

# Discussion of potential sampling strategies for the National Children's Study, Main Study

May 29, 2012

6100 Executive Blvd, Room 5A01, Bethesda, MD

Meeting URL: <https://webmeeting.nih.gov/ncssamplingframe/>

US/Canada Teleconference Number: 1-888-205-5513

Participant Code: 505616

9:00 a.m. -3:00 p.m.

## AGENDA

### Welcome

*Steven Hirschfeld, MD, Ph.D.*

*Director*

*National Children's Study*

*Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH*

### Goals of the Study:

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.

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.

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.

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.

*For more information, please refer to the white paper "Potential Sampling Strategies" under "What's New" on our national website: <http://www.nationalchildrensstudy.gov/Pages/default.aspx>.*

## Meeting Goals and Structure

*Moderator: Enrique Schisterman, Ph.D.*

*Senior Investigator*

*Division of Epidemiology Statistics and Prevention Research  
Eunice Kennedy Shriver National Institute of Child Health and  
Human Development, NIH*

Goal of the meeting:

To discuss in more detail the analytic strategies and potential efficiencies that can be achieved in all stages of sampling (geographic, provider based, and participant). We would like to explore the following questions and concepts in greater analytic detail. Please review and come prepared for discussion on May 29<sup>th</sup>.

## Discussion: Geographic Area Sampling

Desired Characteristics of an NCS Area Frame:

1. Incorporate a wide range of exposures that may be geographically dependent.
2. Effectively sample “unknown” variables through randomization.
3. By defining an area, it provides the ability to compare births in the sampled population to the known births via birth records.
4. Allow for inference between causal mechanisms and outcomes.
5. Ability to produce prevalence estimates of both outcomes and risk estimates, assuming that the sampling within the area is probability based.
6. There is still flexibility within the geographic areas to select secondary sampling units.

Questions:

1. What should the probability sample look like?
  - a. Can we have a purely probabilistic geographic sample (no oversampling, weighting, or stratification)? What is gained or lost by this approach?
  - b. Should the geographic samples be clustered within regions? Would this clustering need to be defined by population density, or could it be defined by environmental characteristics? What is gained or lost by this approach?
  - c. Should the geographic samples be selected using a stratified frame (by population density, demographic characteristics, happiness index, etc.)? What is gained or lost by this approach?
  - d. How could frame deficiencies be identified, and how could they be backfilled?
  - e. How many geographic areas need to be selected in order to generalize the findings of the study?
  - f. What should the area of the sampled geographic units be (i.e. state, county, zip code, census tract, census block group, or census block)?

## Notes on the Provider Based Sampling Experience

Graham Kalton, Ph.D.  
Director  
Westat  
Rockville, MD

### Discussion: Provider Sampling

Desired Characteristics of an NCS Provider Frame :

1. Incorporate a wide range of providers.
2. Allow for the selection of women from diverse socio-economic, race/ethnicity, and educational backgrounds.
3. Allow for recruitment of women in a preconception cohort, as well as early in pregnancy.

Questions:

1. How can providers be enumerated efficiently?
2. How can selected providers that choose not to engage in the study be replaced or substituted in a way that preserves the probability sample?
3. How can the method of selecting providers increase recruitment success (i.e. restricted frame sampling)?
4. Can the demographic characteristics of the provider's practice be determined prior to sampling?
5. Can the sampling method be flexible in this stage in order to allow for regional differences in identification of providers (such as availability of birth certificate records)?
6. Are there features of a provider practice (practice type, size, etc.) that might bias the recruitment of participants?

### Discussion: Sampling within Providers

Desired Characteristics:

1. Include women with diverse backgrounds, environmental exposures, and social contexts.
2. Engage women in the early weeks of pregnancy or prior to pregnancy, as well as later stages in pregnancy.

Questions:

1. What are effective ways to enroll women using systematic sampling? What could the basis of this sample be? Are there additional ways women could be sampled in an effective way that preserves the probability sample?
2. Should women residing outside of the geographic sample, but seeking care from a selected provider, be included? If they are included, what would this do to the comparisons to extant natality or American Community Survey data?

3. Should pre-conception women be eligible in the probability sample, or should they be from a separate cohort?
4. How would the sample of pregnant women be evaluated for frame coverage, or population representation? How could deficiencies be addressed in this stage of sampling?
5. How could women who change providers be handled?
6. How could women who move out of the geographic area be retained?
7. How can additional women, who reside within the geographic sample, but are not in the provider sample be included in a “light touch” cohort (for example, could a provider recruit women from a practice location other than the selected one) as a supplement?

### **Discussion: Supplementary Sampling**

We intend to supplement the enrolled population with women who may be excluded or underrepresented in the above sampling frames. In order to spur the discussion here are a couple of hypothetical examples:

Example 1: Women without prenatal care access; would it be adequate to recruit women from hospitals or birthing centers, excluding those who received prenatal care?

1. Could a sample like this be recruited as a convenience sample?
2. Would the women recruited in this way be considered a sub study, or could they be a part of the larger sample?
3. When should a supplemental sample become a separate sample frame?
4. Could meta-analysis techniques be used to combine a supplemental frame with the larger probability-based cohort? For example, if a hypothesis was posed about left-handedness and an exposure, could the information be pooled from both cohorts with regards to the exposure-outcome relationship?
5. What sources of bias would you anticipate by introducing the supplemental frame?

Example 2: If lower income women are found to be deficient in the frame: could WIC providers or other list frames be added to the provider frame, with exclusion criteria for women who have already seen one of the selected providers?

1. Could a sample like this be recruited as a convenience sample?
2. Would the women recruited in this way be considered a sub study, or could they be a part of the larger sample?
3. When should a supplemental sample become a separate sample frame?
4. Could meta-analysis techniques be used to combine a supplemental frame with the larger probability-based cohort? For example, if a hypothesis was posed about left-handedness and an exposure, could the information be pooled from both cohorts with regards to the exposure-outcome relationship?
5. What sources of bias would you anticipate by introducing the supplemental frame?

### **Discussion: Missingness by Design**

We are considering the use of a core questionnaire for everyone, with additional modules, or datasets for subsets of the study population.

1. Can we have different questionnaire intensities within the larger frame?
2. Can we have different questionnaire intensities between the supplemental frames and the probability sample?
3. What would be the parameters for determining the sample sizes of women receiving low intensity and high intensity instruments?

### **Summary and Adjourn**