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

Appendix A

Advantages and Limitations of Probability-Based Sampling for the NCS

February 2004

White Paper

on

Advantages and Limitations of Alternative Sampling Methods for the National Children's Study

by

Steve Rust

with input from

**Jeff Lehman, Nancy McMillan, John Menkedick,
Colm O'Muircheartaigh, Louise Ryan, Lowell Sever
and Warren Strauss**

**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

Appendices

Tables

Figures

Glossary of Terms

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

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

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

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

Multi-Stage Sampling. Multi-stage sampling methods allow selection of groups of elements from the sampling frame at one stage and then subsequent sampling from the selected groups of elements at a subsequent stage.

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

Probability-Based Purposive Exclusion: A probability-based method for assuring that specified elements of the study population are excluded from the target sample with certainty.

Probability-Based Purposive Sampling: A probability-based method of assuring that specified elements of the sampling frame are included in the target sample with certainty.

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

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

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

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

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

Relative Risk Ratio: The proportion of diseased people among those exposed to a relevant risk factor divided by the proportion of diseased people among those not exposed to a relevant risk factor.

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

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

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

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

Stratified Sampling. Stratified random sampling methods control the sub-sample sizes for subsets (strata) of the sampling frame defined by one or more design variables.

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

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

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

White Paper

on

Advantages and Limitations of Alternative Sampling Methods for the National Children's Study

Steve Rust[1]

with input from

Jeff Lehman[1], Nancy McMillan[1], John Menkedick[1], Colm O'Muircheartaigh[2], Louise Ryan[3], Lowell Sever[1], and Warren Strauss[1]

A-1. INTRODUCTION

The purpose of this white paper is to summarize and assess the advantages and limitations of employing probability-based and non-probabilistic sampling methods when selecting a cohort of children (NCS cohort) for long-term follow-up in the National Children's Study (NCS). As a background for this assessment, we first address a context for design of the NCS sampling protocol.

In the terminology of Cochran (1977), the NCS is primarily an "analytical" survey or study rather than a "descriptive" or enumerative survey or study. The term "descriptive study" is used here to refer to a study having the objective to describe a population of interest in terms of measurable characteristics. The term "analytical study" is used here to refer to a study having the objective to identify differences between subpopulations of a population of interest. For the NCS, subpopulations will be defined by levels of exposure and the differences will be characterized in terms health and developmental outcomes. Often the ultimate purpose of an analytical study is to take action on the cause-and-effect system(s) underlying identified relationships with the aim of improving future conditions (Hahn and Meeker (1993)). While the NCS will necessarily focus on a population of contemporary children, by the time relationships are identified in the NCS data, it will in most cases be too late to take effective action to improve the health and development of the children in this contemporary population.

Figure 1 illustrates a context within which to consider various sampling design options for the NCS. Under any design scenario, data from the NCS cohort will be

[1] Battelle
[2] University of Chicago
[3] Harvard University

analyzed to test multiple hypotheses regarding environmental exposures in the "broadest sense" to identify potential cause-and-effect relationships between environmental exposures and health and developmental outcomes. A long-term objective of the NCS is to influence public health policy and social behavior to bring about the application of effective environmental, behavioral and medical interventions. Such interventions, when applied to a future population of children in the US, should lead to improved health and developmental well-being.

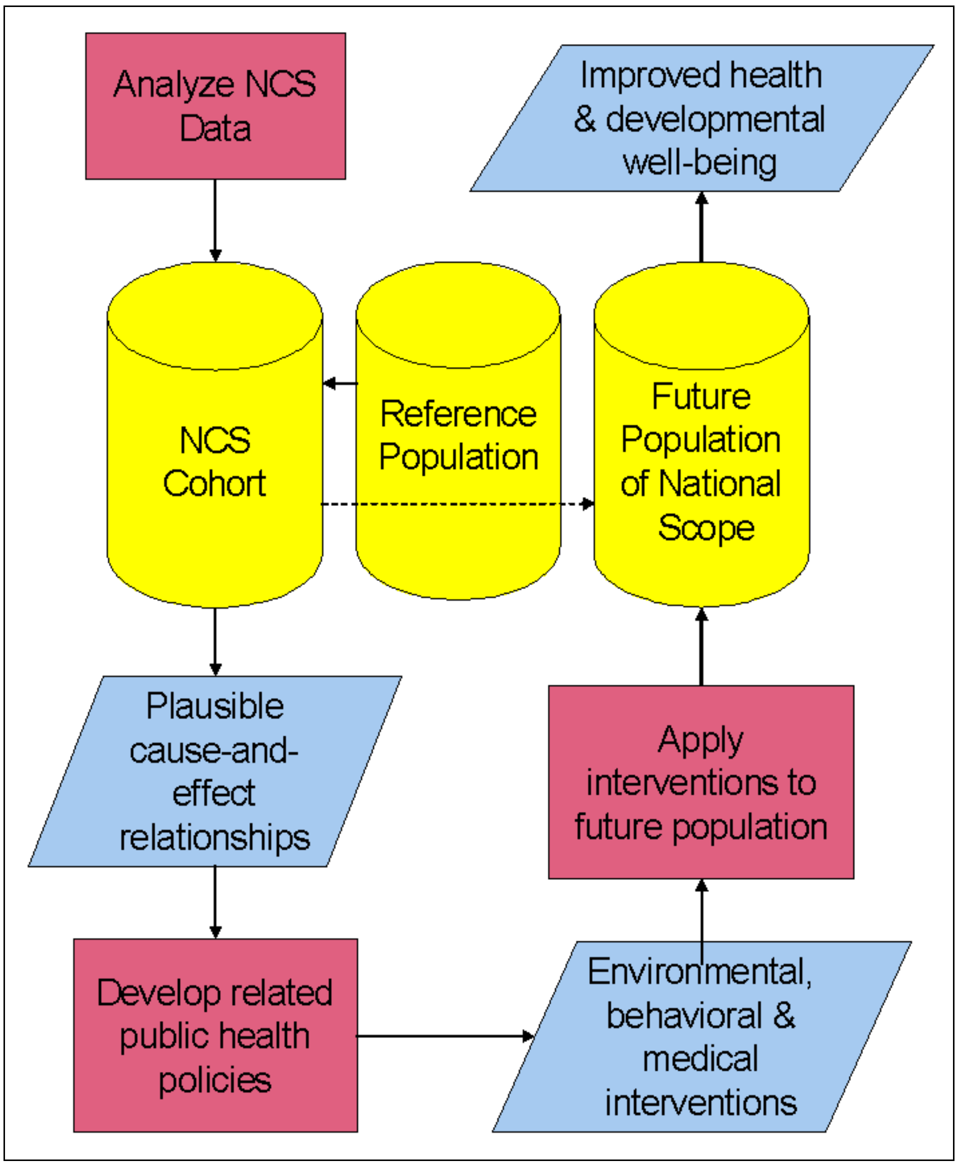

Figure A-1. Context for NCS Sampling Design

Consider the example of characterizing the relationship between pesticide exposure and the presence of autism at a specified stage of development. By the time any firm conclusions have been drawn from the NCS data regarding the relationship between pesticide exposure and autism, all children who had an opportunity to participate in the NCS will have already experienced the pesticide exposure period in question and any increase in the likelihood of autism will have already taken its toll. Thus, it is logical in this example to focus one's ultimate attention on a future population of children for whom intervention to reduce pesticide exposure is possible.

In order to have an impact on the health and well-being of children in the US, the relationships identified in the NCS data will have to be valid for a future population of children in the US for which some form of intervention is possible. But given that this future population of children cannot be studied, it is logical to instead seek relationships that are valid for the current population of children in the US and rely on similarities between current and future populations of children as well as the external validity of models developed from the study to extend the validity of the relationships into the future. Therefore, in this paper, we will consider the population of children born in the US during the NCS enrollment phase to be the reference population for the study.

External validity (Campbell and Stanley, 1963) refers to the validity of relationships identified in the NCS data when extended to this reference population. As illustrated in Figure 1, the reference population serves as a stepping stone between the NCS cohort and the future population for which health and developmental benefits are sought.

Internal validity (Campbell and Stanley, 1963) refers to validity of identified relationships within the restricted context of the NCS cohort. Internal validity derives from several conditions:

- Identified relationships must be statistically significant,
- The cause must temporally precede the effect in the cause-and-effect relationship, and
- The effects of extraneous variables, plausible confounders, and plausible effect modifiers have been properly accounted for.

Imposing conditions of internal validity forces one to consider not just the statistical significance of hypothesized relationships but also the nature of those relationships. The concepts of external and internal validity as they apply to the NCS are explored in Section 2.

There are numerous sampling strategies that could be implemented to select the NCS cohort. In this paper we will first focus on two strategies that have received much attention during the early discussion of design options for the NCS:

- Probability-based random sampling
- Non-probabilistic sampling

In approximate terms, probability-based random sampling strategies select cohort members according to a structured random process that results in every child in the reference population having a known probability between zero and one of being included in the NCS cohort   Non-probabilistic strategies put few or no constraints on the selection of children for the NCS cohort.

In Section 3, we present a historical perspective by first summarizing arguments that have been presented to advocate use of probability-based sampling when selecting the NCS cohort followed by a summary of arguments that have been similarly presented to advocate the use of non-probabilistic sampling. Specific advantages and limitations of the two sampling approaches are discussed in Section 4. These advantages and limitations address issues including the feasibility of constructing a sampling frame, ability to obtain a range of exposures, subject recruitment rates, subject retention rates, data quality, cost-efficiency of data collection, internal validity and external validity.

It is not necessary to restrict the design of the NCS to either probability-based random sampling or non-probabilistic sampling as one of two choices. In fact, practical implementation of the NCS may inevitably result in some mix of the two approaches. In Section 5 we introduce two additional hybrid sampling strategies that can be customized to meet specific requirements of the NCS. The methods are:

- Probability-based purposive sampling
- Probability-based purposive exclusion

Probability-based purposive sampling allows specified members of the reference population to be included in the NCS cohort with certainty. In a complementary fashion, probability based purposive exclusion allows specified members of the reference population to be excluded from the NCS cohort with certainty.

The discussion of hybrid strategies in Section 5 serves as a backdrop for the conclusions presented in Section 6. Finally, references are provided in Section 7.

A-2. VALIDITY

When traditional survey sampling methods are applied during the design phase of an analytical study, the driving force behind decisions is the pursuit of external validity for relationships identified in the study data. Probability-based random sampling methods are employed to select a sample from a sampling frame constructed to include as much of the reference population as possible. If the sample is so selected and response rates are high, then relationships identified in the study data can be validly extended to the reference population based on the random sampling mechanism employed to select the sample. Thus, traditional survey sampling methods are based on the concept of external validity through probability sampling and, therefore, are the methods of choice if external validity is of primary importance. Traditional survey sampling methods have emphasized the statistical significance of relatively simple associations, with little attention to the subject matter nature of the relationships themselves and detailed modeling of the relationships.

In contrast, when traditional epidemiological methods are applied during the design phase of an analytical study, the primary driving force behind decisions is often pursuit of internal validity for relationships identified in the study data. While issues of statistical significance are certainly important drivers of traditional epidemiological studies, much more attention is given the subject matter nature of hypothesized relationships. Does the exposure temporally precede the outcome? Have potential confounders and effect modifiers been measured and ruled out? What is the hypothesized nature of the relationship and how might this affect timing and extent of measurement? Internally valid relationships between exposures and outcomes are more likely to lead directly to intervention concepts. While traditional epidemiological methods embrace external validity as an important objective, external validity is often a secondary objective with the majority of emphasis given to internal validity. Traditional epidemiologists may also rely on consistency with subsequent studies to confirm models developed from an internally valid study.

It is quite conceivable that a team of scientists trained in traditional survey sampling methods and a team of scientists trained in traditional epidemiological methods, when given exactly the same context for designing the sampling protocol for the NCS, would design studies that have vast differences. Survey sampling methods would emphasize coverage of the reference population and considerable study resources might very well be devoted to sampling hard-to-study elements of the reference population, conversion of reluctant participants, and proper handling of non-responders to maintain representativeness. By explicitly devoting considerable study resources in this manner, fewer study resources would be available for data collection, adversely affecting the level of detail possible in the data collection protocol. This effect on level of detail is likely to be somewhat implicit in nature.

In contrast, epidemiological methods would emphasize the level of detail in the data collection protocol because, without that detail it would be impossible to properly assess the role of potential confounders and effect modifiers. By devoting considerable study resources to the data collection protocol, fewer resources are available for assuring the representativeness of the study sample. This effect on representativeness is also likely to be implicit in nature.

To further explore the issue of internal versus external validity, consider a simple linear regression model that explores the relationship between an adverse health effect and some measure of exposure. In a study the size of the NCS, it is certainly possible to identify statistically significant relationships between disease and exposure without explaining a large percentage of the variability in the response variable (as characterized by the R^2 statistic). When attempting to extrapolate these results to a larger reference population, we may feel uncomfortable about the factors left unexplained by this model absent a probability-based sampling design which allows us to assume that the sample is unbiased relative to the reference population. In contrast, it is also possible to explore relationships in which a very large percentage of the variability is explained, perhaps using a more complex model that includes covariates, confounders and effect modifiers. In this situation, it may be reasonable for scientists to conclude, based on the defensibility of the model, that the relationships observed are unbiased relative to the reference population regardless of the mechanism by which participants are recruited into the sample.

A-3. A HISTORICAL PERSPECTIVE

An influential group of scientists involved with the early planning and discussion on an optimal sampling design for the NCS, including the NCSAC, have strongly advocated a probability-based sampling approach for the NCS. Without a significant probability basis to the sampling, they maintain that 1) at the end of the study scientists will be forced to say that they have no idea what population the study results are generalizable to; 2) that this will undermine the scientific credibility of the entire study; and 3) that the unknown and unknowable biases that might be introduced through convenience sampling could lead to false conclusions. Their concern is particularly highlighted for the social environment and behavioral assessments in the NCS where relevant exposure and risk factors are less well-characterized and where unknown biases may be more likely.

There are several positive aspects of probability based sampling that have been stressed, including:

1. Each child in the NCS cohort represents a known number of children in the sampling frame

2. Each child in the sampling frame has a known and positive probability of being selected into the NCS, providing some sense of political fairness to the process of sample selection

3. Probability-based sampling provides a feasible and scientifically defensible mechanism for applying scientific results observed in the NCS to a reference population of children without needing to worry about unintentional systematic biases introduced through the sampling mechanism.

Advocates of probability-based sampling contend that, with the broad range of health outcomes, potential exposures, and critical stages of vulnerability covered in the NCS, it will be operationally infeasible (both from a cost perspective and from the perspective of the burden placed on study participants) to assess all of the important risk factors, covariates, effect modifiers, and confounders across the entire cohort over time. If important factors cannot feasibly be observed within this study, the potential for misleading inferences due to a biased sampling approach become much more likely, increasing the value of a probability-based sample for drawing externally valid inferences.

In addition, while the study design for the NCS is currently hypothesis driven, much of the value of the NCS is based on its ability to support scientific discovery. The NCS will provide a rich source of observational data that will be explored by scientific experts in a variety of disciplines for decades. To support this future research, potentially on topics completely unrelated to the original core hypotheses that currently form the basis for the study design, it is of paramount importance that the NCS sample be generalizable to a reference population with known characteristics.

Finally, advocates of probability-based sampling suggest that limitations in response rates or other sampling deficiencies do not imply that it is acceptable to use a more convenient, less expensive, or less demanding sampling method. It must, however, be recognized that low recruitment and retention rates will make it difficult to calculate meaningful sample weights from a probabilistic point-of-view.

While there has been a influential component of the scientific community that has advocated probability sampling as a requirement for the NCS as discussed above, there has also been an equally influential component, led by, but not limited to, epidemiologists, who have strongly advocated a non-probability approach as the only feasible way to conduct the study. This viewpoint begins with the recognition that the primary objectives of the NCS are related to understanding relationships between risk factors and disease, including understanding the etiology of disease. In order to maximize the likelihood that the study will provide information on etiology, it is necessary to maximize the amount of information on exposure, effect modifiers, covariates, and outcomes that can be collected, maximize the retention rate to observe effects over time, and therefore to maximize the internal validity of the study.

This group feels strongly that a design that is primarily selecting participants at random 1) will not be able to get its participants to agree to the level of burden that will be required to collect the necessary scope of information to reasonably understand the exposure-outcome relationships, and 2) that such a sample will also inevitably lead to such high attrition rates as to jeopardize study objectives related to outcomes or exposures arising later in childhood. In other words, they believe that probability sampling will jeopardize the internal validity of the study. From their perspective, probability sampling leads to a Catch 22. If the study collects sufficient exposure, covariate, and effect modifier data to do a good job of understanding relationships and etiology, then external validity can be achieved through trust in the model and probability sampling was not necessary; on the other hand, if there are associations observed that do not have the exposure, effect modifier, and covariate information well captured, then the statistical validity of the relationship is of little additional value in the overall scheme of the usefulness of study results. They also believe that for most chemical, biological or physical exposures, there is little basis to assume that associations would be significantly biased, if the participants are chosen from well known quotas, or strata, defined by characteristics such as SES, age, sex, and race. There is also a sense that the effective response and retention rates for a randomly-selected cohort will be so low as to result in what amounts to a convenience sample in the long run anyway.

In summary, advocates for non-probability sampling believe that internal validity will be compromised by extensive probability sampling, that external validity can be achieved by collecting better information to construct more defensible models, and that probability sampling will not be able to match the response rates, agreement for burdensome measures, and retention of a convenience based sample, and will cost more. It should be noted that most advocates of this position acknowledge that if their goals of measurement, response and retention could be met equally well with probability sampling, they would recognize the statistical advantages offered by this approach.

A-4. ADVANTAGES AND LIMITATIONS OF ALTERNATIVE SAMPLING STRATEGIES

This section is devoted to describing and examining the advantages and limitations of probability-based random sampling and non-probability based sampling, attempting to capture and place into context an ongoing debate between advocates of each approach. Before addressing each sampling approach on its own merits we describe a conceptual process for drawing conclusions from NCS data so that the process steps can be used as an organizational structure for summarizing advantages and limitations and. we review some assumptions that underlie that discussion.

A-4.1 PROCESS FOR DRAWING CONCLUSIONS FROM NCS DATA

The advantages and limitations of alternative sampling methodologies tend to exist in advantage-limitation pairs. For example, employing probability-based sampling methods has disadvantages with respect to recruitment and retention that result in an advantage when issues of external validity are later addressed. In contrast, employing non-probabilistic sampling methods has advantages with respect to the recruitment and retention that result in a limitation when issues of external validity are addressed. We have found it useful to organize the set of advantages and limitations for each sampling method according to steps in a conceptual process for drawing conclusions from the NCS data. The conceptual process has the following steps:

1. **Sampling Design:** Select the children that will be targeted for recruitment.

2. **Recruitment:** Recruit targeted children into the NCS cohort.

3. **Retention:** Maintain continued participation of the children in the NCS cohort.

4. **Data Collection:** Collect study data for the children in the NCS cohort.

5. **Internal Validity:** Identify cause-and-effect relationships that are valid for the NCS cohort.

6. **External Validity:** Validly extend cause-and-effect relationships identified in the NCS data to the reference population of all children born in the US during the NCS enrollment period.

The second column of Table 1 contains a summary of advantages and limitations of probability-based sampling within the context of the NCS, organized by the conceptual

Table 1. Advantages and Limitations of Alternative Sampling Strategies

Process Step	Probability-Based Random Sampling	Purposive Sampling	Purposive Exclusion	Non-Probabilistic Sampling
Sampling Design	• Advantage: No requirement for a "representative" sample; includes mechanisms for over- and under-sampling specifically defined subsets of the sampling frame – Minorities – Children known to represent extremes with respect to exposure to suspected risk factors for health/developmental outcomes • Limitation: Requires the construction of a sampling frame; could be difficult in the case of pre-natal and pre-conception sampling	• Advantage: Can incorporate effects of constraints imposed by the Federal procurement system • Advantage: Can include – Geographically-isolated areas of exposure – Individuals known to have been exposed (useful in the case of rare exposures)	• Advantage: Can be used to exclude portions of the reference population that are difficult to enumerate	• Advantage: Since there are no constraints on how participants are located, simplifies the process of pre-natal or pre-conception sampling • Limitation: Over- and under-sampling can produce an undesirable sample mix – Mitigator: Quota sampling may be used to control characteristic discrepancies between the NCS cohort and the reference population or to assure sufficient numbers of cohort members in various characteristic classes to support using the associated characteristics as effect modifiers; quota sampling is only useful if all important effect modifiers and confounders are known in advance of sampling; because of low prevalence of many outcomes of interest, there may not be sufficient knowledge to do this – Mitigator: Can post-stratify to adjust for measured effect modifiers
Recruitment	• Limitation: Anticipated high participant burden may lead to a low successful recruitment rate; must contact many more people than you would enroll – Mitigator: Can over-sample children known to have desirable properties (e.g., existing relationships) with respect to recruitment	• Advantage: Sampling design may assign certainty status to subsets of the reference population that are easier to recruit	• Advantage: Sampling frame may exclude subsets of population expected to be more difficult to recruit	• Advantage: Volunteer participants simplify recruitment effort thereby reducing recruitment costs

Table A-1. Advantages and Limitations of Alternative Sampling Strategies (Continued)

Process Step	Probability-Based Random Sampling	Purposive Sampling	Purposive Exclusion	Non-Probabilistic Sampling
Retention	• Limitation: Actual high participant burden will lead to high levels of attrition and high levels of sporadic non-response – Mitigator: If no attempt is made to convert reluctant participants, attrition and sporadic non-response rates may be no worse than for non-probabilistic sampling – Mitigator: Can over-sample children known to have desirable properties with respect to retention and/or reliable response – Mitigator: Can use techniques to maximize retention (e.g., develop inter-personal relationships between staff and participants) • Limitation: Inability to include volunteers and/or convenient participants may result in less community involvement and therefore have an indirect negative effect on retention/response rates for probabilistically-selected participants	• Advantage: Sampling design may assign certainty status to subsets of the reference population that are easier to retain	• Advantage: Sampling frame may exclude subsets of population expected to be more difficult to retain	• Advantage: Allows use of volunteer participants that should have better retention and reliable response rates because of carry-over effects from their initial positive attitude toward study participation • Advantage: Allows use of volunteer participants in cases where the respondent burden is extreme leading to higher retention rates
Data Collection	• Limitation: Use of study resources elsewhere may require that a simplified data collection protocol be employed • Limitation: If cohort is geographically dispersed, may be difficult to maintain data quality and/or cost-efficiency – Mitigator: Provides mechanisms (e.g., cluster sampling) for controlling the geographic dispersion of the sample (can be said for all approaches) • Limitation: Inability to include volunteers and/or convenient participants may result in less cooperation from organizations charged with data collection	• Advantage: Can take advantage of specialized facilities and equipment that are available on a very limited basis; similarly can take advantage of existing environmental data	• Advantage: Sampling frame may exclude subsets of population for which data collection is expected to be more expensive	• Advantage: Allows selection of cohort members that have a relationship with data collection organizations, perhaps leading to very cost-efficient data collection • Advantage: May be easier to schedule volunteers at times that are convenient or available to staff; volunteers may be more willing to travel.

Table A-1. Advantages and Limitations of Alternative Sampling Strategies (Continued)

Process Step	Probability-Based Random Sampling	Purposive Sampling	Purposive Exclusion	Non-Probabilistic Sampling
Internal Validity				• Advantage: Better cost efficiencies may perhaps allow a more detailed protocol to better address – Extraneous variables – Confounders – Effect modifiers • Limitation: Limited ability to empirically check the validity of the assumed statistical model

Table A-1. Advantages and Limitations of Alternative Sampling Strategies (Continued)

Process Step	Probability-Based Random Sampling	Purposive Sampling	Purposive Exclusion	Non-Probabilistic Sampling
External Validity	• Advantage: In theory, inferences may be extended to sampling frame based on the random sampling mechanism employed to select the sample (with assumptions regarding non-respondents) • Advantage: Enables the use of model-based inference procedures by assuring that the sampling method does not select cohort members who represent a biased sample with respect to health and developmental outcomes • Advantage: Provides externally valid prevalence information for exposures and outcomes; this information may be required to – Estimate the number of children affected by policy changes and intervention strategies – Perform a cost-benefit, economic-health impacts analysis • Limitation: Imperfect recruitment and retention complicates extension of inferences to sampling frame – Mitigator: Statistical methods exist for dealing appropriately with non-response; require assumption that responders and non-responders are the same conditional on model covariates and effect modifiers or follow-up with a sample of non-respondents • Limitation: Useful inferences for under-sampled population subsets may be impossible	• Limitation: Useful inferences for under-sampled population subsets may be impossible	• Advantage: Reduced sampling frame is completely defined allowing an explicit assessment of the potential for systematic bias (are exclusion factors likely to be related to associations of interest?) • Limitation: Inferences to the full reference population are subject to systematic bias because it is difficult to say how the exposure-response relationships might differ for excluded groups – Mitigator: If less cost-effective subsets of population have been excluded, benefit of additional information collected may outweigh risk associated with systematic bias	• Limitation: The statistical significance of hypothesized associations cannot be based on statistical inference to a larger population • Limitation: Impossible to characterize the sources of systematic bias introduced by volunteer participants or even the larger population to which one might attempt to generalize study findings because differences between participants and non-participants are not defined; volunteer effect may be particularly important for the NCS because of the social environment and behavioral aspects of the study – Mitigator: Associations, particularly those that are more biological, chemical or physical in nature, may be less subject to external validation problems • Limitation: Does not provide externally valid prevalence information for exposures and outcomes – Mitigator: Perhaps this information can be obtained via another probability-based sampling mechanism (e.g., NHANES)

steps for drawing inferences from NCS data. Similarly, the last column of Table 1 summarizes the advantages and limitations of non-probabilistic sampling.

A-4.2 ASSUMPTIONS RELATED TO ADVANTAGES AND LIMITATIONS

In listing the advantages and limitations for each sampling method, there are general assumptions about the distribution of study resources and the benefits associated with certain types of over-sampling. Sampling designs that employ probability-based methods for identifying potential cohort members are assumed to have two characteristics: (1) they require that more study resources be devoted to sampling design, recruitment, and retention and (2) they will have lower retention rates resulting in data collection resources being devoted to larger numbers of early cohort members that eventually withdraw from the study. Both of these characteristics result in reduced resources being devoted to direct data collection for cohort members that participate throughout the duration of the entire study. Under various assumptions about retention rates and the manner in which probability-based sampling is implemented, the resources diverted from direct data collection can be estimated to be as small as $100 million or as large as $1 billion.

It is assumed that the effect of having reduced data collection resources would manifest itself either in the form of fewer cohort members or a simplified data collection protocol that would be less likely to include the measurement of confounders and effect modifiers associated with hypothesized relationships. Thus, sampling designs that employ probability-based methods for identifying potential cohort members were assumed to have an advantage with respect to extending statistically significant associations present in the NCS data to the reference population but to have a limitation with respect to demonstrating that true cause-and-effect relationships underlie these associations.

It is also assumed that sampling designs that target children with a wide range of exposures will result in better statistical power when assessing the significance of hypothesized relationships. In a similar fashion it is assumed that sampling designs that target children with a wide range of values for potential confounders and effect modifiers will result in better statistical power when attempting to refine statistical associations down to plausible cause-and-effect relationships.

Finally, it is assumed that relationships that have been refined through a process of carefully examining potential confounders and effect modifiers are likely to be more robust when attempts are made to extend those relationships to the reference population. This assumption is strengthened if the differences between the study population and the reference population are characterized and the variables that characterize the difference are considered to be potential effect modifiers.

A-4.3 ADVANTAGES AND LIMITATIONS OF PROBABILITY-BASED RANDOM SAMPLING

Probability-based sampling methodologies are methods for selecting a study sample from a sampling frame such that the probability of occurrence for every possible study sample is a known function of a set of design variables. An important property of a probability-based sampling process is that the probability of inclusion in the study sample is known for each and every element (e.g., child) in the sampling frame. In this section, we focus on probability-based sampling methods for which each element of the sampling frame has a probability of being included in the study sample that is strictly greater than zero and strictly less than one. Probability-based sampling methods that relax this restriction are discussed in Section 5.

There are many forms of probability-based random sampling including:

- Simple random sampling,
- Stratified random sampling,
- Cluster-based random sampling, and
- Multi-stage random sampling.

Simple random sampling methods select the study sample from the sampling frame in a totally random fashion without replacement. Stratified random sampling methods control the sub-sample sizes for subsets (strata) of the sampling frame defined by one or more design variables. Cluster-based random sampling methods allow sample elements to be selected in groups to improve the cost-efficiency of the data collection process. Multi-stage random sampling methods allow selection of groups of elements from the sampling frame at one stage and then subsequent sampling of elements from the selected groups at a subsequent stage.

Probability-based random sampling methods allow a tremendous degree of flexibility in setting inclusion probabilities for elements of the sampling frame. Subsets of the sampling frame can be easily over- or under-sampled simply by adjusting the inclusion probabilities associated with the elements of these subsets. This flexibility could be exploited, for example, to over-sample children with desirable properties with respect to cost-efficiency and/or data quality such as children living within 50 miles of a qualified medical center or children living in a geographical area with existing data on sources of environmental exposure. In a similar fashion, over-sampling can be employed to ensure that sufficient numbers of target minorities are included in the NCS cohort as well as to ensure children with a wide range of exposures are included in the NCS cohort. Over-sampling children with “desirable” properties necessarily results in under-sampling of children with “undesirable” properties.

Probability-based sampling designs derive a large portion of their value from the following characteristic. If the NCS cohort is selected from the reference population according to a probability-based sampling design, then inferences drawn from the NCS cohort will be valid for the reference population as long as a sampling frame corresponding to the reference population can be constructed and the vast majority of children targeted by the probability-based sampling process are successfully recruited into the NCS cohort and retained until required study data have been collected.

Probability-based sampling is often considered to be a potentially cost-inefficient method for selecting a study sample because of expectations of low recruitment levels, high levels of attrition, and prohibitive study costs. In reality, this criticism may be leveled more at the way in which probability-based sampling is often employed rather than at probability-based sampling itself. If the need for the study sample to be "representative" of the entire population is paramount, then a disproportionate share of study resources may be devoted to obtaining equitable coverage of elements of the population that are subject to higher rates of non-response, higher costs for obtaining study information, or lower-quality study information. However, it is quite possible to stay within a probability-based sampling framework and tailor the sampling in a more cost-efficient direction. If response rates can be predicted as a function of the design variables, then elements of the population that are predicted to have higher response rates may be over-sampled. For example, suppose that a dozen major medical centers have agreed to participate in the NCS and leaders of the communities in which these medical centers reside have agreed to sponsor programs designed to emphasize the importance of individual participation in the NCS. A probability-based sampling plan could be designed to over-sample the children in these communities where community involvement is expected to result in higher recruitment and retention rates. While such an approach will increase the cost-efficiency of the study, it requires an up-front acceptance of the fact that other elements of the population will be underrepresented, perhaps to a point where strong statistical inferences cannot reasonably be extended to these under-represented elements.

In a departure from traditional survey sampling practice, it may actually be detrimental to attempt to convert reluctant participants to join the NCS cohort under the theory that converted reluctant participants are very likely to drop out of the NCS cohort before all required study data has been collected. As such, the resources devoted to the successful recruitment of these children as well as the resources devoted to collection of data for these children before they drop out of the study would be largely wasted. A better strategy may be to include only those children among those targeted by the probability-based sampling plan that are enthusiastic about involvement in the study. While such an approach would likely result in a low initial recruitment rate and higher recruitment costs, it is likely to have a very positive effect on the retention rate and minimize the magnitude of resources wasted on children who eventually drop out of the NCS cohort. The overall participation rate (combining recruitment and retention) for this approach may very well be similar to the participation rate that would be achieved via a more aggressive recruitment strategy.

Strict adherence to probability-based sampling would require that volunteers who learn of the study and wish to be included be turned away. Such a policy could have a negative effect on community involvement in the NCS which in turn might have a negative effect on recruitment/retention rates, and the level of cooperation from organizations charged with data collection.

The geographic dispersion of a sample can be a major factor affecting its cost-efficiency. A widely dispersed sample can result in significant increased costs with respect to training data collectors, standardizing data collection methods, maintaining quality control with respect to specimen processing and analysis, and other similar data quality issues. One solution is to over-sample children affiliated with a smaller, more dedicated data collection mechanism. However, within the probability-based sampling framework, cluster sampling provides a mechanism for controlling the geographic dispersion of the study sample and effectively dealing with a number of cost-efficiency issues.

Over the past three decades, the statistical literature contains numerous contributions to an ongoing debate concerning the correct basis for statistical inferences from sampling survey data. For a sample of contributions to this design-based versus model-based inference controversy, see Scott and Smith (1973), Rubin (1976), Smith (1976), Scott (1977), Little (1982), Smith (1983), and Sugden and Smith (1984), Smith (1994), Kish (1995), Valliant, Dorfman and Royall (2000) and Little (2003). It is interesting to note that this entire debate takes place within a context that assumes the sample being analyzed was selected by a probability-based sampling method. Even advocates of model-based inference need to assume that the sampling method or mechanism does not select cohort members who represent a biased sample with respect to health and developmental outcomes. "Otherwise, the sampling mechanism needs to be modeled, and appropriate modeling in such cases is at best difficult. (Little, 2003)" Probability-based sampling methods provide the needed assurance that the sampling method does not select cohort members who represent a biased sample with respect to health and developmental outcomes.

Less than perfect recruitment, retention, and response rates erode the ability of probability-based samples to guarantee external validity. Two options exist for dealing with small amounts of departure from the targeted probability-based sample. The first option is to perform a second near-perfect[4] study of a sample of initial non-responders. This is not likely to be a viable option for the NCS. It may be possible to obtain follow-up information to assess basic differences (e.g., demographics, housing characteristics) of non-responders, but it will likely not be possible to obtain exposure-response information. The second option is to assume that, conditional on any model covariates, the sub-population of initial responders is unbiased relative to the reference population and apply methods such as the multiple imputation (Rubin (1987) and Rubin (1996)) to properly treat the missing data. The rate of recruitment and retention failures associated with the NCS may be too high for such methods to be effective, thereby reducing imputation to a model-based inference approach for improving external validity.

While this paper is focused on the very important NCS objective of identifying cause-and-effect relationships between environmental exposures and

[4] Nearly 100% successful recruitment of targeted participants

health/developmental outcomes, it is important to note that only probability-based samples can provide externally valid estimates of prevalence for specific environmental exposures and health/developmental outcomes. However, prevalence estimates may be alternatively available from other sources such as NHANES. Such estimates may be important to support cost-benefit analyses that must accompany public health policy development.

A-4.4 ADVANTAGES AND LIMITATIONS OF NON-PROBABILISTIC SAMPLING

In straightforward terms, a sampling method is a non-probabilistic sampling method if it is not a probability-based sampling method. Common non-probabilistic sampling methods include:

- Convenience sampling (e.g., recruiting patients as they arrive at a medical facility for otherwise scheduled appointments)
- Volunteer sampling (e.g., recruiting potential participants that respond to an advertised request for volunteers to participate in a study)

It is also worth noting the ways in which a sampling method can be non-probabilistic which include:

- The children in the sampling frame cannot be enumerated, or
- The probability of inclusion in the study sample cannot be determined.

Non-probabilistic sampling methods are characterized by a minimization of constraints on the methods used to identify and select study participants. This minimal level of constraints produces tremendous benefits in terms of simplifying the sampling design process and improving recruitment and retention rates. Study resources that might otherwise have been consumed for sampling design, recruitment, and retention can be diverted to data collection resulting in a more comprehensive data collection protocol. With more resources available for data collection, a more complex data collection protocol could be employed increasing the likelihood that data for potential confounders and effect modifiers would be available.

Non-probabilistic recruitment can emphasize volunteer participants who would be expected to exhibit much higher recruitment and retention rates. The respondent burden associated with the NCS data collection protocol **may** be so great that only enthusiastic volunteers could be expected to remain a part of the NCS cohort for the entire study period of more than two decades. In this case, probability-based sampling might be abandoned entirely in favor of a volunteer sample.

The benefits of non-probabilistic sampling during the sampling design, recruitment, and retention phases result in limitations during the data analysis phase. The primary limitations associated with non-probabilistic sampling are:

- Uncontrolled over- and under-sampling can produce an undesirable sample mix

- Statistical models of relationships may only be valid for the NCS cohort; validity beyond the NCS cohort must be assumed based on some other scientific criteria rather than the known sampling and probability characteristics of the study, and
- Identifying and assessing the potential sources of systematic bias is made difficult by the fact that the true sampling frame and true study population cannot generally be identified; it is difficult to assess differences between participants and nonparticipants given limited (or no) information about those who could have volunteered.

The first limitation may be partially mitigated by employing quota sampling methods. Quota sampling may be used to alternatively (1) control characteristic discrepancies between the NCS cohort and the reference population or (2) assure sufficient numbers of cohort members in various characteristic classes to support using the associated characteristics as effect modifiers. The success of quota sampling depends entirely on the assumption that important confounders and effect modifiers are known. Because of the low prevalence of many of the NCS outcomes, there may not be sufficient knowledge to identify these important confounders and effect modifiers a priori. As an alternative to quota sampling, post-stratification methods can be used after the fact to attempt to adjust for the effects of over- and under-sampling.

Since statistical inferences for non-probabilistic samples cannot be based on a random sampling mechanism for selecting cohort members, assumptions associated with model-based inference procedures may be invalid. Unfortunately, there is also limited ability to empirically check the validity of model assumptions. A key assumption underlying most statistical models in this situation is that the sampling method does not select cohort members who represent a biased sample with respect to health and developmental outcomes. With volunteer and other forms of convenience sampling, this assumption may not be valid and an undetected systematic bias may accompany any conclusions drawn. For example, in the case of pre-natal or pre-conception sampling, women with a prior history of or risk factors for reproductive health problems (including adverse pregnancy outcomes) might be more likely to volunteer for the NCS. Systematic bias related to a volunteer effect may be particularly relevant to the NCS because of the social environment and behavioral aspects of the study. The behavior of volunteering may be correlated with unmeasured aspects of the social environment that have a direct effect on health and developmental outcomes of interest. This correlation is a potential source of systematic bias when relationships identified in the NCS data are extended to the reference population. Relationships that are based on actual physical exposures and biological consequences of such exposures may not be as subject to such systematic biases. However, relationships that have behavioral components (e.g., time-location profiles and activities that affect the extent of contact with contaminants) may be particularly susceptible to systematic bias.

Since one objective of the NCS is to provide a data set that can be used in the future to test hypotheses that are not currently anticipated, the data set would have to include explicit warnings about the degree to which the data from a non-probabilistic sample can be generalized to any population. While one would hope that such warnings would be duly noted and included in reports and publications of findings from the NCS data set, it is quite conceivable that the warnings would be largely ignored by at least a portion of scientists and researchers who use the NCS data set in the future.

A final limitation of non-probabilistic sampling is the inability to provide externally valid prevalence information for exposures and outcomes. Because acquiring such information is only a secondary objective of the NCS, this must be considered only a minor limitation. It is possible that other federally-funded health survey mechanisms (e.g., NHANES) represent better vehicles for obtaining prevalence information.

A-5. NCS REQUIREMENTS AND HYBRID STRATEGIES

In designing the sampling protocol for the NCS, one should consider requirements for both external and internal validity. With limited resources it is generally difficult to simultaneously satisfy strong external and internal validity requirements. Thus, a well-designed sampling protocol for the NCS is likely to strike a balance between external and internal validity.

Applying different scientific perspectives as the basis for drawing conclusions, valid conclusions may be drawn from both probability-based samples (statistically-based) and non-probabilistic samples (statistical and model-based). Probability-based samples offer the very desirable property of basing statistical inferences on the random sampling mechanism employed to select the sample. However, imperfect recruitment, retention, and response rates may limit such inferences. Valid inferences from non-probability-based samples, on the other hand, require the assumption that the NCS cohort is unbiased relative to the reference population. Thus, while inferences based on probability-based samples can draw their validity from the manner in which the sample was selected, inferences based on non-probability-based samples are only as valid as the NCS cohort is unbiased with respect to the relationships of interest.

There are two problems that complicate a sampling design process that attempts to strike a balance between external and internal validity. These are:

- The tendency of probability-based sampling methods to emphasize external validity and de-emphasize internal validity, and
- The emphasis on internal validity in non-probability samples that leads to the abandonment of probability-based methods and a resulting de-emphasis of external validity.

Taken at face value, these problems appear to leave little in the way of middle ground where a compromise might be found. However, it is almost certainly true that a polar application of either probability-based sampling or non-probabilistic sampling that fails to recognize the strengths of and motivations for the opposing approach will fail to address issues that are critical to the success of the NCS. Realistic anticipation of the true magnitude of respondent burden for NCS cohort members places expected recruitment and retention rates for a probability-based sample at low levels. Further, efforts to convert reluctant participants into cohort members, as are traditionally applied in survey sampling applications, may only result in retention and data collection resources being wasted on children who eventually drop out of the study prior to all required study data being collected. Probability-based efforts to obtain a representative sample that result in a geographically dispersed cohort may require unattainable data collection resources and adversely affect the quality of the data that is collected. The expenditure of limited study resources on sampling design, recruitment, and retention activities may require that a simplified data collection protocol be employed to control the data collection resources required.

On the other hand, adoption of a non-probabilistic approach requires faith that systematic biases will not limit the relevance of NCS conclusions to a limited and not specifically identifiable population. The exclusive use of non-probabilistic sampling results in considerable risk that relationships identified in the NCS data may simply not be valid when extended to the reference population of all children born in the US during the NCS enrollment phase. This risk may prevent conclusions drawn from NCS data from being widely accepted and thereby limit the value of the NCS for improving the health and development of future generations of children. The anticipated magnitude of resources that will be invested in the NCS as well as the one-time-opportunity nature of the NCS dictate that actions be taken to mitigate this risk.

There are inherent risks associated with both probability-based sampling methods and non-probabilistic sampling methods. Unfortunately, the raw data required to quantitatively estimate the risks does not exist. Therefore, efforts to choose one set of the methods over the other as optimal are frustrated by a lack of solid information. Within such an uncertain decision-making framework, it is logical to abandon the notion of choosing one set of methods over the other and instead plan for a study that implements both probability-based and non-probabilistic sampling as part of a hybrid sampling strategy. Within such a hybrid strategy, each set of methods acts as a hedge against the risks associated with the opposing set of methods. Motivation for such an approach can be found in the financial investment community where it is not uncommon to package collections of dissimilar investments so that each specific investment acts as a hedge against the risks associated with other investments in the package.

Before drawing final conclusions in Section 6, we present two particularly relevant sampling methods that may be used to provide added flexibility to a hybrid sampling approach. Both methods are completely compatible with a probability-based sampling approach.

A-5.1 PROBABILITY-BASED PURPOSIVE SAMPLING

Probability-based random sampling methods generally attempt to keep inclusion probabilities for all elements of the sampling frame strictly greater than zero and strictly less than one. It is quite acceptable, however, to employ inclusion probabilities of one for some elements of the sampling frame. One can view this as taking the concept of over-sampling to an extreme. For example, consider a two-stage probability-based random sampling process that first selects counties proportional to size from a sampling frame of all counties in the US and then selects a simple random sample of children within each selected county. Suppose that there are a dozen medical centers across the US that have successfully negotiated contracts with NIH to participate in the NCS. It would be quite acceptable, without leaving the confines of probability-based sampling methods, to specify that the dozen counties within which the medical centers reside must be included in the set of counties selected for the NCS. This purposive selection of specific counties does have consequences regarding the external validity of the study results in that the children selected from these specific counties can only represent their own county. Thus, these study subjects will have limited value for weighted analyses conducted for the purpose of demonstrating external validity. However, all the analysis methods that accompany probability-based sampling methods and the external validity that they afford to relationships identified in the study data remain valid in the context of probability-based purposive sampling.

In standard multi-stage applications of probability-based sampling, it is not uncommon for the inclusion probabilities of some primary sampling units to be set to one. For example, this can happen for populous counties when counties as primary sampling units are sampled proportional to population size. Thus, even standard applications of probability-based sampling can involve inclusion probabilities equal to one.

Several aspects of the NCS might lead to the use of purposive sampling. For example, in order to control the overall cost of medical data collection and improve the quality of such data, NIH may choose to solicit proposals from qualified medical centers with the objective of successfully negotiating contracts with a network of medical centers that would collect a large portion of the NCS medical data. This constraint could be accommodated within a probability-based sampling framework by setting the inclusion probabilities for the primary sampling units in which the targeted medical centers reside equal to one. Purposive sampling could also be used to include geographical areas that represent isolated areas of exposure, that represent the extremes of exposure conditions, that contain specialized facilities or equipment, or for which existing environmental exposure data already exists.

The only real limitation associated with purposive sampling is that the extreme over-sampling of purposively targeted elements of the sampling frame necessarily results in the remainder of the sampling frame being under-sampled.

A-5.2 PROBABILITY-BASED PURPOSIVE EXCLUSION

If one takes the concept of under-sampling certain subsets of the sampling frame to an extreme, it leads to setting the inclusion probability to zero for specific subsets of the sampling frame. In this case, rather than these elements of the sampling frame being under-represented, these elements are simply not represented at all. Alternatively, and perhaps more intuitively, one can view this process as defining the sampling frame to exclude certain subsets of the reference population. In this sense the sampling frame represents but a subpopulation of the reference population.

Purposive exclusion methods could be used, for example, to focus the NCS on a subpopulation with desirable properties with respect to cost-efficiency and/or data quality. The concern with this approach is the potential that relationships that are internally valid for the study population will be somehow systematically biased when extended to the larger reference population. Thus, purposive exclusion of elements of the reference population would raise questions about the external validity of relationships identified in the NCS data.

There are most definitely circumstances under which the purposive exclusion of elements of the reference population may be the statistically optimal sampling design approach. The reference population, because of its all-inclusive nature, may contain a sizable number of children that are hard to recruit, hard to retain, and/or more expensive with respect to data collection. Focusing on a study population of children that have desirable properties with respect to recruitment, retention, and cost-efficiency would allow more children to be included in the study or, alternatively, more information to be collected for the same number of children studied. In either case, more information would be available for identifying relationships between exposures and outcomes and, therefore this approach is attractive from the point of view of maximizing the amount of information produced by limited study resources. However, as the study population is narrowed to achieve better cost efficiencies, the exposure-response relationships in the study population may become systematically biased relative to the exposure-response relationships in the reference population. These trade-offs are often navigated as part of sample surveys when the sampling frame is constructed. For example, when household-based sampling frames are constructed for population surveys, homeless people, women living in battered women's shelters, and incarcerated people may be excluded from the sampling frame for practical reasons.

Potential approaches for targeting more cost-efficient sub-populations have varying degrees of specificity. Limiting study participants to those residing within 50 miles of a major medical center might offer some cost-efficiencies while yielding a sub-population that includes a significant percentage of the nation's children. At the other extreme, suppose that study participants were limited to the existing patients of a dozen medical centers that successfully negotiate NIH contracts to conduct portions of the NCS. In this case, the study population includes only the existing patients of the dozen medical centers, a very small percentage of the nation's children. All other things being equal, the smaller the study population relative to the reference population, the greater the potential for bias in exposure-response relationships.

The trade-off between improved cost-efficiency and systematic bias can be formulated quantitatively in terms of total error where total error includes the contributions of both systematic bias and random error. This trade-off is addressed in detail in Appendix A.

The advantages and limitations of purposive exclusion are fairly simply stated. The primary advantage of purposive exclusion methods is the ability to exclude subsets of the reference population that are expected to be difficult to recruit, difficult to retain, or more expensive with respect to data collection. The limitation that these exclusions impose is an inability to extend relationships identified in the NCS data to the full reference population on an empirical statistical basis. Empirical statistical arguments may be used to demonstrate external validity relative to the actual study population but not beyond the study population to the full reference population.

A-6. CONCLUSIONS

There are compelling arguments for the use of both probability-based and non-probabilistic approaches to sampling in the NCS. Probability-based sampling methods add value in terms of protection against unexpected systematic bias. The reality of anticipated low recruitment and retention rates diminishes but does not negate this value. In a similar fashion, the use of volunteer participants adds value assuming that they provide better assurance of continued participation throughout the duration of the NCS. The reality of systematic biases that are inevitably introduced by volunteer participants diminishes but does not negate this value.

It is likely not possible to reach any kind of scientific consensus on the clear superiority of either approach due to the uncertainty that surrounds implementation of a study as unprecedented as the NCS. In particular, there are no definitive data sources that allow the precise prediction of likely retention rates under competing sampling design options. Lacking precise retention rate predictions, it is hard to imagine a clear scientific consensus emerging for either a fully probability-based sampling design or a fully non-probabilistic sampling design

A hybrid sampling design that employs both probability-based sampling and non-probabilistic sampling would allow each set of methods to act as a hedge against the risks associated with the opposing approach. The resulting NCS database would address issues of both internal and external validity resulting in the identification of cause-and-effect relationships that can validly be extended to a population including most or all of the nation’s children.

In order to derive maximum benefit from the probability-based methods employed within a hybrid strategy, it will likely be necessary to take advantage of all the flexibility that such methods provide. Important considerations include:

- It may be advisable to focus attention on a study population (or sampling frame) that represents only a cost-effective subset of the reference population

- Including a wide range of exposures in the NCS cohort may be much more important than having the NCS cohort reflect the demographic characteristics of the reference population

- Over-sampling and perhaps even purposive sampling may be necessary to assure that cohort members have a wide range of exposures

- Purposive sampling may play an important role in allowing targeted resources such as qualified medical centers, specialized facilities/equipment, and existing environmental databases to be employed in conducting the study

- The use of unequal inclusion probabilities to over-sample cost-effective subsets of the study population may be necessary to control data collection costs while maintaining a reasonable level of complexity in the data collection protocol

- Cluster sampling may play an important role in
  - Controlling data collection costs
  - Creating a data structure that is conducive to examining phenomenon that occur at the neighborhood or census tract level

- Unless the hypotheses requiring large sample sizes are related to outcomes that occur early in a child's lifetime, it may be advisable to enroll only enthusiastic participants in the NCS cohort in an attempt to maximize retention rates; active conversion of reluctant participants may be ill-advised

Fortunately, probability-based sampling methods that incorporate elements of purposive inclusion and exclusion have sufficient flexibility to at least partially achieve many of the objectives that motivate the consideration of non-probabilistic methods. That said, the inclusion of some proportion of volunteers in the NCS cohort offers a benefit that probability-based methods cannot provide, that being a self-motivated cohort member that has the highest likelihood of retention until all required study data have been collected.

A-7. REFERENCES

Campbell, D.T. and Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research***.** Chicago: Rand McNally & Company.

Cochran, W.G. (1977), *Sampling Techniques*, New York: John Wiley & Sons.

Hahn, G.J., and Meeker, W.Q. (1993), "Assumptions for Statistical Inference," The American Statistician, 47, 1-11.

Kish, L. (1995), "The Hundred Years' Wars of Survey Sampling," Statistics in Transition, 2, 813-830.

Little, R. (1982), "Models for Nonresponse in Sample Surveys," Journal of the American Statistical Association, 77, 237-250.

Little, R. (2003), "To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling," U. of Michigan Dept. of Biostatistics Working Paper Series, Year 2003, Paper 4.

Rubin, D.B. (1976), “Inference and Missing Data,” Biometrika, 63, 581-92.

Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons.

Rubin, D.B. (1996), “Multiple Imputation After 18+ Years,” Journal of the American Statistical Association, 91, 473-489.

Scott, A.J. (1977), “On the Problem of Randomization in Survey Sampling,” Sankhya C, 39, 1-9.

Scott, A.J. and Smith, T.M.F. (1973), “Survey Designs, Symmetry and Posterior Distributions,” J R Statist Soc B, 35, 57-60.

Smith, T.M.F. (1976), “The Foundations of Survey Sampling: A Review,” J R Statst Soc A, 139, 183-204.

Smith, T.M.F. (1983), “On the Validity of Inferences from Non-Random Samples,” J R Statist A, 146, 394-403.

Smith, T.M.F. (1994), “Sample Surveys 1975-1990: An Age of Reconciliation?” International Statistical Review, 62, 5-34.

Sugden, R.A. and Smith, T.M.F. (1984), “Ignorable and Informative Designs in Survey Sampling Inference,” Biometrika, 71, 495-506.

Valliant, R., Dorfman, A.H., and Royall, R. (2000), *Finite Population Sampling and Inference*, New York: John Wiley & Sons.

APPENDIX A-A

RATIONALE FOR SAMPLING FROM SUBPOPULATIONS

If valid statistical inferences for an entire reference population are sought, the logical sampling approach would seem to be a probability-based sample from the entire reference population. If unbiased statistical inferences are required, this might be the only valid approach. However, unbiased statistical inferences are seldom actually required of a study. Instead, the actual requirement is for statistical inferences that are approximately valid for the reference population. In cases where the cost of obtaining study information is lower for certain elements of the reference population and higher for others, it can be statistically optimal to study a biased portion of the reference population where cost-efficiencies are possible.

For estimation of a statistical parameter, such an approach can be justified in terms of minimizing the total error of estimation. The statistical parameter could be the prevalence of exposure to particular environmental contaminant or the odds ratio between an exposure variable and a health outcome. Thus, the total error of estimation argument applies equally well to descriptive parameters and parameters that characterize the strength of relationships.

Total error has two components, systematic error and random error. Systematic error is that part of the estimation error that derives from (1) differences between the study population and the reference population and (2) systematic biases in the measurement protocols employed to collect study data. Systematic error is unaffected by the size of the sample taken. Random error, on the other hand, derives from (A) differences between the sample and the study population and (B) random measurement error associated with the measurement protocols employed to collect study data. Random error is reduced as the sample size increases.

A widely-employed version of the total error concept is embodied in the mean square error (MSE) of a parameter estimator. The MSE of the estimator is the expected squared deviation of the estimator from the true parameter value. The corresponding systematic error component is represented by the bias of the estimator squared, while the random error component is represented by the estimator variance. With these definitions, we have

$$\text{MSE} = (\text{Bias})^2 + \text{Estimator Variance}$$

or

$$\text{Total Error} = \text{Systematic Error} + \text{Random Error}.$$

Consider the case where the study population is equivalent to the reference population. In this case, systematic error is minimized because there are no differences between the study and reference populations. Because the entire reference population is studied, the sample must include sample elements for which data collection is more expensive. Alternatively, consider the case where the study population is defined to exclude elements of the reference population for which data collection is more expensive. Because the study population is now more cost-efficient, it is possible to observe a larger sample using the same data collection resources. The larger sample size results in a reduction in random error. However, since there are differences between the study and reference population, systematic error has now been introduced. If the reduction in the random error component is larger than the increase in the systematic error component, then it is statistically optimal to study the smaller, more cost-efficient subpopulation because the total error of estimation is smaller.

The preceding paragraph demonstrates that it is possible to justify studying a cost-efficient sub-population as statistically optimal. While it is possible to do so, this justification is rarely formally completed for various reasons. It is generally difficult to quantify the systematic errors and increased cost-efficiencies associated with a proposed sub-population, and therefore difficult to formally compare the increase in systematic error to the expected reduction in random error. Instead, this comparison is made in an approximate manner at a more general level by asking and answering in a general fashion questions such as the following:

- Will statistical conclusions drawn from the proposed subpopulation be approximately valid for the entire subpopulation?
- Will improved cost-efficiencies associated with the proposed sub-population provide an opportunity to make much stronger statistical inferences about the sub-population than would be possible for the entire population?
- Will the increase in strength of the statistical inferences for the subpopulation be large enough to counteract any systematic error associated with the sub-population relative to the reference population as a whole?