A COST-EFFECTIVE AND FEASIBLE DESIGN FOR THE NATIONAL CHILDREN'S STUDY (NCS): RECOMMENDATIONS FROM THE FIELD

PREPARED AND SUBMITTED ON BEHALF OF 28 OF THE 40 STUDY CENTERS ENGAGED IN THE CONDUCT OF THE NCS

INITIAL DRAFT – April 20, 2012

Robert D. Annett, Ph.D., University of New Mexico

Dean Baker, M.D., M.P.H., University of California - Irvine

John Bancroft M.D., Maine Medical Center

Michael Bracken Ph.D., M.P.H., Yale University

Stephen Buka, Sc.D., Brown University

Jane Cauley, Dr.P.H., University of Pittsburgh

Jennifer Culhane, Ph.D., University of Pennsylvania

Dana Dabelea, M.D., Ph.D., Colorado School of Public Health

Dorr G. Dearborn, Ph.D., M.D., Case Western Reserve University

Carolyn Drews-Botsch, Ph.D., M.P.H., Emory University

Nancy Dole, Ph.D., University of North Carolina at Chapel Hill

Donald J. Dudley, M.D., University of Texas Health Science Center at San Antonio

Maureen Durkin, Ph.D., DrPH, University of Wisconsin - Madison

Daniel Hale, MD, University of Texas Health Science Center at San Antonio

Mel Hovell, Ph.D., M.P.H, San Diego State University

Irva Hertz-Picciotto, Ph.D., M.P.H., University of California - Davis

Charlotte Hobbs, M.D., Ph.D., University of Arkansas for Medical Sciences

Mark Hudak, M.D., University of Florida

Claudia A Kozinetz, Ph.D., M.P.H., Baylor College of Medicine

Steven E. Lipshultz, M.D., University of Miami

George Lister, M.D., University of Texas Southwestern Medical Center

Thomas McLaughlin, Sc.D., University of Massachusetts

Jeff Murray M.D., University of Iowa

Nigel Paneth M.D., M.P.H., Michigan State University

James M. Robbins, Ph.D., University of Arkansas

Anna Maria Siega-Riz, Ph.D., R.D., University of North Carolina

Bonny Specker, Ph.D., South Dakota State University

James M. Swanson, Ph.D., University of California - Irvine

David J. Tollerud, M.D, M.P.H.,University of Louisville

Mari S. Wilhelm, Ph.D., University of Arizona

Sharon B. Wyatt, Ph.D., CANP, FAAN, University of Mississippi Medical Center

OVERVIEW

As of January 1, 2012 over $750,000,000 has been allocated to implement the National Children's Study (NCS), much of which has been spent on direct fieldwork at 40 Study Centers throughout the U.S. Since 2005, an experienced group of physicians, epidemiologists, demographers, geneticists, statisticians, health service researchers, behavioral scientists and others have worked individually and collectively to convert the aspirations of the NCS into a concrete and feasible plan for fieldwork.

This document synthesizes the recommendations, honed by field experience, of many of the Principal Investigators who have served as the advance guard for the NCS. We hope it will be received as a constructive response to the document "*Potential Sampling Strategies for the Main Study April 12, 2012*" prepared by the NCS Program Office (PO).[1] That document provided an overview of the NCS and broadly described seven generic design approaches (two of which involve probability samples) which are under consideration by NIH. The document also includes an initial assessment of the strengths and limitations of these design approaches, and invites submission of other design plans, which we are pleased to do.

Our document presents a feasible yet rigorous study design that utilizes existing infrastructure to achieve the original purpose of the NCS as outlined by Congress and endorsed by the Institute of Medicine. We invite the Program Office and the NCS Federal Advisory Committee to review our suggested blueprint for study design and implementation. It is our hope that this document will accelerate promulgation of an NCS Main Study design that will meet or exceed the original objectives in a fiscally responsible manner.[2]

This document emphasizes the following series of collective observations by field investigators:

1. Use of convenience rather than probability sampling in a large cohort study will dramatically impoverish the scientific value of the study (see Attachment A). Hence, our proposed design retains probability sampling.
2. Significant investment has established valuable infrastructure and investigator expertise which should be optimized in the final design for the Main Study. This includes the considerable investment that generated the original sample of 105 counties[3] or Primary Sampling Units (PSUs) as well as the substantial infrastructure developed at the 40 Vanguard study locations over many years.[4]
3. Prenatal care providers constitute an efficient legitimate sampling frame that can generate a nationally representative probability sample (in this respect, we are closely aligned with the *Potential Sampling Strategies* document).
4. A probability-based sample of 100,000 pregnancies can be recruited and followed within the annual level of funding that Congress has most recently appropriated.

[1] We will refer to this as the "*Potential Sampling Strategies*" document throughout this text.

[2] This document summarizes two longer documents which are appended.

[3] In a few instances, the Primary Sampling Units for the NCS were combinations of multiple counties or parts of counties. While the term " 105 study locations" is more precise, we use the simpler phrase "105 counties" throughout this document.

[4] We support the recommendations made in the U.S. President's Budget for FY 2013 regarding the NCS to "… reduce costs by building on existing infrastructure".

A. SCIENTIFIC OBJECTIVES AND DESIGN REQUIREMENTS FOR THE NCS MAIN STUDY

Design requirements for the NCS Main Study are driven by scientific objectives and budgetary limitations. A major challenge in choosing a design for the NCS Main Study has been the need for a clear articulation of the primary aims that the study seeks to achieve. The *Potential Sampling Strategies* document (pp 2-7) describes the research objectives of the NCS. We have distilled these into a set of key aims and essential design considerations for the Main Study derived from the PO document, supplemented by our scientific perspectives and first-hand experiences communicating and implementing these objectives in the field.

1. Enrollment of a nationally representative probability sample including "diverse populations of children";
2. A sample size adequate to study the etiology of relatively rare but complex childhood disorders with high impact (e.g., juvenile diabetes, severe intellectual disability, autism);
3. Enrollment of women early in pregnancy and, if feasible, prior to conception;
4. Follow-up through infancy and childhood to early adult life (21 years or longer) that includes use of standardized and validated tools to assess childhood health and developmental outcomes. Medical records, electronic or otherwise, cannot be relied upon as the sole or major source of such assessments. Consistent with Section 1004 of the Children's Health Act, these instruments should "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";
5. Collection and storage of high-quality biological and environmental samples during the preconception period where possible, but certainly during pregnancy, birth, infancy and childhood, with emphasis on critical time windows that include gestation;
6. Investigation of health disparities;
7. Inclusion of an initial set of *a priori* hypotheses to help ensure the relevance and adequacy of the study design and data collection protocols for hypothesis testing;
8. Refinement of study design through the engagement of individuals who are experts in the conduct of longitudinal studies of children's health and environmental exposures with pregnancy or neonatal enrollment, and in the complex analyses of study results; and
9. Involvement and participation of representatives of the communities in which the NCS operates.

However, a clear statement of aims and objectives alone will not result in final design choices. It will also be necessary to construct a study design that takes account of federal funding levels. Prior designs for the NCS Main Study were elegant from a sampling and design perspective, but when implemented in the real world, they could not be executed with the available budget. The design that we propose can fulfill each of the nine essential requirements listed above within the expected fiscal resources for the NCS (see Section D below). We propose that this list comprise the standard used to compare the multiple alternative designs under consideration. Choosing among a plethora of design options for the NCS (as for most any major scientific study) must maximize the achievement of the principal aims in a scientifically valid method within the anticipated resources available over the lifetime of the study.

B. THE CRITICAL VALUE OF PROBABILITY SAMPLING IN THE NCS

The *Potential Sampling Strategies* document describes several proposed sampling strategies for the NCS Main Study that are based either entirely or partially on convenience samples. Until recently, a national probability sample was the sole design consideration for the NCS. Minutes from the NCS Federal Advisory Minutes from Dec 5, 2006 (page 13) summarize the advantages of such a design:

> *"A national probability sample was chosen because such a sample will:*
>
> - *Ensure that exposure/outcome relationships are valid and apply to the children of the United States*
> - *Avoid selection biases that could lead to invalid inferences*
> - *Capture the diversity of the U.S. population*
> - *Represent the range and diversity of exposures and outcomes and help ensure that key exposures are not missed, given uncertainty regarding their distribution.*
>
> *…The National Center for Health Statistics led development of the multistage probability sample of primary sampling units. The sample includes 105 locations roughly corresponding to counties, or clusters of adjoining counties; 78 sites are metropolitan and 27 are rural. ...*
>
> *A center-based strategy for implementation was chosen because of centers' scientific expertise and facilities needed to carry out the Study's measures. The Study requires flexibility and adaptation of the centers to the scientific design and requires support and guidance by a coordinating center."*

In contrast, the convenience samples proposed in the *Potential Sampling Strategies* document would irreversibly burden the study with at least two major scientific flaws. Each of which will critically impede the original NCS goal of making valid new discoveries that can be accurately generalized to benefit the entire population of US children. First, non-probabilistic convenience sampling inherently restricts the ability of a study to accurately assess a full and representative range of both rare and common conditions and exposures within a population. Hence this approach curtails opportunities to make important new scientific observations and to unmask novel public health issues. Second, and of equal importance, convenience sampling limits the generalizability of study findings and can lead to invalid conclusions. Primary arguments for retaining a national probability sample for the NCS are summarized below and expanded upon in the appended document *"The Critical Value of Geographic Probability Sampling in the NCS*" written by several NCS principal investigators.

B.1. Limited Generalizability and Lost Opportunities

1. Knowing the true burden of disease in children defines public health opportunities: If the NCS Main Study design foregoes probabilistic sampling, we will squander a precious opportunity to secure generalizable incidence and prevalence data on key childhood diseases for which we have little or no national information.
2. Accurate epidemiological data are vital for hypothesis generation: The descriptive epidemiology that emerges from analysis of sound incidence and prevalence data that can be compared across the country - by region, urban-rural status, ethnicity or even neighborhood type - is a necessary resource for generating hypotheses about environmental exposures. To achieve this, detailed environmental exposure data are needed from many different environments around the country. Lacking this, the NCS will

not be able to generate the hypotheses that can elucidate the combinations of exposures associated with specific childhood outcomes.

B.2. Invalid Findings

1. Accurate risk estimates must precede treatment interventions: The absolute risk of disease subsequent to an exposure is a critical parameter to estimate in a cohort study. Non-probability samples are plagued by both under- or over-estimates of disease risk. For instance, patients referred to a major medical center are typically sicker than average. Volunteers for research, by contrast, are generally healthier than average and under-represent the medically disadvantaged. These and other factors interact powerfully with exposure-outcome relationships, and may distort estimates, in either direction, of the absolute impact of an exposure.
2. Identification of accurate associations requires probability sampling. It is often asserted that even though external validity (or generalizability) is compromised in non-probability samples, internal validity remains, especially in regard to associations between exposures and outcomes. Unfortunately, this is true only under the restricted circumstance of the absence of any interactions with the exposure-outcome relationships, and there are few examples of exposure-outcome relationships free of any interaction with a third variable. This is a particular concern in situations of low response, since response may be differential by exposure-disease status, but the absence of exposure-disease information on non-responders renders the potential for bias difficult or impossible to discern.

It will be argued that most studies in biomedicine are not based on probability samples. This is true, and therefore very few smaller studies are definitive or conclusive by themselves. The findings of most studies gain strength only after replication through many other studies on the same topic. The landmark scale of the NCS, in contrast to earlier studies, is such that conclusions will likely be "definitive" with replication unfeasible and unlikely.

C. A PROPOSED SAMPLING MODEL

We propose a multistage sampling model for the NCS Main Study that responds to the primary scientific objectives of the NCS, and takes advantage of the considerable benefits of full probability sampling. Central to the proposed model are the concepts of primary, secondary and tertiary sampling units, as described in the PO document (pg. 8). The initial design for the NCS Main Study obtained a probability sample of 105 counties[5] selected from the 3,141 counties in the US to serve as the Primary Sampling Units (PSUs) for the study. For the Vanguard Study, the Secondary Sampling Units were multiple (up to15) small geographic segments within each county. Tertiary sampling for the Vanguard study was, in essence, a 100% sample of all eligible pregnant and pre-pregnant women who resided within sampled segments.

For the NCS Main Study, we propose to retain the initial 105 counties as the Primary Sampling Units. We further propose to use prenatal care providers as the secondary sampling unit. In smaller PSUs, all pregnant women served by the providers who reside in the county may be the sample; in many larger PSUs, women would be further sampled (tertiary sampling) at the selected providers to reach the required enrollment numbers. However, other secondary and tertiary sampling approaches might also yield a rigorous probability sample within each study county. While we propose that provider-based sampling is the most cost-efficient approach for most locations, specific sampling strategies (especially for tertiary sampling) are best chosen in light of local conditions.

[5] In a few cases, parts of counties or groups of counties

We first detail the fundamental assumptions underlying our reasoning and then summarize the proposed sampling approach.

C.1. Assumptions

1. The objectives of the NCS require a true probability sample at all stages of the NCS design. Hybrid models and convenience samples do not have the same scientific validity. No subsequent national study is likely to have a profile, a level of funding, and a set of expectations as large as the NCS; hence, a full probability sample of the national population of pregnant women is essential. It is important to remember that a random sample of a convenience sample is also a convenience sample.
2. The ultimate goal of the study design is to achieve a probability sample of *pregnancies* (not births or providers, although these may occur as a result of probability sampling).
3. This document focuses on the sampling frame for pregnancies, which is likely to differ from the sampling frame needed for pre-pregnant women. A separate document will consider sampling the pre-conception population.
4. Provider-based sampling can be used to create a probability sample. It will only omit the 2% of women who do not receive any prenatal care, and it is important to ensure that such women are included in the study by sampling them at time of delivery.
5. A prenatal care "provider" must be defined liberally to include non-medical providers.
6. Some geographic and socio-economic dispersion of sampled women across the primary sampling unit is essential to achieve sufficient variance in environmental exposures to permit hypothesis testing. "Environmental" in this document is intended to include a broad range of biological, chemical, physical, socioeconomic, behavioral and lifestyle factors.
7. To reduce the provider burden of participating in the NCS, several sampling schemes that minimize intrusion into the clinical practice are possible. One method would be to concentrate sampling within limited time blocks. At a selected provider, recruitment may be limited to a few weeks (chosen randomly) during which all women seeking their first pre-natal visit are eligible for the study.
8. Tertiary level sampling (selecting women from the provider practice) must be integrated with provider sampling. There are both efficiency concerns and statistical considerations for selecting a smaller number of women from many practices versus a larger number of women from fewer practices. Moreover, local conditions must be considered in making these decisions. For example, in sparsely populated counties, all providers will need to be sampled; in more populous counties a stratified sample will be needed. Similarly, for women residing in some counties, a substantial number of providers and births may occur outside the county.
9. One challenge for the NCS is the cost of conducting assessments at multiple birth hospitals. It is very expensive and inefficient for the NCS to have to make contact with a large number of hospitals, many of which may only deliver one or two NCS babies a year. Provider-based sampling will help achieve this efficiency, because provider groups tend to deliver their babies in only one or two hospitals.
10. The primary level of probability sampling must utilize the 105 initially selected counties. This is not only a scientific imperative (and a main strength of the NCS recognized by the Institute of Medicine) but an economic one. The Main Study should make efficient use of the large amount of funds already invested in the NCS infrastructure, and the considerable efforts spent engaging local communities and healthcare providers in 40 of these counties must not be wasted.

C.2. The Proposed Sampling Model

The model is a multistage probability sample of counties, providers and participants that takes into account the above assumptions.[6]

First Stage: Sampling Counties

The 105 counties presently defined by the NCS is a probability sample of the 3000+ counties in the United States. At this time, we recommend that the existing 105 counties, in total or in part, be the basis for stage 1 sampling.[7]

Advantages:

1. Capitalizes on investment on infrastructure already made. This includes community, provider and hospital relations and university- based structures now familiar with the NCS.
2. Efficiency of retaining the existing 40 Vanguard counties in the final selection of PSUs cannot be ignored, as the NCS has committed to engage participants already enrolled by these counties for 21 years.
3. As the current Vanguard participants will constitute the "pilot" subjects for the NCS Main Study over the next 20 years, there is unquestioned value in pretesting procedures in the same locales that they will eventually be used.
4. The pre-conception cohort can use the same primary sampling frame, hence economizing resources.

Disadvantages:

1. We are not aware of any disadvantage of this choice. It has been suggested that this strategy does not capitalize on access to computerized medical records available at large HMOs. However, these are often administrative data and it has not been shown that HMO records are of "research quality" sufficient to test NCS hypotheses. Moreover, the proposed design does not preclude use of electronic records that are available at many provider offices, which may be appropriate to enhance other forms of study data.

Second Stage: Sampling Providers

The most efficient sampling frame for identifying pregnant women is the prenatal care provider. However, local knowledge must guide construction of a sampling frame that includes all types of providers and organizations who assist with prenatal care and the birthing process: obstetrical physicians, midwives, social organizations (eg. Healthy Start and WIC), and other community sites where pregnant women may receive help (e.g. homeless shelters). Optimal provider sampling procedures will vary by county. In less populated counties, all providers may be required. In heavily populous counties, stratified random sampling of providers is needed which takes into account several factors: the need for an equally weighted sample based on socioeconomic status and related characteristics, optimization of early gestational age at first study contact, achievement of a geographic spread across the county, and setting an upper limit for the number of hospitals in which study subjects will give birth.

[6] Further details of the proposed model are described in the appended paper: Implementing provider-based sampling for the National Children's Study: opportunities and challenges. Belanger KB, Buka SL, Cherry D et al. Draft Manuscript, April, 2012

[7] The 40 Vanguard sites are not a random sample of the 105; nonetheless, they may be representative and this can be readily examined by a comparison of birth certificate data from the 40 counties with national statistics. If it is not possible to conduct the Main Study with the full 105 counties, one alternative is to begin with the 40 vanguard sites and add counties to improve representativeness or for other reasons, this can be done through stratified sampling of the remaining study counties.

Advantages

1. Represents the most efficient and economical method for recruitment.
2. Permits estimation of probabilities of women enrolling in the study.
3. Limits the number of birth hospitals that need to be in the study, because most providers staff only one or two delivery hospitals.

Disadvantages

1. Some providers may refuse participation, but this can be managed by replacement sampling.
2. Some providers may require reimbursement, but the NCS Vanguard Study experiences have shown this to be a very modest cost.

Third Stage: Sampling Participants

Less populated counties may not require a tertiary stage of participant sampling - all pregnant women in the county would be eligible. In populous counties, sampling within providers will be more complex. To avoid this, quota sampling will be needed to limit the number of women sampled from each of the selected providers. Further, sampling in time blocks (e.g days of the week, or weeks of the year) represents a viable option to increase the efficiency of data collection and to reduce provider burden. This strategy would allow concentrated effort by both the provider and the study team. Different providers would be sampled in another time frame, bringing economic efficiencies to the research project.

Advantages

1. Probability sampling of participants within sampled providers preserves the total probability sample.
2. Sampling within time blocks increases study team efficiency and reduces provider burden.
3. Quota sampling within providers allows an appropriate socioeconomic and geographic spread of participants needed to test environmental hypotheses.

Disadvantages

1. In some counties a sufficiently large proportion of women may use providers and/or birth hospitals outside the county. These counties will require modifications to their sampling frame in order to represent these women.
2. In all counties, some women access prenatal care late or not at all. It is important therefore to enroll all eligible women at their first visit whenever it takes place, including at admission to hospital for labor.[8]

Finally, we note that the 1000 participants per county is a “senatorial” design choice that gives equal weight in the final cohort to all counties, irrespective of the size of their population. Our proposal endorses this concept as it enhances the statistical power necessary to test hypotheses based on the prevalence of environmental exposures limited to a few, perhaps under-populated, counties.

[8] We recognize that in some provider groups, the initial NCS interview may not be given until the second visit, to avoid patient burden.

D. COSTS OF THE NCS WITH A PROBABILITY DESIGN

D.1. Why has the NCS been so expensive so far – are dramatic changes needed?

All projections of cost in the NCS have indicated that it will be a very expensive project. The expense is intrinsically driven by the large sample size, the desired density of data collection, and the complexity of arranging collections at times in a woman's life that are unpredictable, and therefore difficult to schedule; in particular, during the first trimester of pregnancy and at birth.

Another set of factors driving costs is the inherently distributed nature of a study that considers itself *national.* A lower cost study could perhaps be performed in a handful of major medical centers, as was done in the National Collaborative Perinatal Project, or in a group of provider systems, as suggested in recent NIH communications[9], but such a solution would be neither probability-based, nor in any sense "national". However, as we discuss below, a probability-based approach does not need to add greatly to the cost of the study. While not appreciably reducing costs, moving from a probability-based study will seriously impair the scientific value and make it impossible to describe as a "national" study.

A second set of factors magnifying costs are those that are not intrinsic to a national study, but are associated with an overly centralized operational model, and, especially, a contract model. It has become clear to the authors of this document, all of whom have conducted NCS field work, that the NCS contract model which involves expensive and time consuming regulatory burdens, along with all strategic and executive decisions being made centrally with no or limited input from the field, has added greatly to the expense of this study.

Obtaining the FISMA[10] certification required by contracts before operations could begin added hundreds of thousands of dollars in expense to each study center, and delayed enrollment by many months. Requiring OMB approval for each change in the study, and for any sub-study undertaken, added further prolonged delays during which it was impossible to enroll subjects. The unnecessarily bureaucratic approach to each step of the study is exemplified by the figure on page 26 of the *Potential Sampling Strategies* document which shows that each study instrument must pass through 21 review steps requiring the input of 3 separate consulting firms in addition to the PO and the field investigators. At the original seven Vanguard Centers, whose five-year contracts began in September of 2005, enrollment began in 2009. During the first year of enrollment, eligibility was restricted to participants who were in their first trimester of pregnancy and knew that they were pregnant, reducing the number of eligible pregnancies by over 60% compared to the total pregnancies in the sampled segments. Active enrollment continued for between 26 and 30 months. For the 30 Vanguard Centers whose five year contracts ran from 2007-2012, no participants could be enrolled until 2011, and then for only 8 to 14 months before they were told by the PO to cease enrollment. In both groups, enrollment was often stopped before a steady state could be achieved. As a consequence, and coupled with the very expensive costs associated with developing 'facilitated decentralized' information management systems at each Study Center, it is clear that the estimated cost per enrolled study participant is grossly inflated. Furthermore, the initial sampling design was based on an assumption that enrollment would continue over multiple years to identify newly pregnant women who lived in the eligible segments. Thus, the cost of implementing the recruitment infrastructure cannot be evaluated based only on the initial short-term yield, which also inflated the estimated cost per enrolled study participant. These expensive and time-consuming factors

[9] For example, current documents on the NCS website ("Congressional Justification FY 2013") state "NIH now proposes that the Main Study sampling frame be based on provider location. One approach for developing such a sampling frame would be to use providers associated with specific health plans. Such an approach would have several advantages in terms of cost and feasibility, but would abandon the geographic based probability sample."

[10] The Federal Information Security Management Act (FISMA) legislates a comprehensive framework to protect government information, operations and assets against natural or man-made threats.

do not operate in cooperative agreements or grants, which rely much more on the expertise of the investigators and staff in the field actually conducting the study.

D.2. Creating an efficient NCS

We show here how savings may be achieved in a large and complex study such as the NCS by modeling operations on existing large observational pregnancy and birth studies similar to those conducted by several NCS investigators. Our financial projections are not theoretical; they are based on actual field recruitment.

In Table 1 we summarize costs/subject of pregnancy or birth enrollment in 16 studies conducted by NCS investigators, most with NIH support. All costs include indirect costs. The "estimate cost per subject" was derived by dividing the total cost for a project by the number of enrolled subjects, prorated to include only the initial period of enrollment, for example pregnancy to birth, or surrounding birth (for neonatal enrollment). For most studies, many costs were included that would not be in the NCS budget. For example, the costs of laboratory tests could not be removed from all studies. We assume that the NCS will not perform laboratory analyses during the period of enrollment, but will defer these to budget periods when the study is less costly. A majority of these studies had biological specimen collections, some at very frequent intervals; more than one-third collected specimens such as placenta or cord blood at delivery. Two had environmental sampling along with multiple home visits. Every study had interviews, as many as five during pregnancy in one study. So while some studies did not collect data in as much depth as the NCS, and others included more, overall, they included many of the features planned for the NCS. All data collection, management and analysis costs, and specimen storage costs are included in these estimates. The range of costs per subject is from $466 to $5,655 per subject, with a mean of $2,076. The variability in costs reflects variation in labor costs, mode of recruitment, as well as the frequency and intensity of biologic or environmental sampling, whether or not home visits were included, and other protocol features.

A major rate-limiting cost in the NCS is enrollment early in pregnancy and a requirement to collect in-hospital data at the time of birth. The high cost of pregnancy enrollment has stemmed partly from the difficulty of assembling pregnancies and births via household recruitment, a problem addressed by the proposed provider-based sampling and enrollment. The provider venue provides several opportunities for cost savings, especially the concentration of pregnant women in a single location, and the linkages of prenatal care providers to a restricted range of hospitals. We assume here that the number of sampled providers in counties should roughly parallel the number of segments, which represent the geographic cluster sample that parallels the clusters of providers. This number ranged from 3 to 15 depending on county population. The cost of enrolling 250 women per year in 3-15 practices should be manageable, and would reduce the number of hospitals involved, a key minimization if effective birth collections are to be managed. A feature of provider sampling is that in large counties, whole practices would have to be sub-sampled if the goal is 250 births per year. Thus, setting the rate of enrollment above 250 births per year might be feasible and efficient in larger counties, allowing faster completion of the enrollment at reduced per subject cost.

TABLE 1: COST OF ENROLLMENT AND INITIAL FOLLOW UP IN 16 PERINATAL STUDIES CONDUCTED BY NCS INVESTIGATORS.										
	INVESTIGATOR	**DECADE**	**SUBJECT ENROLLED**	**N**	**MEDICAL RECORD**	**INTERVIEWS (N)**	**BIOLOGICAL COLLECTION**	**BIRTH COLLECTION**	**IMAGING**	**ESTIMATED COST PER SUBJECT**
1	Paneth	1980's	Newborn < 2kg	1,105	YES	1			YES	$1,600
2	Martinez	1980's	3rd trimester	482		multiple	YES	YES		$5,655
3	Specker	1990's	Post-partum moms	383		3	YES		YES	$522
4	Specker	1990's	newborns	101		5	YES		YES	$990
5	Drews	1990's	newborns	966	YES	1	YES			$1,118
6	Dole,Siega-Riz	1990's	24-29 GA	3,163	YES	1	YES			$2,542
7	Bracken	1990's	< 24 wks GA	2,098		5				$1,667
8	Bracken	1990's	< 24 wks GA	2,478		1	YES	YES		$1,824
9	Bracken	2000's	Age 6 years	1,505		1	YES			$2,601
10	Dole,Siega-Riz	2000's	<16 wks GA	2,006	YES	3	YES	YES	YES	$5,436
11	Siega-Riz, Dole	2000's	Birth	689		2 home visits				$3,274
12	Herz-Picciotto	2000's	Early preg & pre-preg	296	YES	5	YES	YES		$5,000
13	Culhane	2000's	Newborn < 35 wks	1,126	YES	5	YES			$1,065
14	Culhane	2000's	infancy	1,451	YES	3				$959
15	Dabelea	2010's	pregnancy	877	YES	3	YES	YES		$770
16	Paneth	2010's	11-12 wks GA	590	YES	1	YES	YES		$446
TOTAL				**19,316**	**56%**	**100%**	**75%**	**37.5%**	**25%**	**$2,076**

GA = gestational age

D.3. A reasonable data framework for the NCS Main Study

Cost cannot be contained in the NCS without a realistic specification of a minimum protocol for data collection in pregnancy and at birth that fulfills the mandate of the NCS. A priori articulation of a set of hypotheses is essential to facilitate the protocols for data collection, but a general rule could be that only hypotheses pertaining to rare and burdensome disorders need data collection on all 100,000 subjects. Common disorders such as obesity, or continuous measurements such as healthy development, can be effectively studied on sub-samples much smaller than 100,000. Having a nationally representative probability sample from which to sub-sample in order to study such conditions, with funding outside the NCS core allocation will certainly enrich the value of the NCS.

The 2007 NCS protocol included no less than 10 different biological collections on every mother and infant, and up to three study ultrasounds – one in each trimester. Most would agree that such a protocol is overly ambitious and economically non-viable. The protocol framework described broadly in *Potential Sampling Strategies (p. 8)* is as follows:

> *"...data collections are scheduled twice, if possible, prior to about 20 weeks gestation and once later in pregnancy. Data collections for children are scheduled at birth and then every three months for the first year and every six months until five years old for a total of 13 opportunities. Seven of the opportunities will be face to face encounters and include biospecimen and environmental sample collection. The other six are remote data collections, typically by telephone interview. Subsequent data collections have not been scheduled but will be on average about every other year until 21 years old, for a total of eight additional opportunities. In sum, 21 data collection opportunities per child are planned, but that may change."*

Our financial projections (see below) are in broad agreement with these assumptions, although we suggest somewhat fewer in-person assessments during the postnatal years. The *Potential Sampling Strategies* document does not, however, address the content of any visit. We think that minimal content should include enrollment at the first or second prenatal visit with an interview; blood and urine sampling; and house air, water and dust. This should be obtained at least once during pregnancy. In addition, it is very important to collect a placental specimen. The baby should have a study examination at birth or shortly thereafter. Some form of medical record abstraction from mother and baby is required. Thereafter, household visits with environmental collections and examinations at age 2 (disabilities can begin to be ascertained) and age 5 (pre-school development) are needed at a minimum. Although placentas can be set aside for later collection and processing, this is not true for cord blood, which requires either a hospital that routinely collects that specimen (a small minority of hospitals) or one in which there is 24/7 delivery room coverage. Cord blood thus adds greatly to Study costs, or may better be reserved for add-on studies at centers with special interests in the topic.

Additional methods for reducing costs include taking advantage of the ongoing collection of urine and blood during routine prenatal care, and using birth certificate information as an inexpensive source of perinatal variables.

Costs can be reduced by providing local specimen storage for as long as feasible, thus deferring central storage till later years, when the expensive enrollment phase is over. The cost of providing -80 degree freezers to each center is not prohibitive. There is no evidence that liquid nitrogen freezing is needed for the kind of non-cellular biological materials collected in the NCS, and the initial plan to use that very expensive modality for central storage can be revised. At present, storage costs are greatly magnified by the current practice of individual mailing of each and every specimen by express shipment to a central collection site. Batch mailing of specimens after set periods or after collection of a set number of samples would greatly reduce costs but would entail local frozen storage.

D.4. Modeling Costs for the NCS Main Study

A budgetary challenge to implementing the NCS results from the fact that study costs are heavily front-loaded by the expense of the early protocol, which includes in-person assessments with biological samples during pregnancy and at birth. This is heightened by the fixed cost structure of the NCS, in which congress allocates approximately the same funding level each year. After the early phase, the study will include some in-person visits between birth and age 21, but during most of those years, follow-up, including cohort maintenance and data collection, will be conducted by less expensive methods such as phone, mail or internet. Assuming that this could be done for approximately $1,000 per subject (consistent with many previous epidemiologic investigations), in such years the cost for the full cohort would be on the order of $100M per year, which is less than the current annual allocation of $165M. The budgetary challenge, therefore, centers on the early years of recruitment.

In the first cost projections presented below, we assume that the cost to enroll and assess a participant in their first year in the study (pregnancy through birth) would be $5000 per subject. This estimate is twice as high as the average cost per subject for similar investigations (see 16 studies summarized in Table 1, Section D.2.) Because assessment costs beyond the first few years of enrollment are not anticipated to be very cost-limiting, and to keep this initial model simple, we have projected $1000 - $2000 per year for all subsequent years to age 21.

One solution to the high cost of the enrollment phase is to spread this over several years. In Figure 1 we show that even with enrollment spread over as many as six years (i.e. enrolling 16,666 subjects per year), if $5000 per subject is needed per enrollment year, at least one year will exceed the current annual budget level of $165M.

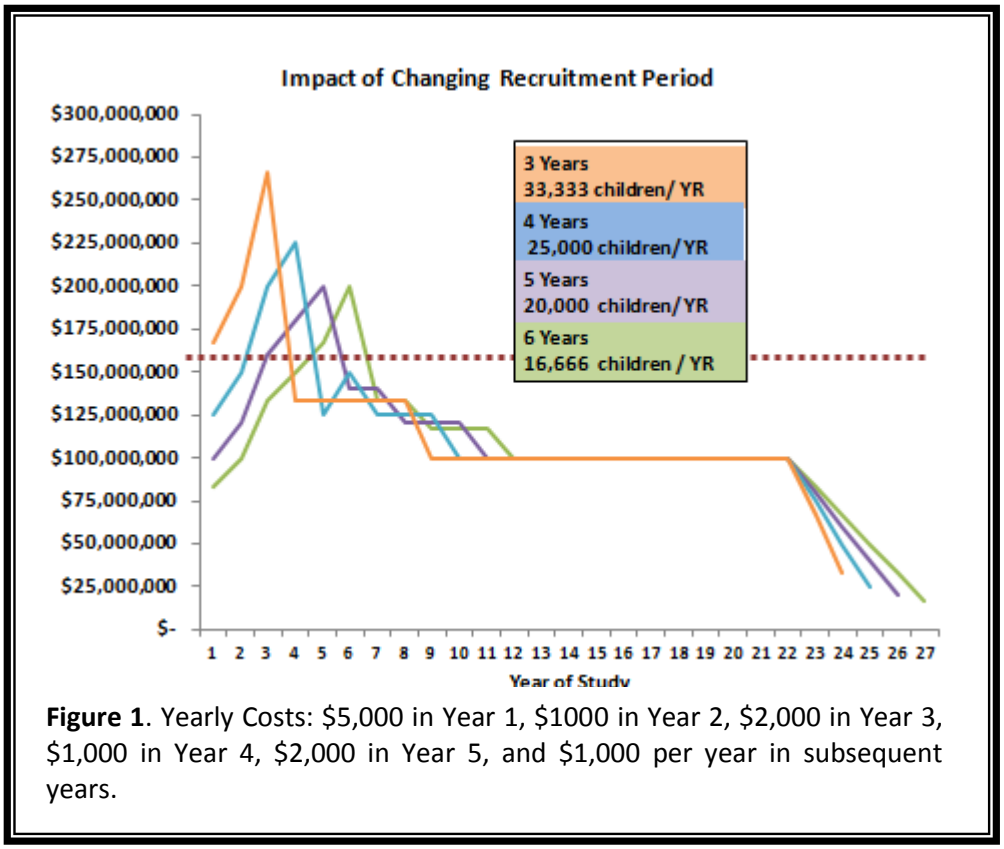

Figure 1. Yearly Costs: $5,000 in Year 1, $1000 in Year 2, $2,000 in Year 3, $1,000 in Year 4, $2,000 in Year 5, and $1,000 per year in subsequent years.

Our proposed solution is to efficiently consolidate recruitment at a given study site within a three year period, while also *lagging or staggering enrollment* by as much as two years between waves of

enrollment. In effect, sub-cohorts within the overall cohort are created that have different starting dates, each separated by two years. In the table below we provide an example, where the entire sample of 100,000 is divided into three sub-cohorts, each enrolled over three years, with the second and third cohorts initiated two years after their predecessors.

TABLE 2. Example of Staggering Enrollment Using 3 Sub-cohorts, Each Lagged by 2 Years.

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	
Sub-cohort 1	11,111	11,111	11,111					33,333
Sub-cohort 2	*2 year lag period*		11,111	11,111	11,111			33,333
Sub-cohort 3			*2 year lag period*		11,111	11,111	11,112	33,334
Total	11,111	11,111	22,222	11,111	22,222	11,111	11,112	100,000

As shown in Figure 2, staggering sub-cohorts in this fashion permits the entire cohort to be enrolled in 7 years while limiting the total cost for any study year. If the cost for the first year of enrollment can be kept at $3000 per subject (dashed tan line) this results in a maximum annual cost of approximately $165M. If the cost of the enrollment year is as high as $5000 per subject, then it will be necessary to have five cohorts, staggered by two years, to stay below the $165M cost ceiling (solid blue line).

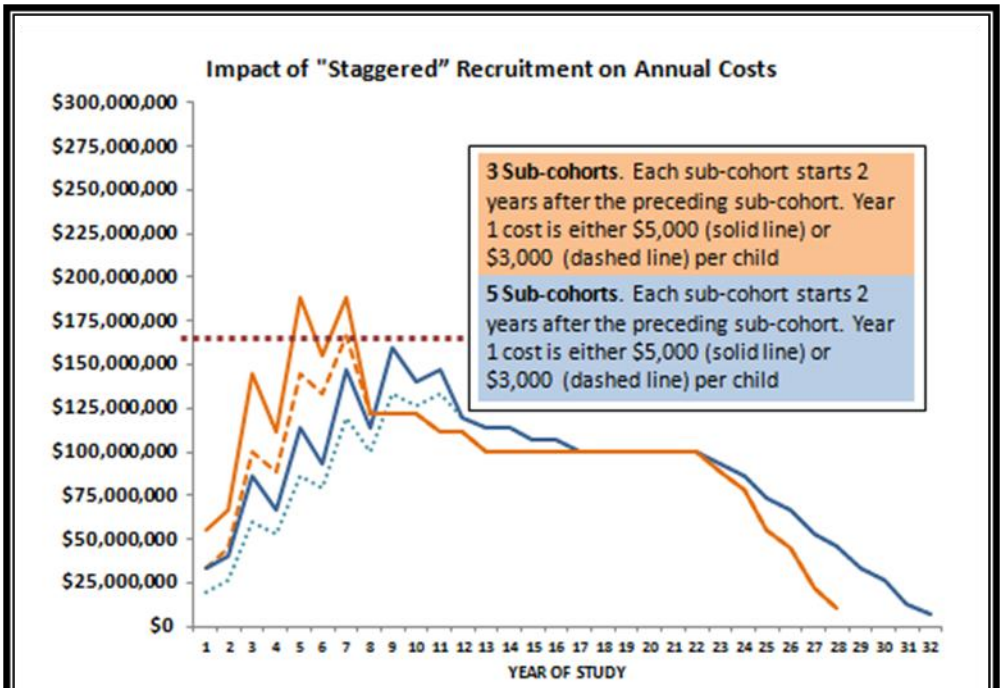

Figure 2. Yearly cost per child: Either $3,000 of $5,000 in Year 1, $1000 in Year 2, $2,000 in Year 3, $1,000 in Year 4, $2,000 in Year 5, and $1,000 per year in subsequent years. Recruitment is by sub-cohort

Detailed year-by-year costs for several models in which the number of years of enrollment, the lag period and the costs of the enrollment year are separately projected are available from Daniel Hale MD

(HALE@uthscsa.edu). In all these models, total costs of the NCS over 25-32 years would not exceed 3 billion dollars.

E) Next Steps and Conclusion

We have identified three aspects of this proposed design that require further consideration. We plan to address these over the next few weeks, and we briefly mention them here.

E.1. Is a provider-based approach feasible and optimal for all counties?

The 13 Vanguard Study Centers that have used provider-based approaches for pilot work are unanimous in their view that provider-based sampling is feasible, cost-effective and will be successful in their regions of the country. These range from large, sparsely populated rural areas where 100% of providers will need to be sampled, to urban regions with hundreds of providers and dozens of birth hospitals where a stratified sample will be required. These also include challenges where a large proportion of pregnant women receive prenatal care outside of the county, as well as those with and without access to electronic birth records from which to develop a sampling frame. As noted above, the blend of secondary and tertiary sampling strategies can and should vary across county, depending on local service delivery configurations.

E.2. Can a provider-based probability sample generate a representative preconception cohort?

The short answer here is “probably not”. Women who have regular contact with health care providers prior to conception are likely to differ substantially from all pregnant women. The stated goal that 15% of pregnant women in the study be from a preconception cohort (p. 4 of the *Potential Sampling Strategies* document) is a drop from the prior target of 25%. At this time there are no clear, cost-effective strategies to enroll a representative sample at even this smaller percentage. While there is great scientific interest in a preconception cohort, our view is that, for now, simultaneously designing an acceptable approach for both the preconception and pregnancy cohorts will compromise the clarity and precision of each. The sampling frame for pregnancies will in all likelihood differ from the sampling frame needed for pre-pregnant women. We favor a sequential planning process, where optimal designs for the pregnancy sample, comprising 85% of the total NCS cohort, are first refined and supplemental approaches for sampling the pre-conception population then added.

E.3. Is the anticipated budget for the NCS sufficient to cover all 105 counties previously sampled?

Our initial cost projections suggest that this is indeed possible. The number of 105 may seem unwieldy, but the NCS already has several coalitions of from 4 to 7 counties, which allow the NIH to incorporate several PSUs into one contract. Although the 105 county sample is the only national probability based sample currently available, if circumstances mandated a smaller set of counties, it would be feasible to conduct a post-hoc stratified sampling of the previously sampled 105 locations to achieve a representative sample. However, the potential cost-savings of reducing the number of study locations is not clear (and needs to be estimated), and both the national scope of the study and the breadth of environmental exposures covered would be reduced. Discontinuing fieldwork and dismantling the already invested infrastructure particularly in the 40 locations actively involved in the Vanguard Study, would be undesirable.

E.4. Conclusion

The authors of this document, 31 experienced Principal Investigators who have each invested 5-12 years to help achieve the historic objectives of the National Children's Study, contend that the scientific benefits of a geographic based probability sample of prenatal care providers far surpass all alternatives considered, each of which involve various forms of convenience samples. Further, we strongly support the language in the U.S. President's Budget for FY 2013 regarding the NCS to *"… reduce costs by building on existing infrastructure".* This should include the time and funds invested in the nationally representative sampling of 105 counties, as well as the millions of dollars invested in fieldwork, community relations, clinical arrangements and other forms of infrastructure in the 40 Vanguard Centers, as well as with several of the remaining 65 counties. The proposed design achieves all of the objectives of the Children's Health Act of 2000 within the anticipated budget and retains a national representative sample which was recognized as a study highlight in the 2008 review by the Institute of Medicine. As noted throughout the document, it is important to emphasize that the proposed sampling model permits considerable within-county flexibility for secondary and tertiary sampling, which is essential to ensure geographic representation at the county level. We appreciate the opportunity to share this plan with NCS leadership, the Federal Advisory Committee and others and look forward to ongoing exchanges, enhancements, and final plans for the design and conduct of the Main Study.

Attachment A:

THE CRITICAL VALUE OF GEOGRAPHIC PROBABILITY SAMPLING

IN THE NATIONAL CHILDREN'S STUDY (NCS)

In a comprehensive review of the NCS in 2008, the Institute of Medicine highlighted its probabilistic sampling methodology as a key strength of the initial NCS design. After generating a probability sample of US counties, further procedures identified household clusters for enumeration within each county, thereby producing a probability-based sample of births across the nation. However, as had been predicted by several perinatal epidemiologists, substantial inefficiencies plagued household-based recruitment efforts at the original vanguard centers during the second stage of sampling. A recruitment strategy designed to obtain a probabilistic sampling frame through partnership with prenatal care providers is now being tested in the field. The success of provider-based recruitment in small counties, where all providers were engaged in enrollment, strongly suggests that the inefficiencies of household sampling could be substantially remediated, while retaining probability sampling.

The NIH has now signaled that the NCS is likely to entirely abandon probability sampling methodology in favor of "convenience" sampling of participants from health systems or HMO's. Such a change would irreversibly burden the study with at least two scientific flaws. Each of these flaws will critically impede the original NCS goal of making valid new discoveries that can be accurately generalized to benefit the entire population of children in the US. Non-probabilistic convenience sampling inherently restricts the ability of a study to assay a full and representative range of both rare and common conditions and exposures within a population. Hence this approach curtails opportunities to make important new scientific observations and to unmask novel public health issues. Of equal importance, convenience sampling limits the generalizability of study findings and can lead to invalid conclusions.

LIMITED GENERALIZABILITY AND LOST OPPORTUNITIES

1. **KNOWING THE TRUE BURDEN OF DISEASE IN CHILDREN DEFINES PUBLIC HEALTH OPPORTUNITIES:** If the Main NCS design foregoes probabilistic sampling, we will squander a precious opportunity to secure generalizable incidence and prevalence data on several key childhood diseases for which we have little or no national information. Existing surveys such as the National Health Interview Survey (NHIS) (4,000 children under age 19) or the National Health and Nutrition Examination Survey (NHANES) (10,000 children) are an order of magnitude smaller than the NCS, and no other national probability sample in the US is large enough to describe with precision the frequency of important childhood conditions not routinely recorded in vital statistics data or in local or national disease registries. Examples of such conditions include asthma, type 1 and 2 diabetes, renal diseases, celiac disease, autism, and seizure disorders. We must ask ourselves: how will we prepare for health care needs in the future without knowledge of the burden and distribution of these and other diseases in children? How will we adequately plan for the range of educational and other developmental services that children with some of these disorders will require for optimal outcome? How can we address health disparities if the children from the defining socio-economic, geographic and ethnic strata are disproportionately represented in the study?
2. **ACCURATE EPIDEMIOLOGICAL DATA ARE VITAL FOR HYPOTHESIS GENERATION**: The descriptive epidemiology that emerges from analysis of secure incidence and prevalence data that can be compared across the country (by region; by urban-rural status; by ethnicity; by neighborhood or housing type) can be a gold mine for generating hypotheses about environmental exposures. For example, a discovery that renal diseases occur in significant excess in a group of rural counties would stimulate careful inquiry into possible agricultural exposure of children to nephrotoxic agents. This type of inquiry highlights the need for detailed environmental exposure data from many different environments around the country, without which we will not be able to generate the hypotheses that can elucidate the combinations of exposures associated with specific childhood outcomes.

INVALID FINDINGS

1. **ACCURATE RISK ESTIMATES MUST PRECEDE TREATMENT INTERVENTIONS:** The absolute risk of disease subsequent to an exposure is a critical parameter to estimate in a cohort study. If this risk can be altered or treated, accurate estimation of the risk:benefit ratio associated with treatment presupposes knowledge of the absolute risk of disease to allow comparison with the absolute risk of the intervention.

Non-probability samples are plagued by both under- or over-estimates of disease risk. For instance, patients referred to a major medical center are typically sicker than average. Volunteers for research, by contrast, are generally healthier than average, and under-represent the medically disadvantaged, an especially important subgroup in the pediatric population. These and other factors interact powerfully with exposure-outcome relationships, and may distort estimates, in either direction, of the absolute impact of an exposure. In turn, these erroneous estimates of risk can lead to over- or under-treatment or remediation of exposures.

2. **IDENTIFICATION OF ACCURATE ASSOCIATIONS REQUIRES PROBABILITY SAMPLING**: It is often asserted that even though external validity is compromised in non-probability samples, internal validity remains, especially in regard to associations between exposures and outcomes. Unfortunately, this is true only under the restricted circumstance of the absence of any interactions with the exposure-outcome relationships, and there are few examples of exposure-outcome relationships free of any interaction with a third variable. If selection of the sample restricts the representation of an important interacting variable, whether genetic or environmental, associations can be distorted in unpredictable ways, and hence conclusions can be misleading. An example has recently been presented of a genetic polymorphism that affects the risk of depression in three ways, depending upon the level of education of the individual. Among the highly educated, the gene reduced depression; among the less educated, it increased depression; and among people in the mid-range of education, it had no effect on depression. The association of this gene with depression in a non-probability sample would depend entirely on the level of education of the sample selected for the study.

It will be argued that most research studies in biomedicine are not based on probability samples. This is true, and therefore very few studies are definitive or conclusive by themselves. The findings of most studies can be assessed only after consideration of many other studies of the same topic.

But the NCS must be held to a higher scientific and operational standard. Congress specifically authorized the NCS; endowed the NCS with significant funding; and mandated that the NCS achieve clearly articulated goals using a scientifically sound study design. That investment was explicitly intended to create a large, longitudinal, and comprehensive study whose results would be highly valid and generalizable to all children across our nation. If Congress intended the NCS to be just another cohort study, why would Congress have allocated special funds for it?

The architects of the NCS must not compromise the study's scientific validity and generalizability by abandoning the one sampling methodology that will provide the most scientifically valid results and set the best foundation for clinical and public health policies that apply to all of our children.

Attachment B is removed. UnderJournal Review.