THE NATIONAL CHILDREN'S STUDY
Potential Sampling Strategies: Main Study

April 16, 2012
DRAFT 3.0

This document is intended to structure discussions about the future NCS Main Study sampling design. The proposed sampling strategies presented are examples and not an inclusive list of all options. Alternative proposals for sampling strategies are encouraged during this period of design evaluation.

Summary

The National Children's Study (NCS or the Study) is a longitudinal observational birth cohort study with a planned 21 years of follow-up for each enrolled child. Initial recruitment began with a two-stage geographic-based probability sample and door-to-door household recruitment. Analyses of the ongoing Vanguard, or pilot phase, of the Study demonstrated that enrollment rates of pregnant women were lower than expected from recruitment assumptions. Therefore, continuation of the initial proposed sampling and recruitment strategy would not be affordable or sustainable and carry some scientific compromise, particularly with regard to the projected duration of recruitment of the cohort. Subsequently, the National Children's Study Program Office initiated additional recruitment strategies using a provider-based model and direct outreach to the public. Based on interim data analysis, a provider-based recruitment model has several advantages. Consequently, NCS is currently exploring options for using prenatal care providers as the basis for future recruitment activities. In this document several alternative sampling strategies based on prenatal care providers are presented for discussion and evaluation. The final section of the document summarizes the administration and operation of the National Children's Study.

Research Objective of the National Children's Study

The National Children's Study was congressionally mandated by the Children's Health Act of 2000 to:

> *"(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;*
> *(2) investigate basic mechanisms of developmental disorders and environmental factors, both risk and protective, that influence health and developmental processes."*
> The Study is required to:
> *"(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 National Children's Study is an integrated system of activities. The current major components are the NCS Vanguard Study, the NCS Main Study, the NCS Substudies, and Formative Research. The NCS Vanguard Study is the pilot phase for methods and runs for 21 years. It was started in 2009 and expanded in 2010 and 2011 with additional locations. The goals of the Vanguard Study are to assess the feasibility, acceptability, and cost of Study recruitment, logistics, and operations. The NCS Main Study is the exposure response phase and runs for 21 years. The planned start for the NCS Main Study is in 2013 with recruitment currently targeted to begin in 2014 and continue for about 3 years. The NCS Substudies are studies within studies. Formative Research is short-term limited studies, often for methods development, to support and inform the Vanguard and Main Studies.

OBJECTIVE

The primary objective of the National Children's Study is to collect information on and investigate the factors that determine children's health and development. These factors include genetic context and environmental exposures with a broad definition of environment. The National Children's Study is not a study in a conventional sense. It will function as a high quality data collection platform for researchers to access and analyze data, biospecimens, and environmental samples.

POPULATION

Prenatal exposures are of interest and significance so the National Children's Study, as the law proposes, will enroll pregnant women with a goal to enroll women as early in pregnancy as feasible. To enrich for a population of pregnant women who are early in pregnancy, the National Children's Study will enroll pregnant women and women who could become pregnant who reside in the United States at the time of enrollment.

SAMPLE FRAME

The sampling frame of the National Children's Study should incorporate a population with diverse racial, ethnic, socioeconomic, educational, cultural, and immigration statuses as well as a geographic gradient of exposures of various types and a range of access to health care services.

EXPOSURES AND OUTCOMES

Examples of outcomes of interest are premature birth, birth defects, growth and development, interpersonal relationships and bonding, inflammatory processes including allergies, asthma, and infections, genetic and epigenetic status, epilepsy and other neurologic disorders, cardiovascular screening and function, childhood cancer, multidisciplinary multidimensional aspects of sensory input, autism and other neurodevelopmental disorders, learning and behavior, and precursors and early signs of chronic diseases such as obesity, asthma, hypertension, and diabetes.

Examples of exposures of interest are exposures to industrial chemicals and byproducts in the air, water, soil, and commercial products, natural products in the air, water, soil, and commercial products, pharmaceuticals used for therapy and in the environment, radiation, and effects of proximity to manufacturing, transportation, and processing facilities. Additional exposures of interest are living with animals, insects, and plants, media and electronic device exposure, noise, access to routine and specialty health care, learning opportunities that are structured and unstructured, diet and exercise, and family and social network dynamics in a cultural and geographic context.

The priority of the mechanisms to be investigated will be informed by the public health impact (based on severity as well as prevalence) on the overall population of children and by scientific opportunity. Examples of conditions of interest are summarized in Table 1.

Table 1: The prevalence estimates per 100,000 for selected childhood illnesses.

Condition	Estimated Prevalence per 100,000
Obese	17,000
Overweight	30,000
Premature Birth	12,500
Learning Disorders	5,000
Asthma	5,000
Birth Defects (aggregate)	3,000
Autism Spectrum Disorders (aggregate)	1,100
Schizophrenia	1,100
Congenital Heart Disease	800
Epilepsy	470
Childhood Cancers	320
Down Syndrome	125
Fragile X Syndrome	50

**Note that the legal federal threshold for a rare disease is a prevalence of about 64 per 100,000.*

The prevalence of many of the conditions in Table 1 is possibly underestimated due to disparities in health and access to health care limiting diagnosis. In addition, the prevalence presented represents only the level of each disease spectrum where formal evaluation and intervention are required. Children with less severe symptoms or with restricted access to health care may have health impacts from these conditions but not rise to a level captured by formal health care records.

SAMPLE SIZE

Based on a two-sided binomial sample size calculation for a condition of interest such as childhood malignancy with a national prevalence of 0.5 percent, an alpha of 0.05, and power of 0.90, our ability to detect small but potentially biologically meaningful changes in the prevalence of the outcome of interest such as an environmental exposure leading to an increase in the malignancy rate of 5 percent or more in the Study population requires a sample size of roughly 100,000 children at the time of analysis. The sample size estimate does not account for attrition over the life of the Study. In addition, in order to assess early exposures *in utero*, meaning less than 8 weeks gestation, we target 15 percent of pregnant women in the Study be from a preconception cohort.

A range of medically important outcomes will be used here to illustrate the ability of the National Children's Study to test exposure-outcome associations involved in a series of hypothetical hypotheses with power of 80 percent. These outcomes exhibit the range of prevalence that NCS outcomes are likely to have. While some outcomes are common, most are uncommon and some are rare. Many of these outcomes are relevant for a single hypothesis, but some are relevant for more than one. For example, several hypotheses address different possible predictors of some childhood conditions, including environmental factors, exposure to bacteria and microbial products, maternal stress during pregnancy, and diet. For each outcome, a set of different exposures is considered. In each case power has been calculated for exposure prevalence of 1.0 percent, 2.5 percent, 5 percent, 25 percent, and 50 percent (this range is based on hypotheses developed for the National Children's Study).

Using cerebral palsy (CP) as an example, the results on power displayed in Table 2 can be interpreted as follows: Since CP has a prevalence of about 0.2 percent in the general population, that is the rate to be

expected in the National Children's Study. Table 1 gives the odds ratio (OR) that can be detected with 80 percent power for exposures or risk factors with a 5 percent significance level and a prevalence ranging from 1 percent to 50 percent. For rare exposures, only those that have a dramatic impact on the occurrence of cerebral palsy (OR $\geq$ 5.0) can be reliably detected in the National Children's Study. However, for more common exposures, such as those with 5 percent prevalence or greater, factors with more modest effects (OR $\geq$ 2.6) can be detected with 80 percent power.

Two simplifications were made in these power calculations. First, the analyses consider only the simple bivariate relationships between the exposures and outcomes without addressing the need to control for confounders. The inclusion of confounders likely results in a reduction in the power for detecting the effects of exposures, but often the reduction will be modest. Second, all outcomes and exposures are assumed to be dichotomous variables. This assumption is again made to simplify the table. In fact, most of the NCS outcomes and exposures will be continuous variables. As a result, the power estimates in the tables are likely to be conservative since dose-response analyses with continuous outcome or exposure variables would likely lead to greater power.

Table 2 displays the magnitude of the minimum odds ratios that can be detected with 80 percent power for the selected outcomes and the range of exposures for analyses. The sample sizes for Table 2 assumed to be the full sample for which data are available. As noted above, the sample available is reduced through attrition and, for some outcomes, by availability of special data required for analysis. As Table 2 shows, the magnitude of the detectable odds ratio depends on the prevalence of both the outcome and the exposure. For a given outcome, the closer the prevalence of the exposed group is to 50 percent, the smaller the detectable odds ratio and the greater the power. Similarly, in general, the closer the prevalence of the outcome is to 50 percent, the smaller the detectable odds ratio; the detectable odds ratios are small when the exposure prevalence is reasonably high. All the ratios are less than two when the exposure prevalence is between 25 percent and 50 percent.

Table 2: Detectable odds ratio when analyzing a total NCS sample of 100,000 children.

Outcome	Age	Prevalence of outcome (%)	Prevalence of exposure				
			1%	3%	5%	25%	50%
Infant mortality*	1	0.7	6.01	3.87	2.95	1.97	1.94
Type I diabetes	18	0.2	5.71	3.72	2.86	1.93	1.89
Musculoskeletal defects	1	0.2	5.00	3.33	2.60	1.80	1.75
Cerebral palsy	1	0.2	5.00	3.33	2.60	1.80	1.75
Nervous system defects	1	0.3	4.09	2.82	2.25	1.62	1.58
Metabolic syndrome	18	0.4	4.03	2.78	2.23	1.61	1.56
Autism spectrum disorder	4	1	2.75	2.05	1.73	1.36	1.32
Heart defects	1	0.6	3.03	2.21	1.84	1.42	1.38
Type II diabetes	18	1	2.75	2.05	1.73	1.36	1.32
Major birth defects	1	3.5	1.76	1.47	1.33	1.16	1.14
Adolescent aggressive behavior	18	4	1.82	1.50	1.35	1.17	1.15
Chronic physical aggression (CPA)	10	4	1.76	1.47	1.33	1.16	1.14
IQ score less than 75	18	5	1.73	1.45	1.31	1.16	1.14
Asthma	4	7.5	1.53	1.33	1.23	1.11	1.10
Neurocognitive development	12	8	1.55	1.34	1.24	1.12	1.10
Depression	18	8.3	1.57	1.35	1.25	1.12	1.11
Asthma	7	8.5	1.51	1.32	1.22	1.11	1.10
Neurodevelopmental disabilities	18	10	1.52	1.32	1.22	1.11	1.10
Preterm birth < 37 weeks	0	12	1.41	1.26	1.18	1.09	1.08
Asthma	18	12.5	1.47	1.29	1.20	1.10	1.09
Adverse pregnancy outcomes	0	15	1.38	1.23	1.16	1.08	1.07
Developmental disabilities	18	17	1.41	1.25	1.18	1.09	1.08
Obesity	12	17.1	1.39	1.24	1.17	1.08	1.07
IQ score less than 100	18	50	1.32	1.20	1.14	1.07	1.06

* The exposure for this hypothesis is a community rather than an individual level characteristic.

To illustrate the increase in the magnitudes of detectable odds ratios for subgroup analyses, Table 3 presents results comparable to those in Table 2, but with the sample size reduced to a 20 percent subgroup. The results in this table could be applied to case-control studies, or other analyses based on subsets of the overall NCS sample. It is assumed that the geographic distribution of the subgroup is proportionate to the general population, which would generally be true in case-control studies and other subset analyses. The detectable odds ratio remains below 2 when the outcome prevalence is 3.5 percent or higher and the exposure prevalence is 5 percent or more, but for rarer outcomes and exposures, it exceeds 2. Many subgroups of interest will comprise less than 20 percent of the population and will thus have larger detectable odds ratios.

Table 3: Detectable odds ratio when analyzing a 20 percent subsample.

Outcome	Age	Prevalence of outcome (%)	Prevalence of exposure				
			1%	3%	5%	25%	50%
Infant mortality*	1	0.7	17.91	10.13	7.08	4.32	5.35
Type I diabetes	18	0.2	16.13	9.42	6.68	4.12	4.99
Musculoskeletal defects	1	0.2	13.44	7.96	5.69	3.50	3.93
Cerebral palsy	1	0.2	13.44	7.96	5.69	3.50	3.93
Nervous system defects	1	0.3	10.23	6.19	4.51	2.81	2.93
Metabolic syndrome	18	0.4	10.04	6.07	4.42	2.76	2.86
Autism spectrum disorder	4	1	5.87	3.78	2.89	1.94	1.90
Heart defects	1	0.6	6.73	4.26	3.22	2.11	2.08
Type II diabetes	18	1	5.87	3.78	2.89	1.94	1.90
Major birth defects	1	3.5	2.96	2.15	1.79	1.39	1.35
Adolescent aggressive behavior	18	4	3.13	2.24	1.85	1.42	1.38
Chronic physical aggression (CPA)	10	4	2.97	2.16	1.80	1.39	1.35
IQ score less than 75	18	5	2.89	2.10	1.76	1.37	1.33
Asthma	4	7.5	2.35	1.80	1.55	1.27	1.24
Neurocognitive development	12	8	2.40	1.83	1.57	1.28	1.25
Depression	18	8.3	2.46	1.85	1.59	1.29	1.25
Asthma	7	8.5	2.30	1.77	1.53	1.26	1.23
Neurodevelopmental disabilities	18	10	2.33	1.78	1.54	1.26	1.23
Asthma	18	12.5	2.21	1.71	1.49	1.24	1.21
Preterm birth < 37 weeks	0	12	2.04	1.62	1.43	1.21	1.18
Adverse pregnancy outcomes	0	15	1.95	1.56	1.39	1.19	1.17
Developmental disabilities	18	17	2.07	1.62	1.43	1.21	1.18
Obesity	12	17.1	2.00	1.59	1.40	1.20	1.17
IQ score less than 100	18	50	1.90	1.49	1.33	1.15	1.13

* The exposure for this hypothesis is a community rather than an individual level characteristic.

CORE ASSESSMENTS

A primary outcome of "children's health," as framed in the Children's Health Act of 2000, is complex and poses several challenges to measure. Without a single unambiguous definition of health, we propose for study design estimation purposes to prioritize a small number of outcomes; specifically, to use the parameters of linear growth rate and body mass as a surrogate for general health; a metabolic screen of serum total protein, blood urea nitrogen, cholesterol, iron, and calcium for nutrition and dietary exposure; frequency and duration of health system encounters for respiratory illness for pulmonary health, and timing of standard neurodevelopmental landmarks for any deviation from adjusted trajectory for cognitive and social development.

For physical environmental exposure we propose analysis of heavy metals, pesticide residues, and semi-volatile organic compounds in household dust, blood, and urine as general surrogates for more specific exposures.

GENERAL DESIGN CONSIDERATIONS

Both the Vanguard Study and the Main Study emphasize data collection early in pregnancy and early in child development because the largest knowledge gaps and perhaps the most critical events occur during those time periods. Consequently, 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.

Multiple modalities for data collection are under evaluation, with the current plan based on a core questionnaire model administered at every childhood visit plus supplemental modules based on events and conditions such as age, developmental stage, and other triggers such as specific exposures or hospitalizations. While the core questionnaire is intended for all participants, supplemental modules may be administered on a missing by design basis to leverage the large study population and extend resources. In addition, the visit schedule is flexible, in that children will not have assessments precisely at a given age but, instead, within a window of several weeks around a particular age to improve compliance and to capture data across a range of specific ages.

In addition to questionnaires, other modalities for data capture such as sounds, images, geographic movements, and mapping of social interactions and networks will be used. The core questionnaire and other questionnaires are essential, however, for calibrating the data from other modalities and for linking NCS data to other data sources.

Vanguard Study Experience

The selection of a sampling frame for the National Children's Study has historically reflected a wide range of options. In 2006, NCS leadership decided on a geographically based probability sample. The sampling frame for the Vanguard Study was initially based on a random selection of about 100 of the approximately 3,000 counties in the United States, using counties as the Primary Sampling Units (PSUs). The counties were then divided into Secondary Sampling Units, or segments, that were normalized to have about 250 live births per year. Some sparsely populated areas required clusters of counties. Recruitment of participants was restricted to those with a residence within the Secondary Sampling Units. Over time, subsequent adjustments were made to the sampling process to address additional criteria, such as increasing demographic and geographic diversity.

A household-based approach for the Vanguard Study began field activities early in 2009 with 7 locations. Later in 2009, NCS leadership recognized that the field data did not support the predictions on which the original recruitment model was based. Consequently the NCS Program Office developed some alternate recruitment strategies using the same sampling frame. The new recruitment strategies were:

- Provider-based, where participants learn about the Study and are referred through health care providers (broad definition of provider, including pediatricians, obstetricians, public health nurses, midwives, etc.) and,
- Direct outreach, where participants learn directly about the Study through media and community outreach and are invited to enroll.

In 2010, the Vanguard Study expanded to an additional 30 locations using 10 locations for the provider-based and 10 locations for the direct outreach strategies, plus a variant on the household-based approach, Enhanced Household, in 10 more locations. This phase of the Vanguard Study was termed the Alternate Recruitment Substudy.

Women who were not pregnant in the screening phase of each recruitment arm were followed to determine whether they became pregnant and eligible to enroll in the Study. In an effort to improve recruitment efficiency from the preconception cohort, probabilities of becoming pregnant were assigned to each woman based on their self-reported intention to become pregnant. During the initial household enumeration arm, an estimated 15 percent of women transitioned from the preconception cohort to a pregnant state within a nine-month period at the end of the Initial Vanguard Study. When followed through the Alternate Recruitment Phase, for a total period of about 25 months, 39 percent of non-pregnant women enrollees became pregnant (Figure 1). Despite the use of a probability algorithm, there appeared to be no enrichment above expected baseline of conversion from preconception to pregnant cohort.

Note that in all tables and figures, values are rounded to the nearest 50 per NCS policy for public display of aggregate data. Consequently, due to the rounding, calculations shown in the tables may not be internally consistent.

Table 4: Recruitment status of Vanguard Study participants, as of 2/23/2012.

	Initial Household (2009 cohort)	Alternate Recruitment (2010 cohort)	All Vanguard to Date
Locations	7	**30**	**37**
Recruitment duration, months	18 active+ 16 follow up	14	
A. Women eligible for contact	35000	49350	**84350**
B. Contacted for pregnancy screen (% of eligible)	34350 (98%)	44000 (89%)	**78350 (93%)**
C. Completed screen (% of contacted)	30900 (90%)	37750 (86%)	**68650 (88%)**
D. Pregnant or trying (% of screened)	2900 (9.4%)	6850 (18%)	**9750 (14%)**
E. Enrolled (% of pregnant or trying)	1900 (66%)	5050 (74%)	**6950 (71%)**
F. Babies enrolled	1100	1750	**2850**

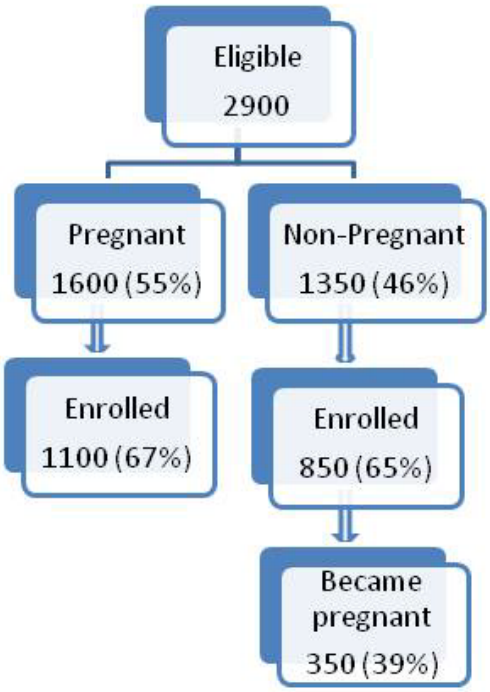

Figure 1: The recruitment experience in the initial household enumeration cohort of 2009, separating the pregnant eligible cohort from the pre-conception cohort. Based on data acquired prior to 2/23/2012, with a total recruitment period of approximately 34 months and a 25-month enrollment period for non-pregnant women.

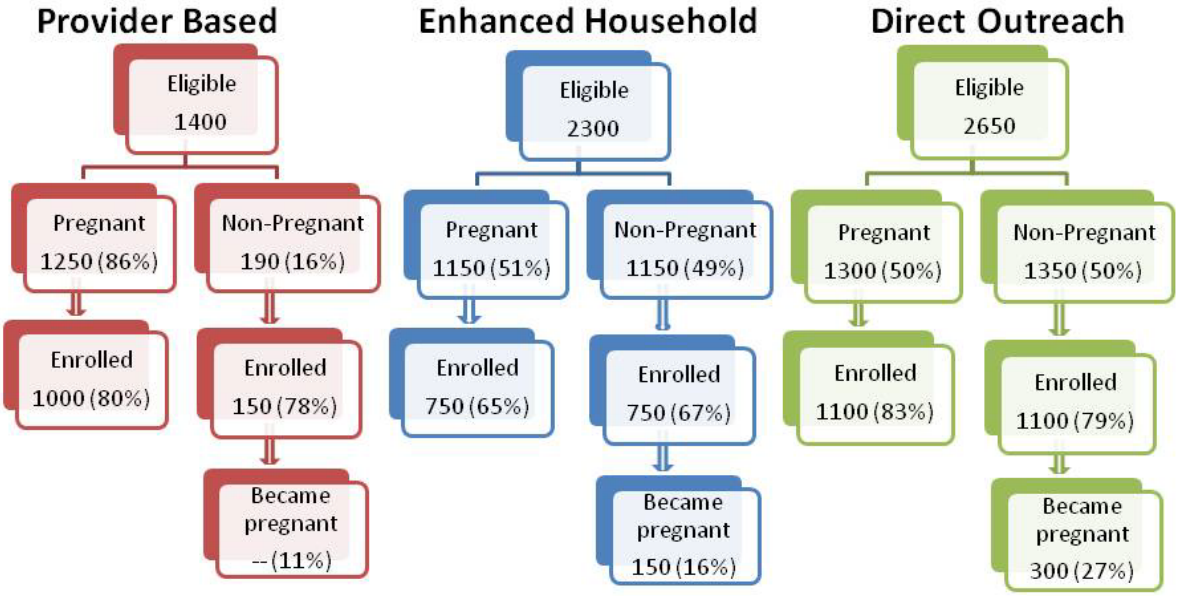

Figure 2: Comparisons of the recruitment strategies, as of 2/23/2012, with a recruitment period of about one year. Some numbers cannot be displayed due to rounding rules.

Table 5: Screening efficiency of the recruitment strategies, as of 2/23/2012.

	Provider Based	Enhanced Household	Direct Outreach
Number of locations x Weeks in field	403	479	471
Mean number of women enrolled per week	3.1	3.2	3.9
Mean number of women screened per woman enrolled	2.6	13.7	8.7

Table 6: The pregnancy related characteristics of women enrolled in the NCS, as of 2/23/2012.

	Provider Based	Enhanced Household	Direct Outreach
Women enrolled	1250	1600	2200
Proportion of enrolled who were pregnant vs. trying at enrollment	87%/ 13%	51%/ 49%	51%/ 49%
Proportion of pregnant enrollees whose pregnancy was <14 weeks of gestational age at enrollment	24%	24%	22%
Proportion of non-pregnant enrollees who became pregnant since enrollment (1 to 12 months of follow-up)	11%	16%	27%

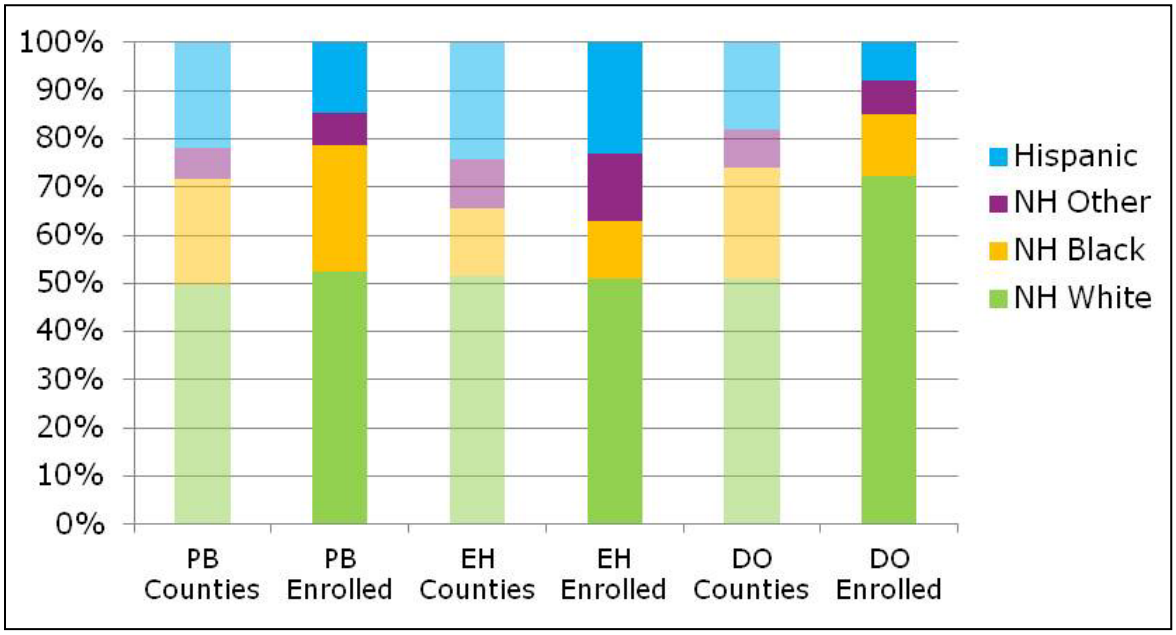

Figure 3: The percent distribution of enrolled women by race/ethnicity and by recruitment strategy, as of 2/23/2012. The Average race/ethnicity distribution comes from the data on 10 counties from 2009 Natality Data.

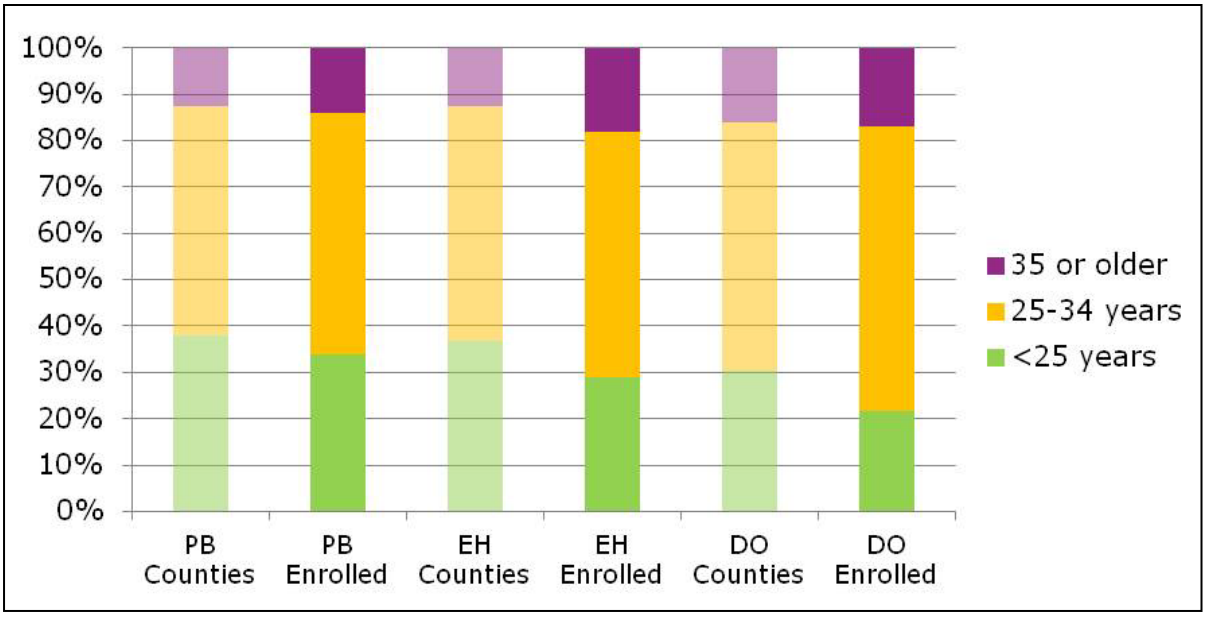

Figure 4: The distribution of enrolled women by age and by recruitment strategy, as of 2/23/2012. The average age distribution comes from the data on 10 counties from 2009 Natality Data.

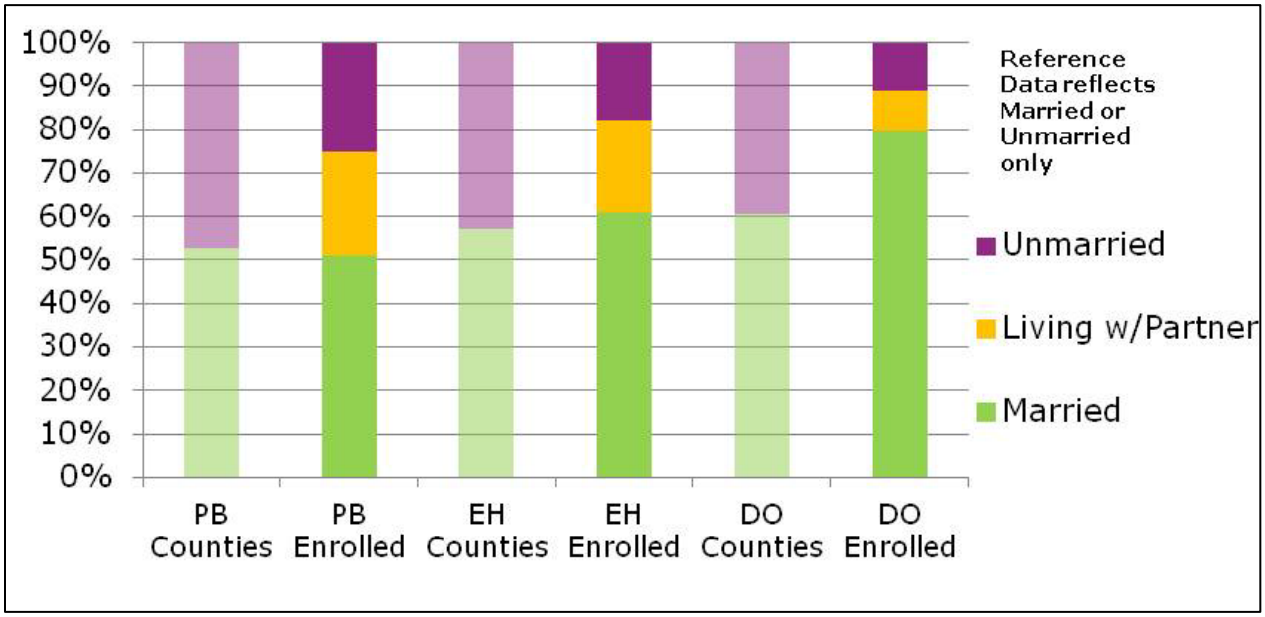

Figure 5: The distribution of enrolled women by marital status and by recruitment strategy, as of 2/23/2012. The average age distribution comes from the data on 10 counties from 2009 Natality Data.

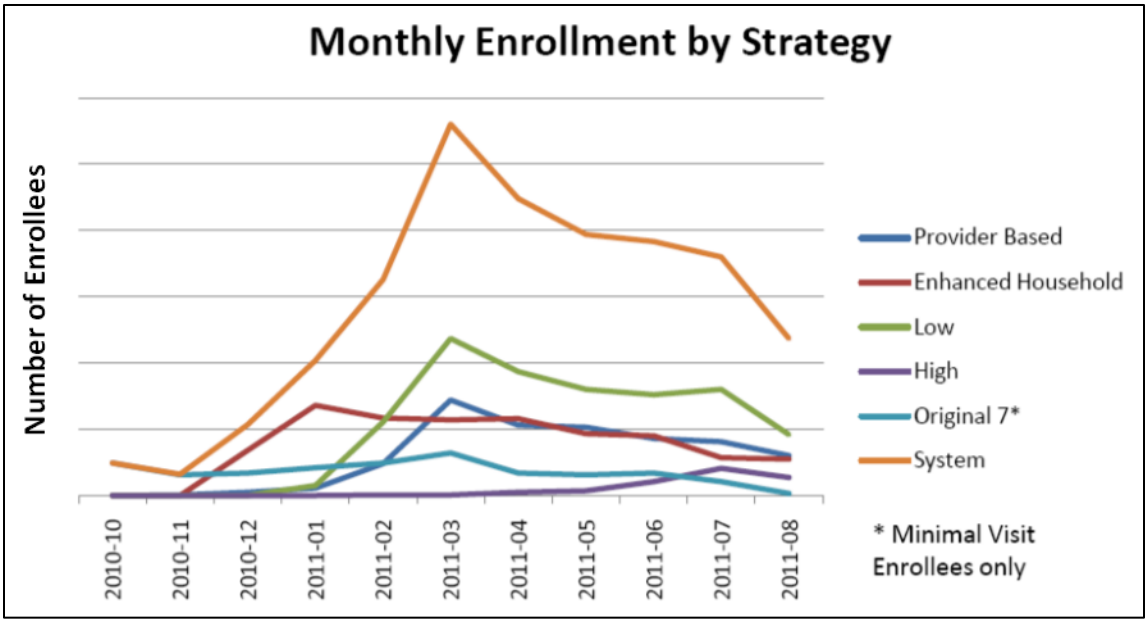

Figure 6: Enrollment kinetics in the NCS by strategy.

The conclusions from these preliminary results indicate that the efficiency of enrollment differed among each recruitment strategy (Table 5). While the baseline demographics for each recruitment strategy locations were generally similar, the demographics of women enrolled for each recruitment strategy differed by varying degrees from baseline and from each other (Figures 1–5). Initial analysis of the kinetics of rate of recruitment in the Alternate Recruitment Substudy showed a peak at about 20 weeks after initiation, followed by a decrease that had a marked slowdown by week 32. Active recruitment for the three arms formally ended in November 2011.

VANGUARD STUDY NEXT STEPS

Based on the preliminary observations of the efficiencies of provider-based recruitment, the next step in the Vanguard Study is to adjust the sampling frame, such that the Secondary Sampling Units are based on health care provider locations, in an approach called Provider-based Sampling. The rationale is that the geographic-based Secondary Sampling Units of the Provider-based Recruitment strategy were so limiting that many pregnant women that visited a selected health care provider could not enroll in the Study because their home address was outside one of the Secondary Sampling Units. In order for the National Children's Study to more fully assess the feasibility, acceptability, and cost of enrollment, we chose to eliminate the geographic restriction of the Secondary Sampling Units. Thus all women who reside in a Primary Sampling Unit, which is usually a county, and receive health care at a selected provider, are eligible for the Study. This phase of the Vanguard Study is beginning activity in the spring of 2012.

VANGUARD STUDY INTERIM ASSESSMENT

The Vanguard Study experience demonstrated that the geographic sampling frame combined with household enumeration for recruitment was neither feasible nor sustainable with the resources of the National Children's Study. Projecting the data from the initial experience, enrollment would take about 9 years. For a variety of scientific and fiscal reasons, we now propose to use provider-based sampling and recruitment for future NCS activities. The focus on provider-based recruitment is due to the relative

efficiency of recruitment, the potential for the trust relationship between patient and provider to increase recruitment and retention success, and the potential to leverage resources such as electronic health records to enrich the data collection effort. Also, while the National Children's Study will still engage in primary data collection, this strategy will allow more efficient use of resources. Using a provider-based recruitment strategy also provides an opportunity to reexamine the sampling frame for the Main Study to ensure that the Study objectives are met.

Options for Main Study Sampling Frame

Desired characteristics in a sampling frame are the ability to address the Study objectives, enroll the target population, collect exposures of interest, and monitor for outcomes of interest, while utilizing the resources available and anticipating the future. While concordance with the target population is always a measure of a successful frame, we need to acknowledge that for the National Children's Study, sampling is always indirect, as the children in the future Main Study have not been born.

Table 7: The Main Study objectives as stated in the Children's Health Act of 2000, along with the design implications of each objective.

Study Objectives	Sample and Study Design Implications
Evaluate the effects of both chronic and intermittent exposures on child health and human development	High retention of children is important to gather chronic and intermittent exposures
Investigate basic mechanisms of developmental disorders and environmental factors	Broad scope of data collection supplemented and informed by formative research program
Gather data from diverse populations of children including prenatal exposures	Need to recruit diverse population groups and capture prenatal exposures
Consider health disparities among children	Ensure sampling of disadvantaged population groups (in terms of exposure, education, socioeconomic status, etc.)

Several sampling strategies are proposed in this document. Other designs are possible and during this period of consultation, the NCS Program Office actively seeks proposals. The designs in this document are proposed for further discussion regarding their potential strengths and limitations.

The nomenclature used in this document may not conform to all expectations but is summarized here with the intent to maintain consistency across different discussions.

Two factors that compromise the interpretation of any data set are uncertainty and bias. The more control the researcher has over the selection method for the target population and the selection method of individuals from within the target population, the greater the opportunity to control uncertainty and bias.

One approach to minimize bias is to select participants for a study from a pool or target population where everyone has the same chance of participation. From the target population a list of potential participants is generated that becomes the source from which potential participants are selected. The list preparation process is termed enumeration. An accepted rigorous approach is to randomly select

potential participants from the list. The study population is defined as the randomly selected participants who agree to enroll in the study. A comparison of the known characteristics of the study population with the same characteristics of the target population should be consistent if there is no selection bias.

If the target population is relatively limited, such as all members of a club or all people receiving health care at a particular clinic, then an accurate list is reasonably simple to generate. However, any analysis and conclusions from the study will generally only be considered valid for the target population and not other populations. For example, if the influenza vaccination rate among a randomly selected group of people from a particular clinic is 85 percent, then estimating that the influenza vaccination rate for all people who receive care at that clinic is 85 percent is reasonable and credible. The estimate will not, however, be considered applicable to people who receive care at other clinics.

If the population of interest is large and dispersed, such as the entire United States, then the situation is more complex. The general approach to select potential participants for a study from a large population such as the entire country involves two stages. First is to select a random sample of the entire population using a proxy such as known geographic locations with boundaries and then prepare an enumeration list of individuals from that sample. Subsequently, potential participants are selected from the enumeration list. The two-stage process of first randomly selecting a sample from the population of the entire country using a proxy and then randomly selecting individuals from within that sample will be considered a probability sample for this document.

To be complete, while the random selection of potential participants from a list may minimize bias, other factors prior to the time of analysis can introduce bias such as errors in the enumeration process, the clarity and comprehension of messages and instructions to potential participants, perceptions of trust, willingness to participate, participant retention, and drop outs. If the target population is accurately enumerated, then sources of bias associated with selection and retention, such as drop outs and non-enrollment, can be adjusted for.

If the target population is restricted beyond the researcher's control, then any sample from that target population is considered for this document a convenience sample. Examples include people served by a particular clinic, members of an organization, or families employed by a national company. Even if the individuals within the target population are selected from a list using probability based techniques, we will refer to the overall approach as a convenience sample.

The term sample frame in this document refers to the listing that should include all those in the target population to be sampled and excludes all those who are not in the target population. The target population to be sampled for the National Children's Study is age-eligible women who could be pregnant and all women who are pregnant who reside in the United States at the time of enrollment.

In this document some models use more than one sample frame. Such models will be referred to as having multiple sampling frames. An example might be a larger population selected using one method and a smaller population selected using another method. The relationship between the two sampling frames would be an important topic of discussion.

Finally, some models use both more than one sampling frame and in addition one sampling frame may be a probability sample and the other may be a convenience sample. In this document the term hybrid model describes a combination of a probability sample frame with a convenience sample frame.

Differential retention rates can be accounted for in a probability based design because the target population can be enumerated and everyone in it had a known probability of selection. This can be contrasted with a convenience sample where the target population may be defined, but careful attention has to be paid to the population excluded to describe the bias inherent in the design. In the convenience sample, differential retention in addition to the limitations associated with selection makes generalizability to a larger population difficult and imprecise.

Some of the potential contrasts between a national probability based design and a convenience sample of prenatal care providers in the context of the goals of the National Children's Study are presented here in Table 8. **Note that even using a provider-based sampling frame or recruitment strategy, the National Children's Study intends to collect primary data directly.**

Table 8: Contrasts between a national probability sample and convenience sample of prenatal care providers for the National Children's Study.

Characteristic	National Probability Sample with NCS primary data collection	Convenience Sample of Prenatal Care Providers with NCS primary data collection augmented with provider health care records
Generalize biological associations	Yes	Yes, with limitations
Generalize psychosocial and behavioral associations	Yes in principle	Limited
Generalize economic associations	Yes in principle	Limited
Geographic dispersion	Yes	Yes
Detailed participant baseline description	Limited and resource intensive	Yes in principle
Operational Feasibility	Limited and resource intensive	Yes
Collect biospecimens	Yes	Yes
Collect environmental samples	Yes	Yes
Inclusion of women outside the health care system	Yes in principle if done by household address	No, but can be supplemented by additional targeted recruitment
Use of woman as sampling unit	Yes	Yes
Use of pregnancy as sampling unit	Yes	Yes
Separation of data collection from health care delivery activities	Yes	Partial, some overlap expected
Use of participant's home as a collection site	Yes	Yes
Allow replication of sampling frame for subsequent studies	Yes in principle	Limited
Linkage to Electronic Health Records	Limited and resource intensive	Yes in principle

Linkage to external databases to describe surrounding community	Yes in principle	Yes in principle
Ability to perform multiple analyses and case control studies	Yes in principle	Yes in principle
Archive data and specimens accessible to researchers	Yes	Yes

The designs presented here for consideration fall into a matrix of sampling designs (Figure 7) that range from convenience to probability, with the potential for single or multiple frames and hybrid designs. For example, in a dual frame design, one frame could be a sample of prenatal care providers, and the other frame could be targeted to capture populations not represented or underrepresented in the provider-based sample based on demographic characteristics or exposures. In another example, a carefully designed convenience sample could have a probability sample nested within it. The rationale for such a design would be to leverage estimates of rarer exposure response mechanisms with a larger sample size, while maintaining the ability to determine prevalence estimates in the probability frame. In addition it may be administratively or logistically advantageous for a group or network of providers to enroll participants in a hybrid design with fewer restrictions than a pure probability design may impose.

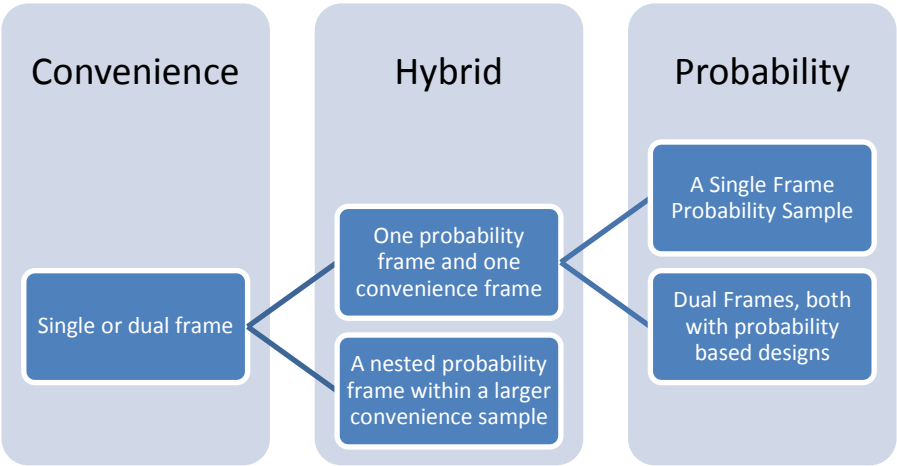

Figure 7: A proposed matrix of suggested design options for the Main Study of the National Children's Study.

Considerations important to design evaluation are the enumeration of the frame population, characterizing the source population, the distribution of providers among the type of specialty, prenatal care provider cooperation, and exposure distribution. Additionally, the ability to combine a dual frame design into a meaningful larger analysis file with appropriate weights is an important factor.

Proposed Sampling Strategies

The hypothetical examples below are offered as models for discussion without prioritization. Each of the models has some potential advantages and disadvantages listed with an expectation that each list will be amended during discussion. Suggestions for additional and different models are encouraged.

PROBABILITY DESIGNS

1. **Geographic based probability sample of prenatal care providers.**
   The first stage would be a probability based geographic sample. The second stage would be prenatal care providers practicing within the probability based geographic sample, which could be enumerated via birth records or a provider survey. From this frame select a random sample of providers, with probabilities proportionate to the size of the prenatal care practice, and engage them for participation. Participants would be selected from the participating providers.

   Advantages:
   - *Where birth records are available the list frame for the larger frame could be replicated.*
   - *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
   - *Prenatal care providers may represent an efficient way to enroll pregnant women.*
   - *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
   - *The base population can be characterized, including health status, through medical records.*
   - *The ability to produce prevalence estimates and the exposure-outcome relationship can be generalizable to U.S. children.*

   Disadvantages:
   - *Engaging providers may be resource intensive.*
   - *There may be variable rates of provider participation.*
   - *This sampling frame requires access to birth records, which could exclude some areas of the country.*
   - *If a list frame needs to be developed via a provider-based survey, the frame may have bias due to non-response, and would be more difficult to replicate.*
   - *This sampling frame is missing individuals with limited access to health care.*

2. **Geographic based probability sample of prenatal care providers supplemented by a second probability sample from an administrative list frame.**
   The first stage would be a probability based geographic sample. The second stage would be prenatal care providers practicing within the probability based geographic sample. From this frame select a random sample of providers, with probabilities proportionate to the size of the prenatal care practice, and engage them for participation. Participants would be selected from the participating providers.

   Supplement this frame with another probability sample of eligible participants from the geographic catchment areas of the selected prenatal care providers using an existing administrative list frame. Potential supplemental list frames could be from the Special Supplemental Nutrition Program for Women, Infants, and Children program (WIC). WIC enrolls

51 percent of pregnant women in their first trimester, and has a 91 percent participant contact rate. Medicaid finances 40 percent of births in the U.S., which represents a large list frame to pull from.

Advantages:

- *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
- *Prenatal care providers may represent an efficient way to enroll pregnant women.*
- *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
- *The base population can be characterized, including health status, through medical records.*
- *The ability to produce prevalence estimates and the exposure-outcome relationship can be generalizable to U.S. children.*
- *Both the WIC and Medicaid list frames could be easily replicated.*

Disadvantages:

- *Defining the catchment areas of the prenatal care providers will likely be difficult and somewhat subjective.*
- *If both WIC and Medicaid frames are used there is likely overlap.*
- *Cooperation of prenatal care providers may be variable.*
- *Engagement of prenatal care providers could be resource intensive.*
- *If a list frame needs to be developed via a provider-based survey, the frame may have bias due to non-response, and would be more difficult to replicate.*
- *Constructing larger analytic files from both frames would require careful sample weights and analytic plans.*

CONVENIENCE DESIGNS

3. **Convenience sample of prenatal health care providers.**
   Construct a pool of prenatal health care providers who are willing to participate. Identify eligible women from the selected providers, with approximately 15 percent in a pre-conception cohort.

Advantages:

- *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
- *Prenatal care providers may represent an efficient way to enroll pregnant women.*
- *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
- *The base population can be characterized, including health status, through medical records, where available.*
- *Recruitment target could be achieved faster than with other models.*

Disadvantages:

- *The environmental exposures of the patients within the catchment area of the providers may be more homogenous than the exposures collected via a probability sample.*
- *There is a probability of drawing inaccurate conclusions due to missing an important range of exposures.*

- *The potential homogeneity in some factors may reduce the power for some estimates of association, and may also limit the ability to assess dose-response relationships.*
- *Effect modification is a concern, because if the full range of the modifier is not collected it may not be caught, and an erroneous conclusion about an association can be made.*
- *A sample from prenatal care providers could be missing those with unattended births or no prenatal care, those that are uninsured or not in the plans and infertile women.*
- *Cooperation of prenatal care providers may be variable.*
- *Engagement of prenatal care providers could be resource intensive.*
- *The preconception cohort recruited from prenatal care providers may be a biased sample compared to age-eligible women in the sampling unit.*

4. Convenience sample of prenatal health care providers supplemented by a second convenience sample.

Construct a pool of prenatal health care providers who are willing to participate. Identify eligible women from the selected providers, with approximately 15 percent in a pre-conception cohort. Supplement with a second convenience sample to capture populations not represented or underrepresented in the provider-based sample based on demographic characteristics or exposures.

Advantages:

- *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
- *Prenatal care providers may represent an efficient way to enroll pregnant women.*
- *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
- *The base population can be characterized, including health status, through medical records.*
- *The supplemental sample will address gaps in populations and areas not represented or underrepresented in the prenatal care provider sample.*
- *Recruitment target could be achieved faster than with other models.*
- *Provides the opportunity to ensure the inclusion of demographics or exposures of interest in the cohort.*

Disadvantages:

- *The environmental exposures of the patients within the catchment area of the providers may be more homogenous than the exposures collected via a probability sample.*
- *There is a probability of drawing inaccurate conclusions due to missing an important range of exposures.*
- *The potential homogeneity in some factors may reduce the power for some estimates of association, and may also limit the ability to assess dose-response relationships.*
- *Effect modification is a concern, because if the full range of the modifier isn't collected it may not be caught, and an erroneous conclusion about an association can be made.*
- *A sample from prenatal care providers could be missing those with unattended births or no prenatal care, those that are uninsured or not in the plans and infertile women—this could be mitigated by choosing a supplemental sample to cover these populations.*
- *Cooperation of prenatal care providers may be variable.*
- *Engagement of prenatal care providers could be resource intensive.*

- *Supplemental recruitment could be resource intensive.*
- *Not feasible to combine the two frames for analysis.*

HYBRID DESIGNS

5. **Convenience sample of prenatal care providers with a supplemental geographic-based probability sample.**
   Targeting a convenience sample of prenatal care providers, allow prenatal care providers to apply to participate in the National Children's Study as the larger frame of the Study. In a smaller sampling frame use a probability based geographic sample to capture environmental exposures, health care access, and ethnic/racial diversity with a random sample from an existing list frame of individuals living within the probability based geographic sampling units.

   Advantages:
   - *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
   - *Prenatal care providers may represent an efficient way to enroll pregnant women.*
   - *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
   - *The base population can be characterized, including health status, through medical records.*
   - *A second frame such as a list frame from the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) is easily replicated, and WIC enrolls 51 percent of pregnant women in their first trimester, and has a 91 percent participant contact rate. Women selected from the WIC list frame could be oversampled geographically for environmental exposures, or by demographic characteristics.*
   - *Introduces less bias in the smaller sampling frame than a convenience sampling approach.*
   - *If the probability sample is of sufficient size relative to the convenience sample, the probability sample can provide additional information about the characteristics of the convenience sample.*

   Disadvantages:
   - *The environmental exposures of the patients within the catchment area of the providers may be more homogenous than the exposures collected via a probability sample.*
   - *There is a probability of drawing inaccurate conclusions due to missing an important range of exposures.*
   - *The potential homogeneity in some factors may reduce the power for some estimates of association, and may also limit the ability to assess dose-response relationships.*
   - *Effect modification is a concern, because if the full range of the modifier isn't collected it may not be caught, and an erroneous conclusion about an association can be made.*
   - *A sample from prenatal care providers could be missing those with unattended births or no prenatal care, those that are uninsured or not in the plans and infertile women.*
   - *A program such as WIC or Title V would need to be successfully engaged.*
   - *If no additional supplemental program can be engaged with an established list frame, a second sampling frame will have to be constructed, which will utilize more resources and be less replicable.*
   - *Cooperation of prenatal care providers may be variable.*

- *Engagement of prenatal care providers could be resource intensive.*
- *Constructing larger analytic files from both frames may not be possible.*

6. **Probability based prenatal care provider sample supplemented by a convenience sample.**
In a hybrid design use a probability based prenatal care provider sampling frame supplemented by a smaller convenience sample. For example, a two-stage approach could be used for the prenatal care provider based probability sample. In this example, the first stage would be a probability based geographic sample. The second stage would be prenatal care providers practicing within the probability based geographic sample. Participants would be selected from the selected providers.

Supplement with a smaller frame, recruiting pregnant women or women trying to become pregnant from outside the probability based geographic sample as a convenience sample.

Advantages:

- *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
- *Prenatal care providers may represent an efficient way to enroll pregnant women.*
- *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
- *The base population can be characterized, including health status, through medical records.*
- *By selecting providers within the probability based geographic samples the geographic dispersion may better represent the gradient of environmental exposures.*

Disadvantages:

- *There could be overlap between the two frames that would need to be de-duplicated.*
- *A sample from prenatal care providers could be missing those with unattended births or no prenatal care, those that are uninsured or not in the plans and infertile women.*
- *Cooperation of prenatal care providers may be variable.*
- *Engagement of prenatal care providers could be resource intensive.*
- *Convenience sample participants may be biased due to self-selection or geographic differences compared to women in the probability based sampling frame.*
- *Constructing larger analytic files from both frames would require careful sample weights and analytic plans.*

7. **Prenatal care provider convenience sample with a nested geographic based probability sample, which could be supplemented by another convenience sample.**
A carefully designed prenatal care provider convenience sample could have a probability sample nested within it. The rationale for such a design would be to leverage estimates of rarer exposure response mechanisms with a larger sample size, while maintaining the ability to determine prevalence estimates in the probability frame. In addition it may be administratively or logistically advantageous for a group or network of providers to enroll participants with fewer restrictions than a pure probability design may impose. This sample could be supplemented by an additional convenience sample to capture underrepresented populations.

Advantages:

- *A growing number of health care providers utilize electronic medical records, which may make some data collection and enumeration simpler.*
- *Prenatal care providers may represent an efficient way to enroll pregnant women.*
- *The relationships that are already built between providers and their patients may also increase the efficiency and efficacy of recruitment and retention.*
- *The base population can be characterized, including health status, through medical records.*
- *The ability to produce prevalence estimates, based on the nested probability frame, particularly for economic and psychosocial associations, generalizable to the larger cohort only.*
- *Sampling weights would not be required in order to assemble analysis files that include the large convenience sample and the nested probability sample.*

Disadvantages:

- *There could be overlap between the two frames that would need to be de-duplicated.*
- *A sample from prenatal care providers could be missing those with unattended births or no prenatal care, those that are uninsured or not in the plans and infertile women.*
- *Analysis files would require special identification of the nested probability sample women and children.*
- *Cooperation of prenatal care providers may be variable.*
- *Engagement of prenatal care providers could be resource intensive.*
- *The main convenience sample would need to be very large in order to nest the probability sample within it.*
- *Constructing larger analytic files from both frames would require careful sample weights and analytic plans.*

THE NATIONAL CHILDREN'S STUDY
ADMINISTRATION

This document provides an overview of National Children's Study operations, administration, and organization.

Administration of the National Children's Study

The National Children's Study (NCS or "the Study") is a prospective national longitudinal study of the effects of environment and genetics on child health, growth, and development. The Study was mandated by the Children's Health Act of 2000 (Public Law 106-310) and is implemented by the National Institutes of Health (NIH) with consultation from a Federal Consortium that includes the U.S. Centers for Disease Control and Prevention (CDC) and the U.S. Environmental Protection Agency (EPA). Funding is provided by a congressional appropriation to the Office of the Director, National Institutes of Health. Within the NIH, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) provides the resources including space, personnel, expertise, and additional funding and support for the administration and conduct of the Study; the National Institute of Environmental Health Sciences (NIEHS) has provided some additional scientific advice. Oversight is provided by the Director, National Institutes of Health, and an Independent Study Monitoring and Oversight Committee. Strategic advice is provided by a federally chartered Advisory Committee. The Federal Advisory Committee meets quarterly and is a major venue for the National Children's Study to share with the public current activities and receive input and advice.

The overall principles of the National Children's Study are:

- Data-driven
- Evidence-based
- Community and participant informed

A Concept of Operations document based on the planned data life cycle is available on the National Children's Study Web site (http://www.nationalchildrensstudy.gov) at http://www.nationalchildrensstudy.gov/about/overview/Pages/NCS_concept_of_operations_04_28_11.pdf.

The National Children's Study is run by contracts to provide the federal government with flexibility in deploying resources and to ensure that the data collected are not the property of multiple awardees. Contracts are awarded for periods of performance that end upon contract expiration. The National Children's Study awards contracts for data collection, data analysis, data and specimen archiving, and for multiple support functions. Both data collection and support contracts are subject to full and open competition.

The National Children's Study Program Office

The National Children's Study (NCS) Program Office has a full time staff of 19 people and is organized on the basis of functional teams. The teams are:

The Planning Team collects information and develops recommendations for the National Children's Study programmatic activities, including protocol development needs and analysis plans. The team conducts gap analyses, identifies risks to Study schedule, and locates additional external resources—projects, programs, and organizations—hat can be engaged to support the mission of the Study.

The Operations Team determines tactical and technical implementation and monitors the ongoing activities of the National Children's Study Vanguard Study, including formative research and supplemental methodological studies. The Operations Team also coordinates functions with the other

National Children's Study Program Office teams to produce deliverables, such as Study Visit Instruments, and manage specimen and sample collection and repositories.

The Analysis and Evaluation Team evaluates the integrity (data quality), feasibility (scientific robustness), acceptability (burden on participants and Study infrastructure), and cost of Vanguard Study data. The Team recommends changes in study protocol and operations based on these evaluations.

The Communications Team, in liaison with the NICHD Public Information and Communications Branch (PICB) and the NIH Communications Office, is responsible for supporting effective communications regarding National Children's Study activities and plans with participant communities, Study Centers, and the general public.

The Regulatory Team manages oversight and compliance with applicable laws and regulations.

The Administrative Team manages NCS Program Office processes, including personnel and resources.

In addition to the process teams, the National Children's Study has content teams.

Study Visit Development is managed by three coordinated teams—the Study Visit Content Team, the Instrument Development Team, and the Forms Development Team. The Study Visit Development Team also coordinates overall protocol development. Each team works in conjunction with support and field contractors to develop the concepts and then the specific elements of each Study visit.

A summary of the workflow process for instrument and visit development is in Figure 1.

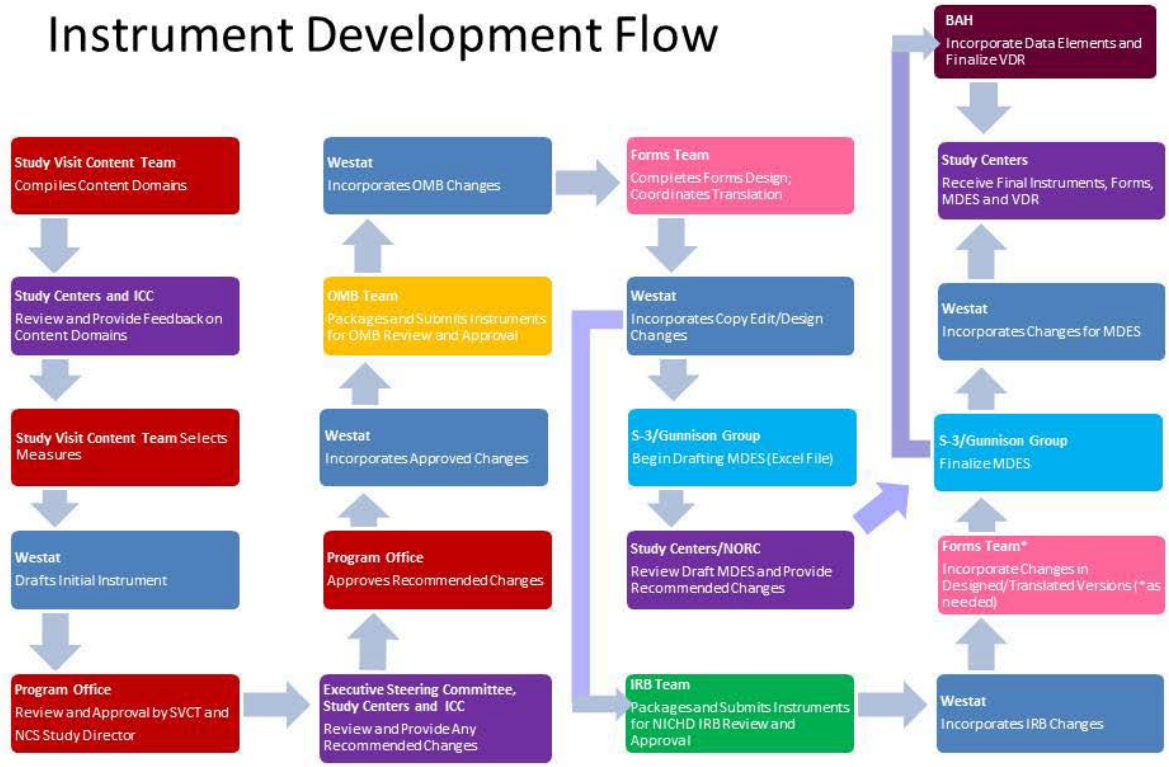

Figure 1: NCS Workflow Process for Instrument and Visit Development with integration of data elements

The Protocol Development Team develops the framework and then the content of the Vanguard Study Protocol and the Main Study Protocol.

The Environmental Team develops formative research projects and contributes to protocol development and analysis.

The Biospecimen Team develops formative research projects and contributes to protocol development and analysis.

The Genetics Team develops formative research projects and contributes to protocol development and analysis, policy and ethics.

The Data Access and Confidentiality Committee sets policy for data access and confidentiality.

The Publications Committee coordinates the publication of NCS-wide publications through the selection and prioritization of topics and organization of writing and analytic teams.

The Partnership Team develops collaborations and partnerships for the Vanguard Study through Supplemental Methodological Studies and will for the Main Study coordinate Ancillary Studies. The Partnership Team also coordinates the NCS Scholars Program.

Supplemental Methodological Studies *(SMS)* pertain to focused studies that take place during the Vanguard (pilot) phase of the National Children's Study. They are geared to inform the Main Study as to the feasibility, acceptability, and/or cost of items pertaining to recruitment, operational and logistic issues, and Study visit assessments. Supplemental Methodological Studies are initiated from outside of the Program Office and are developed outside the Study protocol planning process. They are funded externally; that is, not with the National Children's Study appropriation. The Principal Investigator will be identified by the applicant. Each SMS will have a Study Co-Investigator from either the Program Office or one of the contracted Study Centers. SMS that will be conducted at more than one location will also have a Study Facilitator for each additional location

Supplemental Methodological Studies are integrated with the Vanguard phase. That is, they involve National Children's Study participants and/or laboratory samples. Requests for just data are not SMS. In contrast, Substudies are a type of formative research involving participants and/or laboratory samples, but initiated and funded by the National Children's Study.

Supplemental Methodological Studies will generally be short-term efforts to support the Vanguard pilot goals. For these studies to inform the design of the Main Study, a prompt turn-around time is pertinent.

Please send general inquiries to NCSSuppMethStudies@mail.nih.gov

The ***NCS Scholars Program*** provides federal employees an opportunity to work full time or part time, on site or remotely, on specific projects of mutual interest.

Information Management Systems are coordinated through the Chief Information Officer, NICHD. The Initial Vanguard Study utilized a centralized model of data management, including case management systems and data capture systems. Based on the first year of experience with the centralized model and identification of multiple technical and logistical challenges in planning scale-up, the NCS Program Office implemented a new approach to provide greater flexibility and encourage exploration and innovation to determine preferred methods for case management and data acquisition.

This new approach is termed the "facilitated decentralization" model. In this model, the NCS Program Office develops evaluation questions and plans; data fields, tables and relationships; operational data elements; Study data acquisition instruments; data formatting and transmission standards; a central data archive; and specifications and guidelines for data security, participant confidentiality, and regulatory compliance. This facilitated decentralization model offers distinct advantages over a completely centralized structure: it allows Study Centers under contract with the National Children's Study to select or develop their own case management systems, data acquisition platforms, and as appropriate, data collection modalities to acquire the data. The model builds on local Study Center expertise with existing informatics systems and supports adaptation or development of new systems, with an emphasis on open-source, non-proprietary platforms.

All NCS data systems are certified and accredited per the requirements of the Federal Information Security Management Act of 2002 (FISMA) and related regulations. All NCS data specifications are consistent with international medical research standards, such as those developed by the Clinical Data Interchange Standards Consortium (CDISC).

The new approach to informatics for the National Children's Study is informed by several trends in informatics, including modular architecture, use of standardized terminology with curation, semantic awareness, scalability, defined transmission standards, open source platforms with development communities, vertical and horizontal integration of process, and interoperability. The NCS emphasis on interoperable modular architecture means that any component of a data system can accurately and efficiently communicate with other data systems, while adhering to international data standards. The approach is flexible to support innovation, accommodate evolving technology, and extend functionality. In addition, its components can be reused or adapted for other studies.

Major Initiatives Coordinated by the National Children's Study Program Office

Examples of trans NCS activities that have implications for other research efforts are:

Health Measurements Network – The concept of health is complex and multi-dimensional. Precise quantitative objective age and developmental stage measures for different health dimensions are not available. The National Children's Study, in partnership with other NIH initiatives, particularly the Patient Reported Outcomes Measurement Information System (PROMIS) and the NIH Toolbox for assessment of neurologic and behavioral function, began a formal initiative to develop relevant assessments for the ages and stages of child development. The resulting tools will be tested and validated in the Vanguard Study as well as in other venues.

Terminology – In conjunction with the NICHD and the National Cancer Institute, the National Children's Study began a systematic effort to develop relevant terminology for all ages and stages of development due to important gaps in all the major terminology systems.

Metadata Tagging of Operational and Longitudinal Data – the National Children's Study is required to integrate data collection and data analysis from multiple domains. Due to different data types, structures and formats, the integration has multiple technical challenges. In addition, in order to develop new instruments, consistency and efficiency in the workflow process is essential. The technical solution is to use metadata to describe both structure and content of the data and append or tag data elements with the structured metadata.

Some business practices adopted by the NCS Program Office are described in the following table.

EXAMPLES OF BUSINESS PRACTICES INITIATED FOR THE NATIONAL CHILDREN'S STUDY

Business Practice		Description
1.	Business Plans and Process Mapping	Developing business plans and process maps that outline the comprehensive set of activities along with timeline and resource requirements for the program.
2.	Risk Assessment and Management	Identifying risks and developing mitigation strategies to address them proactively; formalized in a Risk Management Matrix.
3.	Critical Path Plan Development	A plan that sequences a set of tasks (or workflows) with the shortest time required to complete those tasks towards achieving a goal. Any delays in the critical tasks can jeopardize the project if it is not compensated by acceleration of a later task.
4.	Program/Study Retreats	Monthly retreats designed as an open forum where all of the program/Study staff convenes to review program goals, develop plans, discuss operational issues, address and manage risks, obtain input from external experts, and identify implementable best practices from other studies.
5.	Concept of Operations (CONOPS) Development	A manual of operations designed to provide an overall view of the organizational structure, responsibilities, and interactions between the various components of the program.
6.	Decentralized Model for Operations	An operational model that implements field expertise and control at the local level with general programmatic guidance and oversight at the central level.
7.	Continual Program Monitoring	Using project trackers that list the various activities with target dates to periodically monitor progress of the program.
8.	Federated Institutional Review Board (IRB) Model	A model that establishes a shared set of IRB principles and processes for reviewing Study protocols. It permits information sharing across all IRBs and provides an opportunity to facilitate local IRB review by allowing reliance on the NICHD intramural IRB.
9.	Functional Domain-based Organizational Structure	Organizing the program/administrative office based on functional domains of the program/Study and aligning staff based on their expertise into these domains (For example—planning, operations, analysis and evaluation, communication and administration).
10.	Continual Resource Optimization	Assessing and optimizing staff and their roles to meet new or changing needs of the program. This also ensures that the program staff is aligned to activities based on their expertise and interests.

11.	Project Management Training	Training clinical research scientists in the principles and practices of project management to assist them in managing clinical research studies more effectively.
12.	Scholars Program	A program where interested federal employees from various fields, backgrounds, and training join the Study for 6 months to 1 year and contribute *in-kind* to the development of the clinical research/study, while also enhancing their own career goals.
13.	Internal Communication Enhancement • Daily Muster and Activities Dashboard	*Daily Muster* provides up-to-the-minute updates, information, requests, and status of various activities or tasks. *Activities Dashboard* provides a weekly status update of all activities within the program categorized by functional areas (for example —- planning, operations, analysis, communications, etc).
14.	NIH Plain Language Initiative Training	The Plain Language Initiative requires the use of plain language in all new documents written for the public, other government entities, and fellow workers. The Plain Language Initiative training focuses on writing that is clear and to the point so as to improve communication between the government and the public.

In sum, the National Children's Study Vanguard Study reached a critical milestone with preliminary analysis of the recruitment phase that will guide immediate Vanguard activities and Main Study design. The National Children's Study Program Office is continuing improvements in the business model. All operations continue to undergo evaluation with a commitment to adjust based on performance metrics.