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

Appendix B3

Meeting Summary: NICHD Sample Design Project (October 30-31, 2003)

Meeting Summary
NICHD Sample Design Project
October 30-31, 2003

B3-1 EXECUTIVE SUMMARY

On October 30-31, 2003, Battelle hosted a meeting to discuss the sampling design of the National Children's Study (NCS). Participants included individuals from the NCS program office, two expert consultants with significant experience in the field of sampling and study design, and Battelle staff currently working on the sample design project. Specifically, the participants were:

Peter Scheidt and Jim Quackenboss (NICHD / NCS Program Office)
Alan Zaslavsky (Harvard Medical School)
Colm O'Muircheartaigh (University of Chicago / NORC)
Warren Strauss, John Menkedick, Jeff Lehman (Battelle)

Additionally, four Battelle guest speakers (Steve Rust, Nancy McMillan, Nancy Reiches, and Ben Pierce) provided presentations related to topics of particular interest in the design of the NCS (see meeting agenda in Section 2). Over the course of the two-day meeting, a variety of topics were discussed, many questions (some with answers and some without) were posed, and several avenues for further investigation were identified. This report attempts to briefly summarize and document the meeting discussions and any of the resulting conclusions.

The main purpose of the meeting was to continue the process of identifying and specifying promising design options for the NCS through a collaborative effort of the meeting participants. To this end, the meeting (see agenda provided in Section 2) was to begin with a series of presentations relating some of the work already completed in support of this project. More particularly, following an overview of the NCS, a discussion of the criteria for evaluating NCS design options and a discussion of a framework of sampling design options for selecting the NCS longitudinal cohort was scheduled (note that the corresponding Battelle draft documents had been supplied to and perused by the study participants prior to the meeting). This was followed by a series of five presentations on important design considerations for the NCS. The remainder of the meeting consisted of a series of brainstorming sessions on various aspects of this project. While this schedule was not strictly followed (in most cases the presentations themselves turned into brainstorming sessions on the topic of the presentation), it represents a loose outline of the structure of the meeting, and indicates many of the topics that were covered over the course of the two days.

The Battelle draft documents on criteria for evaluation of NCS designs[1] and sampling design options for selecting the NCS longitudinal cohort[2] were meant to serve as starting points for our discussion. Since all of the meeting participants indicated that

they had already perused these documents, the meeting quickly turned to a discussion that hit on topics relevant to both evaluation of design options in light of the requirements for the NCS and to important design considerations that need to be kept in mind. One of the primary themes throughout these discussions was the need to consider all possible options and not summarily eliminate any design options without carefully considering their advantages and disadvantages.

The discussion of the criteria for evaluation of NCS designs focused on two topics, the completeness of the criteria (i.e., are there any other important criteria that were missed) and how to apply the criteria to identify promising design options. The group identified one important missing "given" which calls for the study to include pre-natal recruitment as early in pregnancy as possible (this given will be added to the criteria document), and recognized that "scoring" designs based on the criteria may be difficult since some designs will optimally satisfy some criteria, while other designs will optimally satisfy other criteria. This makes the optimal design depend on which criteria are most important, and, even if "weighting" the criteria based on their importance is feasible, it may still be difficult to identify an optimal design due to difficulties with choosing a loss function for a design (i.e., different people will choose different loss functions) and attempting to optimize a high-dimensional problem. Both of the expert consultants suggested that scoring designs may be an intuitively fruitful exercise, but it would likely not result in an intuitively "optimal" design (in fact, it may result in an intuitively non-optimal design). In addition to a weighting or scoring approach, there was also discussion of the value of designating some criteria or givens as "threshold" or "exclusion" criteria – that is, criteria that must be met to some specified degree for a design or design component to meet minimum eligibility requirements.

It was suggested that perhaps a better means of evaluating designs based on the criteria is to maintain a checklist of the criteria satisfied by the various designs being evaluated. Perhaps designs that do not get a check for certain "important" criteria are determined to be ineligible designs, and designs that do satisfy the important criteria are judged based on their ability to satisfy other criteria.

The meeting participants also discussed the importance of the cost criteria, and it was suggested that with enough dollars all things are possible (i.e., many design approaches, given the financial resources, will have the ability to satisfy the objectives of the NCS). One approach to dealing with cost is to fix the design specifications and then calculate the cost of the design; however, this could lead to identification of a preferred design that is too costly to implement, and thus a less preferred design would be selected. An alternative approach would be to fix the total cost for the study and then evaluate the trade-offs that would be necessary in the various design options to meet the total cost constraint; however, it was noted that we need to be careful in setting the cost to a fixed amount since if a relatively small amount of additional cost could lead to a far superior design, then that design should at least be considered (e.g., if a sub-optimal design results in an initial or start-up cost of $200 million, and an optimal design results in an initial cost of $300 million, then perhaps the additional $100 million cost is justified since it

leads to a far superior design for a small amount of money when considered in relation to the total cost of the study).

In addition to the cost criteria, we spent time discussing the assumptions necessary to generalize the results of the study to the populations of interest. There was wide agreement that there is the potential for multiple populations of interest in this study, including the population of children born in the US while recruitment for this study occurs, as well as future populations of children that might be affected by inferences drawn from the NCS. We discussed the notion of enumerative vs. analytic studies, as outlined by Hahn and Meeker, as well as assumptions that are necessary for generalizing the results of both probability and non-probability-based studies.

Moving from the criteria for evaluation of NCS designs to the design options themselves, not surprisingly one of the topics that the meeting participants returned to many times was the issue of probability-based selection versus non-probability-based selection. Both approaches offer advantages and disadvantages, and both approaches have certain limitations in light of the objectives and constraints of the NCS. In fact, since probability-based sampling may be optimal in addressing some of the NCS objectives/constraints, and non-probability-based sampling may be optimal in addressing other NCS objectives/constraints, the meeting participants began to share the opinion that both of these sample selection methods offer important components to the NCS and may be able to be accommodated in the design. The group recognized that different categories of study users had legitimate scientific objectives that would favor probability sampling in some instances and restrictions on probability sampling to achieve other scientific objectives in other instances. For example, probability-based sampling offers the ability to generalize the results of the study with minimal assumptions, however other types of sampling approaches might offer more flexibility in obtaining previously collected medical history information from a more narrowly defined subset of potential respondents. Therefore, the group recognized a continuum of sampling methods in which a complete non-probability sample is at one extreme of the continuum and a complete probability-based sample is at the other extreme. Somewhere in the middle of these two extremes (i.e., a design that selects some portion based on probability and some portion non-probabilistically) may lie an optimal design that can satisfy *most* (hopefully all) of the objectives of the NCS. In keeping with the Battelle sampling design options paper, this might be called a "hybrid" design, but it may be better referred to as a *family of designs*. In other words, the NCS may not be composed of a single design, but rather a variety, or family, of designs that can be combined to address the multiple objectives of the NCS.

With respect to the concept of a family of designs, we agreed that this type of design would be used to tackle a family of hypotheses and objectives. Thus, different parts of the design would be best suited to service different hypotheses and research demands. Some parts of the design would be essential for measures where data could be collected only in or by major medical centers; other parts of the design would address needs that required inference to the whole population of the US, or where a probability sample base is required to protect against unforeseen circumstances, or where under-

coverage of particular parts of the population would undermine the validity of an inference. In terms of the application of evaluation criteria to the family of designs, we thought that it would be useful to check designs explicitly against these important criteria. By thinking of a family of designs, however, it is quite possible that a particular member of the family may fail a critical criterion, but may contribute enough on other criteria to make its inclusion not only worthwhile, but essential. Considering the array of designs and the array of criteria jointly is what will make the overall design a success.

That said, both of the expert consultants were in strong agreement that the decision of which participants to select for inclusion in the study at the lowest stage of the design where subjects are selected must be specified by the study design team, not left open to the judgment of individual physicians, data collectors, recruiters. This is due to the fact that even well-intentioned recruiters, if allowed to select individuals, will select a biased sample of individuals, and researchers will not have the information they need to make informed judgments about the generalizability of the results. Thus, a strictly convenience sample (i.e., allowing the recruiter to select whomever they like) is very problematic. A random probability-based selection is preferable, but an objective well-specified selection process is mandatory. Of course, this probability-based selection could be specifically constructed so that it is "easy" to conduct, especially if the sample frame is narrow (e.g., all regular visitors to a clinic). For example, one scheme could simply identify a convenient sampling frame (e.g., all pregnant women that attend a specific center, or all pregnant women that have a doctor's appointment during some time period) and select all (100%) of the subjects in that sampling frame. Alternatively, a selection method could consist of simply flipping a coin (or pushing a computer button) to determine whether or not a specific pregnant woman is recruited from within a pre-specified sampling frame. The pre-specified sampling frame could be stratified based on various different selection criteria or the results of a pre-selection screening questionnaire. In summary, there was agreement that control over the selection of participants at the final stage of the design is critical, and recording the success or failure of the selection procedure at this stage is also fundamental. Selection of participants into the study on a probability basis should be the default, with any departures requiring justification.

The discussions also identified other proposed rationales for using a family of designs for the NCS. These rationales are generally related to the size of the study and the ability to propose a design that will meet the objectives of a variety of researchers (medical researchers, epidemiologists, social scientists, health researchers, clinicians, etc.), for whom the values of population descriptiveness, intensity of data collection, exposures, etc. are of differing relative importance. First, since the sample size for the NCS is so large (100,000), the possibility of splitting the cohort into a portion selected non-probabilistically and a portion selected randomly could result in large sample sizes for both groups of individuals (whereas, in most studies that involve a small cohort of individuals, splitting of the cohort would not produce reasonable sample sizes). Second, since there are a variety of opinions as to the appropriateness and limitations of probability and non-probability-based selection for the NCS, incorporation of both types

of sampling through a family of designs may provide a sampling design that can meet the objectives of a variety of NCS stakeholders.

One of the motivating examples that was discussed as a possible approach to a family of designs was a design that selected a portion (as yet to be determined) of the cohort as a national probability-based sample and a portion of the cohort as a Centers-based sample (by Centers-based we refer to the use of university or community medical centers capable of implementing the NCS). The fraction of the NCS cohort recruited using a probability-based sample could be selected as a sample of households (with pregnant women) across the United States, and could involve a variety of data collection organizations. For example, those regions selected with a capable medical center would take advantage of those existing facilities, while regions selected that do not contain an existing qualified center could be covered via some other data collection organization or the establishment of new regional NCS data collection centers. On the other hand, the fraction of the NCS cohort recruited using a Centers-based approach would start with the purposive selection of a set of qualified centers. This would likely involve consideration of the capabilities of the Centers to perform the tasks associated with the study (e.g., data collection, research potential, involve the community, etc.), and could be implemented using some type of RFP-like procedure. Sampling within the selected Centers could be based on existing patient lists – however, there was agreement that there should be some objective mechanism for selecting subjects from these patient lists rather than just a determination by the recruiter or other types of convenience samples. In addition to the above two methods of recruitment, there is a third option that entails conducting a population-based sample around the locations of purposively selected medical centers – which we called center-centered samples. Thus, the family of designs could contain three components – center patients and a national probability sample at the two ends of the continuum, and center-centered samples in between. Of course, there remain a number of questions that would need to be investigated with this type of family of designs (e.g., what proportion of the individuals would be selected on a probability basis, what are the PSUs, how many PSUs will be selected, etc.). In the outline provided in Section 3 we include an approach to investigating this design, and attempt to identify the many questions that would need to be answered in determining the "optimal" design under this type of framework (i.e., in the family of designs framework).

In addition to the above central topics, the following technical issues were discussed throughout the course of the meetings:

- A recommendation from Colm O'Muircheartaigh on conceptually separating frame construction, sampling, recruitment and data collection when considering design options, while still recognizing the interdependence among these elements.
- Consideration of sampling frames other than the household, office, and centers of excellence. In addition, we discussed the notion of using multiple frames options. By doing this, we can integrate the various sampling approaches into a single study design. We can also assess separately the recruitment implications

and the data collection implications of the overall design. This is particularly relevant in thinking about the center-centered part of the design.

- What is the population of interest [target population, sampling frame population, achieved population] for this study? If, for example, the population of interest includes all live births (or, all conceptions) how do we construct a sampling frame and sampling design to approximate this ideal? Also, what is lost if the study fails to approximate this ideal for a subset of participants? What types of measures are truly necessary? (e.g., those that absolutely cannot be assessed retrospectively – such as time-varying exposures, as suggested by Jim Quackenboss).
- Jim Quackenboss raised the issue of trying to consider the implications of different designs on future analyses of the data. How are people going to use the information being collected? We must consider both weighted and unweighted analyses of the data (as both will occur), as well as analyses based on a subpopulation of the cohort. Guidance will need to be developed for different users of the NCS data, based both on the design and the types of inferences that will be drawn from the specific analyses.
- We had some discussion of the criteria for evaluating the relative value of "gold standard" (prospective or concurrent) measures versus information based on retrospective measures (whether based on a physiologic measure or recall). There is a possible (and perhaps common) fallacy of knocking the retrospective measure due to its poor reliability relative to the "gold standard," when the reliability of the "gold standard" itself might be poor. Ignoring this problem could lead to putting excessive weight on the criterion for collecting all measures prospectively. A couple of examples of this sort were mentioned – measures that vary from day to day by substantial amounts, for example. Some evaluations of the reliability of the measures, as well as the accuracy of retrospective (or otherwise more convenient) surrogate measures, might be built into the study or into important pilot studies conducted prior to the NCS.
- Alan Zaslavsky warned of the complications involved when scientists are lured into participating in data-collection centers by the carrot of access to study data. A study of this size cannot be run by consensus, nor is it easy to parcel out research questions among a large group of academic participants. There might be some benefit to local studies that can piggy-back onto the main study, but the parameters of scientific participation in the study must be clearly defined.

Finally, the meeting participants discussed the future steps that Battelle and its consultants should take in continuing the NCS sample design project. Should Battelle proceed to develop discrete and independent alternative design options for the NCS, or should the focus be on developing an integrated family of designs for the NCS? On initial consideration, Jim Quackenboss and Peter Scheidt agreed that it may be better to focus on developing a family of designs for the NCS, and to identify the important questions and issues that would need to be investigated if a family of designs is utilized. Battelle was also charged with developing questions that need to be posed to the

workshop panel, and the Program Office/ICC regarding major decisions related to the design.

The remainder of this document is organized in the following manner. Section 2 provides a copy of the meeting agenda. Most of the topics identified on the agenda were covered at some stage of the two-day discussion; however, as mentioned above, the meeting schedule was not strictly followed. Section 3 provides an outline of the topics discussed, the questions raised by the participants, and any important research directions that were identified over the course of the meeting. Finally, Section 4 provides a list of the next steps that were identified as important avenues of investigation for specifying a promising sample design for the NCS.

B3-2 MEETING AGENDA

Thursday, October 30th
Room: 20-1-82B

Objectives of Day 1		
Criteria for Assessment • Identify missing criteria • Establish hierarchy of importance • Understand inter-relationships between criteria • Determine how to apply criteria	Design Options • Identify missing design classes • Identify advantages and fatal flaws • Specify small number of design realizations within each class	Working Documents • Clarify terminology and definitions • Determine what needs to be added to serve as foundation for all possible design options

8:30 – 9:00 am Greetings and Continental Breakfast
9:00 – 9:15 am Opening Remarks and Overview of Agenda (Warren Strauss)
9:15 – 9:30 am Comments from the NCS Program Office (Jim Quackenboss and/or Peter Scheidt)
9:30 – 10:30 am Overview of Criteria for Assessment (John Menkedick)
10:30 – 10:40 am Break
10:40 – 11:45 am A Range of Design Options to Consider for the NCS (Jeff Lehman)
11:45 – 1:30 pm Lunch – Executive Dining Room 2

Presentations on Five Important Design Considerations
(15 minutes each)

- Advantages and Limitations of Probability-based Sampling for the National Children's Study – Major Themes (Steve Rust)
- The Role of Certainty PSUs in the National Children's Study (Nancy McMillan)
- Utility of Validation Subsamples to Correct for Measurement Error and Bias as part of the NCS (Warren Strauss)
- Overview of Critical Data Collection Requirements for the NCS (Nancy Reiches)
- Overview of Recruitment and Response Rates Achieved in Other Relevant Studies (Ben Pierce)

1:30 – 2:00 Brainstorming Session I – Have we missed any potential design options or sampling frames worthy of consideration?

2:00 – 5:30 Working Session I

Systematic Discussion for Each Design Class

- Review the Advantages and Limitations of the Design Class
- Identify/Specify 1-2 realizations of a design within each Class that would be worthy of consideration

We will attempt to spend approximately ½ hour on each design class during this working session.

6:00 – 7:30 pm Dinner (Optional)

Friday, October 31st
Room: 20-1-82A

Objectives of Day 2	
Narrow the Field of Design Options to Consider	Determine Roles and Responsibilities
• Identify what information is still necessary to make decisions	• Identify what guidance we need from NICHD
• Identify front-runner design options to more carefully specify	• Assignments for key consultants

8:00 – 10:00 am Working Session II - Continental Breakfast

What information is still necessary to gather in order to make a recommendation for a smaller subset (2-3) of design options to fully specify for the consensus workshop?

If we had to choose right now, can we identify a reduced number (5-6) of design options that merit further consideration?

10:00 – 10:15 Break

10:15 – 11:30 Identify areas in which our expert consultants (Alan Zaslavsky, Colm O'Muircheartaigh and Louise Ryan) can help further the process during the month of November

11:30 – 1:30 Lunch – Executive Dining Room 2

B3-3 OUTLINE OF MEETING DISCUSSIONS

In this section we provide an outline of some of the important themes discussed over the course of the meeting. This outline does not represent a chronological record of the discussion, but rather is organized by topic area.

I. Criteria for assessing design options
   a. Criteria overview
      i. Discussed interdependencies of the criteria and what types of things may assist in satisfying the criteria. Some of the criteria group naturally.
      ii. Need to add a "givens" criterion related to the need to include pre-natal recruitment as early in pregnancy as possible. Note that the

list of "Givens" developed by the ICC is attached to this meeting summary as an appendix.

iii. Discussed IRB's and possible difficulties associated with having to obtain approval from multiple IRB's. Additionally, discussed the issue of allowing topics of local interest to "piggyback" on the NCS study, and how this may work in terms of IRB approvals, approval of the NCS program office, additional burden that this may add to study subjects, etc.

iv. Posed the question of how are we going to apply the criteria to evaluation of designs.

b. Evaluating designs based on the criteria

i. Discussed the fact that some designs will optimally satisfy some criteria, while other designs will optimally satisfy other criteria. Thus, the optimal design will depend on which criteria are most important.

ii. Identified difficulties with "weighting" the criteria and "scoring" design options in terms of their ability to satisfy the criteria. Alan and Colm raised the concern that this would be a very difficult means of identifying an "optimal" design, and could lead to a design which is intuitively non-optimal due to the difficulties associated with choosing a loss function (i.e., different people will choose different loss functions), and attempting to optimize a high-dimensional problem. Jim noted that there was still the need to clearly document why and how decisions were made.

iii. John suggested the possibility of a hierarchical decision making process where designs that do not satisfy certain "threshold" criteria are determined to be ineligible and designs that do satisfy the "threshold" criteria are then judged based on their ability to satisfy other criteria.

iv. Similarly, Colm suggested maintaining a checklist of criteria satisfied/not satisfied by the various different elements of design being considered.

c. Cost criteria

i. With enough dollars all things are possible. In other words, any design approach, given the financial resources, will likely have the ability to satisfy the criteria.

ii. Discussed two approaches to dealing with cost

1. Fix study specifications and calculate costs based on study specifications. This will likely lead to a high cost study that is optimal and a low cost study that is sub-optimal. If the high cost study is over the NCS budget, then the low cost study will necessarily be adopted. Thus, since there are limited resources for the NCS, this may not be the appropriate approach to dealing with cost.
2. Fix total cost for the study and evaluate the trade-offs that would be necessary in the various design options.

iii. Colm suggested that we should perhaps consider the total cost for the first few years of the study, and that we need to be careful in setting the cost to a fixed amount since if a relatively small amount of additional resources could lead to a far superior design, then that design should at least be considered (e.g., if a sub-optimal design results in an initial cost of $200 million, and an optimal design results in an initial cost of $300 million, then perhaps the additional $100 million cost is justified since it leads to a far superior design for a small amount of money when considered in relation to the total cost of the study, $2.7 billion).

iv. All agreed that cost can be a difficult thing to estimate and will involve a number of assumptions. Peter and Jim briefly discussed attempts by the Program Office to estimate the cost of the study (i.e., where did the $2.7 billion figure come from). The Program Office will work with the sampling design team to provide input on their cost assumptions that led to the $2.7 billion estimate for the study.

II. Sampling design options

a. Focus is on identifying associations, not univariate analyses to estimate national prevalence and/or national averages. Thus, need to always consider that we are focusing on associations of variables.

b. Different design elements

i. There was a desire to consider the various elements (i.e., selection, recruitment, data collection, etc.) of study implementation separately (e.g., separately consider the sampling method and the organizational structure [who can best recruit may be different than the organization that can best collect the measures]). This would allow better/easier consideration of the various elements and how to best accommodate the various elements. However, all agreed that there are certainly elements that more naturally "group" together. An example of the usefulness of this approach is the center-centered sample, where recruitment could be carried out on a probability-based sample of households by an organization that specializes in this type of recruitment, with participants being referred to qualified medical centers for critical parts of the data collection.

ii. Issue then becomes where to "start" constructing the design. Do we first determine the organizational structure, or the sampling method, etc.? It was suggested that perhaps the best way to begin is to identify the "ideal study," and then all studies can be evaluated based on how much they fail in reaching the ideal.

c. Sampling frame

i. All agreed that we should identify the ideal populations, so that sampling frames could be judged based on their ability to "cover" the ideal populations.

ii. If a probability-based design is utilized, the sampling frame population is the population that can be generalized to from a statistical basis (i.e., in the absence of model-based generalization).
iii. Variety of sampling frames were discussed
    1. Random digit dialing – participants agreed that this is not a promising frame due to the difficulties associated with getting people to not only agree to participate in the study (i.e., poor recruitment rates from phone interviews) but even to simply answer the phone.
    2. Birth certificate registries – this sampling frame would necessarily select children at 6 to 9 months of age and would violate one of the necessary NCS givens. The question was posed as to whether we could use information from the birth registry system to evaluate our sample of subjects. Also, the question was posed as to whether we want to summarily exclude any children that were not enrolled pre-pregnancy. This is a question that may be posed to the Program Office/ICC/Workshop.
    3. All agreed that the Household, Physician's Office, and Centers sampling frames represented the most promising sampling frames that can be constructed for the NCS (i.e., "new" sampling frames that would need to be constructed).
    4. Other "existing" sampling frames were also suggested as possibilities. For example, the Census ACS (American Community Survey) database, the sampling frame used for NHANES, or any appropriate sampling frames used in other government studies or agencies, could all be considered for the NCS [with feasibility and coverage of these frames being critically assessed with respect to their viability for the NCS].
    5. Combinations, or layers, of a variety of sampling frames were also suggested and were recognized as a promising way of attempting to better cover the target population – including those frames suggested by the workshop committee.
iv. Several questions/necessities were identified with regards to the sampling frame.
    1. The primary question is who do we need to enroll. Is it acceptable to enroll some children after birth? Is it necessary to enroll all children prior to conception or prior to birth? What measures are needed / what information is lost given the design choices we make? Peter and Jim gave direction that while we may enroll some subjects at birth so as not to lose important sub-groups of the population, a given of the study was that as many participants as possible

would be enrolled pre-natal and some pre-conception. Alan Zaslavsky suggested the following:

"Consider a two-way table with rows representing parts of the data collection and columns representing parts of the population (and hence of the sample as well). Then place checkmarks for the data collections that are feasible with each part of the sample. On the left, you will have a complete column of checks representing the easiest population to find and recruit due to their own characteristics (willingness to participate, knew they were likely to get pregnant) and organizational connections (patients of cooperating doctors at participating centers). As you move to the right you get less restricted populations but also have fewer checkmarks because they are harder to reach. For example, towards the right you have the population of women who get no prenatal care and are therefore identified only post-natally – whom you don't want to miss because they are a critical group, but some important data-collections will be missed on this population. Now you can think about any study question, select the rows of data collection that are required for it, and identify the parts of the population to which it can be applied. You can then go further to figure out how modeling/validation exercises might enable you to extend those study inferences using statistical modeling combined with retrospective (therefore less accurate) data."

2. What is the target population? Sampling frames can only be judged based on their ability to identify the target population.

v. In order to be able to assess the nature of a center-based sample, we need to accumulate a good deal of information about the patient data-base and the patient flows of the selected medical centers. Without this information, we will have no basis for assessing the inferential strength of such samples. The data we obtain should include information on the number of births in the medical centers, in absolute counts and as a proportion of births in the county and MSA. It may also be necessary to compare the births that occur in the centers to the county/MSA with respect to various different demographic characteristics to understand whether there are any potential biases introduced into the sample by relying on recruitment from center-based patients.

vi. Dual frame and multiple frame approaches were also discussed. Colm O'Muircheartaigh suggested that we envision the family of designs as a multiple frame design. By doing this, we can integrate

the various sampling approaches into a single study design. We can also assess separately the recruitment implications and the data collection implications of the overall design. This is particularly relevant in thinking about the center-centered part of the family of designs.

d. Organizational structure
    i. This aspect of the design was often discussed in relation to the sampling frame and/or the sampling method.
    ii. This aspect of the design is likely the least identifiable at this stage since it can involve a large number of possibilities such as medical centers based data collection in some areas (e.g., major urban areas), data collection by CROs or physician's offices in more rural areas, or establishment of a medical center in geographic regions of particular interest (e.g., highly exposed populations).
    iii. Colm suggested that it is likely preferred to separate the data collection and the recruitment components of the organizational structure since organizations capable of collecting the data may not be the same organizations that are good at recruiting the study subjects.
    iv. Alan Zaslavsky suggested that information related to the organizational levels required to support critical elements of data collection is an important consideration. If almost all of the measures desired for the NCS can be collected using the facilities that are routinely available at any local clinic or some nearby referral area, this is very different from a data collection effort that requires specialized facilities that are only available at major centers. This topic is currently being addressed by one of the white papers that is being developed on this project.
    v. We briefly discussed the possibility and importance of an information management system that would automatically inform the data collection individual/organization of what information needs to be collected at what times. This will likely prove to be an essential piece of the organizational structure to ensure adherence to the data collection protocols of the NCS. We also discussed the notion that all participants will be expected to undergo minimal data collection efforts, with some participants potentially volunteering for more extensive data collection. Therefore, subject-specific data collection requirements will necessarily be part of the information management system.
    vi. Finally, the question was raised as to what are the eligibility criteria for Centers (i.e., what types of "Centers" are appropriate for implementing the NCS).

e. Sampling method – how to select the sample from the sampling frame
    i. Both Alan and Colm strongly suggested that selection clearly specified by the study design (and preferably some sort of random probability selection) should be utilized in the actual subject

selection. In Alan's words, we must define the mechanism of selection, even if imperfect. In this regard quota sampling would not be acceptable as it merely specifies the number of participants in a given category, but leaves who to select up to the individual recruiter. This is unacceptable due to the fact that even well intentioned recruiters, if allowed to select individuals, will select a biased sample of individuals, but more importantly the study will have no information on how these individuals were selected to allow scientists to assess the generalizability of the study. Thus, a strictly convenience sample (i.e., allowing the recruiter to select whomever they like) is likely not ideal. There must be some level of design for which it will be "easy" to incorporate random sampling (e.g., when a pregnant woman walks in the door, flip a coin to determine whether or not to recruit her). So, while the design could entail convenience, or non-probabilistic, selection of Centers to implement the study, the selection of individuals within Centers should really occur in some mechanistic (probabilistic) fashion. Colm wondered what the objection is to selecting probabilistically from a convenient sampling frame (e.g., selecting all, or a random sample of, pregnant women that attend a specific center, or all pregnant women that have a doctor's appointment during some time period), and wondered if there is really significant difficulty in selecting a sample of women in this manner. This would also allow the NCS to be able to better defend the selection process by saying that at least the subjects were selected in some defensible manner even if it was selection from some convenient sampling frame.

We also talked about how a volunteer sample compares to a low response probability sample (the probability sample still represents a fraction of the population, while the volunteers only represent themselves).

ii. All agreed that in the absence of other considerations, some form of a probability-based sample would be desired. The principal of introducing probability-based selection of participants whenever feasible, and always requiring justification of departures from such selection, was discussed as being critically important.

iii. All agreed that a non-probability sample would be "easier."

iv. Thus, the notion of a continuum of sampling methods was identified. At one end of the continuum is a complete probability sample, and at the other end is a complete non-probability sample. All participants agreed that it is likely that somewhere in between these two extremes (i.e., a design that selects some portion based on probability and some portion non-probabilistically) lies an

optimal design that can satisfy *most* (hopefully all) of the objectives of the NCS. We could call these types of designs "hybrid" designs, but it is may be better referred to as a family of designs. In other words, the NCS may not be composed of a single design, but rather a variety or family of designs that can be combined to address the objectives of the NCS.

v. This family of designs would likely involve using different subject selection methods (both non-probability and probability-based selection, different frames, different organizational structures), and would involve varying degrees of data collection requirements and perhaps different periods of recruitment. (Similar to what was referred to as the "multiple cohorts" designs in the draft paper on sampling design options for the NCS).

vi. A discussion of the need to be able to defend the design of the NCS when publishing results of the study also identified that a family of designs may be most appropriate since different designs are preferred in different disciplines. Perhaps the NCS will need to provide materials related to the limitations of the data associated with the NCS (e.g., limitations of the different components of the data) so that all researchers, and the journals they submit papers to, are made aware of the NCS data limitations. Whether or not they heed these limitations will be a separate issue.

f. A possible family of designs in which a portion of the cohort is selected as a national probability-based sample and a portion of the cohort is selected in a centers-based structure. Alan suggested the following outline for how the design might proceed under this type of hybrid approach.
    i. National probability sample
        1. Evaluate the scientific merit of recruitment schemes and concomitant data-collection schemes within each area, defined by organizational structures [or sampling frame] used (Physician's Office, Medical Centers, Households, Other)
            a. Desired capabilities
                i. Ability to identify desired population in timely way
                ii. Ability to do effective recruitment
                iii. Efficiency for project administration
                iv. Potential for efficient ongoing data-collection
                v. Provides coverage of the target population
            b. Consider how this is altered for various kinds of areas
                i. With/Without Qualified Center of Excellence?
                ii. Stratify by Population Density?

c. Consider scientific advantages/limitations of clustering

2. Evaluate per unit (cluster, possibly identified with a recruitment/data collection unit) and per element (individual subject) costs
3. Consider other design considerations such as desired precision for defined subpopulations (rural, regions, etc.) and consequences for stratification and disproportionate sampling.
4. Apply standard methods for design of clustered or multilevel studies
5. Identify certainty Areas (potentially to be covered by centers; within-area design is required)

ii. Center Selection

1. Certainty areas from the probability-based design
2. Access to targeted Populations (Highly Exposed, Important Subpopulations)
3. Capabilities
    a. Access to patients within and outside the center's own system of care
    b. Data collection and management capabilities
    c. Ability to elicit/organize Community Support (from various defined communities)
    d. Scientific research potential
4. Stages of this component of design
    a. Identify criteria
    b. Identify centers that might be explicitly recruited
    c. RFP-like procedure for recruiting additional centers with a priori unknown capabilities

iii. Within-Center Design

1. Access to Patients (within and outside of their own patient populations)
2. Sub-Units relevant to sampling
    a. Designs based on element sampling (individual patients), clinic sampling, physician sampling, etc.
    b. Evaluate based on availability of suitable frames, information systems that will be used, ability to obtain participation by providers at a given level
3. Probability sampling within available frame? (or: take 100% if needed) [question – if take 100% of volunteers – do we still maintain ability to quantify non-response]
4. Responsibilities of center for recruitment and/or data collection follow-up outside its primary population: role as part of probability sampling design

iv. Subsampling of data collection within the recruited cohort

1. Identify data elements that are particularly expensive in general, or needed only for a subsample
    a. Sampling across the entire cohort (example: a very expensive calibration measure when there is a cheap but inaccurate measure being used)
2. Identify data elements that are difficult/infeasible to collect in some segments of populations
    a. Sampling or omission in some defined parts of cohort (e.g.,. some measures can't be collected except for patients very close to a specialized center)

III. Five Important Design Considerations

a. Advantages and Limitations of Probability-Based Sampling for the National Children's Study – Major Themes (Steve Rust)
    i. Discussed various advantages of probability and non-probability sampling (e.g., assumptions required to generalize), and discussed distinction between analytic and enumerative studies. NCS is likely an analytic study.
    ii. Key point: All approaches to the NCS will likely require assumptions in order to make valid inferences. There will be fewer assumptions with a probability-based sampling approach, and more assumptions with a non-probability sampling approach; however, the non-probability approach will likely be more cost efficient. It is scientifically valid to consider a trade off between assumptions that are scientifically palatable and cost efficiency [also feasibility, convenience of non-probability designs].
    iii. Discussed the possibility of intentionally oversampling a probability sample towards the population that is expected to respond well under a convenience approach.
b. The Role of Certainty PSUs in the National Children's Study (Nancy McMillan)
    i. Prompted discussion of: the definition of a PSU, how many PSUs will be utilized, what is a certainty PSU (or strata), how "large" (population size) does a PSU need to be, what are the design effects associated with different numbers of PSUs and different thresholds for certainty PSUs, etc.
    ii. There was considerable uncertainty about the significance of certainty selections. Colm O'Muircheartaigh suggested that we introduce an alternative nomenclature, possibly certainty areas to make it clear that they are strata and not sampling units.
    iii. There was consensus that it will be critical to re-calculate the design effects with alternative designs for the relationships being studied.
c. Utility of Validation Subsamples to Correct for Measurement Error and Bias as part of the NCS (Warren Strauss)

i. Prompted more discussion of the use of a family of designs for the NCS, including the possibility of different groups of people undergoing different data collection protocols, some with a higher degree of subject burden and some with a lower degree of subject burden.
ii. Discuss difference between Validation Samples and use of Matrix Sampling Approaches
iii. Impact on analyses

d. Overview of Critical Data Collection Requirements for the NCS (Nancy Reiches)
    i. Prompted discussion of the subject burden issue and the criticalness of defining the minimal data collection requirements when determining an appropriate design.
    ii. A question was posed that inquired whether there are "critical" data collection needs that must be met in order for the individual to be/continue in the study. For example, is it critical that pre-natal blood samples are obtained? [i.e., there is a trade off between what happens if we don't get this information versus biases introduced if we limit the study to ensure this information is collected] If so, then any individuals recruited at birth are necessarily excluded from the study. On the other hand, is it acceptable for a portion of the cohort to not have prenatal measurements available? If so, then there may be important implications in terms of the sampling options appropriate for the NCS.
e. Overview of Recruitment and Response Rates Achieved in Other Relevant Studies (Ben Pierce)
    i. Prompted discussion of the importance of knowing what other studies have been able to do in terms of recruitment and response rates. Unfortunately, there are really no studies that involve the degree of subject burden that is likely to be a part of the NCS. Thus, it is difficult to apply the recruitment and retention rates seen in these other studies to the NCS.
    ii. Discussed issues and methods for "tracking" study subjects that move during the course of the study. [estimates of within versus between county (area) moves will be provided in the white paper] This will be an important component of the design since a presumably large proportion of the cohort will move at some point during the 20-year study.
    iii. Postulated that study subject burden, or the study subject's perception of burden, will play a major role in recruitment and response rates.
    iv. In thinking about burden and retention, the first few years are probably the most critical – as mobility will lead to considerable dispersion of the sample after this under any design.

IV. Future directions discussed during the meeting

a. What are the future directions that Battelle and its consultants should take in this task? Is it useful for Battelle to develop multiple designs, or would time and resources be better spent in focusing on developing some family of designs, identifying important questions/issues with implementing a family of designs, and attempting to illustrate the usefulness of a family of designs. Alan Zaslavsky suggested that an alternative way of thinking about this is that the designs of interest can be put into a common framework with some parameters defining the differences among them. Some of these are quantitative (fraction of the NCS cohort in probability versus purposive samples, fraction recruited preconceptually versus prenatally) and some are qualitative (recruitment through physicians versus directly from a research organization). This is a different approach from counterposing several complete designs that differ in a number of (confounded) respects. Jim Quackenboss and Peter Scheidt were of the initial opinion that it may be better to focus on developing a family of designs.
b. Discussed how the Battelle products should move forward. Should the current documents be modified and moved forward as progress is made, or should "new" documents be constructed. Initial impressions were that while there can be some minor modifications to the "current" documents, perhaps moving forward with "new" products will be useful so that the process or evolution of the work is well documented. Also, the group mentioned that perhaps time and resources should not be "wasted" in making the documents "perfect," and by moving on to new documents we can avoid allowing any specific work products from growing a life of their own. The current documents can be included as appendices of future reports on this project, with information contained in the main body that discusses process that we went through.
c. Briefly discussed areas in which Alan and Colm can assist in furthering this work. The first thing mentioned was for them to provide any comments they have on the first two Battelle documents. The group agreed that after reflection on the progress made in the meeting, future areas for investigation could be better identified.

B3-4 IMMEDIATE NEXT STEPS FROM MEETING

Based on the outline provided in Section 3 and some of the future directions identified over the course of the two day meeting, a list of further steps and tasks is as follows:

1. Construct an outline of the meeting minutes and distribute to the attendees for their input and concurrence.
2. Solicit comments from consultants on existing documents.
3. Develop an outline for the next document that will present the family of designs.

4. Specify an optimal family of designs for the NCS

   Further work on white paper topics discussed at lunch

   - Advantages and Limitations of Probability-based Sampling for the NCS
   - The Role of Certainty PSUs in the NCS
   - Critical Data Collection Requirements for the NCS
   - Recruitment and Response Rates Achieved in Other Relevant Studies

   Determine an approach or basis for conducting power studies.

   Identify the roles and responsibilities of the key consultants
5. Continue to reflect on the topics covered during the meeting to identify other important areas that warrant further investigation.

B3-5 REFERENCES

[1] Menkedick, J., Strauss, W., Lehman, J., 2003. Draft White Paper on Criteria for Evaluation of NCS Design Options.

[2] Lehman, J., Strauss, W., Menkedick, J., 2003. Draft White Paper on Sampling Design Options for the National Children's Study.

APPENDIX A

"Givens" for NCS Sampling design

1. The study will be national in scope, but not necessarily nationally representative.
   - The sample should be broad-based and include a wide range of population and geographic diversity.
   - The primary purpose of the study is to investigate exposure-response relationships, and not to provide estimates of disease and exposure incidence and prevalence.

2. A large sample (approximately 100,000) -- to allow for evaluation of infrequent outcomes (as low as 2/1,000 in the general population); and of interaction of environmental factors and genetics.

3. Prenatal recruitment should occur as early in pregnancy as possible.

4. Provide for access/collection of biological samples at birth.

5. Stratification -- to obtain adequate ranges of exposures (including social).
   - Ability to have socioeconomic, racial/ethnic/geographic diversity
   - Include population sub-groups of interest (e.g., ethnic/racial groups, urban/rural)

6. Locality-based to
   - maximize recruitment and retention rates;
   - allow for efficient collection of exposure and outcome measures, and measurement of context (physical and social);
   - provide reasonable access to infrastructure needed to support specialized measures; and
   - encourage community engagement.

7. Flexibility to conduct special studies (e.g., special pop'n groups, pre-conception recruitment, or topics of scientific and community interest).
   - these may be sub-samples collected outside of the core design

Revised for ICC -- July 24, 2003