of Excellence	Q		~1 yr	89%	I	Review

Study [Reference]	Identification Method	Types of Measurements[a]	Initial Response Rate	Length of Follow-up	Latest Retention Rate	Figure 7-1, 7-2 Plotting Symbol	Source of Retention Rate Cited in Figures
National Health and Nutrition Examination Survey I	Probability	Q, B, Medical Exam	99%	1 visit	76.4% with incentive; 68.1% with no incentive	P	Review
National Health and Nutrition Examination Survey II	Probability	Q, B, Medical Exam	91%	1 visit	80%	P	Review
National Health and Nutrition Examination Survey III	Probability	Q, B, Medical Exam	83%	1 visit	74%	P	Review
National Household Survey on Drug Abuse	Probability	Q	93% annually	1 visit	78% on average	P	Review
National Survey of America's Families	Probability	Q	Ranged from 77% to 78%	1 visit	84% (child), 80% (adult)	P	Review
Tucson Epidemiologic Study	Probability	Q, B, Medical Exam	55%				
Survey of Income and Program Participation	Probability	Q	92%	2 yrs	72% without incentive 73% with $10 incentive 75% with $20 incentive	P	Review
Mater Misericordiae Mother's Hospital-University of Queensland Study of Pregnancy	Physician/Hospital	Q	99%	14 yrs	62%	1	Strauss et. al. and Review

a. Q = Questionnaire; F = Food Samples; E = Environmental Samples; B = Biological Samples

One difficulty in utilizing retention rates from previous longitudinal studies with extended follow-up is that these studies are generally not initiated prior to birth (i.e., many begin at birth or when a child is 6 months to 2 years of age). For the purposes of modeling, all studies were assumed to begin at "year zero" and the retention rates were then based upon "years of follow-up" rather than a child's age. While necessary, the pitfall of this is that it requires the assumption that the retention rates are the same regardless of age of the respondent.

A second difficulty in applying retention rates observed in previous studies is that in many instances only the response rates and not the retention rates at various stages of the study are available. Therefore, for these studies it was necessary to convert the reported cumulative response rates to retention rates. This was accomplished by assuming that the retention rate at the initiation of the study was 100% and the retention rates at subsequent stages of the study were calculated by dividing the reported cumulative response rate for the stage by response rate reported for the first stage.

7.2.2 Modeling of Retention Rates

Very few longitudinal studies employed a probability-based sampling approach and reported retention rates over time. Thus, the focus of the modeling effort was to develop estimates for a Center-based or hospital-based approach where participants were identified through patient lists. A two-step process was used to model the retention rates. First, a separate

model was fit to each study. Next, a weighted average of the estimated model parameters was calculated and used to estimate retention rates.

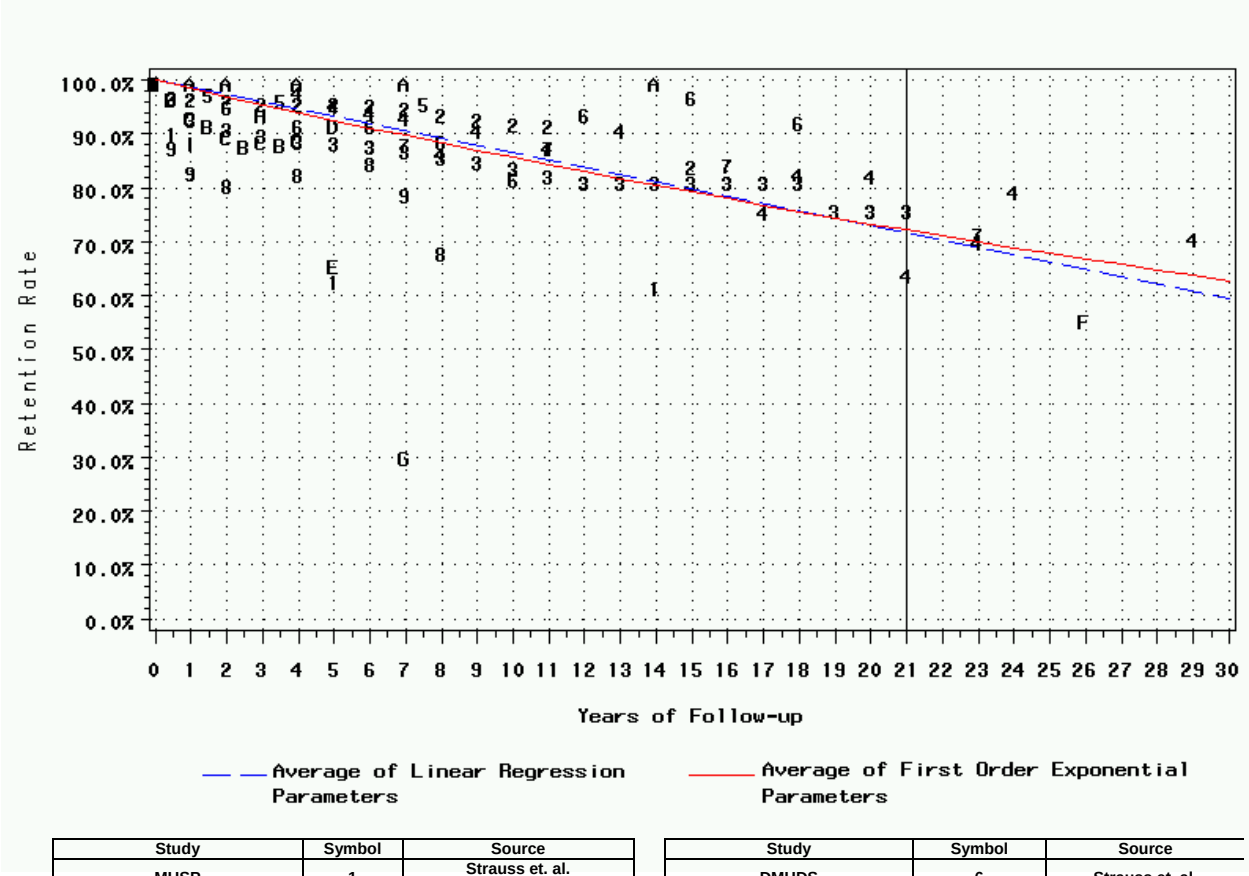
Two different functional models were fit separately to each study. A linear regression model (see Equation 7-1) was fit to each study. The intercept parameter for this model was constrained to be 100% at zero years of follow-up. The linear regression model assumes that the retention rate declines steadily over time with no variation in the rate of decline. That is, the decline in retention rates between 1 and 2 years of follow-up is the same as the decline in retention rates between 20 and 21 years of follow-up. However, this may not be an appropriate assumption because it may be more reasonable to assume that the retention rates will decline quickly during the first few years of follow-up but will then start to level off as the study progresses (i.e., continuing study participants have become invested in the study and are less likely to refuse continued participation). Therefore, a first-order exponential decay model was also applied to each study (see Equation 7-2). Again, the model was constrained to estimate a retention rate of 100% at zero years of follow-up. This model allows for a rapid decline in the retention rate during the initial study years, with a leveling of the decline in retention rates over time. Both models were only fit to studies where retention rates were available for more than one point in time (other than at the initiation of the study).

$$\text{Retention Rate} = 100\% - \beta*(\text{Years of Follow-up}) + \varepsilon \qquad (7\text{-}1)$$

$$\text{Retention Rate} = 100\%* e^{-\beta*(\text{Years of Follow-up})} * \varepsilon \qquad (7\text{-}2)$$

7.2.3 Calculating Estimated Retention Rates

The parameter estimates from the models applied to each study were averaged, and this average was used as the estimated parameter for predicting retention rates for the NCS. However, a weighted average where studies were weighted according to the number of years of follow-up was employed. These weights were deemed to be necessary because an unweighted average places too much emphasis on studies where retention rates were only available for a relatively short follow-up period, which forces a sharper decline in the estimated retention rates than seems appropriate (i.e., an unweighted average places more emphasis on modeling the initial decline in retention rates in the first few years of a study). The results of the modeling and averaging of parameter estimates is presented in Figure 7-1. The dashed line in Figure 7-1 represents the weighted average of the linear regression slope coefficients while the solid line represents the weighted average of the exponential decay model parameters. As observed in the figure, there was very little difference observed between the two different model approaches. Nevertheless, the exponential decay model was utilized because it facilitates estimating a decreasing rate of attrition over time.



Study	Symbol	Source
MUSP	1	Strauss et. al. and Reviewed
CHDS – New Zealand	2	Strauss et. al.
CHDS – Original	3	Strauss et. al.
BBC-1946	4	Strauss et. al. and Reviewed
MSRC	5	Strauss et. al.
ALSPAC-Mothers	B	Strauss et. al.
DNBC	D	Reviewed
BCS - 1970	F	Reviewed
Framingham CS	H	Reviewed

Study	Symbol	Source
DMHDS	6	Strauss et. al.
BBC-1958	7	Reviewed
DBC	8	Strauss et. al.
CPP	9	Strauss et. al.
SLPS	A	Strauss et. al.
ALSPAC-Children	C	Strauss et. al.
Diana	E	Reviewed
Bogalusa	G	Reviewed
NCICAS	I	Reviewed

Figure 7-1. Estimated Retention Rates Based Upon Weighted Average of Model Parameters

The studies used to calculate the estimated retention rates presented in Figure 7-1 were mostly based upon a Center, hospital, or physician-based patient list approach to identify potential study participants. Therefore, these estimates may be reasonable for a Center-Based (Patient List) approach. There were relatively few probability-based studies where retention rates over time were available. A similar approach was attempted with these probability-based studies (i.e., fitting separate models for each study and then calculating a weighted average of the parameter estimates), but this resulted in unreasonable estimates (i.e., much too low) because of the limited amount of available data and the relatively short period of follow-up reported for these studies. Similarly, a modeling approach could not be utilized to calculate estimates for the Probability-Based (Center) approach and Volunteer Sample approach. Estimates of potential retention rates for these three approaches were developed by applying a multiplier to the weighted average parameter estimate (based upon the first-order exponential decay models) for the Center-based patient list estimates. More specifically, the following assumptions were employed:

- Probability-Based (National) – The rate of decline in retention was assumed to be five times greater than the rate of decline in a Center-based (Patient List) approach. Again, exponential decay models were fit to the probability-based rates, but this resulted in unreasonably low retention rates. Even after removing the lowest three retention rates, the estimated decline was still more than five times greater than the rate of decline for the Center-based (Patient List) approach. Because there was some question of the applicability of the retention rates over time for the probability-based studies, a more moderate rate of decline (five times the Center-based (Patient List)) was utilized. Though not supported with data from the reviewed studies, different assumptions could be made for these rates as discussed in Section 7.3, including more of a leveling of the retention rates over time.

- Probability-Based (Center) – The rate of decline in retention was assumed to be four times greater than the rate of decline in a Center-based (Patient List) approach. Although the probability-based (Center) approach is very similar to the probability-based (National) approach, a slightly smaller multiplicative factor was used because the presence of a Center of Excellence may increase the visibility and perceived credibility of the study to participants, which could result in higher retention rates.

- Volunteer Sample – It was assumed that participants of a study under a volunteer recruitment approach would likely be more willing to participate in the study over time than would participants who were selected from a patient list. Therefore, the rate of decline in retention for a volunteer sample was assumed to be 0.85 times the rate of decline estimated for the Center-Based (Patient List) approach.

Figure 7-2 is a graphical summary of the assumed retention rates for the four different approaches. Also included on the graph for reference are the retention rates identified for probability-based studies (denoted by a "P"). Table 7-3 provides a summary of the estimated retention rates for the four different design approaches in tabular form.

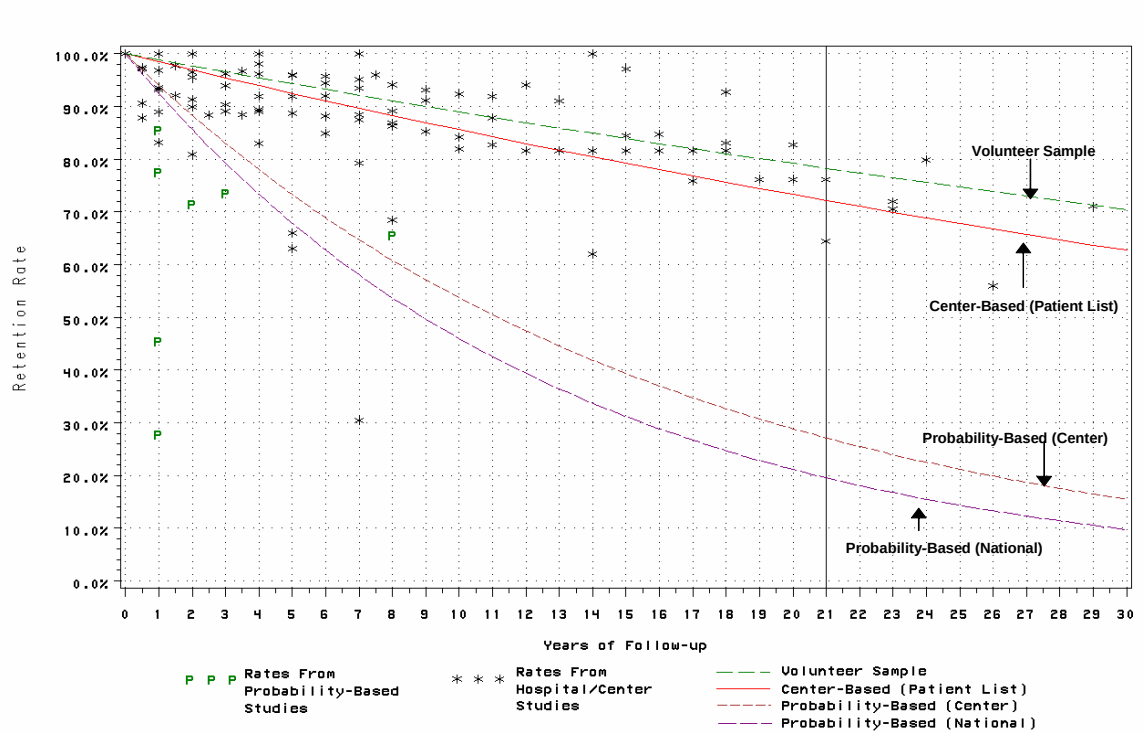


Figure 7-2. Graphical Summary of Estimated Retention Rates for Each Design Option

7.3 CONSIDERATIONS AND LIMITATIONS

As previously discussed, no study has ever been conducted in the U.S. that has the same scope, size, and design as that being considered for the NCS. Therefore, projecting the expected initial response and retention rates that might be observed in the NCS is challenging and inherently relies upon many assumptions that could significantly impact the response and retention rates. For example, the NCS is likely to be more burdensome than previous studies because of the length of the study and the extent of data collection. Without a directly comparable study, the impact of this increased burden on reducing response and retention relative to the impact of other factors that may increase the response and retention (e.g., the size and notoriety of the study) is difficult to concretely ascertain. In addition to the lack of a directly comparable study, there are a number of other significant assumptions that directly impact the calculation of response and retention rates including (a) limitations in the data, (b) alternative modeling approaches, and (c) factors related to study approach.