October 20, 2003

Appendix B2
Draft White Paper

On

Sampling Design Options for the National Children's Study

by

Jeff Lehman, Warren Strauss, and John Menkedick

BATTELLE
505 King Avenue
Columbus, OH 43201-2693

Contract No. 282-98-0019

Dr. Peter Scheidt
Project Officer

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

Table of Contents

Glossary of Terms

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

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

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

Contract research organization: any organization that may be hired through competitive bids (e.g., universities, non-profit organizations, hospitals, commercial research corporations, etc.).

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

Hybrid design option: a design option that results from the combination of two or more simple designs. In this paper we typically refer to the hybrid design options as those options that correspond to a design that samples individuals using both probabilistic and nonprobabilistic methods.

Hybrid sampling frame: A sampling frame that results from the combination of two or more simple sampling frames to provide better coverage of the population of interest. For example, a simple household sampling frame might exclude the homeless and people living in institutions such as prisons, dormitories, health-care and military facilities. A hybrid sampling frame might supplement the household frame with selection of subjects from homeless shelters and these other institutions.

Inference: a conclusion drawn from evidence.

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

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

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

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

Probability sampling: sampling in which each member of the population has a known non-zero probability of being selected in the sample.

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

Quota sampling: similar to stratified sampling except applied in a nonprobabilistic framework; the population is divided into separate strata and within each strata the members are selected using convenience sampling until quotas for each strata have been met.

Random sampling: method of probability sampling in which each member of the population has an equal and known chance of being selected.

Sampling: selecting some part of a population to observe in order to make inference about the whole population; sampling methods can be classified as either probability or nonprobability.

Sampling frame: refers to the mechanism for identifying or enumerating the population from which the sample will be drawn; the sampling frame does not necessarily enumerate the entire population of interest as it may involve some non-coverage of the population.

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

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

Stratified sampling: method of probability sampling that begins by dividing the population into different strata (or groups) and then selects members from each strata using random sampling.

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

Draft White Paper

On

Sampling Design Options for the National Children's Study

Jeff Lehman, Warren Strauss, and John Menkedick

Battelle

B2-1. INTRODUCTION

The purpose of this white paper is to provide the National Children's Study (NCS) Program Office at NICHD a range of design options for selecting the longitudinal cohort into the study. Many of the options included in this white paper are based on previous work from a pilot study conducted by CDC and WESTAT (WESTAT, 2002), as well as ideas generated by the Study Design Work Group and various individuals working with the NCS Program Office on sampling strategies. In a study as complex as the NCS, a variety of design decisions (e.g., advantages and disadvantages of probability and non-probability based sampling) must be considered and evaluated. In another white paper, we discuss the overarching objectives of the NCS, a set of givens for the NCS study design, and a set of criteria for assessing and evaluating possible NCS study design options. This paper builds on that framework for assessing an NCS study design, and presents a range of candidate design options for consideration when determining the optimal and efficient design of the National Children's Study.

The purpose of this discussion is to provide a reasonable range of design possibilities for the NCS, highlight important differences in the various approaches, and present significant advantages and disadvantages associated with each of the approaches. We restrict our attention to options for the organizational structure of the NCS, and the methods for identifying and selecting the NCS cohort. In so doing, the objective is to span the set of reasonable design possibilities for these aspects of the NCS. The hope is that once the range of design options has been identified, and each option has been described, selection of a small subset of the design options that offer the most promising (and most appropriate) means of accomplishing the goals and objectives of the NCS will be more easily accomplished. Then, further investigation of this subset of design options (i.e., a more detailed analysis of the elements, advantages, and disadvantages of each design) will lead to selection of the most efficient and scientifically appropriate NCS study design.

It is important to note that in developing these options the focus has been on the organizational structure of the NCS and the means of identifying, sampling, and recruiting study subjects. In other words, at this stage of analysis of design options we do not specifically consider the means of sampling exposures or health outcomes for the selected cohort, the frequency of environmental sampling, and the other data collection protocols that will most certainly be an important design consideration for the NCS. While this restriction was noted above and appears in the title of this document, it is an important distinction and should be noted as we further discuss NCS design options. Most certainly the required exposure and health outcome measurements, and the frequency of those measurements, will represent a significant portion of the total cost and total burden associated with the NCS. Therefore, these measurements must be considered in any design, and, to the degree possible given that the data collection protocols for environmental and health outcome sampling have not been fully developed, options that appear to be most flexible in providing the required measurements will be highlighted. Also, where there are clear strengths or limitations of a proposed organizational structure or cohort selection method related to the likelihood of successfully measuring exposures and outcomes, these will be noted.

The range of options begins with a set of fundamentally simple sampling design options that result from a choice of whether (1) the Primary Sampling Units (PSUs) are selected via probability based sampling, quota sampling, or some type of convenience sample, and (2) sampling within the PSUs occurs via probability based sampling, quota sampling (to ensure some diversity and representativeness), or convenience sampling. A discussion of the strengths and weaknesses of both probability and non-probability (i.e., quota and convenience) can be found in another white paper to be developed under this work assignment. In addition to these fundamentally simple sampling designs, a class of hybrid design options is described. The hybrid designs include aspects of different simple designs in order to create a multiple-approach design that more efficiently meets the NCS objectives by capturing the strengths (or advantages) associated with a given scheme while attempting to minimize the effect of its disadvantages.

In summary, the overall intention of this paper is to provide NICHD with an objective discussion of several important design elements, and general descriptions of a range of plausible NCS design scenarios. To this end, the remainder of this paper is organized in the following manner. Section 2 describes a set of overarching design elements (selection of a sampling frame; how participants are selected from the frame; and the organizational structure for implementation of the study) and the different approaches possible for each of these design elements. Section 3 provides a description of the range of design options that we have identified, which correspond to a combination of the various design elements. Finally, Section 4 concludes our discussion with several remarks on the importance of this work, and the further steps that must be taken in identifying the optimal sampling design for the NCS.

In subsequent work associated with this white paper, and in consultation with a variety of NICHD advisors, other government officials, and subject matter experts (at a two-day workshop), each of the design scenarios will be evaluated in terms of their

ability to meet the scientific objectives of the NCS. By doing so, a small set of design option "finalists" will be identified as the most appropriate and promising designs that satisfy the NCS study objectives. These design "finalists" will then be more fully investigated, and more fully specified, so that an optimal NCS study design can be identified.

B2-2. IMPORTANT DESIGN ELEMENTS

While there are many elements of sampling design, especially for a study as complex as the NCS, in the following discussion we specifically delineate between three design elements: sampling frame for the population, sample selection of the cohort, and organizational structure of the study.

The first element, the sampling frame, involves selection of the methodology for enumerating or identifying the pool of subjects (i.e., the method of accessing the population) from which the sample will be selected. The true population of interest may be all children who are born in the United States; however, in order to accommodate the NCS objectives of obtaining pre-natal health and exposure measurements (and in some cases pre-pregnancy health and environmental conditions of the mother) it will be necessary to sample either pregnant women early in their pregnancy, women of childbearing age, and/or couples considering or attempting pregnancy. The sampling frame must reflect this necessity, and have the ability to identify these types of individuals. In addition, while practicality considerations will likely limit the sampling frame to a subset of all potential participants, the sampling frame should provide sufficient coverage of the population of potential participants to ensure that no unintentional biases affect the relationships measured in this study. In Section 2.1, several candidate sampling frames and their strengths and weaknesses are discussed.

The second design element is comprised of the methods for sampling the cohort of subjects from the sampling frame (i.e., selecting the subjects that will participate in the NCS). Basically, sampling involves the selection of some member of the population for study participation in order to draw conclusions about the broader population of interest. In general terms, the key to a "good" sample is that the sample be representative of the population of interest for all characteristics of interest. Two basic methods (or classes) for sampling units from a population exist: probability based sampling and non-probability based sampling, with each method corresponding to a set of strengths and weaknesses in terms of their scientific merit, cost, complexity, and assumptions necessary for generalizing to the population of interest with respect to the relationships measured in this study. In Section 2.2 we discuss methods for sampling the cohort of subjects for the NCS.

Finally, the last design element is concerned with the organizational structure for conducting the NCS and implementing the data collection protocols. This involves clearly defining the roles, responsibilities, and interactions between the various

organizations that will combine to implement this study, including the federal agencies, universities, data collection organizations, laboratories, data coordinating centers, etc. As in the other elements, several options are available, however, it should be noted that this design element may be dependent upon the sampling frame, since for some sampling frames, certain organizational structures will be preferred or required. In Section 2.3 we discuss the options for the organizational structure of the NCS, and identify the strengths and weaknesses of each option.

By specifically delineating these three design elements we are not implying that they are entirely independent (i.e., not all combinations of the options for each element should be considered). Rather, discussing each element separately clarifies the strengths and weaknesses of different approaches for each element and helps to later identify which approaches can be most logically combined. In other words, while the design elements must be specified simultaneously for any specific design option (i.e., a design is made up of all the above elements), they can, to some degree, be considered separately as some of the approaches for each element will accommodate several approaches for the other elements.

Even after specifically describing the above three design elements, there remain a variety of the design aspects that may not be defined in detail at this stage of sampling design development. One example of a design aspect that has not been fully specified is the idea of a primary sampling unit (PSU), or the highest level of design clustering in a multi-stage clustered design. Several possibilities exist for defining the PSU. For example, counties, established Centers (e.g., university hospitals), metropolitan areas, or states could be considered the primary sampling units. Once the PSUs are selected, sampling (either probability or non-probability) proceeds within the PSUs, thus the definition of the PSU is an important component of the overall sample. In addition to the method for defining PSUs, the total number of PSUs in the study will also likely have strong influence on the effective sample size (i.e., the design effect resulting from any within-PSU correlations) and power for the NCS to meet certain research objectives. Obviously, the within-PSU correlation will depend on the heterogeneity of the PSU, with larger PSUs (such as states) generally having more heterogeneity, and thereby less within-PSU correlation, and smaller PSUs (such as counties) having less heterogeneity, and thereby more within-PSU correlation.

Another important consideration that is relevant to all aspects and elements of the design of the NCS is the issue of the scientific assumptions that are inherent in each of the sampling methods and in each of the sampling frames. In many cases, these underlying assumptions are ignored, and the limitations associated with a violation of the assumptions are not fully recognized. For example, any form of observational study (be it cross-sectional, longitudinal with probability based sampling, case-control study, etc.) will have limitations. The most notable of these is the danger of confounding, which can only be avoided by random assignment of exposures to subjects in a clinical trial or other experimental approach (obviously not an option for the NCS). Likewise, it is important to note that there are assumptions with all forms of sampling (identifying subjects) that must be met to allow inferences to a larger population than just those sampled. For

probability based designs, one necessary assumption is that the sample of those who *agreed to participate* is representative of the entire population. If all subjects identified in the random sample agree to participate in the study, then the sample can reasonably be considered representative of the population; however, depending on the degree of nonresponse and/or recruitment failure, the probability based sampling of subjects could lead to a non-representative sample. Non-coverage could also lead to bias, both for samples and experimental designs. In the longitudinal setting, the same principles apply to the retention of study subjects, and the assumption that those who drop from the study do so at random. If this is not the case, then again the resulting sample may be non-representative of the original population. There is also an issue of changes in the population over time relative to changes in the NCS cohort (i.e. does the NCS cohort really represent the "population" of interest after more than a few years?) When evaluating design options, it is important to keep these assumptions in mind and weigh the ability to satisfy the assumptions along with all the other criteria for assessing the NCS design options (see criteria white paper). In the following, we highlight some of the important assumptions associated with each of the design options, and point out that if these assumptions are difficult to satisfy, then implementation of mechanisms to test the assumptions may be necessary (if these mechanisms are not available or are impossible to implement, then the limitations that are associated with a violation of the assumptions must be considered). Finally, if it is found that any of the above assumptions are violated in the NCS cohort, methods must be available to adjust important study relationships to correct for any bias introduced by the sampling design.

In the following sections we provide further details of the options for each of these general design elements, the strengths and weaknesses associated with these options, and any scientific assumptions that correspond to these options. Specifically, Section 2.1 provides a more detailed identification of the options for a study sampling frame, Section 2.2 presents the options for the sample selection, and Section 2.3 discusses the options for study implementation and data collection. In Section 3 we will discuss our individual design options that result from combining various aspects of these three general elements.

B2-2.1 OPTIONS FOR SAMPLING FRAME

As mentioned above, the sampling frame for the NCS refers to the method of identifying or enumerating the pool or population of study subjects from which the sample will be obtained. The true population of interest is children born in the United States. More specifically, we can consider children born during 2006-2009 as the population of interest, but we note that there may be other interpretations for the true population of interest (see Section 2.2). Since several of the NCS core hypotheses require pre-natal health and exposure measurements (e.g., those hypotheses related to early neurological development, adverse pregnancy outcomes, development of asthma), and since these types of measurements are best assessed prospectively instead of retrospectively (e.g., see Tingen et al., 2003), it is necessary to sample either women in early stages of pregnancy, women of childbearing age, and/or couples considering or

attempting pregnancy. Sampling from the latter two groups (i.e., women of childbearing age or couples attempting pregnancy) will also provide the opportunity to obtain exposure information prior to conception (e.g., pre-pregnancy health and environmental exposures for the parents) which is also considered a critical window for successful human reproduction and development (Buck et al., 2003). The sampling frame, or the method of identifying a pool of subjects from which to sample, must reflect the need to provide key pre-natal and potentially pre-conception measures, and have the ability to identify both women in early stages of pregnancy and women of childbearing age who are likely to become pregnant over the duration of recruitment for the NCS. The degree to which the sampling frame fails in this endeavor, and the degree to which it fails to "cover" the true population of interest, must be considered when choosing an appropriate sampling frame for the NCS. In the following, several candidate sampling frames and their strengths and weaknesses are discussed.

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

1. **Household Sampling Frame**: This sampling frame consists of all identifiable households in the U.S., and operationally would involve screening a sample of households to identify pregnant women, women of childbearing age, and/or couples attempting pregnancy. The screening and selection of subjects would likely follow a multi-level clustered design that involves the selection of PSUs (such as counties or metropolitan statistical areas), segments (smaller geographical areas) within the selected PSUs, and finally households within the selected segments.

2. **Physician's Office Sampling Frame**: This sampling frame would allow for the selection of a sample of physicians and/or medical offices during a first stage of sampling, and the recruitment of a sample of pregnant women and/or women of childbearing age seen in their practices during a second stage of sampling.

3. **Community or University Medical Center Sampling Frame**: This sampling frame involves selecting a sample of large health centers during the first stage of sampling that have previously demonstrated their ability and interest in conducting the NCS data collection protocol (e.g. through a competitive proposal process). These centers would recruit pregnant women and/or women of childbearing age either in proximity to, or currently being served by their center or associated physician's offices. (We will refer to this model as the Centers model.)

In the following, we highlight several important considerations when comparing these three sampling frames in light of the objectives of the NCS. Note that this

discussion focuses on the choice of a single sampling frame to highlight the differences between approaches.

- All three frames would involve some non-coverage of the population of interest. (Further discussion of this issue can be found following this bulleted list.)

- The non-coverage of the population for the Physician's Office sampling frame and Centers sampling frame may also involve some form of geographic limitations since presumably some regions will not be "covered" by a Center or a Physician's Office that is willing to participate in the study (note that this could be considered a form of nonresponse bias that could be corrected for in the analysis phase of the study results). The Household sampling frame would likely offer a "more" geographically representative sample depending on details related to stratification.

- All three sampling frames offer the possibility of sampling women prior to pregnancy by selecting women of childbearing age (and not just pregnant women). However, it should be noted that if the sampling frame covers all women of childbearing age, then a larger number of individuals is necessary to provide the desired sample size (i.e., a larger number of women of childbearing age versus already pregnant women would need to be selected in order to provide the desired 100,000 children in the NCS cohort). This introduces cost issues into the determination of what proportion of the sample should be women of childbearing age and what proportion should be pregnant women.

- The Household sampling frame may allow for a greater degree of pre-pregnancy measurements than the other two models since it would likely select women of childbearing age; however, this would also require a potentially sizable amount of resources in requiring tracking and measurements for women who never become pregnant. The Physician's Office sampling frame and the Centers sampling frame would provide a less costly means of identifying and selecting women that are already pregnant (and thereby avoid the cost of obtaining measures on women who never become pregnant). These could also be used to obtain pre-pregnancy measures for some participants, with the associated measurement costs, but pre-pregnancy tracking and measurement would not be required for all participants.

- The Household sampling frame may include household screening efforts in order to eliminate households without age appropriate women, and thereby limit the number of households that need to be enrolled.

- All three sampling frames have the potential to foster community involvement, since this aspect of the study is likely more dependent upon

the organizational structure of the study and the mechanisms that will be used for data collection (see Section 2.3). However, we note that the Centers model and Physician's Office model, by their nature, capitalize on local (public health or medical delivery system) infrastructure and relationships with potential study subjects that exist prior to the study and will likely enhance the potential for community awareness and involvement.

- The PSU (primary sampling unit) for the design likely depends on the sampling frame. For the Household sampling frame, the PSUs are most likely pre-defined geographic regions, such as counties. For the Physician's Office sampling frame, the PSUs may also consist of geographic regions, with Physician's Offices being selected at a lower stage of the design. And for the Centers sampling frame, the PSUs could be the Centers themselves, since presumably a center would provide coverage for a large number of possible study subjects, or it could again be geographic regions that correspond to a Center. (Note that not all geographic regions can be assumed to have an eligible Center, thus, defining PSUs as geographic regions would exclude those regions that do not have a corresponding Center.)

- In the Physician's Office sampling frame, a transition from Obstetrics/Gynecology physicians to a pediatric physician will likely be necessary upon birth of the study child. Methods for smoothly transitioning between these two types of physician's offices within the context of the NCS data collection effort would need to be considered and may be better handled through the organizational structure.

As mentioned above, all three of the sampling frames involve some non-coverage of the true population of interest. The Physician's Office sampling frame and the Centers sampling frame will exclude all women who do not visit or cannot be recruited by participating Centers or Physician's offices. This will presumably exclude that portion of the population who do not have access to healthcare. The Household sampling frame will exclude all women that do not live in an identifiable residence, such as female college students living in a dorm, females living in correctional facilities, homeless females, etc. Certainly, unless they are specifically targeted, these types of non-covered women will be difficult to identify in any sampling frame.

Hybrid sampling frames may also be constructed, which combine various pieces of the different sampling frames to provide better coverage of the population of interest. For example, qualified centers could be selected at the PSU level using the Center sampling frame, and at subsequent stages of sampling within the selected centers a household sampling frame could be used to screen and enroll study subjects within the geographic areas represented by the selected centers. Perhaps by combining the sampling frames, better coverage of the true population of interest can be realized in a manner that also enhances feasibility and efficiency.

There are also alternatives that do not include any of the above sampling frames. For example, if it were acceptable to collect pre-natal health and exposure related information retrospectively (probably not the case for the NCS), a sample of infants could be selected from the births reported in the Birth Registration System (see WESTAT report, 2002). Although perhaps not as likely to produce a sample that conforms to the needs of the NCS, other possible sampling frames could involve selecting women of childbearing age based on motor vehicle records or based on participant records from cooperating HMOs or county health clinics. At the present time, and given the necessary components of the NCS, the household, physician's office, and medical centers models (and combinations of these three) appear to be the best options for creating a viable sampling frame for this study.

Finally, although we specifically delineate this design element, it should be noted that the sampling frame is not independent of the other design elements (i.e., some sampling frames will naturally blend with some sample selection methods and some organizational structures). When outlining a set of NCS design options in Section 3, the most likely ("optimal") sampling frame(s) for each design option is (are) proposed.

B2-2.2 OPTIONS FOR SAMPLE SELECTION

Sampling involves the selection of some part of a population to observe in order to make inferences about the entire population of interest. In general terms, the key to a "good" sample is that the sample be representative of the population of interest. More particularly, in a "good" sample, the distribution of the variables of interest and the relationships between those variables observed in the sample are the same as the entire population. When this is the case, the characteristics of the sample can be generalized to characteristics of the population; however, if this is not the case, then we say that the sample is a biased sample, and inference to the general population based on the sample can be misleading. In the design of the NCS and in each of the described design scenarios, this sampling issue plays a key role. In particular, in each of the options described in this report, sampling units are selected using either a probability based method or a non-probability based method. In this section we briefly introduce the strengths and weaknesses of these two approaches to sampling, and identify the assumptions that are required with each approach in order to generalize the results to the population of interest (this topic will be discussed in more detail in other NCS white papers).

The most common method of avoiding sample bias is to use probability based sampling (e.g., random sampling, stratified sampling, etc.) in which each element of the population has a known non-zero probability of being selected. By performing probability based sampling from the population of interest (including stratification, over-sampling and any other sampling components deemed important), sample bias can be removed by incorporating the corresponding sample weights into data analysis exercises (assuming no nonresponse or non-coverage). In other words, by calculating sample

weights for each study subject (i.e., by calculating the proportion of the population that each subject represents), probability based sampling leads to selection of a cohort whose characteristics can be generalized to the entire population in a scientifically justifiable manner and without additional assumptions. Of course, there are other methods of removing sample bias, such as random assignment of experimental factors and removal of all sources of bias through quota sampling; however, these methods are either not relevant to the NCS (i.e., random assignment of exposures is not an option) or require additional assumptions (i.e., assuming that all sources of bias are known and can be accounted for through quota sampling). Thus, from a statistical perspective, probability based sampling is the optimal means of selecting the NCS cohort from the population of study subjects, unless it conflicts with other scientific objectives (e.g., feasibility to collect necessary exposure or outcome measurements). The degree to which it conflicts with these other objectives, and the corresponding importance of these objectives, would then need to be considered (see below).

Of course, there are several difficulties or disadvantages to using a probability based sampling approach for a study such as the NCS. The first, and perhaps primary, disadvantage to using probability based sampling is the difficulty in actually obtaining a "true" probability based sample from the population of interest. Presumably, the scientific understanding that results from this study will have implications for the population of children born in the United States in the future (e.g., the next 50 years), and not just those children born in the years 2006-2009. This may be better understood by considering a specific example. Suppose the NCS cohort consists of children born during the years 2006-2009, and suppose initial interest is in assessing the relationship between pre-natal exposure risk factors and birth defects. This assessment would likely occur sometime after the year 2010 (i.e., once the data is available), and, thus, any scientific understanding of these risk factors and any resulting remedial actions (e.g., warnings against using certain substances) would have no relevance to the children born during 2006-2009. Rather, these remedial actions would have relevance to only those children born after this assessment is available. The same interpretation holds for many other hypotheses of interest in the NCS (e.g., significant risk factors for schizophrenia that are determined from the NCS would be relevant to those children born after this determination is made). Basically, what this implies is that, to some degree, the true population of interest for the NCS is not only children born in the years 2006-2009, but is also all children born long into the future.

Identifying a sampling frame that will allow selection from this "true" population of interest is realistically impossible since identifying and/or selecting a child that will be born 25 years in the future is impossible. Yet for some NCS hypotheses, the risk factors for important health consequences may only become apparent 25 years into the future. If the demographic characteristics, behaviors, or exposures of the population change over that 25 year period, does that mean that conclusions drawn from the NCS are invalid for application to help future generations avoid disease? Not necessarily, however extrapolation of the NCS results to future generations may require additional assumptions and assessment of cohort characteristics that might be related to the observed effects.

In general, this type of noncoverage of the population of interest can be distinguished by considering the difference between enumerative studies, in which all elements of the population can be enumerated and sampled, and analytic studies, in which one is not dealing with a finite, identifiable, and unchanging population (i.e., rather one is dealing with some conceptual population). Hahn and Meeker (1993) discuss the assumptions that are necessary for statistical inference as well as the differences between enumerative studies and analytic studies. They suggest that for analytic studies, as the NCS arguably is, inference to the population based on a probability sample is not assumption free, and depends on an assumption of the representativeness of the sampled population to the conceptual population (further discussion of this issue will be included in a separate white paper).

Disregarding the ability to sample from the true population of interest (e.g., by assuming that the true population of interest is limited to children born during the years 2006-2009), there remains the issue of obtaining a true random sample, which is a sample in which all selected sampling units agree to participate in the study, from this population. In most (or even all) studies involving human subjects, a true probability based sample is impossible to obtain due to nonresponse of study subjects (i.e., selected subjects refusing to participate in the study) and the exclusion of certain segments of the population of interest during the construction of the sampling frame (e.g. homeless women being excluded from a household sampling frame). Thus, while obtaining a true probability based sample requires no additional assumptions in order to generalize results to the population of interest, generalizing the results of any realization of a probability based sampling approach (i.e., a probability based sample that involves any amount of nonresponse) to the population of interest does require assumptions about the randomness of the nonresponse, and/or the amount of nonresponse that is realized in the sample. Some of these assumptions may be testable. In a probability based sample, comparisons can be made between participants and non-participants, given limited information (e.g. from screening questions prior to enrollment) for the non-respondents. For both a probability and non-probability sample, comparisons could be made with census data or birth records.

Of course, if the nonresponse is limited or reasonably considered to be random, then the sample can be assumed to be a probability based sample and the results can be generalized to the population. Additionally, there exist a variety of statistical methods for accounting for nonrandom nonresponse in a probability based sample and for adjusting for things like non-randomness (these are the subject of other NCS white papers). In addition, further research needs to be conducted to determine tolerable rates of non-response and attrition in the NCS.

In contrast to probability based sampling, non-probability based sampling involves the selection of subjects from the population in some nonrandom manner. For example, convenience sampling would select subjects from the population by choosing those subjects who volunteer or in other ways are convenient (or easy) to sample, and quota sampling selects subjects by applying quotas to a convenience based sample, such as limiting the number of subjects that are recruited from a specified category. [Note that

quota sampling can be viewed as the non-probability equivalent of stratified sampling (see glossary of terms).] The main advantage of non-probability sampling is the potential increase in volunteer participant interest and commitment, which is expected to result in the chance for more frequent, and perhaps more accurate, measurements along with increased retention. Additionally, non-probability sampling often results in cost and time savings when compared to a probability based sample. However, the scientific cost of this savings is the inability to generalize the relationships observed within the NCS cohort to the population of interest from statistical analysis of the data alone without making several relatively strong assumptions.

The main assumption that is necessary in generalizing the characteristics of a non-probability based sample to the population of interest is an assumption that the sample is representative of the population. By representative we mean that the distribution of the characteristics of interest is the same in the sample as it is in the population. Certainly, in some cases this assumption may be reasonable (e.g., it may be reasonable to assume that the relationship between exposure and an adverse health effect is the same in the sample as it is in the population of interest); however, it is important to note that 1) for some hypotheses this assumption may be invalid (for example, when diet affects uptake of an environmental toxin), and 2) in the absence of a sample (or a subsample) that is thought to be representative of the population, there is no means to verify the validity of assuming a non-probability based sample is representative of the population. This limitation of the non-probability based sample (i.e., the inability to generalize the results to the population or to generalize the modeled relationships to the population without making strong and unverifiable assumptions) represents the primary objection to this type of sampling. The debate between probability and non-probability sampling will be further considered in another white paper that presents a more detailed discussion of the strengths and weaknesses of these sampling methods.

One possible method for addressing whether a non-probability based sample is representative of the population and/or adjusting for the bias that may be introduced in this nonrandom sample involves the idea of a small validation sample. As mentioned above, in the absence of a sample that is thought to be representative of the population, there is no means of determining whether a non-probability based sample is biased. However, if a representative sample is available, that sample can be used to determine whether the nonrandom sample is biased, and, perhaps, to adjust for the bias that is exhibited by the sample. Thus, if some form of non-probability sampling is to be used (e.g., convenience sampling), then obtaining a small validation sample that is representative of the population (i.e., selecting a small probability based sample) may offer a means of determining the degree of bias in the non-probability sample and adjusting the results to account for that bias. Investigation of the statistical methods for combining and comparing the information in the non-probability sampled subjects and the probability sampled subjects will be necessary in order to further explore and assess this possibility in terms of its feasibility (sample sizes, incidence rates, assumptions and validity of the approach) and any associated design characteristics (e.g., sample weights, proportion of subjects in the validation set, design effects, etc.). (These methods are currently being investigated and a technical Appendix will be prepared that details the

possibility of using validation samples to correct for sample biases introduced through a convenience selected sample if designs using this method are identified for subsequent development.) Several of the hybrid designs and multiple cohort designs described below incorporate this idea by sampling a portion of the cohort on a probability basis and a portion on a convenience basis.

In making a decision between non-probability based sampling and probability based sampling, if all other things are equal, probability based sampling is preferred. It offers the most likely means of obtaining an unbiased sample from the population from which a generalization of the results of the study would be scientifically justifiable with as few assumptions as possible. However, there are certainly other issues, such as controlling for confounders, likelihood of cooperation and retention, cost, difficulty of sampling implementation, and many others, which must be considered in determining the optimal sampling methods for the NCS. The degree to which each of the factors is considered important will certainly impact any decisions on the design of the NCS.

Now that the basic differences between probability based sampling and non-probability based sampling have been identified, we begin discussion of the NCS sample selection options by considering a matrix of nine fundamentally simple sampling approaches that result from a choice of whether (1) the Primary Sampling Units (PSUs) are selected via probability based sampling, quota sampling, or some type of convenience sample, and (2) sampling within the PSUs occurs via probability based sampling, quota sampling (to ensure some diversity and representativeness), or convenience sampling. Table 1 displays the matrix that represents these nine design options, and identifies seven of the nine options that should be considered further (enumerated as 1 through 7 in the table).

Table B2-1. Matrix of nine fundamentally simple study design options.

Sampling PSUs	Within PSU Sampling		
	Probability Based	Quota	Convenience
Probability Based	1	X	X
Quota	2	4	6
Convenience	3	5	7

Note that two of the nine options (the probability based sampling of PSUs followed by either quota or convenience sampling within the PSUs) will not be considered further since the probability based sampling of PSUs (i.e., the higher stage of sampling) is irrelevant if some manner of probability sampling within the PSUs (i.e., the lower stage of sampling) is not implemented. In other words, if probability based

sampling at a higher stage of the sampling design is not followed by some form of probability based sampling at the lower stages of the design, then the advantage of the probability based sampling at the higher stages is lost due to an inability to calculate sample weights for the lower stages. Thus, we will assume that once some form of probability based sampling is utilized, all lower stages of the design should also incorporate some form of probability based sampling. This criterion removes two of the nine options in Table 1 from consideration. On the other hand, the other seven design options in Table 1 represent reasonable, or at least plausible, approaches to the design of the NCS. Specifically, we note that selection of a PSU using non-probability methods does not limit within-PSU sampling options. These options will be discussed in more detail in Section 3.

In addition to these simple design options, we also consider a set of "hybrid" design options that attempt to include multiple aspects of the simple designs in order to create a design that better meets the NCS objectives. These hybrid design options combine probability based sampling and non-probability based sampling by selecting a portion of the sampling units on a probability basis (e.g., a validation set) and selecting all other sampling units on a convenience or quota basis. Table 2 displays the class of hybrid designs where P_1 percent of the PSUs are selected on a probability basis, and P_2 percent of the within-PSU samples are selected on a probability basis. Intuitively, both the probability based sampling scheme and the non-probability based sampling schemes have advantages and disadvantages (see above discussion). By combining these sampling schemes, the hybrid design options attempt to capitalize on the strengths (or advantages) associated with each scheme while minimizing the effect of their weaknesses. This combination of probability and non-probability sampling can occur at a single sampling stage or at all sampling stages; however, it is again important to recall that once some form of probability based sampling is applied, it is intuitively reasonable that all lower sampling stages should also involve some form of probabilistic sampling (so that sampling weights can be calculated).

Table B2-2. Class of hybrid design options. P_1 and P_2 range from 0 to 100 and represent the proportion of the sample selected on a probability basis at each stage of sampling.

	Sampling PSUs	Within PSU Sampling
Probability Based	P_1%	P_2%
Convenience or Quota Based	(100 - P_1)%	(100 - P_2)%

Without loss of generality, we focus on two specific design options from this class of hybrid designs. The first option corresponds to the design for which all within PSU

sampling is conducted on a probability basis (i.e., P_2 is 100 percent), but some proportion of the PSUs (e.g., 80 percent) is selected on a convenience basis. The second option corresponds to the setting where some proportion of the PSUs (e.g., 80 percent) are sampled on a convenience basis, and some proportion of the within PSU sampling (e.g., 50 percent) is also conducted on a convenience basis. Of course, the actual effect of these different proportions and the optimal proportions for sampling PSUs and sampling within PSUs will need to be investigated, and will be the subject of further research if one of these hybrid approaches is selected for more careful specification.

A final, and perhaps most general, hybrid sampling design that we propose involves the selection of multiple cohorts for the NCS. In this sampling framework we would envision selecting two (or more) cohorts to meet the NCS study objectives. The idea is that, while no single cohort can meet all of the NCS study objectives, it may be the case that several cohorts, and an appropriate combination of the information from these several cohorts, will provide a design that satisfies the objectives of the NCS. Similar to the hybrid designs discussed above, one cohort might be selected on a probability basis and one cohort might be selected on a convenience basis; however, in this case the duration of follow-up for each cohort could vary, as could the data collection requirements for each cohort (i.e., the hybrid designs discussed above could be considered a subset of the multiple cohorts designs). As an example, suppose a pre-birth (or pre-conception) cohort is selected on a convenience basis, due to the difficulty of identifying and recruiting women of childbearing age, and suppose that it is determined that only approximately half of the cohort is necessary to provide statistical power to assess all hypotheses related to pre-natal and/or peri-natal exposure. Then, the other half of the cohort could be selected as a post-birth cohort (e.g., children approximately 6 to 12 months of age) using a probability based design of births reported in the Birth Registration System. Presumably, both cohorts would be followed for the entire duration of the study (although alternatives are certainly possible), and pre-natal and early exposure information for the post-birth cohort could be assessed retrospectively. [Note: to assess any bias introduced through the retrospective exposure assessment in the post-birth cohort, similar retrospective assessment could be done in a small validation sample of participants in the pre-natal cohort – thereby providing a basis for the relationship between prospectively collected and retrospectively collected pre-natal exposure information.] This design would provide a convenient sampling frame (the Birth Registration System) for conducting the probability based selection of subjects, and would eliminate the difficulty of probabilistically selecting pregnant women and/or women of childbearing age.

An alternative multiple cohorts design, could select one cohort, for which less detailed information would be collected, on a probability basis, and another cohort, for which detailed information would be collected, on a convenience (or quota) basis. In addition to collecting differing degrees of information on the two cohorts, the cohorts could also be followed for varying lengths of time. Conceivably, this type of design could limit the effect of attrition and/or nonresponse in the probability based sample, since sampled individuals would not be required to undergo a high level of burden in order to participate in the study, or perhaps not be expected to participate in the study

beyond some point (e.g., beyond birth of child). At the same time, for the convenience selected subjects who are more motivated to participate in the study, this design would capitalize on their interest by collecting more detailed information, possibly over a longer period of time, on these subjects.

A final multiple cohort type of possibility (suggested by Frank Speizer in an email discussion) that could adapt to virtually any of the above sampling schemes could involve selection of a "sibling" cohort from an "original" cohort. Basically, an "original" cohort that consists of a large portion (e.g., 90%) of the desired 100,000 children could be selected using any of the sampling methods described above. Then, a subsample of this "original" cohort could be followed (perhaps identified by asking the parents if they intend to have another child in the near future) to enroll a "sibling" cohort which consists of siblings of children in the "original" cohort. This "sibling" cohort would then be recruited prior to conception, and, thus, pre-conception exposure measurements and behavior information would be available. This type of multiple cohort design offers a novel and promising means of obtaining a portion of the cohort prior to conception; however, the design effects, the optimal proportions for each of the cohorts, the expected number of women to needed to provide the desired sample sizes for each cohort, and other logistics problems that affect the study, such as the length of the study recruitment phase to allow for the "sibling" cohort and any corresponding changes to the population of inference, would need to be further considered. There would also be a need to understand the sample bias introduced by this design – e.g., the fact that pre-conception information and measurements would primarily be from women who had already given birth to another child enrolled in the study, and, similarly, that the cohort with pre-conception measurements would not include any first-born children.

This set of sampling methods basically outlines the set of sample design options that will be considered. By combining each of the sampling methods with their likely sampling frames and organizational structures a set of design classes are identified and are discussed in greater detail in Section 3.

B2-2.3 OPTIONS FOR ORGANIZATIONAL STRUCTURE

The options for the organizational structure that could be used for implementation of the NCS is likely somewhat limited, and, to some degree, is determined by the sampling frame and the methods for sampling (see discussion below). We begin by noting two key characteristics that the organizational structure must satisfy:

1. First and foremost, the organizational structure must have the ability to collect all of the desired data, such as environmental samples, biologic samples, and questionnaire information, for 100,000 children across the United States.

2. The data must be collected in a uniform and consistent manner for all study subjects so that it can be merged and combined appropriately. In other words, methods for collecting environmental measurements for children living in the

Northwest should be the same as the methods that are used for collecting environmental measures for children living in the Southeast.

These minimal requirements significantly restrict the possible organizational structures that could be used for the NCS. Three plausible approaches are as follows:

1. Organizations such as University Medical Centers or large Hospitals are contracted to recruit, retain, and collect all relevant data using their own facilities. In this case, the design specifications for sampling study subjects will be developed by study leaders. These standard sampling practices for recruitment, retention, and data collection would be followed by all Centers for research related to the NCS core hypotheses. Some opportunities may be available for center-specific measurements (i.e., of community interest), subject to overall burden and human subjects approval.

2. A single (or a few) contract research data organizations are assigned responsibility for all aspects of the design implementation, recruitment of subjects, retention of subjects, and collection of all relevant data (e.g., environmental samples, biological samples, etc.). These organizations then hire additional personnel and obtain the modern facilities and infrastructure (or subcontract these components out) that will be required to collect the biologic and environmental information.

3. Some combination of the above two structures. For example, a single organization could be responsible for selecting the sample of subjects, Centers could be responsible for conducting the relevant biological sampling, and one or several contract research organizations could be responsible for conducting the relevant environmental sampling.

One organizational structure that fits into option 3 above is a hybrid of the classic Centers model (where a few enduring centers follow the study subjects). After selecting a set of PSUs (e.g., counties, metropolitan statistical areas, regions, etc.), Centers that are in close proximity to those PSUs could be used to follow the study subjects from that PSU. In other words, all study subjects that live in the vicinity of one of the selected centers would be followed by that center. For those PSUs that do not have a corresponding Center in close proximity (i.e., for all individuals that do not live near one of the Centers) an appropriately identified data collection organization(s) (e.g., a contract research organization) could be used to follow the subjects in that PSU (perhaps a single organization would follow all subjects not covered by a Center or a set of physicians offices could be recruited to follow these subjects, etc.). This also allows for coverage of individuals moving out of selected PSUs and not into areas "served" by other Centers.

Another organizational structure that we would like to note is an organizational structure that relies on physician's offices (e.g., obstetrics/gynecology practices, pediatricians, primary care physicians, etc.). While we do not consider physician's offices to be equivalent to large medical centers, if the Physician's Office sampling frame

is selected then, presumably, the physician's offices would be largely involved in the data collection (at least the collection of biological samples). We will consider this type of organizational structure as a structure that fits into option 3.

Determining which of these three options is most appropriate depends on a variety of factors. First, as mentioned above, it depends to some degree on the sampling frame and the sampling methods. If the Centers sampling frame is used then option 2 is likely eliminated as a viable alternative. If the Physician's Office sampling frame is used then option 3 is the most likely organizational structure since the physician's offices would likely be involved in the data collection process (at least in collecting the biological data). Finally, if the Household sampling frame is selected, then any of the three organizational structures could be utilized. Other important factors influencing the determination of an appropriate organizational structure may be: support of the health care and academic communities so that the impetus for the study is not lost, cost considerations, feasibility considerations in terms of the likelihood of providing the necessary infrastructure for meeting the scientific objectives of the NCS, etc. Finally, it should be noted that it is possible, for logistical and implementation reasons, that the organizational structure for the NCS will be selected first, thereby limiting the choice of sampling frame and sampling design.

Generally comparing the three organizational structures in light of the objectives of the NCS we note the following considerations:

- Using some sort of Centers or Hospitals based collection organization provides a more likely atmosphere for community engagement. Additionally, since large Centers and/or hospitals are involved in the data collection, the study would include the infrastructure to support specialized measures (e.g., medical facilities with technologies such as 3D ultrasound and coordination of delivery room biologic samples). Local institutions may also be more sensitive or responsive to topics/issues of community interest.

- Using only one or a very few data collection organization(s) would avoid some of the difficult logistics (e.g., merging data, conformance with uniform data collection protocols, etc.) that will be apparent if multiple data collection organizations are used.

- It is unlikely that a single data collection organization could perform all of the necessary components of data collection for this study. In fact, it may be unlikely that many Centers have the ability to perform all of the necessary components of data collection for this study (e.g., the environmental exposure monitoring aspects of the study). For this reason, the study will necessarily involve multiple data collection organizations, and therefore, it may be necessary to designate a central data coordination agency that is responsible for merging data from the multiple data collection facilities, and ensuring uniform data collection practices.

- Using multiple data collection organizations provides a flexible means of collecting the necessary data for the NCS. For example, if a subject moves to a different area that is already "covered" by one of the existing data collection organizations, then that subject can be assigned to that data collection organization. Alternatively, if a subject moves to a new area (i.e., an area with no other study subjects and/or no data collection organizations), then a new data collection organization can be selected for that subject, or an existing data collection organization could be responsible for managing that subject and all subjects outside "covered" areas.

- In general, in the descriptions provided in Section 3, an organizational structure that involves a variety of data collection organizations (i.e., option 3 provided on page 17) is described as the likely means of implementing the study. However, it should be noted that option 2 also allows for the possibility of involving a variety of data collection organizations as subcontractors to a single (or a few) contract research organizations.

Obviously, there are many yet to be answered questions relevant to the implementation of the NCS and the corresponding data collection. For example, what if the data collection organization, while capable of collecting the necessary biological samples and/or performing the necessary health screening, does not have the capability to collect the needed environmental samples (as might be the case with a university medical center or a hospital)? Presumably, that organization would need to either obtain that capability (through hiring additional personnel and purchasing the appropriate infrastructure), or would need to hire a subcontractor to perform those data collection components that it could not complete. Alternatively, perhaps facilities without the required capabilities would be excluded from the list of eligible data collection organizations? Another open question is how will study subjects be tracked and followed when they relocate, especially if they relocate to a "new" area of the country that was not already represented in the sample?

The list of uncertainties with regards to the organizational structure of the study could likely continue; however, the approaches presented above (in particular option 3), appropriately modified, should accommodate a large range of implementation options (see the more detailed considerations below).

Again, although we specifically delineate the organizational structure design element, note that some sampling frames and some methods for sampling will naturally blend with one (or more) of the organizational structures (e.g., if the sampling frame selects only patients from a set of Centers, than those Centers should be involved in the data collection for the study). When outlining the set of NCS design options in Section 3, the likely NCS organizational structure(s) for each design option is proposed.

B2-3. RANGE OF DESIGN OPTIONS

In this Section we explore a range of NCS design options that correspond to a combination of the design elements outlined above. In so doing, a balance must be struck between detailed descriptions of design options that outline every possible aspect of the study (obviously not the goal here) and ambiguity in which all design options appear the same due to a lack of identification of the key design elements. Our objective is to cover a set of reasonable design possibilities for the NCS, with the hope being that once these options have been identified, and each option has been described in relatively general terms (relatively general given that a full description of any design option would likely require a much higher degree of specificity), selection of a small subset of the design options for further investigation will be more easily accomplished. Then, a more detailed analysis of the elements, advantages, and disadvantages of these promising design options will lead to selection of the most efficient and scientifically justifiable NCS study design.

In identifying the range of design options to describe, we focus on the options for sampling subjects (i.e., Section 2.2), and now identify a set of six design classes. We refer to them as design classes since within each of the classes a variety of specific choices or design details exist. In other words, each of the six described designs really represents a class of design options that correspond to the method of sampling study subjects. As described in the previous sections, once the method of sampling study subjects (e.g., probability based or non-probability based, or some hybrid of these) has been identified, there remain a large number of design issues that must be defined (e.g., organizational structure for implementing the study, means of sampling exposures and health outcomes for the study subjects, etc.). In the following descriptions we attempt to provide a brief overview of (1) the sampling design and sampling frame for selecting study subjects, including the selection of PSUs and the sampling that occurs within the selected PSUs, (2) the implementation and organizational structure options appropriate for this design, (3) the advantages and disadvantages of the sampling design including any plausible difficulties in implementing the design, and (4) similarities of these designs to each other. Future work involves identification of how each of these designs address the givens and objectives of the NCS (see white paper on criteria for assessing candidate designs), and discussion of other prominent studies (e.g., other longitudinal studies or other exposure studies), the designs that were used in those studies, and the rates of recruitment and retention realized as a function of burden of the study protocol and method used to offset that burden (see white paper on recruitment and retention). Again, these descriptions are not meant to provide a detailed account of the design and all its options, but rather they are meant to serve the purpose of evaluating each design option on an overall scale to identify a small set (e.g., two to three) of design options that should be investigated in further detail.

The six design classes for recruiting and retaining the NCS cohort are as follows:

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

Note that options 1 through 3 correspond to a set of relatively simple sampling methods where sampling of PSUs occurs by a single method and sampling within PSUs occurs by a single method. Options 4 and 5, on the other hand, involve some combination of convenience and probability based sampling at either or both the PSU level and within the PSUs. Additionally, we assume that options 1 through 5 require uniform data collection protocols and study subject participation across the entire cohort. Option 6, on the other hand, corresponds to a class of multiple cohort designs where, again, some subjects are sampled probabilistically and some are selected on convenience. The difference with option 6 (making it the most general option) is that it allows for the possibility of differing lengths for the follow up period and/or alternate sampling protocols (and associated burden) for different portions of the cohort. The specific implementation details for a multiple cohort design with differing lengths of follow-up (option 6) would need to be designed to ensure sufficient sample size and power to address the various NCS core hypotheses.

Finally, as discussed in Section 2.2, a potential add-on option that could be combined with any of the 6 design options above is a multiple cohort design where an "original" cohort is selected (via any of the identified sampling methods) and a "sibling" cohort is selected from a portion of the "original" cohort.

Figure 1 provides a visual overview of the range of design options being presented in this report. At the top of the figure is a box that represents the population of interest (i.e. all children born in the U.S. between 2006 and 2009). In each of the design options that will be presented, we recognize that the sampling frame that is constructed will by nature exclude some portion of the population of interest, as demonstrated by the smaller box. Once the sampling frame has been constructed, a sample will be drawn from the sample frame – again with an assumption that the sample will likely exclude

certain segments of the sampling frame due to non-response, difficulties encountered in the field, etc. Underneath the sampling box is an arrow that depicts a variety of options for obtaining the sample, assuming a multi-level clustered design in which a sample of PSUs are first obtained (using either a probability or non-probability basis for selection), and then sampling occurs within the PSU. At the very bottom of the figure is a box suggesting the possibility of designing different data collection protocols for different segments of the study population. These separate protocols may vary in terms of planned periods of follow-up, intensity of data collection efforts as discussed in the multiple cohorts option (section 3.6 of this report). In sections 3.1 through 3.6 that provide a description of each class of design option, we introduce a smaller reproduction of Figure 1 in which the appropriate sampling design strategies (in terms of selection of PSUs and sampling within PSUs) are highlighted for ease of interpretation. As mentioned previously, our aim in this interim report is not to fully specify each of the designs, but rather to provide a description of the design, the many possibilities that are inherent in each design, and any perceived advantages and disadvantages for these designs.

Note: The designs need to be evaluated relative to the "givens," criteria, and hypotheses. This will be provided later, but before any decisions are made regarding the selection of design options to be more fully developed.

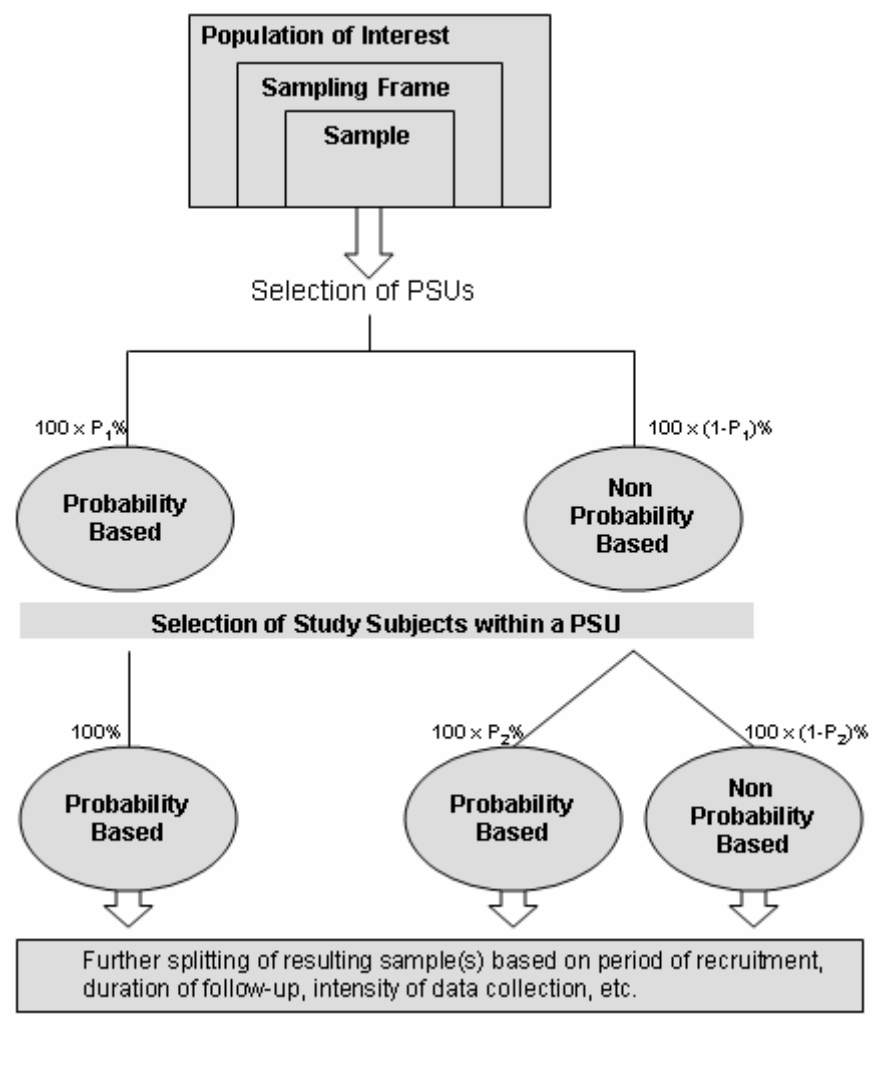

Figure B2-1. A Range of Design Options for the NCS

B2-3.1 COMPLETE PROBABILITY BASED DESIGN

This model calls for the probability based selection of all units at all stages of the sampling scheme. In particular, primary sampling units (PSUs), such as geographic regions, are selected using a probability based design, and all levels of sampling that occur within these primary sampling units are conducted using a probability based design. Figure 2 highlights the sampling strategies that are consistent with a complete probability based design. With this selection method, sampling weights and selection probabilities could be calculated and accounted for, and any systematic biases in the sample should be removed via the probability based sampling and appropriate statistical analyses of resulting data. That is not to say that a single realization of a probability based sample is unbiased since even a random sample could, by chance, be biased

(e.g., hypothesis driven studies are often designed with allowances for the possibility of Type I (α=0.05)and Type II (β =0.20) errors, which are assumed to occur with known frequency because an unintentional bias towards acceptance of the wrong hypothesis was introduced into the sample despite the random selection of study subjects from an appropriate sampling frame). As discussed above, this design class at face value may represent the most scientifically justifiable sampling scheme, in that it potentially allows the procurement of a random sample from the current population of interest. Note that this option allows for the possibility of over-sampling or other methods to increase the precision of estimates for sub-populations of particular interest (i.e., we need to ensure that the NCS has adequate numbers of study subjects to make comparisons between subgroups with different types or magnitudes of exposures).

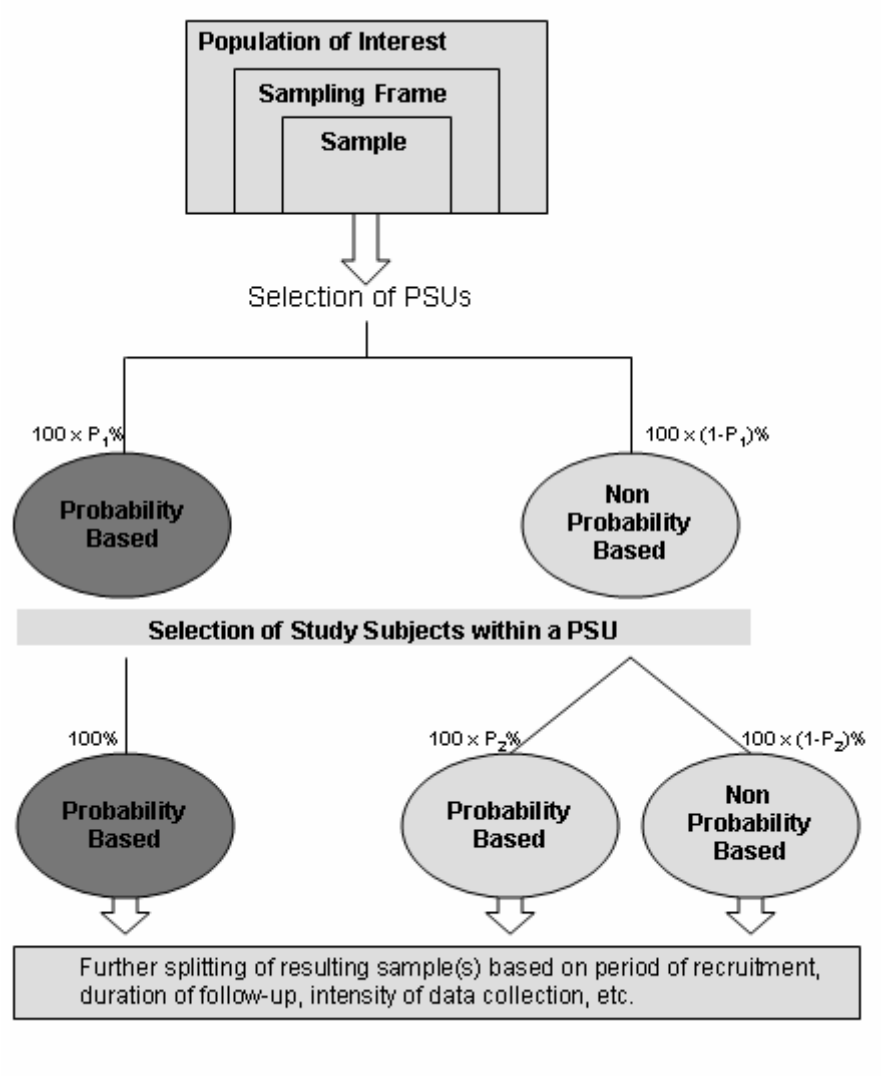

Figure B2-2. Highlighted Sampling Strategies Consistent with a Complete Probability Based Design

From an implementation standpoint, there are a large number of options for the structure of this design and the selection of the study subjects. The sampling methods (i.e., the probability based design specifications such as stratification and over-sampling) would likely be identified and developed in a centralized manner to ensure consistency of the sample selection process across the entire cohort. This design development would necessarily begin with the definition of the sampling frame(s) (or stages of sampling such as PSUs) for which there are again several options. In fact, any of the three sampling frames, or some combination of them, discussed above (see Section 2.1) would be amenable to a probability based design. For example, for the Centers based model a large number of qualified medical centers could be identified as the PSUs (or the regions covered by the Centers could be identified as the PSU), and a random sample of those PSUs could be selected (e.g., using weights proportional to the patient population for each center). Then, within each PSU, a random sample of pregnant women (or perhaps women of childbearing age) that visit those centers could be selected. Of course, a similar procedure could be used in the Physicians office model to select a random sample of pregnant women, and/or women of childbearing age (with the PSU being a geographic region and the Physicians offices being sampled within each of these PSUs).

While these two sampling frames could be adapted to a complete probability based design for the selection of PSUs, they may involve a larger degree of non-coverage of the population than the Household model (see Section 2.1). For example, there are geographic regions of the United States that are not "covered" by a large medical center, and there are segments of the population that do not visit Physicians offices and/or large medical centers prior to giving birth. For this reason, and given the research objectives of the NCS, perhaps the most advantageous sampling frame for a complete probability based sample would be the Household Model in which the PSUs are identified as geographic regions, such as counties, so that a nationally, or at least geographically, representative sample could be obtained and some of the issues with non-coverage of the population can be alleviated (see Section 2.1). (It should be noted that probability based sampling, where the likelihood of selection is proportional to population size and without stratification, would likely result in a sample with a large proportion of urban geographic areas and a large proportion of coastal geographic areas since most of the population lives in these areas.) An example Household model sampling scheme was specified in the WESTAT report (2002) and involves selection of a random sample (with probability proportional to size) of counties in the United States, a random sample of census tracts in each county, and a random sample of women of childbearing age in each census tract (e.g., by randomly selecting households in each census tract and enrolling any women of childbearing age that live in those households). This represents one means of obtaining such a probability based sample – but has raised some concern due to the large number of PSUs recommended (800) and the large number of households that would need to be screened in order to yield the resulting cohort of 100,000 live births (which both might be operationally infeasible).

Organizationally, once the sampling frame and the methods for sampling have been identified, the methods and procedures for following those subjects and collecting the necessary study information can be defined. Of course, if either the Physician's

Office or the Centers Sampling Frames are selected, then the data collection organizations would consist of the Physician's Offices or Centers that were selected in the sample (as well as any other data collection organizations that are necessary in collecting the required information). In other words, it is likely that a combination of medical centers and, if deemed necessary to collect the required information, contract research organizations would be employed if either the Physician's Office or Centers Sampling Frame were selected. Thus, we focus on the implementation scenarios for the case where the Household Sampling Frame is selected. For this case, there are likely two implementation scenarios. In both scenarios the sample of study subjects, or at least the methods for sampling, would be identified in a centralized manner to ensure consistency of the sampling process. Once the set of study subjects has been identified, the data collection for those study subjects could occur by two basic mechanisms. First, a single (or very few) contract research data collection organization (perhaps the organization that identified the subjects) could be responsible for recruiting the study subjects and collecting all the data from those study subjects. Of course, identifying an organization capable of performing this task on a national scale may be difficult (or impossible), and, thus, it is envisioned that if this option were implemented these organizations would necessarily subcontract a variety of agencies (such as the subject's physician office) to assist in implementing the data collection protocols.

Perhaps, a more likely study organizational structure would be to involve a variety of data collection organizations in the study. For example, if PSUs are defined as counties, and the sampling of PSUs occurs with probability proportional to population, then a large proportion of the sampled PSUs will coincide with urban areas that are likely in close proximity to a major medical center (i.e., a community or university medical center). Those PSUs that coincide with a Center could be assigned (for the data collection aspects) to their corresponding Center, and those PSUs for which there is no Center in close proximity could be assigned to some other data collection organization, such as a Physician's Office, or another contract research organization that is capable of conducting the required data collection. Alternatively, competitive bids could be used for Centers to be established in these presumably rural areas (e.g., such as the Children's Health Center operated by UC Berkeley in the Salinas Valley Farm Community). Logistic issues with data combination, standardization of data collection and storage methods, and timely data availability would all need careful thought and consideration in order for this type of structure to work efficiently; however, the utility of having such a community-based study may overweigh the logistic difficulties with using these multiple data collection organizations.

Statistically, this type of design *potentially* allows the selection of a random sample of the *currently available* population of interest. As discussed previously, a random or probability based sampling approach is the primary means of selecting a sample that is representative of the population, and is free from scientific criticism of the selection mechanisms (of course there may be criticisms of the method for accomplishing the probability sampling, such as the number of sampling stages, the clustering effects, etc.). In other words, in a probability based sample, the distribution of characteristics in the sample (appropriately weighted) should reasonably match the distribution of

characteristics in the population from which the sample was drawn. These characteristics include (but are not limited to) the relationships between important health outcomes of interest and the range of exposures and risk factors. Since the probability based design allows calculation of the sample weights for each study subject (i.e., all study subjects represent a known portion of the population), generalization of the characteristics of the sample to the entire population can be accomplished in a scientifically and statistically rigorous manner. This highlights one of the major advantages to probability based sampling, namely the ability to generalize inferences to the entire population in a statistically justifiable manner; however, recall that this generalizability is not assumption free (i.e., assumption of no noncoverage and no nonresponse in a probability based sample) – yet, these assumptions are testable, given limited information available about nonrespondents and checks for noncoverage of the population.

Actually implementing this design is likely to be somewhat difficult and costly in order to ensure the selection of a valid probability based sample (i.e., that allows us to capitalize on the whole point of performing this type of sampling). The primary obstacle involves the difficulty of recruiting and retaining individuals who may not be interested in participating in a study as large and burdensome as the NCS. In other words, by conducting probability based sampling and essentially asking "random" people to participate in the study, there is a higher likelihood of high nonresponse and attrition rates, or at least higher nonresponse and attrition rates then a convenience sample would entail (this is especially the case with the Household model). Of course, incentives can be given and experienced interviewers can be used in recruiting these individuals so that the nonresponse rates and study subject attrition rates are minimized. However, these types of measures will affect the overall cost of the study, and, thus may affect other important components of the study such as the ability to collect expensive but important environmental exposure information. (See white paper on retention and recruitment for additional discussion of this issue.)

Additionally, it should be noted that if a valid probability based sample is not obtained, then the advantage of conducting probability based sampling may be lost. In other words, if the probability based sampling results in both high nonresponse rates and high study subject attrition rates, and the nonresponse and subject attrition are not random, then the resulting cohort may not be able to be considered random and may end up being similar to a convenience sample. So, the question becomes, what are "high" nonresponse rates and "high" attrition rates, and what kinds of response and retention rates are necessary to ensure a valid probability based sample. Unfortunately, there is little guidance on this issue at the current time.

There are statistical methods for adjusting and accounting for nonrandom nonresponse and study subject attrition (e.g., see Chambers and Skinner, 2003) in a probability based sample. Certainly, some of these methods can and should be applied to the data collected from the NCS if a probability based sampling scheme is utilized. Additionally, it may be the case that some of these methods can be utilized to adjust and account for the nonrandomness associated with a purposive or convenience sample,

especially if a small validation sample that is probabilistically sampled from the population is available (see discussion in Section 2.2 on validation subsamples).

The discussion of a probability based sample has focused on the use of a multi-stage design in which a stratified sample of PSUs are selected at the highest level of hierarchy using probability proportional to size sampling. An important requirement of this type of design is the identification of "certainty PSUs" from within each strata, when the candidate PSUs within a strata are not relatively uniform in size. Certainty PSUs enter the sample with a probability of 1, and represent the largest PSUs, which, using probability proportional to size sampling, have the highest likelihood of entering the sample. Certainty PSUs are identified using an iterative sequential selection process that considers the size of the largest PSU remaining within a stratum relative to the average stratum size and the total number of PSUs that are to be sampled from that stratum. As certainty PSUs are identified, they are conceptually removed from their respective strata, with the next largest PSU being subsequently considered for certainty selection. Once the selection of certainty PSUs is completed, the remaining PSUs for the sample are then selected using probability proportional to size sampling. Once all the PSUs are identified, it is assumed that the sampling that occurs within each PSU is approximately uniform (with a relatively similar number of segments and subjects sampled at random within each PSU). Obviously, there are many significant and labor intense steps that need to be implemented to ensure appropriate within-PSU sampling occurs in such a uniform manner.

NHANES is an example study that followed a multi-stage design, in which counties served as PSUs, and 13 of the 81 counties selected into the sample were designated as certainty PSUs. Following the initial guidance outlined in the Westat (2002) report, a similar design for the NCS involving the selection of 800 counties would have over 280 certainty PSUs based on the use of 24 strata (4 regions, 3 levels of socio-economic status, and 2 levels of urbanization).

In terms of the implication for sampling design, the concept of certainty PSUs within a multi-stage design is very important because it provides for the selection of multiple PSUs into the sample with probability one (including methods for weighting and analyzing the resulting data). In this case (complete probability based design), the certainty PSUs are selected on the basis of size. However, in later sections of this report that introduce hybrid design options, we will explore the potential for other reasons for the selection of certainty PSUs (e.g. creating certainty PSUs out of those areas that contain a qualified center(s) for excellence). This expansion of the rationale for selecting certainty PSUs may also involve relaxing the assumption of uniform sampling within a PSU to make this a viable option for these hybrid design options (e.g. designing a disproportionate amount of the sample being contained within the certainty PSUs). Further research will likely need to be conducted to determine the impact that these changes will have on the ability for these hybrid design options to meet the scientific objectives of the NCS.

B2-3.2 QUOTA/CONVENIENCE SAMPLING OF PSUS AND PROBABILITY SAMPLING WITHIN PSUS

In this design, a probability based sample is obtained within each of the selected PSUs in the study, and the PSUs are selected on a quota or convenience basis (in this case we envision the PSUs being geographic regions). For example, a convenience-based selection of a set of counties around the nation (e.g., counties that correspond to a Center) could represent the PSUs in the sample. Then, study subjects in each PSU (or each county) would be sampled on a probability basis (e.g., by selecting households in each county or selecting women of childbearing age that are in proximity to or have a relationship with the Center or associated physicians). Basically, this design attempts to capitalize on the gain in implementation feasibility that could be realized by selecting PSUs that correspond to a Center (i.e., selecting PSUs that are convenient to sample) while maintaining some aspect of probability based sampling. Figure 3 highlights the sampling strategies that are consistent with this design option.

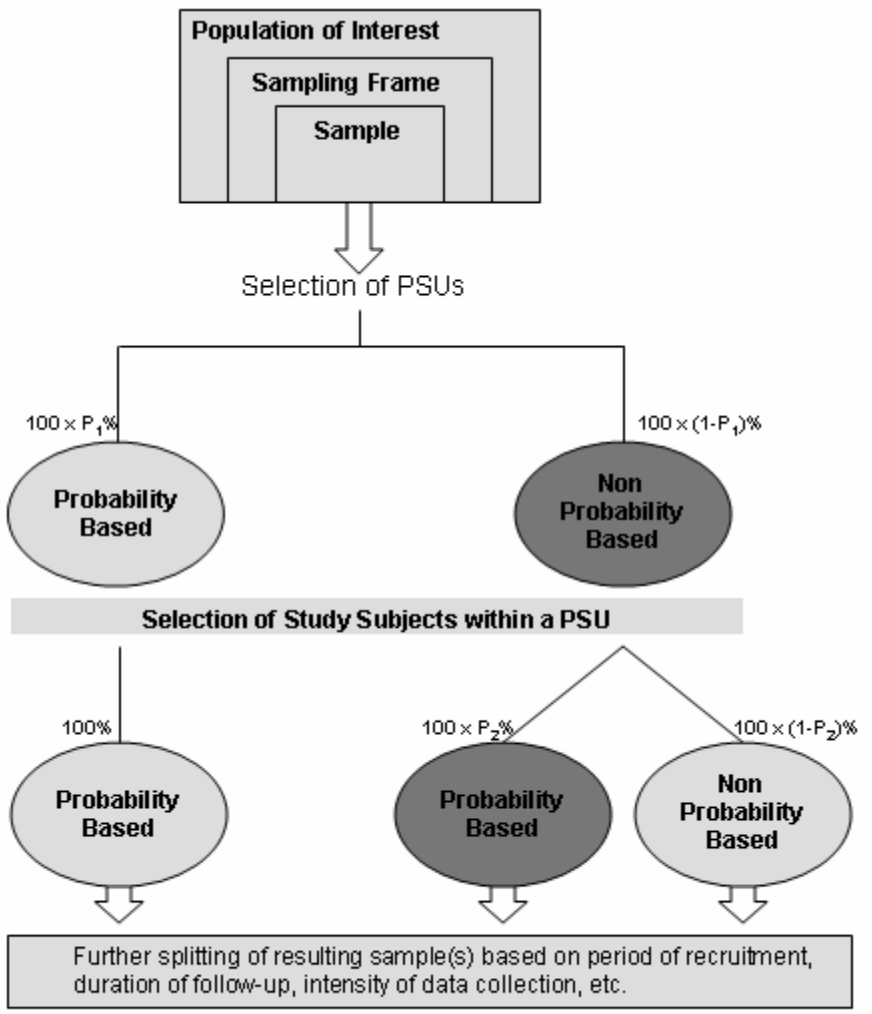

Figure 3. Highlighted Sampling Strategies Consistent with Quota/Convenience Sampling of PSUs and Probability Sampling with PSUs (P_1=0, P_2=1)

From an implementation standpoint, again a variety of options for the structure of this design and the selection of the study subjects could be utilized. The sampling methods (i.e., the convenience selection of the PSUs and the probability based selection of subjects within these PSUs) would likely be identified and/or developed in a centralized manner to ensure consistency of the sample selection process across the entire cohort. As before, this would begin by identifying the most likely sampling frame. Any of the three sampling frames discussed in Section 2.1 would be amenable to this sampling method; however, we envision that if this sampling method were utilized it would be done so in order to select PSUs that geographically coincide with preferred Centers or organizations (since this would be a convenient way of selecting PSUs). In other words, the likely reason for selecting PSUs on a convenience or quota basis would be so that the selected PSUs correspond to a qualified Center. The corresponding Centers would then be responsible for implementing the study in their PSU. Of course, the PSUs would also

be selected so that geographic and demographic diversity exists in the study, and so that any desired stratification of the sample and any quotas of interest could be satisfied. When regions that do not correspond to a Center are selected to meet a desired quota, either a "new" Center would need to be established (e.g., see discussion in Section 3.1) or an alternate contract research organization could be contracted to complete the data collection protocols for that region.

The probability based selection of study subjects within each PSU would likely occur by one of two means. The first option, and the most straightforward option, would be to randomly select women of childbearing age and/or pregnant women from the patient lists for the Centers that correspond to the selected PSUs. This random selection could be done in a stratified manner to ensure inclusion of different ethnic groups, different socio-economic groups, and any other important strata that should be included in the study; however, the degree to which the Centers patient population does not "match" the population of the selected PSU may be an issue and will introduce non-coverage of the PSUs population (i.e., anyone in the PSU that is not on the Centers patient list would be excluded from the sample). For this reason, perhaps a more rigorous sampling frame would be to adopt the Household model within the PSUs by selecting households in each PSU, and recruiting the women of childbearing age in each selected household into the study. The Center would then be responsible for following the women selected within their PSU, and collecting the necessary questionnaire, biological, and environmental data. While not a "hybrid design", this option could use a combination of sample frames to supplement the Center patient-based population with other clinics or household samples.

Organizationally, and as mentioned above, it is envisioned that either a Centers based data collection organization or a combination of a variety of data collection organizations would be the means of collecting the necessary information for the study. Presumably, to the degree that the Centers are not qualified to collect the needed data (e.g., not qualified to collect household environmental samples for toxics analysis), other data collection organizations would need to be involved (either through direct contracts or subcontracting mechanisms).

Statistically, this design has some of the nice characteristics of a complete probability based design in that it involves a probability based sample from each PSU so that sample weights for each subject can be calculated (within a PSU that is). In other words, this design potentially offers a representative sample for each of the PSUs. However, note that it does not involve probability based selection of PSUs, and, thus, may not provide a nationally representative sample as this depends on the representativeness of the selected PSUs. From a rigorous statistical standpoint, this implies that inferences would only be generalizable to the population of subjects in the selected PSUs, and would not be generalizable to the population in all PSUs. As is the case with any convenience or quota-based sample, if the selected PSUs are reasonably assumed to be representative of the entire population of PSUs then generalization to the entire population, while not rigorously justifiable, could be considered reasonable. Study leaders will need to assess whether the utility of being able to select the PSUs on a

convenience basis (and thereby select the Centers), and the potential that a convenience based selection of PSUs could be representative of all PSUs, overcomes the necessity of obtaining a probability based sample of PSUs.

As in the previous probability based design (see Section 3.1), actually implementing this design may share some of the same difficulties. First, it may be difficult and costly to ensure the selection of a valid probability based sample within each PSU (i.e., that allows us to capitalize on the whole point of performing this type of sampling). Again, the primary obstacle involves the difficulty of recruiting and retaining individuals who may not be interested in participating in a study as large and burdensome as the NCS (i.e., asking "random" people to participate may result in high nonresponse rates and/or high attrition rates). However, perhaps the involvement of a community-based Center, with the doctor-patient relationships that will occur naturally under this model, will encourage higher response rates and better retention rates. If the Household model is used, a second possible obstacle involves determination of the appropriate organization to conduct the within PSU selection of a probability based sample. For some Centers, and perhaps for many Centers, a probability based sampling of Households in their corresponding PSU may be a challenging endeavor, and, thus, it may be necessary to provide guidance, expertise, and/or experienced personnel to assist Centers in performing such a task.

3.3 COMPLETE CONVENIENCE OR QUOTA SAMPLING

In this design, all aspects of the sample selection process are done on a convenience or quota basis. The PSUs (again assume the PSUs are counties) could be sampled based on convenience with quotas ensuring geographic representativeness and regional demographic representativeness (e.g., urban and rural counties should be selected). Within each PSU, study subjects are identified also based on convenience with quotas again ensuring representativeness of several person-specific demographic characteristics such as socioeconomic status and/or ethnicity. This design option, like the previous option, is likely a Centers type of design where the convenience sampling of PSUs would select a set of geographic regions around the nation that correspond to a Center. Then, within each PSU, the study subjects would be selected on a convenience basis, perhaps by recruiting pregnant women or women of childbearing age that visit the Centers and volunteers for the study. Figure 4 highlights the sampling strategies that are consistent with a complete convenience or quota sampling design, and volunteers for the study.

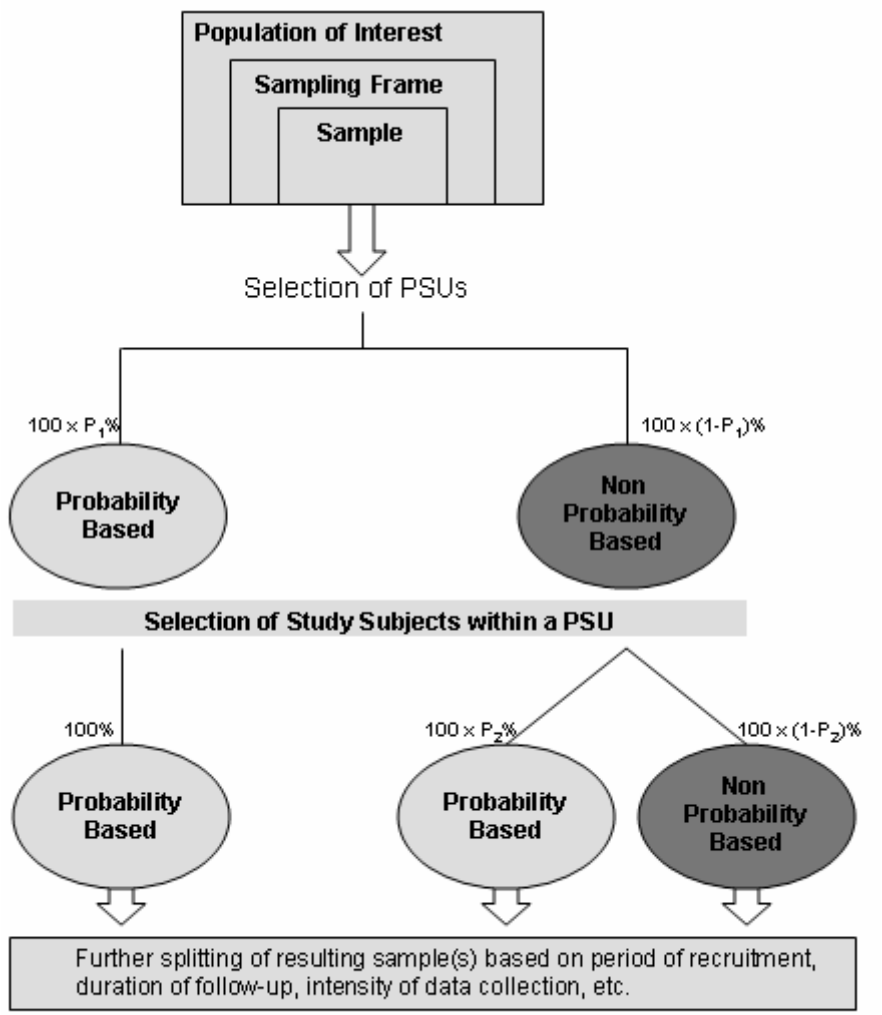

Figure B2-4. Highlighted Sampling Strategies Consistent with Complete Convenience or Quota Sampling (P_1=0, P_2=0)

From an implementation standpoint, this is likely the easiest design to implement since all selection is made on a convenience basis (with perhaps some quotas to ensure geographic and demographic diversity). The Centers that correspond to the selected PSUs would be responsible for implementing the study, recruiting their own convenience sample of study subjects, following those study subjects, and collecting the necessary data (if the Center was not qualified to collect all the needed data, for example environmental data, then another data collection organization could be identified to conduct this type of data collection). Since the sample is simply a convenience sample, the sampling frame would presumably be all women of childbearing age (or pregnant women) that visit the selected Centers or visit and volunteer. Additionally, only subjects that are relatively willing would be convinced to enroll in the study (or only those who volunteer) and presumably this would reduce the degree of study subject attrition without

implementing the costly incentives that might be necessary in a complete probability based design.

From a statistical viewpoint, however, this design does not offer the statistical rigor that is associated with a probability based design and would not allow the explicit calculation of sample weights for each study subject (i.e., the number of people in the larger population that are represented by the study subject). In other words, generalization of the characteristics of the study population to the entire sampling frame population would necessarily involve stronger assumptions than if a probability based sample had been obtained (see discussion in Section 2.2). However, one approach to generalizing results from a convenience sample back to the population of interest involves a thorough assessment of key characteristics of the sample, including demographic characteristics, covariates of interest, potential confounders, and other factors that might differ between the sample and the population of interest with respect to the scientific hypotheses of interest in the NCS. Based on this assessment, a predictive model may be generated to account for any differences between the sample and the population of interest, so that the results of the NCS could be generalized to not only to the original population of interest for the study (i.e. children born in these areas or in the U.S. between 2006 and 2009), but also other populations of interest (e.g. children born long into the future, when the demographic, social, and exposure characteristics of the U.S. population is likely to be vastly different). In concept, this approach is not dissimilar to developing post-stratification weights for the individuals that enter the convenience sample – however, the use of a convenience sample still suffers from a valid criticism that study subjects are not chosen at random, and not everyone in the original population of interest (or sampling frame) has a known positive probability of being selected into the NCS sample. Unlike designs with some component of probability sampling, no comparisons can be made between participants and non-participants, given the lack of information on those who did not volunteer. Similarly, there is no way to compare how exposure-response relationships differ between the participants and the population.

One prominent study that has used this type of design very successfully is the Framingham Heart Study, which is directed by the National Heart, Lung, and Blood Institute (NHLBI), and has the goal of identifying the common factors or characteristics that contribute to cardiovascular disease (CVD). This study enrolled individuals between the ages of 30 and 62 that lived in Framingham, Massachusetts and had not yet developed any symptoms of CVD. While based on a cohort of individuals that are primarily white and live in Framingham, Massachusetts (i.e., cohort is not nationally or ethnically representative), the utility of this study has far exceeded its goal of identifying factors that contribute to cardiovascular disease (e.g., identified high blood pressure, high blood cholesterol, smoking, obesity, diabetes, and physical inactivity as CVD risk factors) and claims to have revolutionized the way that the medical community thinks about, treats, and prevents CVD. The major CVD risk factors identified by the FHS have been shown in other studies to apply almost universally among racial and ethnic groups. While Framingham illustrates the potential success of such a design, it should be noted that

there are many examples of studies with this design that could be used to illustrate shortcomings as well as successes.

B2-3.4 MIXTURE OF CONVENIENCE AND PROBABILITY SAMPLING OF PSUS AND WITHIN PSU PROBABILITY SAMPLING

The first hybrid sampling design option that we consider corresponds to a design for which all within PSU sampling is conducted on a probability basis (i.e., P_2 in Table 2 is 100 percent), and some proportion of the PSUs (e.g., 80 percent) are selected on a convenience basis while the remaining PSUs are selected on a probability basis (e.g., P_1 in Table 2 is 20%) Essentially, the idea of this hybrid model is to select the majority of the PSUs on a convenience basis, and then conduct a probability based selection of the other PSUs so that all individuals in the sampling frame have a non-zero probability of selection and to "fill-in" those areas/regions that are not represented in the convenience based sample of PSUs. The rationale is that a large portion of the sampling frame population could be covered using a convenience based selection of PSUs (e.g., by selecting those PSUs that correspond to the largest populations), and those PSUs that are not represented could be included using a probability based selection. Presumably, since all PSUs, and therefore all subjects, in the sampling frame, would have a known non-zero probability of selection, sample weights for each study subject could be computed and utilized in any appropriate analyses (see below). Figure 5 highlights the sampling strategies that are consistent with this hybrid design option.

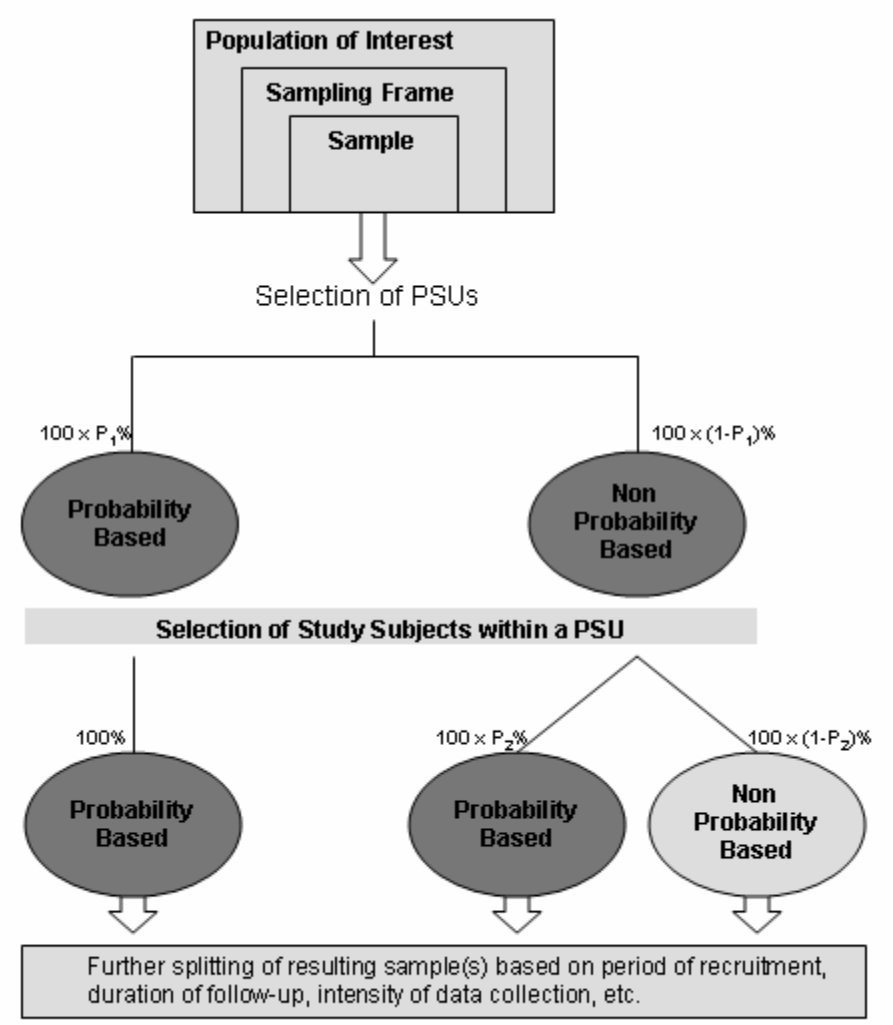

Figure B2-5. Highlighted Sampling Strategies Consistent with a Mixture of Convenience and Probability Sampling of PSUs and Within-PSU Probability Sampling ($0 < P_1 < 1$, $P_2=1$)

The options for implementing this design would be very similar to the options described for implementing the complete probability design (Section 3.1) and/or the design that selects PSUs on a convenience basis and within PSU samples on a probability basis (Section 3.2). As described in those sections, any of the three sampling frames could be utilized and the organizational structure of the study would likely occur using a combination of Centers and other data collection organizations (e.g., Centers for some PSUs, and other data collection organizations, such as Physician's offices or other capable institutions for other PSUs). The convenience selected portion of the PSUs would correspond to regions or PSUs deemed important, such as large metropolitan areas, high population PSUs, and/or PSUs that coincide with one of the premiere Centers (again, this depends on the definition of a PSU); whereas, the probability based selection of PSUs would be conducted in order that all PSUs have a non-zero chance of selection

(and therefore all study subjects in the sampling frame have a non-zero probability of selection).

As in Sections 3.1 and 3.2, the probability based selection of study subjects within each PSU would depend on the selected sampling frame and on the definition of the PSU. If the Household model sampling frame is utilized, then PSUs are likely defined as geographic regions, and for each PSU a probability based sample of households would be obtained in an appropriate manner (see Section 3.1). On the other hand, if the sampling frame is either the Centers model, or the Physician's Office model, then the probability based sample of pregnant women (or women of childbearing age) is obtained from the patient lists for those Centers or Physician's Offices.

Statistically, this design retains some of the preferred characteristics of a complete probability based design in that it involves a probability based sample from each PSU so that sample weights for each subject can be calculated. However, an open question with this design is what proportion of PSUs should be sampled probabilistically, and what is the effect (in terms of the design effect and the sample weighting) of allowing a portion of the PSUs to be selected on a convenience basis? Perhaps, the PSUs selected on a convenience basis can essentially be viewed as certainty PSUs (i.e., have probability of selection one), and, thus, this design would be similar to a complete probability based design that has a fixed number of certainty PSUs (certainty PSUs are discussed in more detail in Section 3.1). Further research into these questions is necessary.

As in the previous probability based designs, actually implementing this design will share some of the same difficulties (see Section 3.1 or 3.2); however, the convenience based selection of a significant portion of the PSUs may alleviate some of the difficulty in conducting a complete probability based design (e.g., by limiting the number of disjoint sampling regions and/or by selecting some PSUs that correspond to a Center capable of developing support for, and implementing, the study in their geographic area) while still retaining the attractive properties of a probability based design (e.g., unbiased samples, ability to calculate sample weights, etc.). In other words, this hybrid design represents a design that attempts to capitalize on some of the advantages of the design described in Section 3.2 while maintaining the statistical properties of a complete probability based design (Section 3.1). As mentioned above, statistical properties and the degree to which the convenience sampling of a portion of the PSUs affects the "optimal" characteristics of the design would need to be further explored for this design.

Finally, as in the above models, a plausible organizational structure for capturing the data in this model involves a combination of data collection organizations, such as medical Centers, Physician's Offices, and/or other institutions that have the ability to conduct the necessary data collection protocols. This design would most likely involve the use of Centers for data collection in the convenience PSUs and the use of contract research organizations or physician offices for data collection in the probability PSUs with other subcontracting or contracting arrangements possible similar to those discussed in Section 2.3.

B2-3.5 MIXTURE OF CONVENIENCE AND PROBABILITY SAMPLING OF PSUS AND WITHIN PSUS

A second hybrid sampling design that we discuss corresponds to the setting where some proportion of the PSUs (e.g., 80 percent) are sampled on a convenience basis, and some proportion of the within PSU sampling (e.g., 50 percent) is conducted on a convenience basis (see Table 2). All other sampling is conducted on a probability basis. As in the previous hybrid design, this design attempts to capitalize on the advantages associated with probability based sampling (e.g., unbiased samples, ability to calculate sample weights, etc.), while also capitalizing on some of the advantages associated with convenience sampling (e.g., lower nonresponse rates, presumably higher retention rates, lower cost, ability to select PSUs and subjects that are "easy" to follow, etc.). Figure 6 highlights the sampling strategies that are consistent with this second hybrid design option.

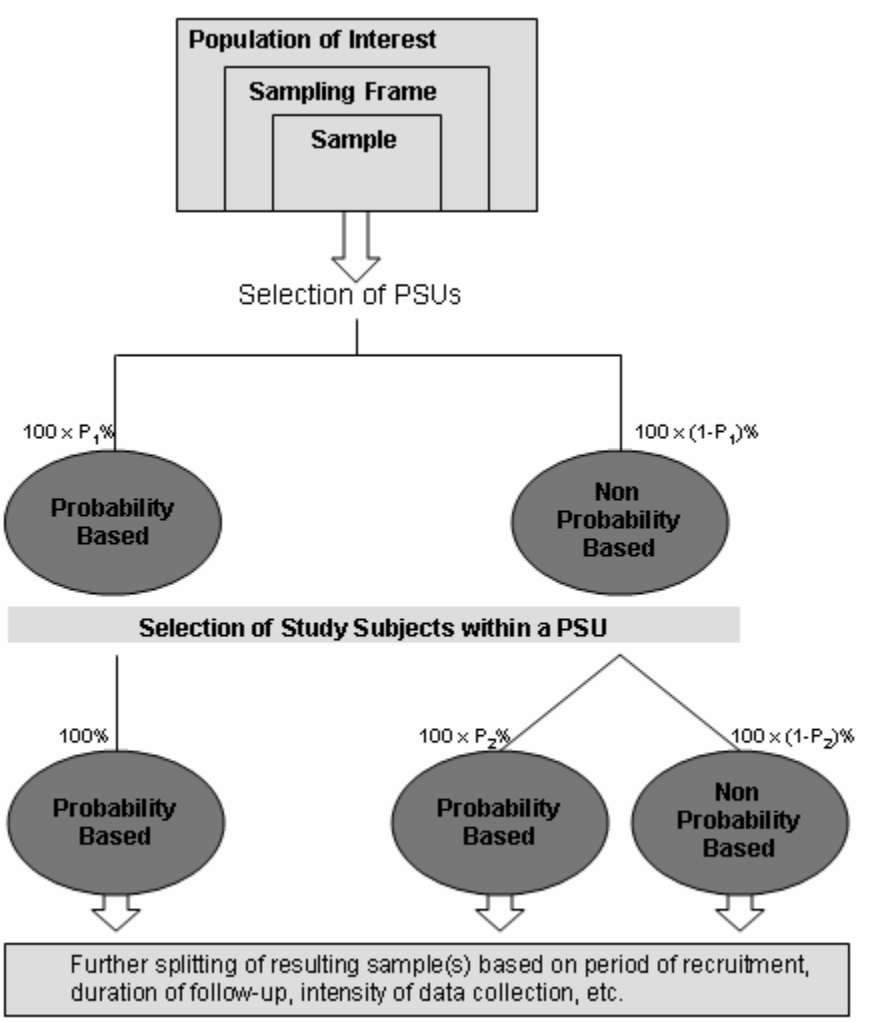

Figure B2-6. Highlighted Sampling Strategies Consistent with a Mixture of Convenience and Probability Sampling of PSUs and Within PSUs ($0 < P_1 < 1$, $0 < P_2 < 1$)

One argument for this type of design involves the idea of a validation set of individuals that could be used to "adjust" for sample biases resulting from the convenience sampled individuals (see discussion in Section 2.2). Essentially, within each PSU the idea would be that those subjects selected on a probability basis could be compared to the convenience selected subjects to determine if the convenience selected subjects are a "biased" sample from the population. If there is a bias, then that bias could be estimated and accounted for appropriately. Of course, the statistical methods for combining and comparing the information in the convenience sampled subjects and the probability sampled subjects would depend on a variety of factors (e.g., the hypothesis of interest, possible confounders of interest, etc.), and would need to be further explored, reviewed, and developed before we can recommend the appropriateness of this type of validation sampling for the NCS.

It should be noted that a number of questions would need to be answered before the "optimality" of this design is determined. For example, what is the optimal proportion of PSUs that should be selected probabilistically, and what are the optimal proportions for sampling within each PSU (i.e., what is P_1 and P_2 in Table 2). Certainly, there is some tradeoff, in terms of the design effect, power for statistical analysis, and sample weighting, for being able to select a portion of the sample on a convenience basis. In other words, the balance between cost, feasibility, and ability to meet the scientific objectives of this study if this type of design is used would need to be further explored (as is the case with all of the design options).

Implementation options for this design correspond to the implementation options described previously in Section 3.4, and, thus, are not repeated here. No known studies have used this specific type of design, where some elements are selected probabilistically and some are selected on a convenience basis; however, since this design can be thought of as a type of probability based design (with the convenience selected elements having probability one of being sampled), any study involving probability based selection of subjects could be referenced as an example.

B2-3.6 MULTIPLE COHORTS DESIGNS

The multiple cohorts model attempts to select two (or more) cohorts to meet the NCS study objectives. In concept, none of the independent cohorts on their own will be able to meet the scientific objectives of the NCS, however, by combining information across the multiple cohorts, all objectives of the NCS would be met in perhaps a more efficient manner. Essentially, the idea for the multiple cohorts design is a generalization of the ideas presented in the hybrid designs of Sections 3.4 and 3.5, where a portion of the cohort is selected on a probability basis and a portion is selected on a convenience basis. In those sections, once the selection of study subjects was accomplished, presumably all subjects would be followed for the same period of time and would undergo the same data collection protocols. In the multiple cohorts design we relax this assumption by allowing two additional options: 1) allow some subjects to undergo more detailed sampling (i.e., provide more detailed exposure, questionnaire, and behavior information) than other subjects, and 2) allow some subjects to be sampled and/or followed during different time periods while still ensuring sufficient sample size and power to meet all NCS scientific objectives. The main goal of these options would be to limit the effect of attrition and nonresponse in the probability sampled subjects (e.g., by requiring a lower degree of subject burden) so that a valid probability based sample is obtained, while capitalizing on the interest and motivation to participate in the study for those subjects sampled on a convenience basis (e.g., by obtaining more detailed information over a longer time period for these subjects). Basically, the idea is to use probability based sampling on a cohort that is in some sense either "easy" to sample probabilistically, or is "easy" to follow, and use convenience based sampling on a cohort of subjects that have a higher interest in the study and will be willing to undergo more detailed sampling and subject burden. Figure 6 highlights the sampling strategies that are consistent with multiple cohort design options.

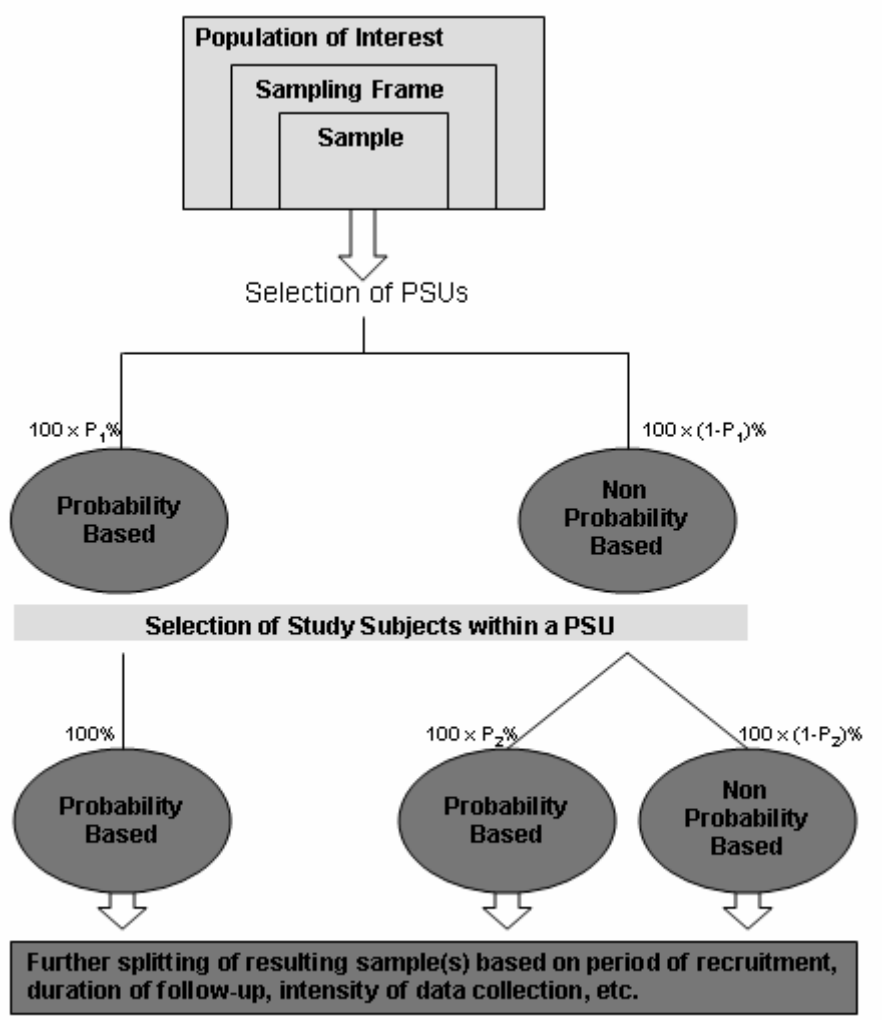

Figure B2-7. Highlighted Sampling Strategies Consistent with Multiple Cohort Designs ($0 < P_1 < 1$, $0 < P_2 < 1$)

In this sampling framework, one possible approach would be to enroll a pre-natal cohort primarily via convenience sampling (with a small validation subsample enrolled via probability based sampling to account and adjust for any sampling bias) and enroll a post-natal cohort[1], perhaps with pre-natal exposure information assessed retrospectively, via a probability based sample of recent live births (e.g., births reported in the Birth Registration System). The fraction of the population enrolled in each cohort would be optimized to ensure maximum power across NCS core hypotheses, with the post-natal

[1] While we recognize that at face value, the idea of a post-natal cohort appears to violate the spirit of the NCS in terms of providing pre-natal (or pre-conception) data, it could provide useful information in a resource efficient manner as a probability based sample for hypotheses that do not rely upon prospectively collected measures prior to birth, for general assessment of these hypotheses or as part of a validation sub-sample.

cohort providing a much more resource efficient mechanism for achieving a probability based sample.

On the implementation side for this option, the issues of selecting a sampling frame are somewhat alleviated since the pre-birth cohort is selected on a convenience basis and the after-birth cohort is selected from the Birth Registration System, which is assumed to cover the population of interest. However, for the after-birth cohort, there remains the difficulty of subject nonresponse and subject attrition that is presumed to be more prevalent in a probability based sample. Additionally, there remains the determination of what type of organizational structure would be most promising for implementing this design. For the convenience selection of a pre-birth cohort, we envision a set of Centers selecting pregnant women or women of childbearing age from their lists of patients, and following those patients to collect the necessary information (using subcontractor help as needed). For the after-birth cohort the range of options for the organizational structure matches many of the options described previously.

Alternatively, if it is determined that all (or most) subjects should be identified in the pre-natal period (i.e., the distinction between a pre-natal cohort and a post-natal cohort is not desired), a multiple cohorts structure could involve the probability based selection of a "low burden" cohort, for whom less detailed information would be collected, and the convenience (or quota) based selection of a "high burden" cohort, for whom detailed information would be collected. Conceptually, this multiple cohort design differs from designs that consider how the burden of detailed data collection (e.g. invasive biological samples, time consuming activity diaries, collection of household environmental samples, etc.) can be spread across the cohort at random, so that every participant completes a "core" data collection protocol, with only a few being severely burdened (and incented) to complete the more demanding protocols. In the multiple cohort design, the more burdensome data collection activities would not be assigned at random. Rather, they would be assigned in some purposive manner to those subjects who are most likely conform to the protocol without dropping from the remainder of the study, such as those in a volunteer sample.

Conceivably, this type of design could limit the effect of attrition and/or nonresponse in the probability based sample, since sampled individuals would not be required to undergo a high level of burden in order to participate in the study (or perhaps not be expected to participate in the study beyond some point), while retaining the ability to collect any desired detailed information for a portion of the cohort. Of course, we note that this design could be combined with the previous idea of selecting a pre-natal cohort and a post-natal cohort (e.g., by selecting four cohorts consisting of pre-natal and post-natal cohorts for both the high burden and low burden groups).

For this option, the sampling frame likely consists of either the Centers sampling frame or a combination of the Centers sampling frame and the Household frame (e.g., with subjects selected on a convenience basis selected from the Center patient lists and subjects selected on a probability basis selected using the Household model).

Additionally, the data collection mechanisms for this option also likely consist of a combination of qualified medical centers and other contract research organizations.

From a statistical viewpoint, this class of designs offers a promising avenue for addressing and generalizing the results of the study to the sampling frame population, since at least a portion of the cohort would be sampled probabilistically, and, thus, all individuals in the sampling frame would have a non-zero chance of selection in the study. For those individuals that are sampled on a convenience basis, the probability of selection is one, and for those individuals sampled on a probability basis the chance of selection is some non-zero value. In this sense, these designs are similar to the hybrid designs described in Sections 3.4 and 3.5, or any design that incorporates some form of probability based sampling. As indicated previously, the main difference, and attractive property of this design is that it 1) allows some subjects to undergo more detailed measurements (i.e., provide more detailed exposure, questionnaire, and behavior information) than other subjects, and 2) allows some subjects to be tested and/or followed during different time periods. Both of these additional options may prove to be efficient and important methods for conducting this study in order to meet the scientific objectives.

Of course, this flexibility will likely incur some cost in terms of implementation difficulty. One of the primary difficulties is in defining the different characteristics of varying measurement methods, and in developing appropriate methods (statistical methods) for combining these different measurement methods across the cohort. For example, it may be the case that for some individuals (i.e., the low burden individuals) only questionnaire information related to a particular hypothesis is available, whereas, for other individuals, both questionnaire information and biologic information (such as a biomarker of exposure) is available. To perform the appropriate, accurate, and most efficient statistical analysis of the data, the full information across all subjects would necessarily be combined through statistical modeling methods. Statistical methods for adjusting for different measurement methods in the different cohorts to allow combining data across cohorts will need to be explored, but precedents exist (e.g., Bayesian methods, latent variable methods, imputation techniques, etc.); however, application of these methods to this situation would benefit from further investigation so that appropriate design decisions can be made. In particular, this could (should) involve the following:

- Investigation of relevant statistical methods for combining information from various sources.
- Investigation of the appropriate types of information for the two cohorts.
- Determination of optimal sample sizes for the multiple cohorts.
- Investigation of effective sample size for each cohort.
- Investigation of power across the multiple cohorts to address the NCS core hypotheses.
- Any other relevant topics that may be identified through the course of these investigations.

Like the hybrid designs described in Section 3.4 and 3.5, another potential difficulty in this design is the determination of the appropriate proportion of the cohort that should be sampled on a probability basis and the proportion that should be sampled on a convenience basis. These "optimal" proportions will certainly influence the power and efficiency of the design, the sample weighting of the design, and may affect the cost, feasibility and ability of the design to meet the scientific objectives of the study. These issues would need to be further explored for this design.

B2-3.7 SIBLING COHORTS

A final design consideration that was discussed in Section 2.2, and is revisited here, entails the selection of what we refer to as a "sibling" cohort. Essentially, the idea is to select an "original" cohort, using any of the design options specified above, that consists of a large portion of the desired 100,000 children in the study. Then, from this "original" cohort, a cohort of siblings is selected that consists of *younger* siblings of children in the "original" cohort. In other words, a set of families with children in the "original" cohort (and their parents) would be followed, and if an additional pregnancy occurs in that family (perhaps limited to a certain time span that corresponds to the study enrollment period), then the resulting sibling is enrolled in the study. This would result in a "sibling" cohort of subjects for which pre-conception exposure measurements and behavior information would be available, and provides a promising means of recruiting at least a portion of the full cohort prior to conception.

In terms of the statistical properties of this design, further work is necessary. As described above, any of the previous design options could be used in selecting the "original" cohort of individuals, and, therefore, the statistical properties of those options would pertain to this "original" cohort. However, the statistical properties of the full cohort in this case would need to be further investigated to determine the design effect of selecting two subjects from the same family (i.e., presumably highly correlated subjects), to determine the proper proportion of individuals in each of the cohorts, and to determine the approximate number of women needed to provide the desired sample size for the "sibling" cohort. Additionally, there would be the need to understand the potential bias in the cohort of individuals with pre-conception measurements given that this cohort would include a very small proportion (if any) of first-born children, and would only include children of mothers who had already given birth to children enrolled in the study.

Since this idea is a potential add-on option that could be combined with any of the design options described above, the implementation options and organizational structure options would correspond to the design option that is selected for choosing the "original" cohort. Thus, we do not describe these options here, but rather refer the reader to the previous sections.

B2-4. CONCLUSIONS

In this paper we have attempted to describe a variety of design elements and design classes that should be considered when planning the NCS. As mentioned previously, the objective of this paper is to cover the reasonable range of design possibilities for selecting and recruiting the longitudinal cohort into the NCS, to highlight important differences in the various approaches, and to describe general advantages and disadvantages associated with each approach. The importance of carefully considering and evaluating these design options, and their corresponding technical and logistical challenges, when implementing a study as large and complex as the NCS is fundamental to identifying an efficient design that will lead to the ultimate success of the National Children's Study.

Table 3 presents a summary of the design classes and corresponding possibilities described in the previous sections. Admittedly, there are certainly other options that could be implemented (e.g., other sampling frames, other and more complex hybrid sampling methods, etc.); however, we believe that these design classes capture many of the most promising avenues for this aspect of the design of the NCS. In fact, it should be recalled that although we describe only six classes, as well as an add-on option that could pertain to any of the classes, a large number of design possibilities exist in each class, and determination of the "optimal" design in each class will require significant effort.

For example, consider the design class described in Section 3.1, which involves a complete probability based selection of the cohort. This design has a number of possible sampling frames, a number of possible organizational structures, and a large number of options in terms of stratification possibilities and other design characteristics (e.g., number of PSUs, number of levels of sampling in a multi-stage design, oversampling, etc.). While we have highlighted a few of the general advantages and disadvantages of this design, the detailed specification of several specific design options within this class can result in designs that are significantly different in terms of their effective sample size, power to assess the core hypotheses of the study, and cost. For example, a design with probability based sampling of 800 PSUs will likely have significantly different characteristics than a probability based sampling design with 100 PSUs. Thus, we note that for this design class, as well as all the other design classes, there are varying degrees of study design optimality that will need to be further considered and evaluated as the more detailed design descriptions are developed.

Table B2-3. Table of design classes with sampling methods, options for sampling frame and options for organizational structure.

Option #	Sampling Methods	Sampling Frame	Organizational Structure*
1	Complete Probability Based Design	Household	Any of 1), 2) or 3)
		Physician's Office	3)
		Centers	Either of 1) or 3)
		Combination of Centers and Household	Either of 1) or 3)
2	Quota/Convenience Sampling of PSUs and Probability Sampling Within PSUs	Centers	Either of 1) or 3)
		Combination of Centers and Household	Either of 1) or 3)
3	Complete Convenience or Quota Sampling	Centers	Either of 1) or 3)
		Physician's Office	3)
4	Mixture of Convenience and Probability Sampling of PSUs and Within PSU Probability Sampling	Centers	Either of 1) or 3)
		Physician's Office	3)
		Combination of Centers and Household	Either of 1) or 3)
5	Mixture of Convenience and Probability Sampling of PSUs and Within PSUs	Centers	Either of 1) or 3)
		Physician's Office	3)
		Combination of Centers and Household	Either of 1) or 3)
6	Multiple Cohorts	Centers	Either of 1) or 3)
		Combination of Centers and Household	Either of 1) or 3)
+	Sibling Cohorts		

* Organization structures are: 1) Medical centers contracted to conduct study in their geographic areas, 2) Single, or relatively few, contract research organizations contracted to conduct study with subcontractors as needed, and 3) Combination of medical centers, contract research organizations, and physician's offices.

The next stage in developing a promising NCS design involves further consideration of the above design classes, and perhaps other design options if they are subsequently identified, in light of the criteria for assessment of NCS design options (presented in another white paper). Using the criteria, which were identified through consideration of the necessary components of the NCS, the scientific objectives of the NCS, and cost considerations, the design classes and corresponding options will be evaluated with the goal of identifying a small set of "finalist" designs that can be further specified and investigated. This effort will involve the following:

- Identification of other design options through consultation with NICHD officials and other subject matter experts. This will occur at a two-day working session held in late October.

- Further investigation of statistical methods for utilizing validation subsamples to correct for possible biases introduced in a convenience sampled cohort.

- Further investigation of the statistical properties (e.g., design effects, sample weights, etc.) associated with designs where a portion of the sample is selected probabilistically and a portion is selected based on convenience.

- Investigation of the appropriateness of using sampling weights (or excluding sampling weights) when assessing the relationship between health outcome and exposure.

- General discussion and evaluation of all proposed design options in light of the criteria for assessment.

- Further specification of the proposed design options as needed for this general evaluation.

- Determination of a set of "finalist" design options that most satisfy the criteria for the NCS design.

- Further consideration and full specification of the "finalist" design options.

In successfully completing these steps we hope to provide NICHD with an objective assessment of several NCS design options that will assist in determining the appropriate design for the NCS.

B2-5. REFERENCES

Buck, G.M., Lynch, C.D., Stanford, J.B., Sweeney, A.M., Schieve, L.A., Rockett, J.C., Selevan, S.G., and Schrader, S.M. (2003). Prospective pregnancy study designs for assessing reproductive and developmental toxicants. Environmental Health Perspectives, accepted.

Chamber, R.L. and Skinner, C.J. (2003). Analysis of Survey Data. John Wiley and Sons, England.

Hahn, G.J., and Meeker, W.O. (1993). Assumptions for statistical inference. The American Statistician 47:1-11.

Tingen, C., Stanford, J.B., Dunson, D.B. (2003). Methodologic and statistical approaches to studying human fertility and environmental exposure. Environmental Health Perspectives, accepted.

WESTAT (2002). *Sampling Strategies for the Proposed National Children's Study*. Prepared for National Center for Health Statistics, October 25, 2002.