October 20, 2003
Appendix B1

Draft White Paper

On

Criteria for Evaluation of NCS Design Options

by

John Menkedick, Warren Strauss, and Jeff Lehman

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

Draft White Paper

On

Criteria for Evaluation of NCS Design Options

John Menkedick, Warren Strauss, and Jeff Lehman

Battelle

B1-1. INTRODUCTION

The purpose of this document is to present and discuss proposed criteria that can be used to choose between different study design options for the National Children's Study (NCS). The intent is to present the proposed criteria as succinctly as possible, providing general discussion of the rationale for, and key features of, each criterion, while referencing other documents for further details and specificity.

NCS Objectives and Guiding Principles

The Children's Health Act of 2000 has authorized the National Institute of Child Health and Human Development (NICHD) to conduct the NCS. The language in the legislation calls for "a national longitudinal study of environmental influences (including physical, chemical, biological, and psychosocial) on children's health and development." The additional direction in the legislation is sparse but critically important. It calls upon the Director of NICHD to "establish a consortium of representatives from appropriate Federal agencies (including the Centers for Disease Control and Prevention, and the Environmental Protection Agency) to –

(1) plan, develop and implement a prospective cohort study, from birth to adulthood, to evaluate the effects of both chronic and intermittent exposures on child health and human development; and

(2) investigate basic mechanisms of developmental disorders and environmental factors, both risk and protective, that influence health and developmental processes.

Finally, the legislation requires that the study shall:

> "(1) incorporate behavioral, emotional, educational, and contextual consequences to enable a complete assessment of the physical, chemical, biological and psychosocial environmental influences on children's well-being;
>
> (2) gather data on environmental influences and outcomes on diverse populations of children, which may include the consideration of prenatal exposures; and
>
> (3) consider health disparities among children which may include the consideration of prenatal exposures."

The language in the legislation provides the guiding principles and overall objectives for the NCS. The distinguishing features of the NCS – what makes the study an unusual if not unique research opportunity – are its size (100,000 children), its duration (pre-natal to adulthood) and its comprehensive charge to assess multiple effects on diverse populations. ***Together, these guiding principles and objectives provided by the legislation and the distinguishing features of the NCS underlie evaluation criteria proposed in this document, and should guide all subsequent decisions requiring difficult tradeoffs between different study design options.***

Criteria are identified during the course of the discussion in the document, and as each is identified it is highlighted in bold and italics and labeled for ease of reference. Table 1 at the end of the document includes a listing of all 21 criteria that are proposed in this document, four related to study givens, fifteen related to scientific merit, and two related to cost.

Categories of Criteria

There are three primary types of criteria that will be used to judge any proposed design. These are:

1. Ability to satisfy "Givens" (constraints) required by the legislation or the government
2. Scientific merit (validity and value)
3. Cost.

Each will be discussed separately below. Section 2 discusses those criteria related to satisfying the study "Givens", Section 3 presents criteria related to the scientific merit of the study, and Section 4 reviews criteria related to cost of the study. Included under scientific merit are criteria to assess each design for its ability to meet the core hypotheses of the study, ability to serve as a general resource for future studies, and

feasibility. While there is obviously a relationship between feasibility and cost (some might maintain that anything is feasible if the cost is unlimited), the two concepts are distinguished here. Criteria related to cost involve estimates of direct expenses that can be planned and controlled and that would represent a budget for the study. Feasibility criteria, on the other hand, represent estimates of the degree to which the study will work – for example, will people participate and remain in the study or will measurements be effectively standardized across all participants. In this way, cost criteria can be assessed relative to budget constraints decided upon by the government. Feasibility criteria on the other hand will be assessed relative to the probability of impacting the overall scientific merit of the study results.

B1-2. GIVENS

Following is a discussion of "Givens" or constraints that have been imposed by the legislative requirements or government study leaders. These represent requirements for the NCS that each design option must address. Some translate directly into criteria for evaluating study design options (such as the community involvement mandate), while others have an indirect effect, for example influencing the technical details of criteria used to evaluate scientific merit (such as the requirement for an observational study).

(1) The study will be observational in nature and will address multiple environmental influences on a diverse population.

The fact that the NCS is charged with looking at a wide range of environmental influences and their potential interactions, combined with ethical issues, mandates that the NCS be conducted as an observational study. This has significant implications. The fact that exposures cannot be randomly assigned to the study population introduces a host of additional complexity into the process of drawing inferences about observed associations. For example, two critical questions for an observational study are:

> Is it possible that the association is due to some other factor that is confounded with the exposure of interest? In observational studies with the objective of assessing causal relationships, this is perhaps the most significant concern that can lead to erroneous inferences. Therefore, in an observational study, the degree to which a design allows for collection of information on potential confounding factors that might be related to any associations of interest is one of the most important factors in determining the ultimate scientific value of the study.
>
> Has the study observed a sufficient range of exposures? Since exposures are not controlled, only observed, this can significantly affect the expected power of a study to detect effects.

Inferring causality in an observational study will inevitably require scientific assessment beyond statistical analysis of the collected data. Design evaluation criteria that take into account these issues associated with the constraint of an observational study design are presented in the Scientific Merit section (Section 3) below.

(2) The study will be national in scope, broad-based, inclusive of a wide range of populations and geographic diversity, and as representative as possible given tradeoffs with other features of scientific value to the study objectives.

This constraint will be translated into multiple criteria based on technical details discussed under the Scientific Merit section below.

(3) The study will include a large sample (approximately 100,000) – to allow for evaluation of rare exposures and outcomes; and of interaction of environmental factors and genetics.

All proposed design options will meet this given. Criteria to evaluate differences in designs relative to their ability to retain the cohort over time will be presented in the Scientific Merit section.

(4) The study will include pre-natal recruitment, as early in pregnancy as possible. The study will consider recruitment of some, or all, participants before pregnancy. The study will include clustering of samples to allow for efficient collection of exposure and outcome measures. The study will consider stratification to obtain a) an adequate range of exposures (including social), b) socio-economic, racial/ethnic/geographic diversity, and c) population subgroups of interest.

The criteria for assessment of how these givens are achieved in each design are presented along with other technical details as part of the discussion in the Scientific Merit section.

(5) The study will be locally based to foster community involvement.

This constraint leads to the first criterion for evaluating design options.

Givens Criterion 1: ***To what degree does the study engage the local community? To what degree does it create awareness, infrastructure, outreach, and a sense of community? To what degree does it leverage community involvement to maximize recruitment and retention, and in turn, is leveraged to enhance community respect and help local community public health efforts?***

With respect to the relationship between the NCS and local community public health efforts, it is important that the NCS does not interfere with any type of public health interventions pursued by a local community. However, it is also critical that the NCS captures information regarding how these intervention efforts might affect the NCS measures and hypotheses. This criterion attempts to capture these two important elements.

(6) The study will include infrastructure to support specialized measures (e.g. medical facilities with technologies such as 3D ultrasound).

The purpose of this constraint is primarily to ensure that the study is designed in a way that it can take advantage of medical technology and can be responsive as a resource for future studies and assessment by having access to specialized measures.

Givens Criterion 2: ***To what degree does the study provide access to infrastructure that allows specialized measures?***

(7) The study will provide access/collection of biological samples at birth.

This constraint is related to, but not identical to the constraint above.

Givens Criterion 3: ***To what degree does the study provide for ease of access/collection of biological samples at birth?***

(8) The study will provide flexibility to conduct special studies (e.g. special population groups, pre-conception recruitment, or topics of community interest).

These may be sub-samples collected outside the core design, may be coordinated with the local community interests, and may be non-probability based,

Givens Criterion 4: ***To what degree does the study provide flexibility to conduct special studies, particularly related to topics of community interest?***

B1-3. SCIENTIFIC MERIT

As discussed in the Givens section above, the criteria for scientific merit are affected and shaped both by the study objectives and the constraints placed on the type of study to be conducted. The most important aspect of the study objectives affecting evaluation of scientific merit is that the study is intended to understand ***associations and causal effects*** of environmental influences on child health and development. The most important constraints on the NCS study design affecting scientific merit, as discussed in the Givens section above, are that the study will be ***observational in nature***, and must be

designed to assess ***multiple environmental influences and outcomes on diverse populations*** of children. These three features, taken together, have a profound impact on what the study requires in terms of inferences, representativeness, and generalizability. Figure 1 below illustrates the primary factors that affect the ability to draw inferences about causal effects given an observational study assessing multiple environmental influences and their interactions. Each element of the figure leads to specific criteria that can be used to assess the strength of each design option, and will be discussed separately in the following sections.

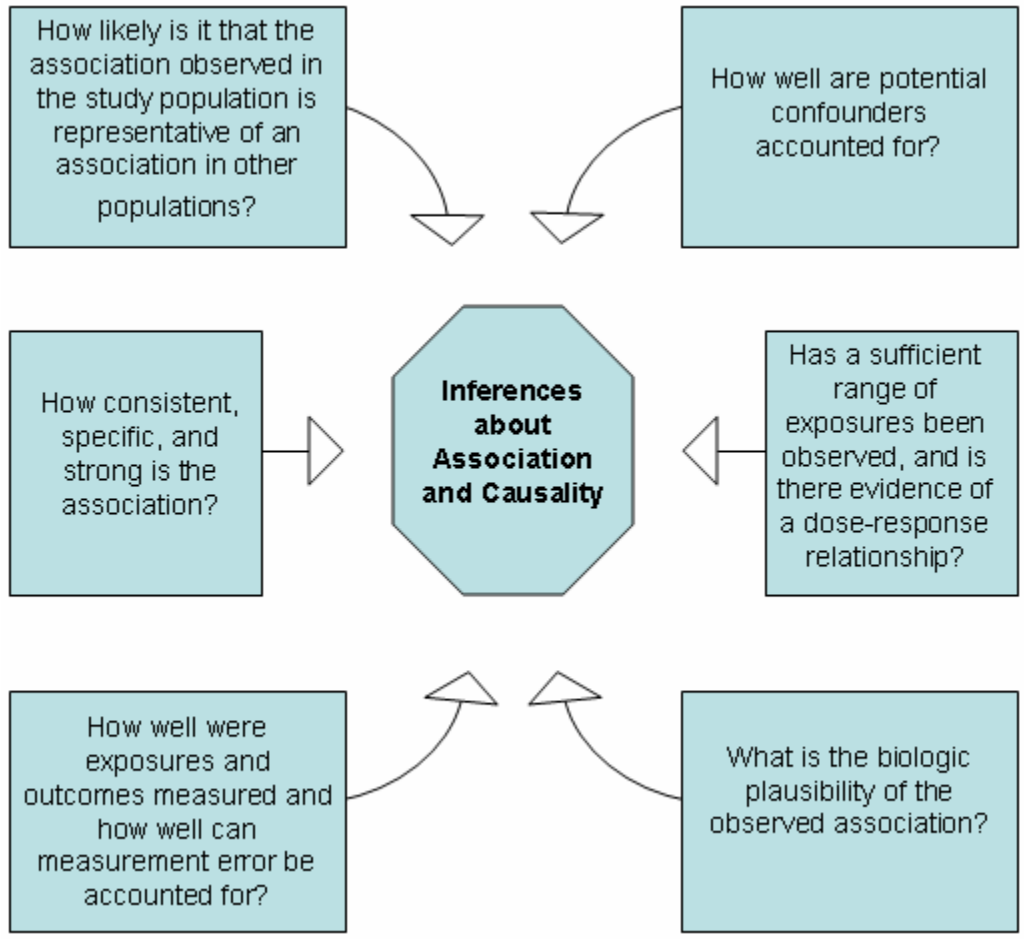

Figure B-1. Primary factors affecting causal inferences in the NCS

B1-3.1 REPRESENTATIVENESS AND GENERALIZABIITY

The question of representativeness and generalizability when applied to associations and causal relationships is more complex than when just applied to prevalence or incidence statistics. While it is not the purpose of this paper to provide an extensive discussion of "representative" sampling, the appropriateness of "probability" or

"selective" methods of sampling for different objectives, and statistical methods for adjusting for things like non-randomness – these are the subjects of other NCS white papers – there are several key points on this topic that might help clarify the choice and wording of the proposed criteria that address the degree of "representativeness and generalizability" of the different NCS design options.

The first point concerns the ultimate population about which inferences will be drawn. The terminology of Deming, as discussed by Hahn and Meeker (Hahn and Meeker, 1993), in which a distinction is made between an enumerative and an analytical study is useful. An enumerative study (Deming, 1975) is one in which "action will be taken on the material in the frame studied". For example, a study that is intended to estimate the prevalence of an exposure or health outcome in a specific timeframe could be considered an enumerative study. An analytic study, in contrast, is one in which actions taken as a result of the study will be applied outside the sampling frame of the study. In the case of the NCS, while the sampling frame will likely include some subset of children born between 2006 and 2009, the relevant populations about which inferences will be drawn will certainly be much broader than that frame. For example, inferences from the study will certainly be used to inform public health practices applied to children born well after 2009. Therefore, the first important point is that the NCS represents, in Deming's terminology, an analytic study and thus brings with it the requirements of an analytic study, including:

> Involvement of subject matter scientific expertise to extrapolate inferences beyond the sampled population. (This is reflected in Figure 1 in the factors affecting inferences other than "representativeness".)
>
> Need to design sample selection and data collection in a way that provides the information needed to determine whether results are applicable to populations beyond those in the sampling frame.

In light of the ultimate population to which inferences will be applied, the concept of what might constitute a "representative" sample is important. There is a history of literature dating back to a series of articles by Kruskal and Mosteller (Kruskal and Mosteller, 1979) on representative sampling that address what might be considered representative in light of different scientific objectives, and this is the subject of a separate white paper. Different concepts of a representative sample include:

> Being random or absent of selective forces
> Being a miniature of the population
> Being typical
> Having coverage of the heterogeneity of the population
> Being drawn by a specific sampling method
> Permitting good estimation of parameters defined by study objectives
> Being representative of a typical biological response, dose-response relationship, or other relationship specific to a study's objectives.

Thus, representative can imply more than just unbiased – requiring representation of certain characteristics of the population.

Of particular note is representativeness relative to the relationship between exposure and a health outcome in an observational study. In general, it is often postulated that statistical requirements for random sampling to eliminate selection bias is less critical in studies of association, as opposed to studies of prevalence or enumeration. This may explain why many epidemiological studies assessing the etiology of disease do not include random selection of participants, choosing instead study designs that are intended to maximize the amount of measurements that can be taken from participants and the retention of participants over time, rather than minimizing selection bias.

It may well be the case that certain core hypotheses in the NCS will seem less susceptible to concerns about representativeness than others. For example, a core hypothesis that assesses the effect of a chemical exposure on a specific target organ in the study population may be considered more easily generalizable to other populations than one that addresses the effect of a few specified parental psycho-behavioral factors on the frequency of injury, where there might be more concern about extrapolating the relationship outside the study population without a sound statistical basis or extensive investigation of ways in which this population might not be representative. In most cases, the most significant issue with selection bias is that it can exacerbate the confounding problem, since a selective sample that shares certain traits may make it more problematic to detect and account for the confounding factors that are behind the biased relationship in the selective population. It should be noted that in observational studies the effect of eliminating selection bias through probability sampling can aid in reducing the effects of potential confounding factors.

Finally, based on preliminary work on the likely sampling frames and response rates for the NCS, the criteria proposed relative to potential bias introduced into the sample due to the sample selection method is assessed assuming a continuum of bias, and a continuum of statistical adjustment for such biases. By this we mean that we assume that any sample selected for the NCS will have a component (likely significant) of non-response and therefore non-randomness, and will require statistical adjustments, and that statistical adjustments will be available for any design that includes some subset of the population which is randomly selected from a known frame.

The following four criteria address the merit of design options relative to representativeness, generalizability, and the ability to draw inferences from study results.

Merit Criterion 1: ***Does the design clearly specify a sampling frame from which the study population is drawn, and can statistical methods be used to generalize results and characterize uncertainty for the sampling frame population?***

Merit Criterion 2: ***What statistical adjustments are necessary and available to account for non-response, non-random selection, or other deviations from specified probability selection from the sampling frame for the design?***

Merit Criterion 3: ***To what degree does the design allow recruitment of a diverse population – geographic, ethnic, socio-economic, and other factors determined by study leaders. To what degree does the design ensure that target populations or exposures are represented in the study, and that heterogeneity in the study population is captured?***

Merit Criterion 4: ***Are there aspects of the design that will aid in collecting information or understanding characteristics of observed associations or characteristics of the population in which associations are observed that will help in extrapolating results beyond the sampling frame subset of children.***

B1-3.2 CONFOUNDERS

The critical importance of measuring factors that might be confounded with the hypothesized associations of interest has been discussed above and cannot be overstated. It is particularly important in meeting the objective of using the NCS as a resource for future studies where a reservoir of information on potential confounders – particularly on exposures – may prove critical to investigating new hypotheses. The degree to which any particular design can accommodate the collection of information on confounding factors may be influenced by cost constraints, availability of infrastructure (e.g. neighborhood clinics or health centers), and willingness of subjects to incur extra burden. The ability of a design to cover the range of values expected for each of the important confounders could be aided by a variety of approaches, including stratified sampling and/or oversampling of specific segments of the population.

Merit Criterion 5: ***To what degree does the design support collection of information on a wide range of potential confounding factors that might influence associations of interest? This includes both a) ability to measure a significant number of potential confounding factors; and b) ability to obtain a range of values for important confounders.***

B1-3.3 RISK CHARACTERISTICS, EXPOSURES, AND OUTCOMES

The first criterion related to the design's impact on measurement of risk characteristics, exposures, and outcomes is the degree to which the sample design helps ensure that the sample population will result in observing the type and range of risk characteristics, exposures, and outcomes of interest.

Merit Criterion 6: ***To what degree does the design maximize the probability of observing the type and range of risk characteristics and exposures of interest, or maximize the probability of observing an outcome of interest?***

The second criterion related to exposures and outcomes recognizes that the study fundamentally involves three sampling components: sampling the population, sampling risk characteristics and exposures, and sampling outcomes. Sampling risk characteristics and exposures is certainly as important as sampling the population. Historically, in many studies the exposure assessment has been the weak link in attempting to understand risk factors associated with health outcomes. While outcome measures are not usually associated with sampling, some outcomes of interest to the NCS, such as injury and neurobehavioral effects, will involve a sampling assessment. Although this paper is focused on the population selection, an important criterion is whether a given design is more conducive to effective sampling of exposures or outcomes. NCS core hypotheses suggest the testing of very specific relationships between health outcome and exposure, however the specific mechanisms and metrics of exposure related to the health outcome may not be well understood at the current time. Obtaining representative, accurate, and effective measures of exposure may be problematic in terms of the original data collection effort. Therefore, the degree to which a design offers the flexibility to more fully characterize exposure or to achieve this goal through statistical adjustments should be assessed, and designs that offer significant advantages in terms of ability to adequately characterize exposure and the uncertainty in exposure be identified.

Merit Criterion 7: ***To what degree does the design help in obtaining representative, accurate, and effective measures of risk characteristics, exposures, or outcomes on the sampled population? To what degree does it minimize measurement error or allow for statistical adjustments to account for measurement error?***

B1-3.4 ADDRESSING CORE HYPOTHESES - MINIMIZING ERROR AND MAXIMIZING POWER

When assessing differences in study designs relative to their ability to address the core hypotheses of the study, it is helpful to conduct the assessment in terms of the probability of a given design to minimize both Type 1 statistical error (the chance of

concluding there is an effect when in fact there is not) and Type 2 error (the chance of concluding there is no effect when in fact there is). Type 2 error is usually characterized by examining the power of the study. Power represents the probability of correctly concluding that there is an effect when an effect of specified size is present.

From a statistical standpoint, Type 1 errors are controlled by the choice of alpha level for statistical tests conducted as part of the data analysis. Related to Type 1 errors are errors resulting from concluding there is an effect when the effect may be due to confounding, or when measurement of the effect has been biased due to model misspecification or biases in measurement. Scientific Merit Criteria 2, 4, 5, and 7 all measure aspects of the design that could contribute to erroneously concluding there is an effect when none exists.

Type 2 errors are generally assessed by examining the power of the study to detect a specified effect level (for example an odds ratio of 1.5) for a key study hypothesis. For a study like the NCS, with multiple hypotheses and multiple inferences of interest, there are many ways to assess power and the results can be quite dissimilar.

First, power is dependent upon the statistical model chosen to characterize the relationship. In particular, power may be greatly increased when the data and underlying biological model support estimation of a continuous dose-response relationship between an exposure and outcome (as opposed to estimation of a categorical effect such as an odds ratio). Some hypotheses will effectively define the appropriate statistical model while others will allow for different statistical model options. Therefore, power will be assessed for several key hypotheses with both categorical and continuous statistical analysis models.

Power is also heavily dependent on the inference space – whether inferences are being drawn for only the study population (those people who are actually participating in the study), or for a broader population that could expand up to the entire sampling frame from which the study population was drawn. Conceptually, this can be thought of as doing separate power analyses associated with a weighted or un-weighted analysis of the data (i.e., use or no use of sampling weights, if they are calculable). As discussed above, the ability to detect statistically significant relationships within the study population alone will likely be of interest, and may be combined with other scientific assessments to generalize to other populations. Likewise, when possible, the ability to estimate the statistical significance of a hypothesized relationship for a broader population within the sampling frame is also extremely valuable – providing a basis for inferring significant relationships based on the statistical analysis alone. Therefore, power will be assessed, when possible, for both inferences that only apply to the study population, as well as inferences that can be generalized to the wider sampling frame population. [Inferences beyond the specified sampling frame (e.g., children of non-institutionalized pregnant women with known addresses during the years 2006-2009) of the study can not be based on statistical analysis alone.]

Merit Criterion 8: ***Assuming confounding factors have been appropriately accounted for, what is the power of the study design to detect the target associations of interest for:***
multiple hypotheses associated with both categorical and continuous models;
for inferences that are limited to the study population of participants, as well as, when possible, for inferences to a broader population.

In addition to the power to detect specified core hypotheses in the study, there is also reason to address, in a qualitative way, the question of the power of the study to serve as a resource for future studies and as a data source for assessment of hypotheses that will arise in the future. While not a statistical concept, the power of the study to meet this objective will depend on such things as the quality of the exposure assessment, the representativeness of the sample population, the standardization of measures, the retention of participants over time, and the flexibility to incorporate new measures if necessary. Examples of future use of the data in addition to testing hypotheses that are unrelated to original NCS core hypotheses, include identification of major issues for additional research funding, and use of the data subject to scientific, legal and public scrutiny for supporting policy and regulatory decisions by federal agencies.

Merit Criterion 9: ***What aspects of the study design make it particularly strong in terms of serving as a resource for future studies and assessment of hypotheses identified in the future?***

B1-3.5 FEASIBILITY

The feasibility criteria assess the degree to which the proposed study design is likely to be implementted without significant compromise or problems.

The first measure of feasibility is the estimated response rate associated with the study design. Note that the cost implications associated with the estimated response rate are covered as part of the Cost criteria section below.

Merit Criterion 10: ***What is the estimated response rate for the design option? How will this affect inferences and what statistical adjustments can be applied to account for the non-response?***

The second measure of feasibility is the estimated rate of attrition of study participants from the study. While this must be assessed from a qualitative viewpoint in terms of how it will limit the usefulness of the study as a resource over time, it can also be assessed quantitatively by examining the effect on the power of the study to assess the core hypotheses.

Merit Criterion 11: ***What is the estimated rate of attrition for the design option? How does this rate of attrition affect the power of the study as characterized in Scientific Merit Criterion 8?***

An extremely important characteristic of the NCS design is the ability to collect the same measure in the same manner across all study participants. While statistical adjustments for differences in measurement methods and for missing measures are possible, significant variability across the study population in what was measured and how it was measured can have a devastating effect on the ability of the study to assess relationships, especially in the case of rare outcomes or exposures.

Merit Criterion 12: ***To what degree does the study design option maximize the chance of obtaining standard measures across the entire cohort throughout the life of the study? To what degree does the study design option support incorporation of QA/QC training and monitoring?***

All estimates point to significant mobility of the study cohort over the life of the study. A study design's ability to handle moves is critical.

Merit Criterion 13: ***How well does the study design option account for mobility in the study population – in terms of retention of study participants and collection of all required measures?***

The ability of the study design option to obtain human subjects approvals, to avoid any ethical show-stoppers, and to obtain approval for standard measures across all subjects is critical to successful implementation.

Merit Criterion 14: ***How will human subjects approval be handled by the study design and what complications might be expected?***

Finally, another important aspect of the study design is the degree to which it facilitates successful compilation and use of the data in a timely fashion.

Merit Criterion 15: ***What aspects of the study design allow for effective and timely access to the study database by a variety of researchers?***

For example, a design that involves multiple data collection organizations may provide immediate access of local data to local investigators, while delaying the assembly of a comprehensive national data source. On the other hand, data collection by a single (or few) organization(s) may streamline the assembly of a comprehensive data source, while failing to allow most researchers interim access to subsets of the data because of their lack of involvement in the implementation of the study. In addition to the organizational structure for implementing the NCS, other issues that may affect access to the study

database and must be considered during the design phase include confidentiality of data and local IRB concerns.

B1-4. COSTS

Costs for each study design option will be developed based on the best estimate of implementation details and reasonable expectations of level of effort associated with the design option. When there are reasonable alternatives relative to cost, both will be presented. For example, if a design option is expected to incur a specified amount of costs to achieve a 50% response rate, but twice the costs to achieve an 80% response rate, both options may be presented. However, if ethical concerns over incentives would make implementation of the plan for an 80% response rate very unlikely, then costs would only be estimated for the likely implementation scenario for the proposed option. In this way, costs are kept as independent of scientific merit as possible.

Cost Criterion 1: ***What are the estimated costs of the study for the following categories:***

Study design and inception
Training and standardization
Recruitment
Data collection
Study management and operations
Information management
Retention
Community involvement

[this list needs comment and revisions]

Cost Criterion 2: ***What are the opportunities of the study design to obtain cost sharing from other organizations, such as local communities, health departments, or medical centers?***

B1-5. CONCLUSIONS

We conclude our discussion by providing a Table that summarizes the 21 criteria for evaluating NCS study designs (see Table 1). As noted previously, future work involves utilization of these criteria to evaluate and rank the options for NCS study designs. To that end, one necessary component of these criteria that has not been defined is the relative importance of each criterion (referred to as a weighting factor in Table 1). Certainly, some design options will adequately address some criteria, and fail to address other criteria, while other designs will adequately address these other criteria. Thus, the optimal design will depend on which criteria are the "most important" criteria. For example, a design that involves a convenience sampled cohort may offer a large cost savings over a design that involves a probability based sampling scheme; however, the

convenience based sampling design may not have the generalizability characteristics that are desired for the NCS cohort. If cost is the primary concern, then perhaps the design that calls for a convenience sample should be utilized, but if generalizability and sample representativeness are the primary concerns, then perhaps a probability based sampling design should be used. Of course, this is a simple example that only considers two of the criteria, and these types of determinations become much more difficult when we consider all of the criteria presented above.

For this reason, it is necessary to determine the relative importance of each of the 21 criteria. In Table 1, we include a "weighting factor" column to indicate the importance of each criterion. Those criteria that have a higher weight would be considered "more important", while those criteria that are not as important would be given lower weights. As indicated in the table, once the weighting factor is determined for each criterion, the candidate designs can be "scored" for each criterion, and a total score for the candidate designs can be obtained (e.g., by calculating the weighted average of the scores over all the criteria). Based on these total scores for each of the candidate designs, promising designs can be identified and further evaluated, while poor designs can be "weeded out."

To meet these goals, future work involves the following:

Determine the relative importance of each criterion through consultation with the NCS Program Office and the Study Design Work Group members.
Determine appropriate metrics for measuring the degree to which a candidate design satisfies a given criterion.
Evaluate candidate designs.

Table B1-1. Criteria Worksheet for Evaluating NCS Study Design Options

Criteria				Score		
Category	**Criterion #**	**Description**	**Weighting Factor**	**Des A**	**Des B**	**Des C**
GIVENS	1	To what degree does the study involve the local community? To what degree does it create awareness, infrastructure, outreach, and a sense of community? To what degree does it leverage community involvement to maximize recruitment and retention, and in turn, is leveraged to enhance community respect and help local community public health efforts?				
	2	To what degree does the study provide access to infrastructure that allows specialized measures?				
	3	To what degree does the study provide for ease of access/collection of biological samples at birth?				
	4	To what degree does the study provide flexibility to conduct special studies, particularly related to topics of community interest?				
SCIENTIFIC MERIT	1	Does the design clearly specify a sampling frame from which the study population is drawn, and can statistical methods be used to generalize results and characterize uncertainty for the sampling frame population?				
	2	What statistical adjustments are necessary and available to account for non-response, non-random selection, or other deviations from specified probability selection from the sampling frame for the design?				
	3	To what degree does the design allow recruitment of a diverse population – geographic, ethnic, socio-economic, and other factors determined by study leaders. To what degree does the design ensure that target populations or exposures are represented in the study, and that heterogeneity in the study population is captured?				
	4	Are there aspects of the design that will aid in collecting information or understanding characteristics of observed associations or characteristics of the				

Criteria				Score		
Category	Criterion #	Description	Weighting Factor	Des A	Des B	Des C
		population in which associations are observed that will help in extrapolating results beyond the sampling frame subset of children.				
	5	To what degree does the design support collection of information on a wide range of potentially confounding factors that might influence associations of interest?				
	6	To what degree does the design maximize the probability of observing the type and range of risk characteristics and exposures of interest, or maximize the probability of observing an outcome of interest?				
	7	To what degree does the design help in obtaining representative, accurate, and effective measures of exposure or outcome on the sampled population? To what degree does it minimize measurement error or allow for statistical adjustments to account for measurement error?				
	8	Assuming confounding factors have been appropriately accounted for, what is the power of the study design to detect the target associations of interest for: multiple hypotheses associated with both categorical and continuous models; for inferences that are limited to the study population of participants, as well as, when possible, for inferences to a broader population.				
	9	What aspects of the study design make it particularly strong in terms of serving as a resource for future studies and assessment of hypotheses identified in the future?				
	10	What is the estimated response rate for the design option? How will this affect inferences and what statistical adjustments can be applied to account for the non-response?				
	11	What is the estimated rate of attrition for the design option? How does this rate of attrition affect the power of the study as characterized in Scientific				

Criteria				Score		
Category	**Criterion #**	**Description**	**Weighting Factor**	**Des A**	**Des B**	**Des C**
		Merit Criteria 8?				
	12	To what degree does the study design option maximize the chance of obtaining standard measures across the entire cohort throughout the life of the study? To what degree does the study design option support incorporation of QA/QC training and monitoring?				
	13	How well does the study design option account for mobility in the study population – in terms of retention of study participants and collection of all required measures?				
	14	How will human subjects approval be handled by the study design and what complications might be expected?				
	15	What aspects of the study design allow for effective and timely access to the study database by a variety of researchers?				
COSTS	1	What are the estimated costs of the study for the following categories: • Study design and inception • Training and standardization • Recruitment • Data collection • Study management and operations • Information management • Retention • Community involvement *[this list needs comment and revisions]*				
	2	What are the opportunities of the study design to obtain cost sharing from other organizations, such as local communities, health departments, or medical centers?				

B1-6. REFERENCES

Deming, W.E. (1975). On probability as a basis for action. The American Statistician 29: 146-152.

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

Kruskal W. and Mosteller F. (1979). Representative Sampling, I: Non-scientific Literature. International Statistical Review 47: 13-24.

Kruskal W. and Mosteller F. (1979). Representative Sampling, II: Scientific Literature, Excluding Statistics. International Statistical Review 47: 111-127.

Kruskal W. and Mosteller F. (1979). Representative Sampling, III: The Current Statistical Literature. International Statistical Review 47: 245-265.

Kruskal W. and Mosteller F. (1980). Representative Sampling, IV: the History of the Concept in Statistics. International Statistical Review 48: 169-195.