Final Report from the
National Children's Study Sampling Design Workshop
March 21-22, 2004
Arlington, Virginia
May 9, 2004

Introduction

The National Children's Study Panel on Sample Selection was charged with 1) Providing an approach to the sampling design that would reconcile competing priorities, needs, and limitations; 2) Assessing the background papers provided by Battelle for addressing the design decisions; 3) Addressing the strengths and weaknesses of selected design options; and 4) Identifying options that require pilot testing to reach a final decision. The panel consisted of nine researchers with a diverse range of disciplinary backgrounds and research experiences, listed at the end of this report. We were provided with the detailed Battelle "Draft White Paper on Evaluation of Sampling Design Options for the National Children's Study" along with appendices. The panel met for two days, March 21-22, 2004, in Arlington, Virginia, with the first day devoted to hearing from selected key individuals involved in the planning of the study from both within and outside the federal government. The second day was set aside for panel deliberations and an oral summary to Dr. Alexander and other leaders of the planning effort for the study.

Those who helped to prepare us for the workshop, particularly Drs. Quackenboss and Scheidt, were extremely responsive to our needs, offering candid insights before and during the panel meeting. The presenters gave succinct, informative talks on a range of issues bearing on the approach to sampling and were able and willing to respond to all the questions that we posed. While we prepared and deliberated over a relatively short period of time, and cannot claim the depth of knowledge of those who have been engaged over several years, we believe we can offer a useful perspective of informed outside experts free of entrenched, longstanding positions regarding the study. The panel was chosen to have research backgrounds that would enable them to appreciate the goals and methods of the study, but there was little previous involvement of panel members in the study, and we were able to approach the issues objectively.

Points of Agreement Regarding Sampling Plan

We discussed a number of issues that bear on the approach to recruiting participants into the National Children's Study, which set the stage for more detailed consideration of two competing selection plans, a national household probability sample and a center-based design in which recruitment is conducted by academic medical centers working in targeted communities. The panel agreed unanimously on the following points:

1) A national probability sample is preferred to other sampling approaches based on a number of specific reasons as described in detail below. All panel members recognize the challenges in implementing this approach successfully, with varying views regarding the feasibility for such an approach to generate acceptably high participation and retention proportions and its feasibility relative to a center-based design. However, we are all in agreement that it would offer distinct benefits. Such a national probability sample would call for incorporating extensive biomedical and clinical detail into the design, well beyond simple

biospecimen collection, which has become common in such surveys. The alternative, a center-based model, would require extension in the other direction, moving from the traditional convenience sample based solely on recruiting patients towards a more complete community representation through collaboration and outreach, which would include women outside the medical system, some of whom would be recruited prior to conception. Both approaches would seek to integrate the strengths of biomedical and population research, and each poses real challenges in deviating from the ways such studies have been done in the past.

2) We do not see advantages in allocating proportions of the study sample across recruitment approaches, unless there are explicit goals regarding what can be learned from each subset. While having a larger cohort that provides core information and a subset that is followed more intensively should be considered, simply recruiting individuals through different mechanisms into the overall cohort does not offer any apparent advantages over expanding the best approach to include the entire sample.

3) Under any approach to sampling participants, many of the key activities of the National Children's Study will need to be centralized in order to maintain standard methods and quality control and to ensure that the most capable groups are performing key tasks. The formulation and conduct of interviews, the collection of environmental samples, specimen receipt, processing, storage, and assays, and follow up of children over the extended study period will require central planning and management regardless of whether the pregnancies are initially identified for the study through a national probability sample design or independently by multiple centers. The continued follow up of children, however they are initially recruited, will occur throughout the country (given the mobility of the population) and require ongoing decisions regarding the data to be collected and hypotheses to be tested. Contrary to the citations provided in the Battelle report, several long-duration national probability samples (e.g., the National Longitudinal Survey of Youth) have generated high recruitment and retention proportions through such an approach.

4) A mechanism is needed to fully engage the energy and intellectual resources of the research community, balanced against the need for a centrally managed study. The strengths of central planning are consistency and quality control, but the potential weakness is the loss of the creativity, energy, and full buy-in of the broad research community concerned with children's health. There would clearly need to be early access to the data that are generated for public use and a way to entertain proposals for use of biospecimens. At an earlier stage, it would be desirable to consider ways to solicit and evaluate competing ideas for specific research proposals that might be incorporated into the study.

5) Both social and biomedical aspects of children's health are of central importance to this effort, with a need to consider how social factors affect behavior and biological pathways, as well as discover more basic mechanisms of disease causation. The approach to sampling needs to accommodate both themes.

6) All pregnancies, and all the fetuses and infants resulting from a given pregnancy occurring during the recruitment period to a given woman should be included to

optimize recruitment and retention, and to allow for study of siblings. This would facilitate special studies of genetic and gene-environment effects.

7) The potential for including a fraction of women who would be enrolled and monitored prior to conception was seen as highly desirable and can perhaps be done by monitoring non-pregnant women of reproductive age for the onset of pregnancy. Inclusion of such women would permit study of infertility and pregnancy loss. In principle, such preconception enrollment would also allow for specimen collection to be done before or very early in pregnancy, in the period during which structural malformations and perhaps other pediatric health problems have their origins. The importance of this effort depends on the priority given to such outcomes as congenital malformations, for which the rarity of individual types may make even a sample of 100,000 marginally adequate.

8) To address some of the study goals, it may be necessary to overrepresent selected geographic locations and subgroups, such as locations with certain environmental exposures of interest, e.g., specific forms of air and water pollution, and participants of certain race or ethnicity or with specific socioeconomic conditions. This overrepresentation is attainable under either sampling approach. Views of panel members vary on the extent to which such decisions to optimize one goal may compromise other study goals.

9) A streamlined approach is needed for design and implementation decisions in order to allow for our suggestions regarding design and further pilot work to be of value in moving the study forward. While the benefits of having multiple agencies, diverse sources of scientific input, and careful deliberations regarding the conduct of this study of unprecedented size, cost, and importance are clear, a mechanism needs to acknowledge competing considerations, reach decisions, and move forward. We came to quickly appreciate the strong views held regarding how this study should be done as well as the magnitude of challenge it poses. Given the complex array of committees in place, we have real concern that there may not be a clear, widely understood plan for exactly who will make the hard decisions required and ensure that the benefits of outside influences can be realized without preventing progress. We would hope that our panel offers useful insights to accelerate progress towards firm decisions about the design and not result in any unnecessary delays. We propose below specific, limited pilot activities that should provide key information for making a decision regarding the design and may even warrant a prespecified decision algorithm in response to the data that are generated in order to avoid indecision at the end.

Option 1: National Probability Sample

The national probability sample design we considered calls for a full national probability sample of households, recruitment of reproductive age potentially fertile women residing in those households, and prospective monitoring of those women over some period of years for pregnancy and births. There would be a clear need for geographic clustering in the sampling strategy, but the approach would involve many such geographic units, well over 150, and be widely dispersed. The sampling plan could be weighted to achieve the desired diversity of geographic location and ethnic composition for estimating prevalence and attaining sufficient precision for measuring associations of

particular interest, but allowing for generation of weighted nationally representative estimates.

The reasons for preferring this approach were notably diverse across panel members, not all of who value each benefit similarly. Nonetheless, the multiple perceived advantages build on one another and collectively make a case that the panel was unanimous in supporting. The key points are as follows, not necessarily in order of importance:

1) The National Children's Study needs to be able to contribute to understanding of the impact of social, economic, and environmental factors, not just biomedical factors, providing guidance to public policy decisions affecting the health of children. Given the desire to address both individual-level and population-level effects on behavior and ultimately on health, the use of a probability sample offers substantial advantages. While biomedical influences may also be vulnerable to biases as a result of recruitment using non-probability based sampling, social and economic influences are likely to be more directly linked to access to and selection of health care providers. Therefore, associations between such factors and child health outcomes are particularly susceptible to varying in relation to the source of participants. Furthermore, a true national probability sample is likely to enhance the perceived value of the study's findings among the public at large as well as policy makers, including those who will be called upon to consider continued funding for the research effort. Links to agencies concerned with education, housing, and a range of other policy arenas are more likely to appreciate and support the study if it is a national probability sample.

2) The environmental and social influences on child health that are of interest vary both within and between geographic locations, and it would be advantageous to be able to describe such variation in explicit and quantitative terms as can be done in a probability sample that is appropriately geographically dispersed.

3) As influences on children's health are identified through this research, the information to generate estimates of attributable fractions, which require population-level information on the distribution of determinants and relative risks, would be available directly from a study with a probability sample provided such sampling has sufficient recruitment and retention rates.

4) Having access to medical care or seeking medical care would not be prerequisites for enrollment in the study, allowing inclusion of women who do not seek prenatal care or who only seek care late in pregnancy. While other designs could sample within households to achieve this goal, a probability sample provides a scientifically valid means of doing so. This is particularly so when the sampling unit is the woman and not the pregnancy, since probability-based sampling of prevalent pregnancies would be subject to length bias (in which pregnancies that persist for longer periods are more readily identified and included).

5) A household probability sample would involve sampling of women, not pregnancies, and allow preconception enrollment to be done in an unbiased way and fully within the context of the study of pregnancy outcome and subsequent

children's health. Recruitment through prenatal clinics, for example, would not include all such women and would not enroll women at the beginning of their pregnancy, and risks biases associated with the many characteristics influencing enrollment in prenatal care. Under a center-based model, preconception recruitment would have to be done in a separate arm of the study.

6) With relatively few exceptions, the desired data and specimens can be collected in the home, not requiring collection within the setting where prenatal care is obtained. Home-based rather than clinic-based collection would likely be more convenient for the participants.

7) A probability sampling plan would be rigorous and explicit, allowing replication of the sampling process and the analytical results, at least in principle, presuming recruitment and retention rates are sufficiently high. Bias due to non-participation could be assessed by comparing the study participants to population-based data from the census and from birth certificates.

8) Separation from the medical care system by sampling households may offer advantages in marketing the study and recruiting women, freeing the study of any perceived negative aspects of medical research, sometimes an issue of particular concern in minority and poor communities. In addition, such a separation may help to facilitate the needed standardization of methods and centralized follow-up through an experienced survey organization.

9) The desire to ensure a unique contribution of the National Children's Study, above and beyond what is already being done within medical centers and the large cohorts that have been assembled in other settings. Current efforts in Norway and Denmark do not involve probability samples, and only some of the studies in England have had this characteristic. None of the studies in North America have previously attempted to develop a true population-based probability sample.

10) A geographically clustered probability sample of households may facilitate collection of data on the social, physical, and chemical environment in which the participants reside since the data collection would be done at home. Studies based in health care facilities have a physical separation of the data collection site (clinic) and the environment of interest (home).

While the assessed *desirability* of such an approach was universally supported, the assessed *feasibility* of this approach was judged differently among panel members. The lack of precedent for a study of this size and complexity at least in the United States dictate the need for pilot efforts to assess the feasibility of this approach. The pilot study needs to be carefully planned and implemented so that the most effective methods of recruitment and data collection are identified.

We were able to identify two key concerns, that, if answered affirmatively, would persuade even the most skeptical of panel members that study participants can be identified and recruited as a household probability sample of the population and that the required array of data can be collected. While not all issues are unique to the probability sampling design, all are important considerations in assessing the feasibility of this approach. They are as follows:

1) Feasibility of identifying and recruiting women in a household survey and identifying and recruiting those who become pregnant: There are several steps required, with the most experience in obtaining respectable response rates for household surveys to the point of enumerating the household members and identifying any of them who are currently pregnant. Proceeding from that point, the study would require the initial and continued involvement of reproductive age women not currently pregnant and not known to be sterile (regardless of whether they report being sexually active or using contraception) in order to identify new pregnancies as they occur. The issues of unplanned pregnancies, induced abortion, and general sensitivity surrounding pregnancy would all pose a challenge in that the study would be focused on women who have not self-identified by seeking prenatal care, for example. The feasibility of recruiting pregnant women in this manner could be tested by determining the yield, in terms of response rates and numbers of pregnancies that can be identified for a given cost, through a pilot study conducted in a small number of carefully chosen diverse geographic locations.

2) Feasibility of obtaining access to hospitals serving recruited women: Once enrolled in the study, it is generally believed that the desired data can be collected without active involvement of prenatal care providers. What is desired, however, is access to the woman and her infant at delivery in order to collect cord blood, the placenta, and conduct research-quality neonatal examinations. This would require cooperation at every hospital or other location at which the sampled women deliver with either hospital staff, recruited and trained to collect these data, or study staff attending each delivery to collect the needed specimens. There are both access issues, in the sense of requiring a high degree of cooperation of health care providers, and logistical issues, in having an appropriate person collecting the right information and specimens at the needed time. It is recognized that obtaining the desired extent of access and assistance would not be easily obtained in the center-based approach either, but there would be greater familiarity with the investigators seeking such material, making this a special challenge for the national probability sample design. The degree to which this is needed for all study participants rather than a subset is not entirely clear.

In addition, three other important issues could be usefully examined in such a pilot study, addressing issues important to but not necessarily unique to a national probability sample:

3) Feasibility of and need for fostering a community commitment in the randomly selected areas: Views of the panel varied on how important they thought it was to have the National Children's Study be a visible, collective activity on the part of the community versus simply identifying and recruiting individual participants in isolation. If in fact community engagement is advantageous, it was unclear how feasible it would be to foster this commitment in a widely dispersed array of randomly selected locations, i.e., locations not chosen for making this aspect of the study feasible. While it is not clear that this issue can be effectively resolved through pilot studies, perhaps there would be an opportunity to compare across at least two communities, one of which could have the study promoted energetically in the media, through local civic groups, etc., in a manner that

would be feasible for any chosen location, and the other half pursued without such efforts in order to assess the impact on response.

4) Feasibility of collecting reproductive tract specimens outside medical settings: While there is abundant evidence that most biospecimens can be collected in the home, such as blood, urine, saliva, or hair, there would be an interest in having some reproductive tract specimens collected. Further work would be needed to determine if self-collection would be adequate for these purposes, and second, whether women would be compliant with doing so in their homes. It is presumed that a complete pelvic examination as required to collect cervical specimens, for example, would not be feasible under this design, unless special arrangements were made with all prenatal care providers to collect the specimens within prenatal care settings or trained clinicians (nurses or physicians) were sent to the home.

5) Feasibility of collecting biological specimens very early in gestation: With the planned identification of some fraction of pregnancies prior to conception, there is the potential for collecting biological and environmental specimens in the first weeks of gestation, a period of special interest. In practice, what is unclear is just how burdensome this approach would be with regard to participant tolerance and cost, and how early such specimens could in fact be collected. For even a small number of such women identified before conception, it would be helpful to undertake specimen collection to more accurately weigh the feasibility of this desirable component of the study.

Option 1A: Probability Sample with Investigator-Initiated Components

A potential limitation in the fully centralized national household probability sample approach was seen as the limited opportunity to engage fully the ideas and talent of university-based investigators. While there would be a need for the medical care community to support in-hospital data and specimen collection, there is not under this approach a direct mechanism for investigators with promising ideas to have the opportunity to compete for their incorporation into the National Children's Study.

One way for this to be done would be to set aside some resources, including funds, interview time, specimen allocation, and respondent burden more generally, that would be open for competition to the research community. This would need to be done *before* finalizing the study plans to allow for the most promising ideas to be incorporated. The scientific promise would need to be evaluated in balance with the burdens imposed on study participants and staff, but the process should generate more useful information than would be provided by having the entire set of hypotheses and data needs determined centrally.

Option 2: Center-Based Recruitment

The center-based approach that we considered is based on academic medical centers working to identify and recruit participants within targeted communities. In this model, the natural base of these academic centers would need to be expanded in most cases in order to be more broadly representative of the population residing in defined

communities. This extension could be done through health care providers and through active outreach to the population. Potential centers would be invited to compete for this opportunity and judged on such criteria as their ability to accurately reflect pregnancies in the population residing in a defined geographic area and inclusion of the full spectrum of socioeconomic and ethnic groups, as well as the ability to provide a sufficient number of participants. If such centers had a means of generating probability samples of the catchment area, that would of course be most appealing and competitive for selection as a site, but it seems more likely that they would instead seek to generate populations reflective of the composition of the community in a less formal manner, working with other health care facilities. Birth records would allow for making comparisons of those included with live births in the area, at least with regard to those characteristics available from the birth certificate. In addition to centers being selected in part based on their ability to meet specified desires for special populations targeted by the study planners, e.g., agricultural workers or ethnic minorities, optimal geographic locations could be sought out for encouragement to compete. There would still need to be centralized planning and administration of the study and a standardized approach to interviewing, biological specimen collection and processing, environmental measurements, and follow-up of the children. A sizable number of sites would be needed to generate the desired study size of 100,000, with a tradeoff such that more sites provide for greater geographic diversity and more extensive involvement of the clinical and research community, but also are more challenging and expensive to coordinate.
The key strengths of this approach include:

1) The scientific community would be fully engaged through the development of proposals and competition for participation in the National Children's Study. The most creative approaches of the nation's best researchers would be brought to bear from beginning to end under this model, within the practical constraints of cost and respondent burden. Supplementary research would undoubtedly evolve from the centers and networks of collaborating centers.

2) An active community engagement would be attainable under this approach in that centers could be selected in part based on having such a component. An identity would be established for the study locally, with a committed leader or set of leaders, and a carefully fostered sense of the value of this study. This model has worked successfully in selected locations and would be applied for the first time in a large number of sites with central coordination and assistance.

3) By building out from existing research centers into the community, a beneficial change in the perspective of these research centers would result, more fully embracing a population perspective on the health of the local community.

4) Collection of specimens and medical examination data at the time of delivery may be enhanced by the involvement of those medical centers at which many of the study participants would deliver.

With this model, there are also real concerns that would need to be addressed to ensure its success. While there is little doubt that academic medical centers can do what they have already been doing through the collaborative research networks, in which common goals and protocols are pursued, there are features of expanding this model for the National Children's Study that are unproven.

1) Sufficient response from needed number and scope of centers to conduct a national survey of the desired size: While a handful of academic medical centers with strong records of funded research of this type would undoubtedly respond to an invitation to develop proposals, it is unclear whether a sufficiently large number of well-qualified centers covering a diverse enough population would in fact step forward. Many academic medical centers are based in urban areas and may have less access to and experience with other populations in their region. It is difficult to resolve this question with much confidence in advance, but perhaps by generating selected locations or populations of interest and surveying candidate center leaders, informative insights could be gained about the scope of candidate sites and whether there would be a sufficiently broad menu from which to select.

2) Whether academic medical centers are capable of expanding beyond their traditional patient clinical base is uncertain. There is much competition for delivery of clinical services, and relatively little tradition for such clinical centers to extend beyond their traditional boundaries. In many cases, local providers compete with one another rather than cooperate. In addition, economically disadvantaged populations, an essential component of the planned study, have particular considerations with regard to where they obtain medical care and whether they can be successfully enrolled in the study. The ability to enlist the support of other health care providers in their communities to obtain a sufficiently broad patient base would need to be evaluated. To establish the feasibility of a center-based approach, there would be a need for a pilot study comparable to that proposed for the probability sampling strategy and ideally done at the same time. A small number of centers would need to implement the process from identifying women from the community at large, not just relying on their patient base, obtaining a sufficient response rate for enrolling such women, and collecting required interview data and specimens, up to and including the collection of placenta, cord blood, and standardized neonatal examinations at delivery.

3) Relying on specific centers would imply relying on the institution to follow through for a sustained period, more than 20 years, even if the principal investigator were to change. The most intensive involvement would be early in the recruitment phase, with more and more centralized work needed in the follow-up period, but nonetheless, some sustained linkage to the medical center would be required over the life span of the National Children's Study, at minimum for Institutional Review Board purposes. For follow-up purposes, the population's mobility would result in a need for national efforts and any initial advantages in having center-based recruitment would diminish over time.

4) Centers may not be universally capable of generating desired recruitment rates or fully complying with a standardized recruitment protocol. Although some centers would have experienced research teams in place to achieve the desired rates of recruitment of eligible participants, many would not. Lacking a centralized mechanism of recruitment through a survey research organization, there would be significant variability in the level of success, with some centers falling below desirable rates. Furthermore, such centers would need to be capable of recruiting women early in pregnancy, collecting the required biological specimens and environmental samples, and arranging for newborn

examinations. While there would be some transition to the centralized research organization following recruitment, the early components would have to be facilitated by the individual centers. Another concern with the decentralization to multiple centers as opposed to a national survey research organization is that the desired standardization with regard to defining eligibility and recruiting and enrolling participants would be more difficult to sustain.

Conclusions

Both the national probability sample and center model have successful precedents to draw upon, but in neither case has there been a study that combines the scope, size, and detail intended for the National Children's Study. In fact, it is probably appropriate that the study set its sights on making a contribution that is far more ambitious than could ever be done through ongoing mechanisms of developing research proposals. Expanding the scope of either approach into uncharted territory will be very challenging, and for that reason, we would encourage the simultaneous pilot testing of key issues affecting the feasibility of both approaches. We recommend proceeding rapidly with targeted, efficient pilot efforts to better inform this key decision regarding sampling design.

We recognize that the product of the pilot data collection effort is certain to be informative, but unlikely to be definitive unless one approach or the other (or both) is shown to be completely infeasible. Assuming that instead the relative feasibility of each is measured and quantified, a panel such as ours or some other appropriately constituted group should be configured to balance the strengths and limitations and quickly put forward a plan to conduct the best study possible and recommend a specific sampling approach for implementation.

National Children's Study Sampling Design Workshop Panel Members

David A. Savitz, Ph.D., Panel Chair, Professor and Chair, Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, North Carolina

Scott Zeger, Ph.D., Panel Vice Chair, Professor and Chair, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Jean Golding, Ph.D., Professor of Paediatric and Perinatal Epidemiology, Scientific and Executive Director, Avon Longitudinal Study of Parents and Children (ALSPAC), Department of Community Based Medicine, University of Bristol, Bristol, England.

Barry I. Graubard, Ph.D., Senior Investigator, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

Graham Kalton, Ph.D., Senior Vice President and Senior Statistician, Westat, Rockville, Maryland

Michael S. Kramer, MD, M.Sc., Professor of Pediatrics and Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada, and Scientific Director, Canadian Institutes of Health Research, Institute of Human Development, Child and Youth Health

Anne R. Pebley, Ph.D., Professor of Population Studies, UCLA School of Public Health, University of California, Los Angeles

John Spengler, Ph.D., Professor of Environmental Health and Human Habitation, Department of Environmental Health, Harvard School of Public Health

Clarice R. Weinberg, Ph.D., Chief, Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina