Design of the National Children's Study: A Workshop Summary

ISBN
978-0-309-28840-8

94 pages
6 x 9
PAPERBACK (2013)

Nancy Kirkendall, Rapporteur; Committee on National Statistics; Division on Behavioral and Social Sciences and Education; Board on Children, Youth, and Families; National Research Council; Institute of Medicine

Design of the National Children's Study

A WORKSHOP SUMMARY

Nancy Kirkendall, *Rapporteur*

Committee on National Statistics

Division of Behavioral and Social Sciences and Education

Board on Children, Youth, and Families

Institute of Medicine

NATIONAL RESEARCH COUNCIL *AND*
INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract/Grant No. HHSN263000019 between the National Academy of Sciences and the U.S. Department of Health and Human Services. Support of the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (No. SES-1024012). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-28840-8
International Standard Book Number-10: 0-309-28840-1

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu/.

Printed in the United States of America

Suggested citation: National Research Council and Institute of Medicine. (2013). *Design of the National Children's Study: A Workshop Summary.* Nancy Kirkendall, *Rapporteur.* Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Board on Children, Youth, and Families, Institute of Medicine. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

STEERING COMMITTEE
A WORKSHOP ON DESIGN OF THE NATIONAL CHILDREN'S STUDY MAIN STUDY

Sara McLanahan (*Chair*), Center for Research on Child Wellbeing, Princeton University
Barbara Lepidus Carlson, Mathematica Policy Research, Inc., Boston
Steven B. Cohen, Agency for Healthcare Research and Quality
Naihua Duan, Columbia University Medical Center
Greg Duncan, School of Education, University of California, Irvine
Elena Fuentes-Afflick, Department of Pediatrics, School of Medicine, University of California, San Francisco
Irwin Garfinkel, School of Social Work and Population Research Center, Columbia University
Carol J. Henry, Consultant
Marie C. McCormick, Department of Social and Behavioral Sciences, Harvard School of Public Health and Harvard Medical School
Edward J. Sondik, National Center for Health Statistics

Nancy Kirkendall, *Study Director*
Agnes E. Gaskin, *Administrative Assistant*

COMMITTEE ON NATIONAL STATISTICS
2012-2013

Lawrence D. Brown (*Chair*), Department of Statistics, The Wharton School, University of Pennsylvania
John M. Abowd, School of Industrial and Labor Relations, Cornell University
David Card, Department of Economics, University of California, Berkeley
Alicia Carriquiry, Department of Statistics, Iowa State University
Constantine Gatsonis, Center for Statistical Sciences, Brown University
James S. House, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor
Michael Hout, Survey Research Center, University of California, Berkeley
Sallie Keller, Department of Statistics, University of Waterloo, Ontario, Canada
Lisa Lynch, The Heller School for Social Policy and Management, Brandeis University
Sally C. Morton, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh
Ruth Peterson, Criminal Justice Research Center, Ohio State University
Edward H. Shortliffe, Columbia University and Arizona State University
Hal Stern, Donald Bren School of Information and Computer Sciences, University of California, Irvine
John H. Thompson, NORC at the University of Chicago
Roger Tourangeau, Westat, Rockville, MD

Constance F. Citro, *Director*

BOARD ON CHILDREN, YOUTH, AND FAMILIES
2013

Reviewers

This workshop summary has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published summary as sound as possible and to ensure that the summary meets institutional standards for clarity, objectivity and responsiveness to the charge. The review comments and draft manuscript remain confidential to protect the integrity of the process. We wish to thank the following individuals for their review of this workshop summary: Elena Fuentes-Afflick, Department of Pediatrics, School of Medicine, University of California, San Francisco; Roderick J.A. Little, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor; Marie C. McCormick, Department of Social and Behavioral Sciences, Harvard School of Public Health and Harvard Medical School; Melissa McPheeters, Vanderbilt Evidence-based Practice Center and Emphasis Program Area on Healthcare and Public Health Research and Management, Institute for Medicine and Public Health, Vanderbilt University Medical Center; and Samuel H. Preston, Population Studies Center, University of Pennsylvania. Although the reviewers listed above have provided many constructive comments and suggestions, they did not see the final draft of the workshop summary before its release. The review of this summary was overseen by Shari Barkin, Department of Pediatrics, Monroe Carell Jr. Children's Hospital at Vanderbilt University. Appointed

by the National Research Council, she was responsible for making certain that an independent examination of this summary was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this summary rests entirely with the author and the institution.

Contents

1

Introduction

The National Children's Study (NCS) was congressionally mandated by the Children's Health Act of 2000. Section 1004 focuses on the National Children's Study stating, "it is the purpose of this section to authorize the National Institute of Child Health and Human Development (NICHD)[1] to conduct a national longitudinal study of environmental influences (including physical, chemical, biological, and psychosocial) on children's health and development."[2]

The NICHD Office of the Director has primary responsibility for planning and coordinating the National Children's Study, the largest (100,000 children) and longest (before birth to age 21) study of environmental effects on children's health ever conducted in the United States. The NCS will examine the effects of the environment, broadly defined to include factors such as air, water, diet, noise, family dynamics, community and cultural influences, and genetics, on the growth, development, and health of children across the United States, following them from before birth until age 21. The goal of the study is to improve the health and well-being of children and contribute to understanding the role various factors have on health and disease.

Between 2005 and 2007, the research plan for the NCS was developed in collaboration among the Interagency Coordinating Committee, the

[1]Congress passed Public Law 110-154 on December 21, 2007, renaming the Institute the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

[2]Children's Health Act of 2000, Public Law 106-310, 114 Stat. 1101.

NCS Advisory Committee, the NCS Program Office, Westat, the Vanguard Center principal investigators, and federal scientists.[3] A review of this plan can be found in National Research Council (NRC) and Institute of Medicine (IOM) (2008). The current design uses a separate pilot (or "Vanguard Study") to assess quality of scientific output, logistics, and operations and a "Main Study" to examine exposure-outcome relationships. The house-to-house recruitment strategy endorsed with qualifications by the 2008 panel of the NRC/IOM was tested in the Vanguard sites, resulting in recruitment of fewer pregnant women and births than originally estimated. The issues associated with sampling were further studied in the additional Vanguard sites. As the results from Vanguard sites became available, there was much discussion about the most effective sampling approach among the NCS Program Office, federal and non-federal sampling experts, Vanguard principal investigators, and others. Considering this input, the NCS proposed the use of a multilayered cohort approach for the Main Study, which was one of the topics for discussion at the workshop that is the subject of this publication.

In the fall of 2012, NICHD requested that the Committee on National Statistics (CNSTAT) of the NRC and the IOM convene a joint workshop, to be led by CNSTAT. The statement of task was as follows:

> An ad hoc steering committee will organize a public workshop on issues related to the overall design (including the sample design, participant recruitment, and framework for implementation) of the congressionally mandated National Children's Study (NCS). The NCS is intended to follow a cohort of children identified at or before birth through age 21 years. The study consists of a pilot or Vanguard Study, currently under way, focused on the feasibility, acceptability, and cost of study implementation and logistics to run in advance of, and parallel to, a Main Study. Based on Vanguard Study experience to date, the National Institute of Child Health and Human Development (NICHD) is proposing a multilayered approach to enrollment, using a national probability sample of hospitals and birthing centers for a birth cohort, supplemented with additional cohorts of pregnant women and preconception women, as well as special samples of population groups that may be underrepresented in the main cohorts. NICHD will provide a background document on the proposed design that will be discussed by workshop participants. The committee will develop the agenda for the workshop, select and invite speakers and discussants, and moderate the discussion. Following the workshop, a designated rapporteur will prepare an individually authored summary of the presentations and discussion. NICHD will also be provided with a verbatim transcript of the event.

[3]The National Children's Study has a website with a wealth of information concerning the history of the study and the activities that are under way. See http://www.nationalchildrensstudy.gov/Pages/default.aspx [June 2013].

In preparation for the workshop, NCS provided a background paper to the steering committee (National Children's Study, 2012).[4] The steering committee, collaborating by email and telephone, discussed the paper and the challenges to be addressed at the workshop, determined the agenda, and selected potential speakers. In addition, NCS and CNSTAT collaborated on the preparation of a more abbreviated workshop background paper that was distributed via the CNSTAT website in advance to all meeting attendees. The paper was referred to during the workshop as the Background Paper (and is cited in this publication as Kwan et al., 2013). The purpose of Kwan et al. (2013) was to explain the NCS and the status of the program and to pose specific questions for consideration at the workshop.

The Workshop on Design of the National Children's Study Main Study took place at the National Academy of Sciences on January 11, 2013. The three main topics addressed at the workshop were the collection of environmental exposure measures, the distribution of the sample among cohorts, and statistical issues associated with the sample design. These discussions, as outlined in Kwan et al. (2013), were intended to inform the NCS Program Office on specific design questions to guide the NCS Main Study, including considerations related to prenatal exposures, alternative approaches for collecting information, costs of such collections, and their value to analysis. The workshop was organized around four sessions, each with a specific set of questions to discuss. (The workshop agenda can be found in Appendix A, and a list of workshop registrants can be found in Appendix B.)

Each of the following four chapters is dedicated to one of the workshop sessions. The chapters begin with the information about the topic provided in Kwan et al. (2013) and the detailed questions posed by NCS. The chapters then summarize the main points of the panelists' remarks and the ensuing discussions. The overarching questions from Kwan et al. (2013) covered in each chapter are as follows:

- Chapter 2: Given the challenge as stated in the Children's Health Act of 2000 to "perform complete assessments of environmental influences on children's well-being," does the proposed visit schedule and sample collection balance the complex requirements?

[4]In addition to this paper, the NCS homepage for the NAS Workshop has links to the workshop agenda, the transcript, and many other background resources. See http://www.nationalchildrensstudy.gov/research/workshops/Pages/nationalacademyofsciencesworkshop.aspx [June 2013]. The white paper for the steering committee is titled *The National Children's Study Institute of Medicine Workshop Steering Committee Briefing Document*. The background paper for workshop participants is titled *Background for Discussion at the Workshop on the Design of the National Children's Study.*

- Chapter 3: What should be the criteria for the cohort allocation decision and what evidence is available to support an assessment of each criterion? What should be the allocation of sample cases among the various cohorts?
- Chapter 4: Given the study design proposal described in Kwan et al. (2013), and using the example cohort proportions proposed in the Chapter 3 questions, what enhancements can be made to address estimation and imputation challenges?
- Chapter 5: From today's discussion, can you synthesize the tradeoffs among factors, issues, and values that need to be balanced and considered by NCS leadership?

This report was prepared by a rapporteur as a factual summary of what occurred at the workshop. The steering committee's role was limited to planning and convening the workshop. The views contained in the report are those of individual workshop participants and do not necessarily represent the views of all workshop participants, the steering committee, or the National Research Council/Institute of Medicine.

2

Environmental Measures

This chapter begins with information on environmental measures provided in advance to workshop participants via Kwan et al. (2013, pp. 2-4), followed by two questions that panelists were asked to address. For clarity, the content from Kwan et al. (2013) is shown in block quote text throughout this workshop summary. The third section of the chapter provides highlights of the panel members' remarks and open discussion with the audience about the two questions.

BACKGROUND ON CURRENT PLANS

> The primary objective of the National Children's Study (NCS) is to examine the relationships among exposures and outcomes that affect children's health and development. While the NCS is considering a broad array of exposures, including characteristics of the family and neighborhood, this discussion will focus on a few exposure and outcome examples to probe some specific design questions. Current plans include, but are not necessarily limited to, collection of the following samples:
>
> - Household dust
> - Blood
> - Urine
> - Questionnaires on exposures and the social environment
> - Placenta and cord blood at birth

These samples could be tested for heavy metals, pesticide residues, semivolatile organic compounds, and pharmaceuticals. Outcome measures include but are not limited to

- Linear growth rate and body mass index as a proxy for general health
- A metabolic screen of serum total protein, blood urea nitrogen, cholesterol, iron, and calcium as a proxy for nutrition and dietary exposure
- Frequency and duration of health system encounters for respiratory illness as a proxy for pulmonary health
- Timing of standard neurodevelopmental landmarks and any deviation from adjusted trajectory as a proxy for cognitive and social development

The NCS also plans to use general exposure data collected at the municipal or neighborhood level (water quality, air quality, known industrial pollution) by either direct specimen collection or extant data collection.

The current data collection plan is based on an approach that uses a core questionnaire administered at every childhood visit, plus supplemental modules to be administered to specific participants or subpopulations based on events and conditions such as age, developmental stage, and other triggers such as specific exposures or hospitalizations.

Modules may be administered on a "missing by design" basis. There are at least two aspects of this missing by design approach: modules triggered by age, exposure, or specific events; and a "validation sample" approach. Triggered modules based on age, for example, make most sense where either scientific evidence indicates that exposures only at certain ages are likely to cause health concerns or a knowledge gap exists.

A validation approach may be useful, for example, when there might be two ways to measure a specific item, one inexpensive, the other expensive but more comprehensive. A smaller random sample may be assigned to have the expensive measure taken, but all respondents would provide the inexpensive measure. The data from the samples with both data might be used to establish a model to provide a correction to the inexpensive data that are available for the entire sample.

This only makes sense if there is knowledge that such a model exists or could feasibly be developed and could provide an improved estimate based on the inexpensive data. In addition to questionnaires, other modalities for data capture such as sound recordings, images, geographic movements, and mapping of social interactions and networks will be used. The NCS emphasizes data collections early in pregnancy and early in child development because the largest knowledge gaps, and perhaps the most critical events, occur during those time periods.

Pregnancy data collections are scheduled, if possible prior to about 20 weeks of gestation and once later in pregnancy. Data collections for

children are scheduled at birth and then every three months for the first year and every six months until five years old, for a total of 13 opportunities. Seven will be face to face, including biospecimen and environmental data collection. The other six are remote collections, typically by telephone. Subsequent data collections have not been specifically scheduled but will be on average about every other year until 21 years old, for a total of eight additional opportunities. The visit schedule is flexible in that children will not have assessments precisely at a given age, but within a window of several weeks around a particular age.

The NCS *Examples of Potential Exposures amd Outcomes* table[1] indicates that biospecimens of blood and urine will be collected from the mother prenatally, at birth, and when the child is 6 months and 12 months old. From ages 2 years to 5 years, blood and urine will be collected from the child. At each of these opportunities, except birth, the mother would complete a questionnaire, and household dust, among other samples, will be collected.

The table *Potential Environmental Exposures of Interest*[2] lists environmental exposures of potential interest to be measured in all NCS participants (general) and in a subset of participants (selective). Selective sampling will be based on the principle of enriching for a population more likely to have a risk of a particular exposure. For each exposure, corresponding examples of target analytes are listed. The rationale for biospecimen or environmental sample collection examples of target analytes, proposed and alternative measures, and potential health outcomes of interest are provided for each exposure type. Another column lists the preferred data sampling modality method to be used by the NCS with optional approaches for the Committee's review and consideration in adjoining columns.

The intent of the NCS is to have the highest quality biospecimen or environmental sample available, but the NCS may not have the resources to analyze each specimen or sample for each analyte in real or near time. Consequently, processing and storage of the specimen or sample are important considerations that will be based on analyte stability. The table does not represent all the specimens and samples the NCS intends to collect but is limited to those specimens and samples targeted to assess selected environmental exposures.

[1]To view the table, see http://www.nationalchildrensstudy.gov/research/workshops/Pages/nationalacademyofsciencesworkshop.aspx [June 2013]. Click on *National Children's Study Examples of Potential Exposures and Outcomes.*

[2]To view the table, see http://www.nationalchildrensstudy.gov/research/workshops/Pages/nationalacademyofsciencesworkshop.aspx [June 2013]. Click on *Potential Environmental Exposures of Interest*.

QUESTIONS ON ENVIRONMENTAL MEASURES

Given the challenge as stated in the Children's Health Act of 2000 to "perform complete assessments of environmental influences on children's well-being," does the proposed visit schedule and sample collection balance the complex requirements? Specifically comment on the proportion of different types of data collection—primary environmental sample collection, use of biological specimens for biomarkers of exposure, and use of secondary sources including retrospective analysis for environmental exposures. Considerations may include the following:

1. Are the proposed measures (biomarkers, questionnaires, physical measures) the most appropriate to assess exposures of interest? If not, what measures should be taken?
2. How should the NCS prioritize decisions regarding exposure assessments?
3. Some examples of factors to consider are
    a. potential public health impact of the outcome,
    b. technical feasibility including timing of data collection with regard to potential developmental vulnerability,
    c. scientific opportunity to address knowledge gaps and illuminate developmental pathways.

KEY POINTS OF THE DISCUSSION

Marie McCormick (Harvard University School of Public Health and the Harvard Medical School) moderated a panel that consisted of Antonia Calafat (Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention [CDC]), Nicole Cardello Deziel (Occupational and Environmental Epidemiology Branch, National Cancer Institute), Melissa Perry (Department of Environmental and Occupational Health, George Washington University), and Linda Sheldon (National Exposure Research Laboratory, Environmental Protection Agency [EPA]). The panel members focused their discussion of environmental measures and the timing of those measures by introducing the audience to the results of a workshop held in 2010 sponsored by EPA and the National Institute of Environmental Health Sciences (NIEHS) to consider exposure metrics for the NCS (Environmental Protection Agency, 2010). After Sheldon described the EPA-NIEHS workshop, a discussion among panelists about the proposed measurements and schedule ensued, followed by open discussion with the audience. The panelists then addressed prioritizing decisions, again followed by open discussion with the audience.

Proposed Measurements and Schedule

Description of EPA-NIEHS Workshop

Sheldon said that the charge given to the EPA-NIEHS workshop participants was to develop innovative exposure metrics and to look at the minimum amount of exposure data to collect to be able to assess three specific health linkages with exposures: air pollution and asthma, insecticides and neurological development, and endocrine-disrupting chemicals in reproductive endpoints. She explained that workshop organizers felt it was important to consider the minimum exposure data needed, rather than looking at all possible ways to measure exposure.

Three separate workgroups conducted their work before the EPA-NIEHS workshop. Each workgroup had an epidemiologist, a toxicologist, and two people who worked in exposure, ensuring cross-discipline coordination about the chemicals, the time periods of susceptibility, and, knowing that information, the feasibility of collections. About 50 people attended the 1.5-day workshop. Each workgroup presented its findings on chemicals of interest, kinds, routes and pathways of exposure, critical time windows, biological samples, environmental samples, non-measurement approaches, protocol recommendations, and research recommendations.

Sheldon noted the EPA-NIEHS workshop provided background and justification for its proposals. Two important topics addressed beforehand were the definition of an exposure metric and criteria for a good exposure metric. EPA-NIEHS workshop participants defined an exposure metric to include direct measurements as well as measurements combined with other data via a model that would provide the ability to estimate an exposure. They defined a true exposure metric as one that would indicate biologically relevant exposure during the entire time window of susceptibility. It was further expressed that an exposure metric might be a biomarker in urine if it were related to exposure and that a biologically relevant exposure metric is one where concentration in urine provides information about the concentration at the biological target where the effect would take place. Blood lead was noted to be an excellent biologically relevant exposure metric.

Sheldon said defining exposure metrics and what makes a good metric provided the standard for assessing potential metrics at the EPA-NIEHS workshop. The two important criteria used were whether the measurement leads to biological relevance and whether a sample collected on one day provides information about exposures over the entire period of susceptibility. She said this second criterion is extremely important, especially when considering prenatal exposures. The EPA-NIEHS workshop participants selected five time windows of exposure: first trimester, third trimester, first year after birth, years 1-4, and puberty (for endocrine

disruption). She said they also recognized that, though important, collection of measurements in the first trimester may not be possible for all women, and they considered how measurements collected during the third trimester might relate to exposure during the first trimester.

EPA-NIEHS workshop participants also discussed biological matrices. Blood for the mother was considered important for the first trimester and third trimester. For the child, while collecting blood may be important, only a limited amount can be collected from a newborn, and it is important to use a "blood spot" wisely. She noted the EPA-NIEHS workshop also viewed the collection of urine as important because it is easy to collect and is a good exposure metric, and that breast milk can be a source for measuring many exposures.

According to Sheldon, EPA-NIEHS workshop participants gave house dust the highest priority among environmental samples. A single sample can be used to measure not just the concentration per square meter in the dust, but also the dust loading can be collected with a standard collection mechanism during a visit. Vacuum methods appear to be most feasible, and a protocol can address the collection of multiple analytes from a single dust sample. A dust sample can yield measurements of organics, metals, and biological agents.

Shelton concluded by noting many of these recommendations are similar to the protocol proposed in Kwan et al. (2013).

Discussion Among Panelists About Measurements and Schedule

Perry noted minimizing the use of questionnaires, especially to collect environmental exposures, is important because of the burden associated with questionnaires, although a small number of questions can be used to collect information for specific key exposures that cannot be collected using a biologically relevant exposure metric. Deziel said very specific well-designed questions about pest treatments (e.g., "Do you treat for termites?"; "Do you treat for fleas and ticks?") provide good correlation with actual measurements of the expected active pesticide ingredients in dust. However, some information cannot be collected on a questionnaire because people do not know whether, for example, they have polychlorinated biphenyls (PCBs) in their home or if they have polybrominated diphenyl ether (PBDE) flame retardants in their televisions or couches. Dust can provide useful exposure information for those types of chemicals. Deziel echoed some of the benefits of dust measures, which she said can be very useful in providing information for chemicals for which there are no good questionnaire items.

Deziel said researchers at the National Cancer Institute (NCI), University of California at Berkeley, EPA, and elsewhere have shown that even

non-persistent chemicals tend to be rather stable over time once they are in a residential environment. For example, some of the work that NCI has done resulted in observed interclass correlation coefficients of 0.6, 0.7, 0.8, and 0.9 for pesticides, polycyclic aromatic hydrocarbons, PCBs, and even some PBDE flame retardants.[3] Hence, if these are the analytes of interest for the NCS, dust measurement could provide useful information about exposure during critical time periods, such as preconception and the first trimester, when the study is unlikely to actually have samples collected.

Deziel said the tables and documents provided by NCS do not make the intended method of dust collection clear. Possible methods include a vacuum bag, a high-volume small surface (HVS3) sampler, a subtle dust plate, a dust wipe, and an air sample. Almost all of these methods have been considered as part of the NCS sampling protocol in the past. She stated collection of some sort of bulk dust sample is important. She noted in some of her research at NCI, she and her colleagues have compared concentrations of chemicals from a participant's vacuum bag or vacuum canister with a more standardized vacuum approach, the HVS3, with very good correlation between the two methods for a range of chemicals, like pesticides, PCBs, and polycyclic aromatic hydrocarbons (PAHs). Hence, she said, a vacuum bag may be a feasible way to obtain useful exposure information using a method that is not too burdensome on the data collector or the participant.

Calafat said, referring to the EPA-NIEHS workshop, a good measurement plan would strive for the minimum data that would provide useful information. Since not one approach will fill all information needs, it is very important to get a minimum set of environmental measures, a minimum set of questionnaire information, some residential data, and, last but not least, biological specimens. Different biological specimens—blood, urine, and breast milk—are appropriate or most appropriate for certain chemicals. Persistent chemicals are measured in blood. Because many of these persistent chemicals are also lipophilic and partition into fat, breast milk would be an excellent matrix for assessing postnatal exposure to some persistent chemicals.

Calafat said monitoring the impact of non-persistent chemicals is a great challenge, but unfortunately the market is moving from persistent into non-persistent chemicals. Non-persistent chemicals metabolize quickly and, for the most part, people are exposed to them through episodic events. When a non-persistent chemical is combined with episodic exposures, such as those that might be encountered through diet, and not through regular exposures, such as through the use of per-

[3]Later in this chapter, Deziel notes that correlations above .7 have been observed in samples taken months or even years apart.

sonal care products, the issue of variability arises. The concentrations of these chemicals as measured in urine have tremendous variability, and a measure taken on a given day may not reflect exposure during the critical window of susceptibility or reflect exposure in the future. She said no single solution currently exists to identify episodic exposures to chemicals that metabolize quickly. One approach would be to collect as many samples as possible, as well as information about when the sample was taken and the time of the last urination. All could be important when analyzing the data.

Calafat said her laboratory is analyzing data from the initial phase of the Vanguard pilot consisting of samples for about 500 women. It was a nationally representative sample, and the urine samples were analyzed for a suite of different chemicals including some phenols, phthalates, and metals. Despite the variability, she said her lab is already seeing very important differences in concentrations of some of these chemicals, depending on the demographics of the population. Just because these measures are variable, she noted, does not mean that they are not useful. In some cases, these chemicals are very prevalent in the environment and could be introduced into the sample through the materials used for collection of the samples; thus, prescreening of materials might be considered to make sure this is not an issue. She noted prescreening would be very important if the goal is to measure metals.

Perry stated the proposal in Kwan et al. (2013) for the collection of biological matrices, to include urine, blood, blood from the infant, and cord blood, seems to be well put together and well timed to the extent that windows of susceptibility are understood. She acknowledged the challenges associated with identifying and collecting information from the prenatal cohort, although the prenatal period represents a critical window of exposure. She said the proposed protocol does not prominently feature collection of breast milk, which, although challenging to collect, is important to consider. She agreed with Sheldon that the collection of a single blood spot for a newborn and planning how to use it are important, and she also concurred that house dust is a very important biologically relevant matrix of exposure.

Perry noted the proposed sequencing chronology of measurements—possibly first trimester, third trimester, at birth, and then going forward—looks like a logical and a well-thought-out trajectory of sampling over time. However, Deziel said assessments of the appropriateness of the timing and method of sample collection usually depend on the research question of interest, while the current NCS proposal contains no specific research questions. She characterized the NCS proposal instead as a broad-based approach to collect a lot of detailed information to address numerous future research questions. Given this model, the repeated dust,

blood, and urine measurements proposed seem appropriate and strong metrics to collect.

Deziel also observed that Kwan et al. (2013) did not mention GPS measurements or obtaining a good residential history, both of which she said are important. She said with GPS coordinates at a residence and answers to a few questions about how long the individual has lived in that residence, the growing number of rich publicly available databases can provide useful information about exposure. Examples she cited included EPA's Toxics Release Inventory (TRI), National Air Toxics Assessments (NATA) databases,[4] and pesticide-use databases maintained by some states. These databases may provide a way to get exposure information during critical time windows when samples were not collected.

Open Discussion About Measurements and Schedule

Naihua Duan (Columbia University) asked how well first-trimester exposure can be assessed through retrospective recall to fill in missing data if needed. Perry replied recall depends on the contaminant. She reminded the audience of Deziel's comments that, for some more ubiquitous invisible compounds, no one is fully aware as to when he or she is being exposed. At the same time, one can imagine remembering a pesticide event or using a paint or solvent during the first trimester of pregnancy. Questions of this type might provide information about prenatal exposures during a very specific period such as during the first trimester, especially if a woman is asked immediately postpartum. However, questionnaires may not assess persistence of exposure.

Deziel added many of the studies in which she is involved look at samples collected months or even years apart and still see interclass correlation coefficients of 0.7 or higher for many pesticides, PAHs, and PCBs. The repeatability of these samples will depend on the physical and chemical properties of the chemical and consistency of use. Over a several-year timeframe, she and her colleagues observed a single sample may be representative of a period of months or years for some chemicals. They are also looking at questions to determine how well people recall pesticide use during different time periods of pregnancy and how well that recall correlates with dust collected. However, she thinks relying on recall could be challenging.

Sheldon noted one issue brought up in the EPA-NIEHS workshop was relocation: Every time a person moves, samples need to be recollected at the new site. At a minimum, it is important to know whether (and how long) the person was in the environment where samples are collected.

[4]These databases can be found at the following links: http://www.epa.gov/tri and http://www.epa.gov/nata [June 2013].

Duan noted the next workshop session would include remarks by Irwin Garfinkel (Columbia University) about the importance of the first-born. Duan asked whether much is known about mobility, especially as a family prepares for the birth of the first child, and any information on how mobility might compromise the persistence of an exposure. McCormick replied, in her experience, a young couple, particularly for the first-born, is likely to move into a house because they have a child. Deziel added this question highlights the importance of taking a residential history to learn the relevant timeframe of the sample. Michael Bracken (Yale University) said his studies show 30 percent of families move within seven years after a birth. Greg Duncan (University of California, Irvine) noted in one of his national studies, mobility rates were just under 20 percent per year. He said in general, rates are higher among low-income than high-income families and are higher among younger families. Most mobility is local within a county but still is a change of household residence.

Duncan said one of the design options under consideration is to recruit some of the children from a hospital or prenatal care providers, and then have subsequent births to the original mothers become part of the sample. The advantage of including subsequent births is that preconception and very early prenatal information on exposures can be obtained for these subsequent births. A problem is that they are all second and higher priority births, so there is no preconception and early prenatal information about first births. He asked the panel about different hypothesized effects of exposures for first births versus subsequent births and the importance of the distinction. He also asked about the importance of obtaining exposure information on first births quite early in the preconception period.

Perry noted even though collecting pregnancy and prenatal data is challenging, the data are critically important for generating information about in utero exposure related to new findings about prenatal bases of adult disease. She said one flaw she sees with foregoing prenatal sampling in anticipation of collecting information on the second- or third-born is the fact that individuals who have children with health problems may not go on to conceive and reproduce again. If the study relies solely on second births for prenatal and pregnancy information, there would be no information to study first births affected by an immediate or chronic disease. She went on to say that there has been active discussion about preconception and how challenging it is to collect preconception data, but perhaps "missing by design" approaches or validation studies, where the NCS could collect preconception data from subsamples of women, would help fill the data gap. She stressed the importance of information about exposures in infancy and early development as they relate to predicting adult disease. Noting, however, that if the NCS can collect early expo-

sure data, the study would be positioned to accomplish something very significant, because few studies have realized a large sample with early exposure information.

Deziel said by recruiting the second child, not only is there unbalanced exposure information on these siblings, but also the exposures of these siblings will be correlated. As a result, they would have to be analyzed separately or through different statistical techniques. She expressed doubt this approach would provide sufficient statistical power to analyze prenatal and preconception periods.

Calafat said there could be differences between the first-born and subsequent children in terms of exposure to persistent pollutants. Jean Kerver (Michigan State University) said the interpregnancy interval would also have to be considered. A woman could have decreased nutrient stores after her first birth if she does not have time to build them up before the second birth. Decreased nutrient stores would have an impact on prenatal biomarkers such as vitamin D or other fat-soluble vitamins, which Kerver said would be a big consideration in going to a design that would eliminate first births by design.

Garfinkel asked the panelists to sharpen the question about the effects of exposure. He said he understands the prevalence of exposures may differ and there may not even be second births if the first child has health issues. He noted the key question for him is if the biological effects of exposures differ for first and second births.

Nigel Paneth (Michigan State University) emphasized much of what is important in child health, particularly in many neurodevelopmental disabilities, birth defects, and preterm births, has been determined by birth. He said postnatal environmental measurements are irrelevant to the causation of those central components, which puts the onus squarely on the prenatal period. He described an analysis in which he is involved of MOBAND, a combined Danish and Norwegian birth cohort that totals 200,000. All have prenatal collections with prenatal blood and urine, but none has prenatal environmental house exposures. The possibility of collecting prenatal exposure environmental information is unique to the potential NCS.

Kerver said she would add diet back into the conversation and expressed concern that Kwan et al. (2013) stated that "a metabolic screen of serum total protein, BUN, cholesterol, iron, and calcium" may be used "as a proxy for nutrition and dietary exposure." She said those measures are not valid estimates of, and cannot be used as a proxy for, nutritional status or dietary exposure. Additionally, she said it is important to collect dietary exposure, like other exposures, in real time and prenatally.

Sheldon identified two reasons to collect dietary information: nutrition and exposure to contaminants. Her work has shown much variability

day to day in dietary exposure to contaminants for any particular individual and shows a study can only get the extremes.

How to Prioritize Decisions

The panel went on to address the second question: "How should the NCS prioritize decisions regarding exposure assessments?"

Discussion Among Panelists About Prioritization

Sheldon started the discussion by saying when she looked at the current NCS strategy she thought hypotheses to test were needed. As one starts to look at different hypotheses and what is known about environmental pollution, she said, key elements of the strategy are reflected in the sampling scheme, measurements, and operational feasibility. This leads to almost the same strategy as the current proposal.

She said the health outcomes being targeted should be kept in mind. The important questions become what can be measured and how well measures indicate exposures that are persistent or persistent for the time window of interest. A sample collected in one day has to represent an entire period. She said it is important to focus on groups of chemicals that are going to have a reasonable correlation or interclass correlation with different time periods.

She further observed that there are new models for air pollution and three relevant environments: the home, the community, and the ambient environment. For some air pollutants, available databases and modeling can retrospectively determine a person's exposure over various time periods. Knowing the time window of interest and the exposures to monitor during that window are absolutely critical for prioritizing decisions. Whether measuring exposure for epidemiology, exposure analysis, or another purpose, there is often a tendency to focus on things already looked at and with existing knowledge about how to measure them. New analytical techniques in the future may allow for screening of tens of thousands of chemicals in matrices like house dust. As technologies improve, they may help in the analysis of archived samples.

Perry expanded on the idea of using new technologies, referring to active progress in the United States and other countries in determining a way to consolidate multiple exposure measurements within one small sample of blood or urine. She reminded the audience that future analytic opportunities are dependent on success in collecting blood samples and cord blood samples. She suggested the use of personal monitoring. It is now feasible for individuals to wear personal monitors, monitor their physical activity, upload it online or in real time, and see personally cus-

tomized data. She suggested NCS might consider inexpensive monitors that participants could wear to monitor indoor air quality or nutrition intake. She underlined the importance of being fully aware of technology trends in order to use them at an appropriate time.

Sheldon followed up, saying many groups with National Science Foundation support are looking at this topic. As an example, a group with which she is working at North Carolina State University is using nanomaterials to generate power for monitors to collect the pollutant and a physiological response. When this technology is fully developed, a person would get a monitor on a patch and no further visits by staff would be needed, which would save on the cost of repeated visits. The key is having electronic transmission for long-term monitoring.

Open Discussion About Prioritization

Dorr Dearborn (Case Western Reserve University) commented on the availability of personal monitoring devices. He said with NCS funding, his group developed eight different residential air quality monitoring parameters that could be wirelessly downloaded and could detect when a cigarette was lit or gas cooking stove turned on. These devices are not the size that comes to mind when the term "personal monitors" is used, but they can also collect air particulates with laser light scattering. The researchers are almost ready to put them into some NCS participant homes to gain field experience with how they work. They could easily add a microphone and collect sound to get some sense of the source and nature of the sound on a continuous basis or put photocells into a child's bedroom to track nocturnal light exposure. These ideas are feasible now, and they are seeking input on the value of adding the sound and/or light monitoring.

Perry observed the technology for small cameras and other technology exists, but privacy issues are very real. It is not clear to what extent participants would be willing to use these devices. She noted people now seem to be willing to wear monitors for physical activity and have their data uploaded, and perhaps NCS respondents would also be willing to use these devices. The technology may be relevant to monitoring environmental contaminants and exposures as well as nutrition and physical activity.

Bracken noted the NCS has not commissioned systematic reviews of the use of different technologies. He suggested systematic literature reviews on what has been done in other cohorts might provide more information than relying on the limited Vanguard data. He provided examples from his work. In one of his studies, they gave women monitors to wear in pregnancy during three different weeks, all using randomized

nested subgroups. Although monitoring an entire cohort might be too expensive, he said it can be done in randomly selected subgroups. In one study cohort, subgroups of women wore monitors to measure environmental tobacco smoke; in a second cohort, subgroups wore monitors to measure electromagnetic field exposure, as documented in papers that have been out for 15 years now. A wealth of data is available on how to actively monitor pregnant women throughout their pregnancies.

Duan called personal monitoring a very promising technology. He described statistical issues associated with personal monitoring data. First, a stream of personal monitoring data over a period of time would be the ideal way to address the variability question that Calafat raised earlier. It might provide information about variation in the short term and variation over a period of time. Second, personal monitoring provides an automated way to sample a person's exposure across different activity patterns, rather than measuring the environment at a certain place or the residence.

Deziel said although she supports consideration of these new technologies, specifying how they would be used and the research questions they would address would be important to achieve a balance between costs and feasibility. Duan added the technologies are not all entirely new. The EPA has conducted a variety of personal monitoring studies for several decades, including the Total Human Exposure Study. The technology is advancing, but there is a history of its use. Learning from what has been done is a good idea. Perry agreed the technologies are not new, but people seem to be in an era of greater ubiquity and involvement in embracing them.

Sheldon observed the technology has to be inexpensive and easy enough to use so that repeated visits to the home are not required. One exciting idea is a self-powered, bandage-type monitor, which would make personal monitoring more affordable. If a device is cheap enough, a participant can throw it away when the measurement is complete. She said it is important to consider the cost of deploying and undeploying instruments in a large study.

Calafat noted the NCS will collect a large number of specimens, both environmental and biological, and store them for future analysis. A key activity will be to document exactly how the samples were collected. In the future, these samples may be used to test new chemicals (or biological or genetic information). She said the tradeoff between the cost of collecting and storing samples versus future use must be kept in mind. She noted if the cost becomes very high, one potential option would be to collect and/or store samples for a subset of the participants, much as is done within the National Health and Nutrition Examination Survey (NHANES). NHANES is a survey conducted by CDC that examines the

general population, but most chemicals and some other biomarkers are only measured in a subset of the population that can be made representative of the whole United States.

Bracken asked about sample storage, collection, and stability, particularly at the point of collection, saying samples are almost certainly going to be looked at for assessing environmental exposure in terms of gene-environment interaction or epigenetics. He stated it is crucial these samples are preserved for later decades, although many are not stable and deteriorate over time.

Deziel said exposure metrics might be prioritized based on simple descriptive statistics, such as the percent detection or the range of variability in the population, so there is adequate statistical power to look at questions of interest. She pointed to the 1.5 to 2 years of pilot data available to inform the Main Study, but there is now a very short turnaround between the pilot and the Main Study. She urged the NCS to mine those data as quickly as possible so the pilot can really inform the prioritization scheme and related decisions.

Sara McLanahan (Princeton University) asked about the tradeoff between the first trimester and the third trimester of pregnancy and the relative importance of measuring exposures during those two periods. Perry responded the answer depends on the outcome of interest, because the fetus goes through various stages of development during the three trimesters. If chromosomal abnormalities are of interest, data from the preconception phase are most important. If neurodevelopmental outcomes are most important, third-trimester data are most important. Even knowing the outcome of interest, she said, the precise timing at which measures are needed is not well known.

Paneth commented one of the questions the panel was asked to address concerned potential public health impact of the outcome and noted some of the conditions that might be related to prenatal exposures are not very frequent in the population. Congenital heart defects, cerebral palsy, and type-1 diabetes, for example, are prevalent at less than 1 percent. According to the data in National Children's Study (2012), the power in a sample of 100,000, even with 25 percent exposure, will barely pick up an odds ratio of 2. If the sample is cut down to 40,000 exposures and does not include measurements on all subjects, then the public health impact of what could be done with the NCS would be proportionately reduced. McCormick added NCS ought to be very explicit about which analyses (conditions) its data will and will not support. There may be conditions of high salience, such as autism, that may not be able to be addressed. Such limitations are important to state upfront.

Carol Henry (George Washington University) said she was not sure that the panel has yet come to grip with priorities. The EPA-NIEHS

workshop described by Sheldon recognized the NCS cannot be all things to all people, instead choosing three health outcomes to look at and to better understand the correlations between environmental exposures and health outcomes. If those three outcomes were emphasized in the NCS and data collection started, there could be progress in understanding them. Monitoring those three outcomes would be critical to the success of the NCS. Perry said in her comparison of the current plan to the EPA-NIEHS workshop, she determined that based on the chronology, including the prenatal with attempts at preconception, and the sampling plan over time, the NCS would have the matrices necessary and would be prepared to analyze the exposures of interest as they pertained to those three specific outcomes.

Perry also noted that she did not want to dismiss the question about the prevalence of outcomes and whether or not the study will be adequately powered given rare events. In fact, she noted, there may not be an adequate number of actual diseases or precursors to disease to be able to study them. As an example, she cited her own studies on chromosomal abnormalities in sperm, which are potential precursors to congenital abnormalities if that sperm is successful in fertilization. A number of predisease indices from DNA adducts[5] to chromosomal abnormalities could be identified. At the same time, there are genetic and epigenetic mechanistic studies that blood in particular is going to afford and that would provide insight into mechanisms without having the critical mass of required cases.

McCormick said she is uncomfortable when people talk about neurodevelopment as a specific outcome because neurodevelopment covers a fairly large number of relatively rare conditions. The study is unlikely to have the power to look at individual conditions.

Edward Sondik (National Center for Health Statistics) asked about geographic diversity in the sample: using the original 105 primary sampling units (PSUs) versus using a smaller number of PSUs. He asked about the tradeoffs associated with a more clustered sample in terms of the diversity of geographic experience in the country. Calafat said the response depends on the chemical of interest. If the chemical is an agricultural pesticide, for example, the study would want to ensure agricultural areas are covered in addition to non-agricultural areas. She noted NHANES samples about 15 localities every year, yet they obtain representative data for the whole U.S. population. The NCS will have a tradeoff among number of PSUs, cost, and the impact on outcomes and exposures

[5]From Wikipedia, "a DNA adduct is a piece of DNA covalently bonded to a (cancer-causing) chemical. This process could be the start of a cancerous cell, or carcinogenesis. DNA adducts in scientific experiments are used as biomarkers of exposure."

of interest. She added there may be differences in the use of some particular chemicals by demographics, such as phthalates. If a chemical related to exposure in a residential use is driven by socioeconomic status and there are not enough localities to cover that particular chemical, it could be a problem. However, a more ubiquitous chemical may make a larger number of sites less relevant.

Sheldon commented that community-level environmental exposures are important: air pollution, water pollution, and soil. Perry said in environmental exposure assessment, a wealth of information shows that environmental exposures are not uniformly distributed and different subpopulations are affected and exposed in different ways. People living in public housing, for example, are more likely to be exposed to pesticides and a variety of other chemicals and fumigants, as are people living in proximity to Superfund sites; thus, the importance of representative sampling when it comes to patterns of exposure is obvious. Kerver also expressed concern that reducing the number of PSUs would result in a reduction in the variation in food intake, which would reduce variation in estimates of both nutrient intake and exposure to pesticides through food among sampled women.

Roderick Little (University of Michigan, Ann Arbor) commented that one way of increasing power is potentially to increase the variability of the predictive variables of interest. He noted the proposed design includes a sample of about 10,000 births for "something else," the something else not yet specified. He asked the panel about the promise in focusing some of the 10,000 births on areas where it might increase the variability of the predictive variable, which might improve the power.

Perry noted that Kwan et al. (2013) covered two useful ideas: missing by design and the validation approach. She said the sample of 10,000 could come into play, for example, to study children in the Salinas Valley or children of farm workers who may be excessively exposed to ambient pesticides and food residue pesticides. A missing-by-design study could take advantage of the fact that exposure is much higher in that subpopulation. Considering two different measurement mechanisms—one more expensive, more invasive, and perhaps more precise; the other easier, cheaper, but perhaps not as reliable—would be a perfect scenario for a validation study. She also said these ideas might be used for preconception opportunities given the challenges of recruitment.

Calafat noted exposure for some chemicals is quite homogeneous, depending on how exposure is defined. If exposure is coming from use of personal care products that are also environmental chemicals, exposure is much more ubiquitous around the whole nation. She also noted many of the exposures proposed to be collected in the NCS are from indoor environments. There may be differences in exposure between populations in

urban and suburban environments, because the population in suburban areas is more likely to be spread out versus an urban population where people share more exposures. In an urban environment, even though a person is not using a particular product or chemical, a close neighbor may be.

3

Sample Design—Consideration of Multiple Cohorts

This chapter begins with information on sample design that was provided to workshop participants in advance via Kwan et al. (2013, pp. 4-7), followed by the questions panelists were asked to address on this topic.[1] The third section of the chapter provides the individual viewpoints of the four panelists, followed by open discussion.

BACKGROUND ON SAMPLING AND COHORTS

Geographical Area Sampling The original National Children's Study (NCS) plan called for 105 primary sampling units (PSUs) consisting of whole counties or groups of counties, with each PSU expected to generate about 2,000 births over the recruitment period based on 1999-2002 birth statistics, and with stratification by county size, percent black, Hispanic, Asian, and low-weight births. After sampling of segments (groups of census blocks), and door-to-door household screening, the PSUs were expected to generate about 1,000 births for inclusion in the NCS over a 4-year recruitment period.[2]

The NCS has abandoned the use of house-to-house screening methods due to projections based on the unexpectedly high and unsustainable resources that were expended in the initial phase of the Vanguard Study. The NCS is still planning to base the Main Study on a probability-based

[1]Subsequent to the workshop, the sampling plan for the NCS is undergoing revision.
[2]The duration of the recruitment period is still under consideration.

sample with current plans to start with a probability-based geographic area sample, though other probability-based options are under consideration. The optimal balance between number of PSUs, number of births per PSU, and environmental variety is not yet known. Different contract teams are working on scenarios, but this will not be resolved prior to the workshop and therefore will not be presented or discussed. The important point for the remaining discussion is that the NCS is currently planning an area probability sample of PSUs that is expected to generate 100,000 live births for participation in the NCS.

The Birth Cohort consists of births collected from a sample of hospitals and birthing centers and a subsample of women giving birth at those selected centers. This is tentatively planned to be about 45,000 participants of the overall sample. A 2-year initial recruitment period is proposed.

Within the current proposal, for each selected geographic PSU, a list of all hospitals and birthing centers will be prepared as a sampling frame for the birth cohort. Based on data from 2006, roughly 98 percent of all births in the United States take place at hospitals or birthing centers. A random sample of hospitals and birthing centers will be selected, with probability proportional to the number of births, and recruited to participate in the study. All women who give birth at the selected hospitals and birthing centers during specific times within the planned 2-year initial recruitment period will be eligible to be sampled while at the hospital, regardless of whether they live within the selected PSU or not. A systematic sample of women giving birth will be selected.

The NCS has documented multiple studies that recruit new mothers (and fathers) in the hospital and some that collect specimens (the Fragile Families Study is one of these). The acceptance rate is high and in some cases over 90 percent. The NCS has several strategies for collecting the relatively few specimens of interest (maternal blood and urine, cord blood, placenta, and perhaps an infant second dried blood spot following newborn screening), including collection from all sampled women during the recruitment windows and then discarding specimens from women who do not consent. The NCS is also piloting a few methods in the Provider-Based Sampling Vanguard sites to give some empirical data on acceptance, logistics, and costs. It is also possible in the birth cohort to attempt to collect medical records not only from the hospital or birthing center, but also from the sample member's prenatal care provider (if any).

The birth cohort will be a nationally representative sample of births in the United States. It can include stillbirths as well as live births.

While the recruitment of a relatively unbiased sample at acceptable cost is attractive, a knowledge gap that needs to be addressed is prenatal ex-

posure data. An essential question for the NCS is what is the scope and integrity of data that can be captured indirectly that would inform the prenatal history for a child recruited in the birth cohort.

The birth cohort mothers can be followed over time and subsequent children added to the sample. This provides an opportunity to include a sibling cohort and to collect both preconception and prenatal measures for some births. Information from the Fragile Families Study indicates that about 4.5 percent of women who have a child have another within 18 months, and 25 percent have another within 3 years.

Applying a similar analysis to the NCS would project the following scenario. For a sample of 45,000 births recruited over a 2-year period, allowing for about 350 stillbirths (1 in 150 of pregnancies) and, say, 4,150 attriters after the hospital interview, would leave 40,500 in the sample (assuming that women at the hospital would be oversampled to allow for refusals to participate). Of these 40,500, about 10,000 would be expected to have another child within a 3-year follow-up recruitment period. The subsequently born children could have prospective documentation of preconception and prenatal exposures.

The Prenatal Cohort is a sample of the prenatal care providers that are linked to the sampled hospitals or birthing centers from the birth cohort and a subsample of women who visit a prenatal care provider and expect to deliver at one of the selected hospitals or birthing centers. This is tentatively planned as about 45,000 births.

The primary purpose of a prenatal cohort is to obtain prospectively collected exposure data. There is some evidence that exposures within the first 8 weeks of pregnancy are the most critical. However, a prenatal sample enrolled from community care providers is unlikely to recruit very many women this early in their pregnancy. It has been estimated that at 8 weeks, only about 10 percent of pregnant women may have sought prenatal care, and these are likely to be those seeking fertility assistance, or those who are trying to get pregnant or have preexisting medical conditions and are monitoring. The prenatal cohort could, however, provide a reasonable sample of women in their third trimester of pregnancy.

The NCS will work with hospitals and birthing centers selected into the sample to identify the prenatal care providers including clinics, family practitioners, midwives, etc. that refer women to the hospitals and birthing centers. A sample of these prenatal care providers will be selected, using probability proportional to number of births. All women who are expected to give birth in one of the selected hospitals or birthing centers are eligible to be selected into the sample.

The NCS is currently exploring several options in the field with the

Provider-Based Sample Vanguard sites, including working with county medical societies and other professional societies and licensing bureaus, as well as using birth records (where available) to construct a list of prenatal care providers in the PSU. Birth records are available in some sites but not all sites. The logistics and resources required to prepare sampling frames of prenatal care providers as documented in the NCS Vanguard Study experience combined with information from other studies and the desire to work with selected hospitals and birthing centers for collecting birth information led to the approach described above.

The NCS's Vanguard Pilot Study data indicate that the proportion of women that providers inform about the study and that actually enroll is between 35 and 50 percent. In other words, for the most efficient providers, about 1 in 2 women enroll, and for others it is about 1 in 3. Prenatal cohort mothers can also be followed over time and subsequent children added to the sample. This provides an opportunity to include preconception measures and additional prenatal measures for some births as described in the birth cohort example. See above for further discussion.

It is not clear what population a prenatal care cohort would represent on a probability basis. If the prenatal cohort were limited to women visiting the sample of prenatal care providers within a specified time period who were in their third trimester, then the enrolled population would likely cover close to the entire population of pregnant women who receive prenatal care. However, the NCS would not be able to obtain measures of exposures earlier in the pregnancy except to the extent medical records contained relevant information. If the cohort were extended to include all women visiting the sample of prenatal care providers within a specified time period, then its representation of women in their first or second trimesters would be incomplete and could be biased given that women vary in the stage of pregnancy at which they seek prenatal care. In either case, the prenatal cohort is likely to have the measures most uniform for women in the third trimester. Although the proportion of women who receive prenatal care is relatively high, there are women who for multiple reasons do not receive prenatal care. Women who do not receive prenatal care could only be enrolled into the Study at a hospital or birthing center.

Preconception and Special-Purpose Cohorts are currently undefined, but may include a preconception group or a special-purpose group. About 10,000 sample cases are reserved for these, as yet unspecified, cohorts.

Neither the birth nor the prenatal care provider cohort can obtain information on preconception exposures for first-order births (except for what may be available in medical records for some sample members), nor will they necessarily include geographic areas of special interest (e.g., environmental "hot spots," such as areas where natural gas fracking is under way). The originally proposed design of household screening was intended to generate a preconception sample but proved infeasible on

grounds of excessive costs and time for recruitment. Following women in one or both cohorts will generate samples of subsequent births that occur within a window (of 2-3 years) that will provide both preconception and prenatal exposures. In addition, it might be possible to consider some special cohorts that could be sampled purposively or on a probability basis. For example, it might be useful to identify a small number of known environmental "hot spots" and seek to enroll all or a large sample of women of child bearing age at these locations. This cohort will be a convenience sample in addition to the other enrolled participants and is not intended as a topic of discussion for the workshop.

QUESTIONS ON ALLOCATION AMONG COHORTS

1. What should be the criteria for the cohort allocation decision and what evidence is available to support an assessment of each criterion? Examples include the following:

a. Recruitment costs, which include the costs of constructing the frame and the relative costs and efficiency of enrolling a participant.
b. Generalizability. What population is being represented?
c. Extent of exposures and other information that can be gathered. By definition, a birth cohort will have more limited data on prenatal exposures than a prenatal cohort, while a prenatal cohort will have less information on prenatal exposures (and much less information on preconception exposures) than the cohort of subsequent births to already enrolled mothers or a separate preconception cohort.

2. What should be the allocation of sample cases among the various cohorts? Assume that 10 percent of the sample is reserved for preconception and special studies; then, the allocation involves the remaining 90,000.

a. One option is the current proposal, which is about a 50-50 split or 45,000 participants in each.
b. Another option is something like an 80-20 split allocated between birth and pregnancy, with the pregnancy sample used to form the basis for imputing prenatal exposures (after using medical records for the mothers to get as much prenatal information as possible).
c. Yet another option is like an 80-20 split allocated between pregnancy and birth, with the birth sample used to form the basis for providing generalizability to the data analysis.
d. One extreme could be the entire initial enrollment allocated to the birth cohort, with studies of prenatal and preconception exposures using primarily the 25 percent cohort of subsequent births to originally enrolled mothers.
e. At the other extreme, most of the sample could be allocated to the prenatal cohort with a small birth sample consisting of women who did not receive any prenatal care and are enrolled at the hospital.

KEY POINTS IN THE DISCUSSIONS ON SAMPLE DESIGN

The moderator for this panel was Barbara Carlson (Mathematica Policy Research) and panelists were Michael Bracken (Center for Perinatal, Pediatric, and Environmental Epidemiology, School of Public Health, Yale University), Naihua Duan (Division of Biostatistics, Department of Psychiatry, Columbia University Medical Center, Columbia University), Irwin Garfinkel (School of Social Work, Columbia University), and Nancy Reichman (Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey and Department of Economics, Princeton University). Carlson introduced the session saying that it was set up as a debate with each panelist stating his or her own views about the questions. The individual viewpoints of the four panelists, discussion among the panel, and open discussion are summarized below.

First Viewpoint

The first viewpoint was presented by Irwin Garfinkel (Columbia University). Garfinkel reminded the audience that the NCS is likely to be the most important birth cohort study in the United States for several decades, and, in retrospect, the protracted struggle and very expensive pretesting over how to conduct it is not surprising. Reconciliation of different objectives and different disciplinary traditions will be important. In particular, it is not obvious how to collect prenatal and preconception exposure data from a population-based probability sample of births at reasonable cost. He characterized the recent evolution of the study design as very positive and on the brink of reconciling conflicting objectives.

Garfinkel observed that he and fellow panelist Bracken were on opposite ends of the spectrum about the appropriate balance between the birth and prenatal cohorts, but they have a fundamental area of agreement: Probability sampling is essential to the quality of the NCS.

He noted that in Kwan et al. (2013), the proposal was to enroll 45,000 mothers and children at birth from separate probability samples of hospitals and prenatal providers. He characterized the 50-50 split between hospitals and prenatal clinics as a huge step forward from sampling only from prenatal clinics in terms of cost and scientific value. He argued that neither the prenatal nor a birth cohort would produce excellent prenatal data, and that only sibling data could produce them at reasonable cost. He said if NCS used an almost 100 percent birth cohort that enrolled subsequent sibling births, this would save even more costs than the current 50-50 split and would immeasurably increase the scientific value of the study when it is completed 21 years from now.

Garfinkel said he and Bracken also agree that collecting prenatal data is a critical component of the NCS and that his own understanding

was reinforced by the discussion of the first panel (see Chapter 2) that first-trimester data are the most valuable part of prenatal data for many questions of interest and, for some questions, preconception data may be equally important. He stated that data produced by the prenatal sample would fail on these grounds because first-trimester data would not be collected from a sufficient percentage of sampled women.

Garfinkel offered a potential design and the data needed to implement it. He suggested an area probability design with a sample of hospitals in each selected primary sampling unit (PSU), and a sample of births within each selected hospital. The ultimate probability of selection would be known. He assumed that 60,000 mothers and their children would be enrolled in hospitals at birth, and that placentas and cord blood would be collected, as would breast milk. For these selected women, all subsequent sibling births over the course of the 21 years of the study would be enrolled in the study. Women would be appropriately monitored to determine when they become pregnant. He stated that this design would provide nearly as large a sample of children with prenatal data as the proposed 50-50 design. Further, if first births were oversampled in the birth cohort, the sample of siblings with prenatal data would be as large as a pure prenatal sample. He said these sibling prenatal data would be superior to the prenatal data provided by the prenatal sample because the sibling prenatal data can be collected earlier during the first trimester and may also include data on preconception conditions as well as data on a previous birth.

Garfinkel said if the fundamental biology of harm from environmental exposures is the same for first and subsequent births (observing the first panel provided no evidence to the contrary) and early prenatal data and preconception periods are critical, his suggested design is nearly optimal. He noted even if this assumption were not true, his proposed design enormously simplifies what is otherwise an extremely complex sampling problem. A third virtue is that the design points to the importance of finding out what is known about this assumption. Since it is possible that the biology of exposures is different among first-born and siblings, a small prenatal sample of first births may be justified.

Enrolling sibling births from a birth cohort has enormous virtues, he said, because it is the most cost-efficient method of sampling births during preconception and very early pregnancy. Within 3 years of all births, nearly 30 percent of mothers have a subsequent birth. Within 5 years, the figure is about 44 percent. Within 21 years, the overwhelming majority of mothers would have completed their childbearing. Assuming that completed fertility is about 2 children, a birth cohort of first births would have sibling births with preconception and prenatal data on about the same number of births as a 100 percent prenatal cohort.

Further, according to Garfinkel, each observation generated by the sibling sample would be superior to the prenatal sampled observation because it would contain data not only on preconception, but also earlier prenatal data and data on a mother's previous births, including placentas and cord blood. This information would be invaluable for imputing missing exposure for the first-birth prenatal period. He argued that as long as the biology of exposure is the same, the best data, not just on preconception but also the prenatal period as a whole, would come from siblings and not from births sampled prenatally.

He described two other advantages of the sibling sample. First, although sibling-based estimates would be less precise than corresponding non-sibling estimates, the sibling sample would allow researchers to control for or rule out confounding from genetic and environmental circumstances shared by siblings. Second, collecting sibling data would be cheaper from start to finish than collecting data from two children from different mothers and different household circumstances. Each birth enrolled in a prenatal sample cohort would be de nova. Each sibling enrolled from a birth cohort would be enrolled from a mother who was previously recruited and is a loyal member of the study.

He further stated that a birth cohort would be superior to a prenatal cohort in terms of cost and sample size for two reasons. First, enrollment costs of a birth cohort would be smaller than enrollment costs of a prenatal cohort because of economies of scale. Second, the prenatal data collected from the first births enrolled in a prenatal cohort would be very expensive. NICHD estimates that the cost per child of prenatal enrollment and collection of prenatal data is at least three to four times, and may be as much as 10 times, the cost of enrolling a child at birth. He illustrated how these ratios could be so large, assuming a cost of $1,000 to enroll a mother in either a prenatal sample or birth hospital sample and another $5,000 to collect prenatal data from the mother. The ratio of total cost would be 6 to 1. If enrollment costs were $2,000 and prenatal data collection costs $18,000, the ratio would be 10 to 1. In other words, he said, for every child enrolled in a prenatal cohort, 3 to 10 children could be enrolled in a birth cohort for the same cost.

Although costs would be incurred for every sibling birth enrolled, he said total costs are lower for four reasons. First, enrollment costs of already loyal members of a longitudinal study would be lower than enrollment costs of de nova prenatal mothers. Second, the costs of collecting data on family circumstances would be lower for siblings. Third, the siblings would be followed for a shorter period of time. Fourth, the enrollment and data costs of siblings come later than the enrollment costs for a prenatal sample, which means they are lower because the later-incurred costs would be discounted.

He noted the only parts missing from a birth cohort with siblings are the prenatal and preconception data on first births. The prenatal missing part could be efficiently filled in by a relatively small sample of first-time pregnant mothers drawn from prenatal providers. He estimated that roughly 10,000 would suffice and might indeed be too high. Every additional birth to a first-time pregnant mother drawn from the prenatal providers would reduce the number of sibling births that could be enrolled in the study by between 4 and 10 children.

Garfinkel concluded that his analysis identifies the key scientific questions underlying the choice between the size of the prenatal and birth cohorts: How important are early prenatal data? How important are preconception data? Is the fundamental biology of harm different for environmental exposures for first and subsequent births? The key operational questions all relate to cost. Is the ratio of the cost of enrolling the prenatal sample as opposed to the birth sample 3 to 1 or 10 to 1? How costly will it be to collect placenta and cords on the first births? Finally, time is important. The birth-cohort sibling design would collect prenatal data in later years than would a prenatal cohort. Once these issues are clarified, a formal sample design would provide a precise optimal allocation.

Second Viewpoint

Naihua Duan (Columbia University) supported Garfinkel's suggestion about incorporating the sibling cohort into the study and thanked the workshop's first panel for laying the groundwork on samples to be collected and important time periods. He said he agreed with the suggestion by some Vanguard investigators that objectives are a good basis for the design of a study, noting the NCS has the potential to go beyond being a descriptive study. Specific hypotheses will help everyone understand how design decisions are made. He concurred with Roderick Little's comment during the open discussion in the previous panel (see Chapter 2) that to the extent possible, maximizing the dispersion of potential exposures in the sample to get both high- and low-exposure measurements is a good idea.

Duan observed that a large, complex, and multifaceted study like the NCS needs to balance across multiple study objectives, and the theory of optimal design may shed some light on this issue. The optimal design literature mainly started with Kiefer (1959) and has evolved into a major literature in statistical methodology, mainly in experimental design. The main idea is to use the methodology to balance across multiple study objectives. He described how he has used this approach in several studies. For example, the Human Immunodeficiency Virus (HIV) Cost and Services Utilization Study (HCSUS), initially sponsored by the Agency for

Healthcare Research and Quality, recruited HIV-positive patients through care providers, somewhat similar to the way the prenatal sample is proposed for the NCS. Another example is the National Latino and Asian American Study, one of the surveys sponsored by the National Institute of Mental Health Collaborative Psychiatric Epidemiology Surveys (CPES). Duan noted the potential for wider applications of this methodology in sampling applications.

He also provided several examples of using the theory of optimal design. In his first example, he used what he termed is a naïve simplistic model for what the study might want to accomplish. Here, Y represents the outcome, perhaps the cognitive function for children at age 5, and E_1, E_2, E_3 represent measures of exposure at different time periods.[3] E_1 might represent exposure in the first trimester, E_2 might stand for exposure in the third trimester, and E_3 might stand for postnatal exposure. The relationship among these variables is given in the regression model

$$Y = b_0 + b_1E_1 + b_2E_2 + B_3E_3 + e.$$

The magnitudes of the estimated regression coefficients b_0, b_1, b_2, b_3 reflect the relative importance of each study objective in explaining Y. Duan noted that Kwan et al. (2013) listed five candidate designs under consideration, which he labeled D_1, D_2, up to D_5. An alternative, D_6, might be to take 40,000 from a birth cohort, 40,000 from a prenatal cohort, 10,000 from a sibling cohort, and maybe 10,000 from hot spots. Other designs might allocate the cohorts differently. An exercise in optimal design would specify performance criteria for each design and then, somewhat similar to doing power calculations, calculate those criteria for each design to determine which has the best performance according to the specified criteria. The key is to specify common criteria across multiple study objectives.

One simple performance measure or criterion in the optimal design literature is called the "A-Optimality Criterion." It is the sum of the variances of the parameters associated with each design:

$$P_1(D_k) = V(b_0; D_k) + V(b_1; D_k) + V(b_2; D_k) + V(b_3; D_k).$$

He noted this might not be a very good criterion for the NCS because it does not take into account the relative importance of the different study objectives.

[3]Exposure measures are standardized to have zero mean and unit standard deviation. Hence, the effects are standardized, and b_0 has the interpretation of the population mean.

His next example was a weighted version of the performance metric. Each *W* stands for the relative importance or weight the investigators want to attach to each study objective:

$$P_2(D_k) = W_0V(b_0; D_k) + W_1V(b_1; D_k) + W_2V(b_2; D_k) + W_3V(b_3; D_k).$$

Duan stated this example does not really answer the question, instead transforming it to a question that might be more tangible for the investigators to think about, namely asking how important it is to reduce uncertainty in the estimated parameters.

In his final example, he stated that instead of considering the variances that are usually used in the optimal design literature, it is conceivable to use a performance measure like the mean square error—the variance plus the square of the bias—to incorporate both the sampling error and also the non-sampling error:

$$P_3(D_k) = W_0MSE(b_0; D_k) + W_1MSE(b_1; D_k) + W_2MSE(b_2; D_k) + W_3MSE(b_3; D_k).$$

Duan said non-probability sample strategies might be able to be incorporated into this framework to assess how bias and variance tradeoff, and suggested the study recruit a statistician who is familiar with the issues involved to work on the design.

Duan argued lifetime costs are a more useful basis for decision-making than only recruitment (or up-front) costs. The ultimate product in 21 years or so is what the study overall has accomplished and what it cost. He suggested taking follow-up costs into consideration in choosing sampling strategies, with future costs discounted in today's dollars. He also observed the multicohort study uses a multiframe sampling strategy. The multiframe sampling strategy does not require each cohort to be representative of the entire population. Instead, the combination represents the entire population. It would be possible, for example, to include the sibling cohort that does not cover the entire population but gives good coverage for an important part of a population as long as the rest of the population is covered otherwise.

He suggested integrating special cohorts into the overall design. The 10,000 special cohort may be from hot spots; however, it will likely be analyzed together with a main cohort comparing hot spot exposure to exposures among the general population. It would be advantageous, he argued, to have a single probability sample that covers all special populations, with perhaps other special populations useful to include. For example, the first panel noted the potential higher exposure to pesticides in multifamily housing. Integration of the sample might allow NCS to be more flexible in thinking about sampling strategies, Duan concluded.

Third Viewpoint

The next viewpoint was presented by Nancy Reichman (Princeton University). She said that as a research associate at the then-new Center for Research on Child Wellbeing at Princeton University in 1997, she was involved in the new birth cohort study called Fragile Families. She described how Fragile Families involved interviews with new parents and medical record data collection at 75 hospitals in 20 cities across the United States with a success rate of over 90 percent of sampled hospitals. Reichman was responsible for gaining hospital access for the Fragile Families Study.

In her opinion, it would not be harder to get hospital access today than for Fragile Families. In the late 1990s, institutional review board (IRB) policies were in flux. Some of the biggest problems Fragile Families had were in hospitals that initially had the easiest application procedures but then retroactively decided that the approved study was not acceptable. The issue of changing IRB procedures is likely to be much less of a problem today because procedures have universally become much more formalized. However, substantial resources and a well-chosen team would be needed to get through the necessary processes. She noted that for Fragile Families, whether it would be logistically possible to collect placental material and cord blood when mothers give consent after they give birth was not an issue, although it may be now.

She said she and Garfinkel, who consulted with doctors and hospital administrators, think it would be possible to collect placental material and cord blood when sampling is done in hospitals for several reasons. First, there are no risks to the mother from collecting the needed materials. Second, it is apparently not unusual for mothers to take the initiative to have their placentas preserved and banked. She said they have been told by hospital administrators and research deans that if fairly compensated, and the burden to the hospital minimized, the hospitals would likely agree to a system in which placentas that might be needed for the study are preserved and stored, at the study's expense, with those of non-consenting mothers later destroyed. She suggested storage of the placental material and cord blood could be an incentive to consenting mothers, made available to them if needed in the future.

Reichman agreed that the question about optimal allocation cannot be answered without a clear accounting of projected costs and benefits of each type of sampling. She asked about administrative approval or outside institutional approval at prenatal care providers. She suggested that if the provider is in the hospital, such as at a hospital clinic, the hospital IRB approval is needed. She asked about the different types of prenatal sites, the average number of mothers expected to be recruited per site, and the costs to maintain quality control and standardization of protocols

across possibly a large number of small sites, noting that keeping track of case dispositions and response rates, particularly the denominators, could be a logistical challenge. She observed that it is difficult to compare the cost of securing institutional access and running the study without having better information on these aspects.

Sampling done at hospitals, she noted, would require a hospital encounter. If, on the other hand, sampling is done at prenatal care providers, mothers could request placental material, cord blood, and medical records from the delivery hospital for purposes of the NCS, hence eliminating the need for a hospital encounter. If a hospital encounter would still be needed for some reason, sampling from prenatal care providers would be enormously expensive compared to sampling from hospitals because of the added cost of the prenatal data collection encounters, access to both prenatal care providers and hospitals, and the logistics of coordinating the study across so many sites. She said key pieces of information are missing, including the cost of access and recruitment for prenatal care providers and hospitals, cost of obtaining placental material under both options, and detail about whether women in a prenatal cohort require a hospital evaluation. Once the relevant information is available, the two approaches could be compared via full cost benefit calculation that includes the sibling cohort.

Reichman added that while the NCS will be truly pioneering by collecting prenatal and preconception data in addition to birth data and beyond, she urged obtaining additional information, such as information on the health or death of the mother's and father's parents. Examples of information that might be collected from death certificates include parent's name, cause of death, education, date of birth, and year of death. Brief information about the parent's lifetime smoking and drinking could also be collected from mothers, since, as animal and human studies increasingly demonstrate, determinates of health can originate well before the parent's generation.

Fourth Viewpoint

Michael B. Bracken (Yale University), the final panelist, said his remarks went in a different direction based on his experience and biology background. He proposed recruiting the majority of the NCS participants (85-90 percent) in pregnancy through the prenatal sample. He explained that fetal origins of disease are dominant issues in studying both childhood and adult disease and only very large pregnancy cohorts could provide the information to study them. As examples, he cited Herbst et al. (1971) as a crucial paper showing how female fetuses exposed to diethylstilbestrol tended to develop vaginal adenocarcinoma when they

grew up. Antibiotic use in pregnancy is known to increase risk for asthma. Five percent of all pregnancies result in children with mental birth defects and physical defects. Pregnancy cohorts are needed to study all of these.

Pregnancy cohorts could also provide data to answer such questions as why people born at low birth weight have higher risks of adult cardiovascular mortality or seem to have lower risks of cancer mortality, and the influence of many drugs used by millions of women in pregnancy on their children's physical and mental disabilities remains uncertain. Other public health concerns can only be understood by studying pregnancy cohorts. As examples, what are the effects of exposure in pregnancy to antidepressants, antiepileptics, antiemetics, or pesticides on the developing fetus? Causes of autism, cerebral palsy, attention deficit hyperactivity disorder (ADHD), and many other so-called perinatal conditions actually have origins earlier in pregnancy, but they are not understood. He emphasized that all of this research could be supported by pregnancy cohorts.

He said exposure data are poorly recalled in questionnaires, even when women are asked at birth about exposure during pregnancy. Further, infant mortality in the United States ranks 34th in the world and is becoming worse (compared to 12th in 1960 and 23rd in 1990), but the causes will not be found in birth cohorts. Rather, he said, they are due to associations in pregnancy, including disparities in prenatal care, and only the prenatal cohort would provide the data to study these issues.

Referring to an earlier publication he wrote, Bracken noted, "We know that the vicissitudes in our own uterine existence may profoundly influence the rest of our lives, both physically and behaviorally" (Bracken, 1984). He said nothing has changed since then; moreover, pregnancy itself merits study. Miscarriages occur in about 15 percent of clinically recognized pregnancies, and fetal death and stillbirth are outcomes of great concern. Again, he stressed, only the pregnancy cohort would provide the information to study their causes.

He suggested the NCS could make a real contribution by looking at pregnancy cohorts, as many birth cohorts are being completed around the world. In his view, the proposed mixed cohort, the layered sample, is too cumbersome, unnecessary, and misses the real scientific goals.

Bracken said there is no evidence that a pregnancy cohort is more expensive or more costly to recruit in provider practices. The table in Annett et al. (2013) showed evidence from 16 cohorts, three of which were Dr. Bracken's. Most collected biospecimens for an average cost, including indirect costs, of $2,000. Even with inflation, the costs could not possibly exceed more than $5,000, still two orders of magnitude less than the NCS Vanguard costs.

He explained that within a PSU, a list of providers and the hospitals in which their patients deliver is developed into a cluster. These clusters

are then sampled to form a probability sample. There is no cost to the sampling process itself, and recruiting sampled providers in a hospital would be no more costly than recruiting a convenience sample of providers in a hospital. Sampling fractions and denominators could be obtained from birth certificate data.

He said a blood sample collected prenatally carries the same cost as a blood sample at birth but contains more valid pregnancy information. There may be additional costs in prenatal exposure assessment versus estimating prenatal exposures at birth, but these costs are related to sample collection, rather than subject recruitment. To him, they are costs worth bearing because they relate to collecting more valid data.

Bracken posed the question about how early gestations could be studied in a pregnancy cohort. He said he has had four Yale cohorts, a total of almost 17,000 pregnancies. In one, where the researchers restricted gestational age to week 16, they recruited 30 percent at 8 weeks and 91 percent by 12 weeks. In another cohort restricted to 22 weeks gestation, it was almost the same at 8 weeks—29 percent—and at 12 weeks, 76 percent. Extrapolating to a cohort of 100,000 pregnancies, as many as 30,000 women could be assigned for interview by 8 weeks, and 75,000 to 90,000 by 12 weeks. He stated collecting first-trimester exposures in pregnancy cohorts is well documented and is not particularly complicated.

Bracken stated collecting prenatal information on first births, not just the subsequent births in women already enrolled, is crucial because first pregnancies are biologically different from subsequent pregnancies. For example, preeclampsia is a first-pregnancy disease and fetal growth restriction is more severe in first pregnancies; in addition, as children are followed up through childhood, birth order becomes important. He stated that collecting prenatal data only for children who already have a sibling would be a detriment to the NCS.

He noted that biological exposures may not differ between first and subsequent pregnancies, but the scientific interest is in the interaction between these exposures and the fetus, and the fetus changes from one pregnancy to another. There are important biological effects that are already being studied in gene environment studies and using epigenetics. He said these are areas with more hypothesis than fact but warned against never being able to study these questions because of assumptions made at the sample design phase. An assumption-free strategy for sampling and recruitment places fewer constraints on the way pregnant women are sampled so they are representative of all pregnancies in the United States.

He labeled the preconception cohort as a particularly interesting group because many hypotheses concern exposures at the time of conception or before. It is also a difficult cohort to recruit. Women in fertility clinics who may know and plan the exact date of conception are highly

selected exactly by virtue of their infertility. Preconception probability samples are almost impossible to obtain and likely not worth the effort. He viewed the sibling cohort as a natural way to obtain data to support preconception studies. It uses women who are already recruited to the NCS and is based on an original probability sample. Although it has the significant disadvantage of including only preconceptions after a prior pregnancy, it may be the NCS' only feasible alternative to a preconception cohort before first pregnancies.

Bracken emphasized that he sees no advantage to the birth cohort because, to him, it misses the unique opportunity offered by the NCS to study the most important scientific questions. He stated that recruitment at hospitals is only worthwhile for women who receive no prenatal care, an important group of women who are often at high risk for poor health and have problems in child rearing. He supported development of special recruitment strategies for these women.

He said, in his experience, recruitment is easier in a prenatal clinic than at a hospital. One has to consider the provider, the hospital, and the research subjects. Providers are easy to recruit and do not have IRBs (although, in answer to a question from Reichman, it was noted that the researcher's own IRB would have jurisdiction). Providers are also more homogeneous than hospitals in the way they deliver care. Bracken said he has had more hospitals refuse to join research than providers. He noted refusal by a hospital to participate eliminates many more women from a sample than does refusal of a private practice, because all of the associated practices are eliminated.

Bracken said consent is more readily obtained from subjects prenatally, and it is unethical to try to obtain consent when a woman is in labor. In hospitals, after a mother has delivered, either she or the child may be indisposed. Twelve-hour discharges are very common in hospitals, which would mean missing a large number of women. Regarding obtaining consent after labor to get cord blood and placentas, he said it may be possible to get consent, but the cord blood and placenta would probably have disappeared. In addition, the presence of families and the excitement of postbirth are other barriers to obtaining consent after labor. He noted in contrast, when recruiting in a prenatal practice, the medical records of study subjects are flagged when they go into the hospital so the delivery room staff know to keep the placenta and cord blood.

In conclusion, he stated the most sophisticated sampling design will fail utterly unless the practical details of how obstetrical care is delivered in this country are taken into account, both in provider offices and in hospitals.

Discussion Among Panelists About Sample Design

Garfinkel noted he and the other three panelists in the session agreed that NCS needs to get prenatal data, and first-trimester prenatal data are important. He suggested one approach might be to compare which approach would result in the most first-trimester prenatal data. He expressed his opinion that when all up-front costs are considered, NCS could get as many or more women with early prenatal data from the birth cohort with sibling follow up, because collecting prenatal data on the first birth is so expensive.

Bracken replied that it is a matter of the scientific questions, not just cost. Prenatal information on first births is important; if the study resulted in no prenatal data on pregnancies to women delivering for the first time, important scientific questions would remain unanswered.

Garfinkel referred to the previous panelists (see Chapter 2), who stated they are not aware of evidence that the biology of exposures differs by first and subsequent births. Bracken replied that the biology of exposures is only half of the question: The question remains how these exposures interact with a developing fetus and said many examples may indicate that the developing first-pregnancy fetus is not identical to subsequent developing fetuses.

Reichman provided another argument for the prenatal sample, stating collection is structured as part of the prenatal care of the mother during her regularly scheduled visits to the provider. With subsequent siblings, she noted, there is no connection to the provider, which might complicate collections.

Duan reiterated that he and the other three session panelists agree that exposure data as early as possible are important. He suggested a prenatal sample and sibling sample are not mutually exclusive, with the question how to combine and make the best use of them. He also noted that missing first births is an important question to address. The prenatal cohort might offer the best solution, unless there is a practical way to get a preconception cohort. Strategies may be combined, instead of trying to use one or the other.

In response, Bracken stated complex designs are more difficult to manage operationally. Managing the schedules of women in different subsamples is complicated, increasing the chance of mistakes. He suggested a straightforward sampling strategy where women are only recruited during pregnancy. This would simplify the study and remove errors that might occur in the field in trying to implement sophisticated subgroup study designs.

Duan stated that he appreciates the argument for simplicity, but noted advances in information technology may make some approaches more feasible. He noted in a provider sample, the appropriate design needs to

take into account both the response rate at the provider level and at the individual patient level; when a provider refuses to participate; all its patients are automatically non-respondent to the study. He pointed to the American Association of Public Opinion Research (AAPOR) definitions of response rates and cooperation rates, saying that measuring and monitoring them is likely to be an important component of the NCS.

Carlson noted she had an opportunity to listen to one of the weekly calls among the Vanguard principal investigators during the week before the workshop. She related that they are giving prenatal care practices four choices for the prenatal cohort, and each practice selects the one that works best for it as a way to sample and recruit women. Most practices are choosing a temporal type of sampling, although there are cases where they feel that they may not be completely keeping track of the denominator. The denominator is probably harder to track in prenatal providers than in a hospital.

Bracken noted that birth certificates will eventually provide the data. He went on to say that the document provided by the Children's Study on power calculations (National Children's Study, 2012) uses very broad categories of defects and shows how well they could be estimated. Included are nervous system defects, major birth defects, neurocognitive development, neurodevelopment disability groups, and developmental disabilities. However, this is not the way people study malformations. For birth defects, the important issue is congenital heart malformations; even then, there are numerous subgroups. When these are studied in the proposed pregnancy cohort, by going from a sample size of 100,000 children to 45,000, even the bit of (inadequate) power presented in National Children's Study (2012) has been reduced.

Open Discussion About Sample Design

Dorr Dearborn (Case Western University) noted that Vanguard recruitment was not limited to first births and asked if the dataset would support a comparison of first births and subsequent births. Garfinkel agreed the question is worth testing. To compare first births to subsequent births, good early data from a small prenatal cohort would be important. Garfinkel asked how big that cohort has to be, saying that he doubts it has to be more than 10,000. He added that determining the size of the cohort is a scientific question.

Bracken noted that with the birth cohort plus siblings, no real-time pregnancy data on first births are obtained, which he termed a dangerous position for the study going forward. In contrast, he noted that the pregnancy cohort would advantageously include first, second, third, and all other births. Further, the ability of analysts to examine the effects of

covariates on child and adult health would be severely limited if the study data are confounded by the lack of a representative sample of first births. This is one more reason, in his opinion, not to rely on the birth cohort.

Nigel Paneth (Michigan State University) reminded the audience about Bracken's experience in enrolling 17,000 pregnancies in four cohorts, 30 percent of them as early as eight weeks, for under $2,000 a person. He related his own experience in seven such studies, six of which were funded by the National Institutes of Health. His research has concentrated on enrolling either births or pregnancies, with four birth cohorts and three pregnancy cohorts. He said it is much easier to recruit in pregnancy than birth, and it is administratively simpler to deal with prenatal care providers because they do not have IRBs. In his experience in Wayne County, he had one refusal of a prenatal care provider in some 70 different prenatal care settings. In contrast, working for two years in Wayne County, he could not get 25 percent of 28 hospitals to agree to even a protocol where the woman consented in advance to placenta collection. He said a random sample of hospitals would be unlikely to agree to alter their protocol in the delivery room to do something different with the placenta and cord blood. Some academic hospitals may participate, but he said he doubted many others would.

Duan noted the sibling cohort does not necessarily have to come from a hospital birth cohort but could very well come from a prenatal cohort. NCS could recruit a prenatal cohort and then go on to recruit the siblings. The advantage of the sibling cohort is that the mother has already agreed to participate in the study so there may be some economy of scale in recruiting her for the next child. In some sense, he said, it is not a question about hospital versus prenatal care, but, rather, once the first child is in the sample, what can be done to recruit additional children.

Graham Kalton (Westat) described the provider-based sampling (PBS) methodology now being implemented in three NCS Vanguard sites. An argument that has been made for the birth cohort is that a prenatal provider sample alone does not have complete coverage. In fact, the PBS as it is currently being implemented with a hospital component gives marginally better coverage than a birth cohort. The current approach lists as many prenatal providers as possible within an area, and it recruits women from a sample of those providers. The women are sampled at their first prenatal care visits.

Kalton said existing data indicate that approximately 70 percent of women report that they have their first prenatal visit during the first trimester. The question for the NCS is how quickly it can enroll and interview the women once they have been sampled. Very few women have no prenatal care, but the design is such that they are covered by treating their birthing hospitals as their "first prenatal care" visits. Thus, women

who have had no prenatal care are picked up at the hospital. The current approach also provides coverage for any deficiencies in the prenatal provider frame, because women who use only prenatal care providers that are not on the frame are sampled at the hospital. Thus, this approach provides virtually complete coverage. It also enrolls women as early as possible when using a provider-based frame.

Kalton noted that in comparing prenatal and birth cohort approaches, an important question to be addressed is which methodology is more acceptable in practice: Is it better to recruit through prenatal practices or is it better done through the hospitals? He noted Fragile Families obtained a very high response rate in the sampled hospitals, but the study did not collect biospecimens and did not cover situations where the woman or the baby was ill. Recruiting in the hospital may not work in these situations. Enrolling sampled prenatal care practices is also challenging because they are generally very busy. Enrolling the pregnant women presents additional challenges. Kalton endorsed the idea of a sibling cohort, but identified some missing operational details, particularly concerning how to efficiently collect preconception data. The plan for the NCS is to collect data on the child fairly frequently after the birth, every three months in the first year, then every six months, and finally less frequently. For the sibling cohort to be effective in collecting data on the women very early in their pregnancies, the women will need to be identified shortly after conception. The logistics associated with accomplishing this would need to be worked out. Carlson said she thought the original NCS had a pre-pregnancy data collection plan and suggested examining the previous strategy for collecting this type of data under the house-to-house recruitment plan.

Duan noted that PBS, a combination of the prenatal cohort and what has been called the birth cohort, sounds like a very good approach. For the operation of the sibling cohort in the detection of pregnancy, he said some of his colleagues make use of information technology, such as mobile devices, to encourage or invite the participants to send feedback to the study when an important event occurs. He said with careful planning and technology, it is possible to get close to desired event timing.

Reichman asked about the incentive for prenatal care providers to participate in the study. Nina Markovic (University of Pittsburgh) said her institution has an NCS site and, in her experience, providers like to have a plaque on the wall as recognition. She said her study featured providers and their children in their brochures, providing public recognition that they were supportive of the study. Providers felt that they were affiliated and contributing to good science. She also noted she has participated in studies in which recruitment occurred at hospitals and found buy-in at the hospital was top down. They did not get good cooperation in labor

and delivery until they placed a 24/7 research staff team in the hospital to collect the samples. With her current cohort, they pay the woman and/or her significant other $25 to call them when the woman is headed to the hospital, so her staff can collect the samples, much less expensive than 24/7 staffing.

Markovic commented on the first-born issue by noting that from a woman's perspective, many significant changes occur during the first pregnancy. For example, she may continue to work or may be smoking or drinking or have other exposures during the periconception time that do not occur with a second or third pregnancy because there is a toddler in the house. Duan speculated that the issue is not just the biology of the exposure health outcome but potentially also the sociology. The first-born's parents are getting on-the-job training, and later, children are exposed to more experienced parents who might be better able to cope with child-rearing issues. He added, based on his experience with various provider-based studies, one approach is to compensate providers for the time and resources they had to devote. Some studies pay for a staff member to help with the recruitment or offer providers the opportunity to be collaborators as part of research teams. In a sense, he said, this is an extension of recognition such as a plaque, and many collaborators were genuinely interested in the topic of the study. He noted that to acknowledge local collaborators, the author list for the study's papers included "HCSUS Research Team" (a long list of all the study's local collaborators).

Bracken agreed that providers do contribute to research, and how a study manages providers varies. This is why local knowledge is useful in working with providers. He said it is going to be difficult for NCS contractors to come in from the outside to manage this process, because of the role of personal relationships. Providers are more likely to be receptive to a colleague talking to them about research than they would be to an outside group. He considers this is an area, one of many, where losing the local academic centers will be a real detriment to NCS recruitment. Duan agreed, saying that in his experience with the HIV study, prominent HIV providers in each sample geographical area were recruited to serve as site captains. This was a collaborator model, and the site captain helped identify and recruit the other providers.

Sara McLanahan (Princeton University) asked how well the two sampling cohorts generate good representative samples based on actual cooperation and response rates. To her, the most important things are the overall response rates and representativeness of the sample. She shared her sense that many stories about success with providers and people's own hospitals are based on convenience samples and asked whether there is a difference between the provider-based cohort and the hospital-based cohort in terms of representativeness and overall participation

rates. Bracken said most of the provider examples that he has worked with are from convenience samples, but, since he had 100 percent acceptance, he finds it hard to believe that a random model would result in large numbers of defections.

McLanahan asked about the importance of eight-week first-trimester measures. If the provider sample can do as well on response rates, this provider cohort might be preferable because it would result in more data on prenatal care. But if it turns out that the most important data are in the first six weeks of pregnancy, then there is a question about whether it is worth the extra cost to get very early data. Information from the scientific community would help to make the decision about how to allocate the sample. Bracken said that the importance of getting data at eight weeks entirely depends on the hypothesis. Some exposures, such as cigarette smoking and the outcome of low birth weight, exert a lot of their effects in the third trimester, so third-trimester exposures are very important, but it is crucial to be able to measure early exposures as well. He stated restricting this massive study to look only at late trimester exposure is unnecessary. According to his data and estimates, he said, if the study is conducted efficiently, they can expect about 30,000 women to be recruited by eight weeks of pregnancy in the all-provider cohort.

McLanahan said Fragile Families found that most mothers received prenatal care but not always in the first trimester. She observed, however, that there is a big difference in access to early prenatal care by race and ethnic minorities. She urged NCS to consider the ability of the design to address disparities. She asked whether starting with a provider sample would produce consistency across race and ethnic groups, income groups, and other subpopulations, in terms of representativeness and response rates.

4

Imputation and Estimation

This chapter begins with the questions that panelists were asked to address on imputation and estimation provided in advance to workshop participants in Kwan et al. (2013, p. 8). Unlike previous chapters, no additional background information was provided. The second section of the chapter summarizes the discussion among panel members on specific issues related to the questions, followed by open discussion with the audience.

QUESTIONS ON IMPUTATION AND ESTIMATION

Given the study design proposal[1] above, and using the example cohort proportions proposed in the Session 2 questions, what enhancements can be made to address the following estimation and imputation challenges:

1. How can the data from the two cohorts be combined to increase the effective sample size?

   a. What should the parameters for the sampling procedure, for example, using the same PSUs, be in order to enhance data combination?
   b. What sampling protocol deviances could impact the ability to combine data?
   c. What considerations (if any) for sample weights need to be taken into account in the sample design? Specifically when certain groups

[1]The study design proposal was described at the beginning of Chapter 3 and was presented in Kwan et al. (2013).

may be oversampled in one cohort (such as women receiving no prenatal care who would only be present in the birth cohort), should any special considerations be made for the sampling probability in order to construct appropriate weights?

2. How can data imputation be used effectively, particularly for prenatal exposure?

a. What threshold level of imputation of prenatal exposure data is acceptable?
b. What should the proposed trigger for the more expensive comprehensive sampling look like—should this be a random sampling, event-based trigger, or a validation subset or some combination?
c. How should the sample be recalibrated in the future to account for attrition?

KEY POINTS OF THE DISCUSSION

Steven Cohen (Agency for Healthcare Research and Quality) moderated this session, and the panelists were Graham Kalton (Westat, Inc.), Colm O'Muircheartaigh (Harris School of Public Policy and NORC at the University of Chicago), and Richard Valliant (Joint Program in Survey Methodology at the University of Maryland and University of Michigan, Ann Arbor). The panel discussion below is organized by the topics covered during the session, which included the desirability of a unified design, allocation, weighting, missing data, and the special population sample.

Unified Design

In responding to the questions above on combining data from different cohorts and imputation, Kalton, O'Muircheartaigh, and Valliant agreed that a unified design with a clearly defined population of inference, as proposed by Kalton, has many advantages over an approach with separate prenatal and birth cohorts. In the unified design outlined by Kalton, the population of inference is defined as all births in a specified enrollment period. In that approach, a sample of women who would give birth in the enrollment period would be selected from a list of prenatal care providers that includes hospitals and birthing centers. Women who do not receive prenatal care, or receive it from a provider that is not on the list, would be sampled at the hospital or birthing center (at the birth). The integrated design has significant analytic advantages over a design with separate prenatal and birth cohorts, as well as leading to important simplifications in weighting, point estimation, and variance estimation. Some of the details and complications of implementing a unified design are described below.

O'Muircheartaigh noted the first task in a sampling problem is to think about what one is trying to represent in a study and what makes the study different from a convenience study. The special characteristic of the National Children's Study is that it could allow for a population-based inference that is substantively much deeper and more intricate than most studies. Kalton proposed defining the target NCS population as all births in the United States during a given enrollment period. The enrollment period may be two years (the shorter the better, he noted), but it should be multiples of years to cover seasonality. With a two-year enrollment period, the sample should be representative of all the births during that two-year period. This implies that to collect prenatal data, some pregnant women would have to be recruited and enrolled prior to the beginning of that period, and births associated with these women would be included in the sample only if the births occurred during the two-year period. The same issue arises at the end of the period—pregnant women are included in the sample only if the birth will occur before the end of the two-year period. He noted births may also be picked up at hospitals and birthing centers, but all would be births that occur during the two-year period. This design has the advantages that it is a single integrated design with a clearly defined temporal definition of the population of inference, and benchmark data from birth certificates can be used to support assessment and adjustment. A disadvantage of this approach is the operational complexity introduced by the specified time frame.

Kalton then described two potential variations. The first is the provider-based sampling approach he described earlier (see Chapter 3) that is now in the field at the Vanguard sites. In this design, he said, a sample is selected from a frame of prenatal care providers (including hospitals), with hospitals the provider of last resort. With this approach, women who receive prenatal care but deliver at home are included in the survey population, which provides additional coverage for the use of a hospital frame alone (although this is not of great significance since about 98 percent of births occur in hospitals). The downside to this design is that a list of prenatal care providers has to be compiled in the sampled areas, with the sample of providers then being drawn from that list. Another downside is that the biospecimen data collection at birth may be spread across a large number of hospitals. This is an important practical issue.

He said an alternative approach selects a sample of hospitals at the first stage of sample selection. For each selected hospital, the prenatal providers associated with that hospital are identified and a sample of these providers is selected. This approach concentrates the data collection at birth in just the sampled hospitals. Two issues to be addressed are that some prenatal care providers are not linked to just one hospital, and some may not have any hospital linkage (depending on how "linkage"

is defined and operationalized). Both of these approaches can be viewed in the unified approach framework once the target population of interest has been determined. Issues of practical implementation in the recruitment of prenatal providers and hospitals and in the enrollment of women are important factors to consider in making comparisons between them.

O'Muircheartaigh further explained the unified approach, suggesting that it be viewed as covering all parts of the population and that it is possible to cover a fairly large part of the population through a provider-based sample. However, some providers may not be included on the frame, and some mothers do not seek prenatal care. He acknowledged some non-response by providers, but the remainder would be captured through sampled providers. The rest of the population that has not been covered is due to failures of non-coverage and non-response. He said this would suggest supplementing the sample with coverage for the cases that are missed, and birthing centers or hospitals are the right places to go. If the hospital is considered as a part of the sample design rather than as a separate venture unrelated to the prenatal care providers, then there is a unified, stratified approach to the sample design, and no problem in accumulating data across the two.

O'Muircheartaigh stated his default option would be to have equal probabilities of selection for each birth in the defined inferential population. This design takes advantage of structured hierarchies, and the size of the strata (cohorts) would be determined by the empirical reality of data collection. In regions with many births to mothers who do not receive prenatal care, a large number of providers who refuse, or many providers are missing from the frame, the hospital/birthing center sample would be larger because there would be more eligible births at the point of delivery. He said it is not necessary to decide now how big the cohorts/strata are. Rather, they will define themselves because they are strata in the population rather than a predetermination about relative sizes.

He said he does not view a conflict between the idea of cohorts and the idea of a unified design. Instead, he said it is viewing the cohorts as non-overlapping strata rather than as possibly overlapping units with joint probabilities that is difficult to estimate. In the single design approach, each birth will be classified into one stratum. When patients arrive at a hospital from a prenatal care provider who refused to participate, they would be eligible for selection in the hospital. Births to a woman who did not receive prenatal care would also be eligible for selection in the hospital. The combination provides essentially a probability sample of all births.

O'Muircheartaigh said the big problem with the cohorts is not with the concept of different ways of collecting data. Instead, if done in an unorganized way, without advanced consideration of how to put the

pieces together, there will not be a good analytical product at the end of the process and specialized methods would be needed to combine cohorts for estimating descriptive statistics, such as variances and fitting statistical models. In contrast, with a single unified sample design, there is no problem with dual estimation, multiple frame estimation, or figuring out joint probabilities between two cohorts because it really is only one design with multiple components. In a unified design, variance estimation becomes straightforward. Valliant said the idea of the hospital being the last-resort selection is important in the sample selection in this unified design because sampling with probably proportional to size (number of births) would tend to select really big hospitals with high probability. But if the hospital is the last resort for picking up women who did not receive prenatal care, an adjusted measure of size will reflect the number of births per year to women who did not get prenatal care. He termed this a sticky technical detail, but said it also avoids the issue brought up in an earlier session (see Chapter 3) about potential refusals by large hospitals and the loss to the sample.

Kalton agreed that determining a measure of size for the PBS selection of hospitals is not straightforward. It is not the total number of births in that hospital, but rather the total number of births that would not have had a chance of selection from a listed prenatal care provider. Women may not have had a chance of selection from a prenatal care provider because the woman had no prenatal care or because the provider was not on the provider list frame. The measure of size is difficult to estimate but to control the sample size selected from the hospital, it is important to estimate it as well as possible. He said another alternative for the sample plan would be to develop a sampling frame of hospitals based on a list from the American Hospital Association of 6,000 hospitals and the number of births in nearly all of them. It would be straightforward to sample from that list (with some likely need for clustering for undersized hospitals that would not support the required sample size). The sample would still be a clustered sample design, with the hospital as the cluster as distinct from the geographical area.

Allocation

Valliant noted in terms of optimal design, even a unified design approach has allocation issues, including questions about the number of geographic primary sampling units (PSUs), the number of providers per PSU, and the sample size in each provider. The usual solution is to estimate variance components associated with each of these steps, which would require identifying one or more important statistics. They could be descriptive statistics: How many women had underweight babies? How

many women were exposed prenatally to something? What is the relative risk of a certain condition? He explained if the estimator is complicated it could be linearized and written in such a way that variance components for PSUs, providers, and women are calculated. He said it is likely that there are insufficient data directly related to the variables that the NCS is going to collect, and it will be necessary to piece together available information to make somewhat informed decisions about allocations. He pointed to the somewhat related datasets from the Vanguard sites, with about 4,000 or 5,000 births, and that the NHANES data are health-related with many physical measurements. He also noted the American Hospital Association publishes hospital data.

Valliant reminded the audience that the problem in deciding how many providers to sample per PSU boils down to thinking about how much alike the women are with a particular provider. He suggested another way to think about it is how much the providers differ in size. The way the math works out, the variance between providers depends on how many women they serve. Hospitals can serve hundreds or thousands of women in a year's time, while individual doctors' offices are much smaller. This built-in disparity in size will push toward sampling more providers rather than more women per provider, and there are cost implications to going to many more providers. Probability-proportional-to-size sampling of providers is efficient if the measure of size is the number of women served. However, selecting an efficient sample is complicated by the fact that the counts of women served by each provider may be inaccurate. He said with enough data, it would be possible to follow Duan's advice (see Chapter 3), resulting in a mathematical programming problem. The general idea is to determine an allocation of the sample to optimize an objective function subject to a set of constraints. With a set of statistics of interest, variances can be weighted according to their importance to the survey to form an objective function to be minimized. A fixed budget may be a constraint, and other additional constraints may also be needed, such as a minimum number of providers and women per PSU. To do this allocation properly requires a lot of data.

Weighting

Kalton noted that with the design he proposed, the study has benchmark data from birth certificates that can be used to adjust the sampling weights to account for some births that are missing due to nonparticipation or non-response. As in other standard panel survey designs, weighting adjustments are typically used to account for attrition (children who drop out of the study and cannot be followed) as the study moves forward. An initial weighting adjustment based on vital records is

intended to make sure that at the outset, the sample is representative of the defined population of inference.

Kalton noted the sample design could include oversampling of certain groups. Oversampling is usually done for groups that may be different in some way, such as low socioeconomic status, race, or ethnicity group. Valliant observed the fact that groups are oversampled is prima facie evidence that the survey manager thinks that they are different, which leads to weights that are different for the oversampled groups. He said the question about which vital records to use to create calibrated weights may be a modeling problem. There are at least two reasons to use calibrated estimators. First, if the study undercovers or mis-covers different parts of the population—for example, if it is known that too few lower socioeconomic status women are being recruited, which he said is typical in U.S. household surveys—calibrating can rebalance the sample to make sure it better reflects the population. He noted calibration only works if the women in the sample are a good representation of the non-sample women. However, if there is a skewed representation of lower socioeconomic women, then calibrating will not help.

The second reason he gave to calibrate is to reduce variances. This process requires covariates that are related to coverage and to the key variables being estimated. Birth certificates contain a lot of information: birth weight, APGAR score, whether the infant required assisted ventilation, and whether he or she was admitted to the intensive care unit. Many other potential variables, in addition to the mother's characteristics, could be tabulated and used for calibration or control totals. He called all of these variables fair game for a research project to determine the most appropriate control totals.

Missing Data

Kalton noted the problem of missing prenatal data for some births is similar to the attrition problem, except looking at time in the reverse direction. He suggested the data can be considered to be geared around the birth with incomplete responses in both time directions, and missing data might occur before birth (the prenatal data) or after (attrition). While complex to analyze, he said, conceptually, it provides a framework in which to think about approaches to the problem of missing data. When the study is viewed as two separate cohorts, one can perform analysis on each cohort separately, but it is very unclear how the cohorts can be combined. The unified approach provides a framework that supports joint analysis.

O'Muircheartaigh noted the issue of replenishment is difficult. He agreed with Kalton that the strength of the NCS as a longitudinal study

is that people are measured very early on and over time. A replenishment using Kalton's approach could be conceptualized, making inferences backward from a sample boosted by replenishment. It could be thought of as a parallel cohort, for example, as a sample of adolescents added to the sample. They would be followed forward, and some of their characteristics would be tracked in relation to the earlier panel. This might be a possibility in 10 or 15 years if a scientific question about this age group becomes apparent and the original NCS panel has become too small, although this approach would not provide any birth or early childhood data for the new sample.

Valliant noted that even in the unified design, some women will be recruited only at the hospital. They will not have prenatal covariates except to the extent that they can be collected by consulting medical records or auxiliary measurements (e.g., dust, EPA databases). One option would be to use the sampled women for whom prenatal covariates, exposures of different kinds, are available and use them as donors to impute for women who are missing the prenatal covariates. Multiple imputation is one approach. A valid imputation model would be as correct as possible and could be evaluated, possibly using simulation.[2]

A statistic that could be used to assess the impact of imputation is the fraction of missing data. If there are too much missing data, he said imputation may do more harm than good. This fraction of missing information has a "between" and a "within" component for a multiple imputation variant and is a measure of the variability being injected by imputation. One question is whether imputation represents too large a proportion of the total variance. This, too, could be measured in a simulation study. He said if it were possible to put together a pseudo-population based on the Vanguard data or NHANES and then divide that population into women with and without the prenatal covariates, for example, a simulation might inform the study about the impact of missing data.

Valliant noted the University of Michigan conducts longitudinal surveys, including the Health and Retirement Study and the Panel Survey of Income Dynamics (PSID), with a number of cohorts, recruiting a new panel every five years or so. In practice, the big losses occur immediately at the first interview with people who do not want to cooperate, while the people who cooperate on the first interview are likely to continue. The Health and Retirement Study collects information about older people, and he said he thinks respondents like to have somebody to talk to periodically. Attrition is very low after the first few interviews. He posited that

[2]During the open discussion later in this chapter, Roderick Little (University of Michigan, Ann Arbor) noted that this type of imputation has its limitations when the data that are missing (e.g., the prenatal data) are to be used in causal analysis.

once people are convinced the NCS is important and sign up for it, they will continue to participate.

O'Muircheartaigh added that continuation may be particularly difficult for the NCS. He agreed almost all longitudinal surveys have most of their non-response in the first wave, and conditional response rates to later waves are quite high, often from 95 to 99 percent. This argues for minimal intrusion at the earliest stage to maximize the initial response rate. Unfortunately, in the NCS Vanguard sites, the opposite has been the practice, with collection of as much data as possible at the first visit. He said not all the data are necessary at the earliest point, and collecting only the necessary data at early points would maximize the initial response rate. He noted after 20 years, the overall attrition from the PSID was about 50 percent, with about half in the first wave. Planning carefully how to maximize the initial response to the NCS will be important.

Special-Population Sample of 10,000

Cohen asked the panel to comment on the set-aside special-population sample of about 10,000 and the impact of attrition over time. He noted Kalton talked about weighting back to the original sample, adding that after 15 years with all the different levels of non-response, overall representation might be fairly low. Cohen asked if the 10,000 might be used as a replenishment sample.

Kalton suggested that the special-population sample could be reserved for studies of rare populations, such as births that came about from assisted technologies. He said if the study were designed in this way, that group would not contribute to the national estimates in any way, because it would involve oversampling a miniscule population at a very high rate. But if there are special interests, this methodology could be used for a benchmark comparison.

O'Muircheartaigh said the special sample would not contribute to the overall NCS, but it would be possible to do some linkage to the NCS because it would be contemporaneous and share some characteristics. But if it is not linked to the design, then it does not give any strength in terms of inferences to be made about the national sample. He questioned the desirability of setting money aside to tackle a specific problem that cannot be tackled within the framework of the NCS.

Open Discussion About Imputation and Estimation

Irwin Garfinkel (Columbia University) asked how the sample can be weighted without data on the proportion of births served by each of the prenatal clinics. Kalton said in the current PBS, the prenatal care providers

are sampled with probability proportional to estimated size, where the measure of size is the estimated number of first prenatal visits in the past year. If the sampling scheme within a sampled location were to select one week in four and take all eligible women during the selected weeks, the selection probability for a sampled woman would be four times the location selection probability. He noted that the efficiency of the sample design and the spread of the workload across the sampled locations depend on the quality of the estimated measures of size.

To O'Muircheartaigh, the options are to fix the probabilities or fix the sample size, both of which can be fixed only with a lot of information. Fixing the probabilities is not difficult, but fixing them while simultaneously controlling the sample size requires information. If there is good prior information and the provider data are accurate, then the sample size will be more or less as planned. If they are completely wrong, the sample size will fluctuate.

Garfinkel noted both O'Muircheartaigh and Kalton are proposing to use the hospital only as the last resort and asked whether it might be simpler with better data and fewer assumptions to use the information on the number of hospitals and number of births. Kalton said that, in the context of a sample of pregnant women, the hospital sample is designed to provide coverage for women who had no prenatal care or who had prenatal care only from a prenatal care provider that was not included in the provider sampling frame. The number of these women has to be guessed, although sometimes the number can be based on birth certificate data for the hospital from the past year. Although a problem, it could be dealt with, referring to O'Muircheartaigh's observation that if a probability is misestimated or underestimated, then there will be a sampling fraction that will result in the sample including more or fewer births than planned.

O'Muircheartaigh added it is unfortunate to use the term "last resort" when referring to the hospital, as it is the appropriate place to select certain women. It is saying that if the birth has not been covered by the sample of providers, the hospital will be the stratum that generates the birth. This group has several categories, and estimates are needed for the numbers of women in each category. Some of these numbers are available only through field activities in the location. One category is births to women who have no provider, and a second is births to women who use prenatal care providers that choose not to cooperate with the study. The operation will have uncertainty, no matter how sampling is done, and the exact number of births in the sample will not be known in advance. It is something that is empirically determined by the population and not by some presupposition.

Valliant noted not having complete control over the sample size in the survey is fairly standard. Another unknown besides provider cooperation

is cooperation of women, who are at a stressful time in life and may not want to participate. Even well-founded advanced estimates will result in some inaccuracies in the number of people who agree to participate. It might be possible to do what household surveys typically do by creating replicates of sample units. If the sample is smaller than expected after recruiting for six months, a replicate of the provider sample could be released.

Roderick Little (University of Michigan, Ann Arbor) made clarifying comments about imputation. He said in a regression, there is a Y, some Z variables that are observed, and an X variable that is the early pregnancy variable that is missing for some cases. If the values of the X's are imputed purely based on Y and Z, the imputed values provide no information about the association between X and Y given Z, the topic of interest. The only way to get information on the association between X and Y given Z is by having auxiliary data to help with the imputation. Those auxiliary data could be from a questionnaire, from a dust measurement, or in an auxiliary database. It is important to realize that multiple imputation only helps if there is additional information to be recovered in the data that are used for imputation. If the relationship between Y and X is the topic of interest, some other variables are needed. Kalton noted that birth certificates might also provide valuable information to use in imputation.

Nigel Paneth (Michigan State University) thanked the panel for clarifying that one cohort perhaps with different strata is the sensible approach and raised a question about statistical prioritization. He asked whether it depends on the questions being asked. He stated the NCS as currently described is a study about every childhood outcome and every potential exposure. With that as a framework, Paneth said it is impossible to decide whether prenatal data are more important than delivery data, or other important tradeoff questions. The struggle over sampling strategy and design reflects the absence of any closure concerning prioritization of public health relevant outcomes, key exposures to be investigated, and their relationships. Some of these considerations were subsumed by the hypotheses the NCS once had, he noted, but now the NCS does not have hypotheses. He said answers to important sampling questions cannot be answered until relevant health outcomes and exposures are determined.

Kalton noted the integrated design comprises 100,000 births to follow from birth forward. Prior to birth, there is the critical issue of how many women will be sampled from prenatal care providers and how early they are enrolled. As has been noted, there will likely be some subgroups of women, such as the socially disadvantaged, who are underrepresented in the prenatal sample and will be covered mainly through the birth stratum. He said issues about the effectiveness of this strategy remain

to be examined, but the unified approach seems to be the best provider-based approach.

O'Muircheartaigh also argued that if the problem is defined as obtaining a representative sample of 100,000 births, then the unified design with equal probabilities of selection would be the best design. The result would be a representative sample of births with as much information, including prenatal data, as possible. If pre-pregnancy data for later siblings are also collected for as many cases as possible, the result will maximize the amount of information contained in a representative sample of births, which he termed a noble ambition and a fine achievement whatever the outcome. Before the sample is selected, it is possible to debate about whether urban areas, inner urban areas, poor rural areas, or areas with high environmental risk should be oversampled, which is possible within the structured design. He noted that by taking a population representation approach and maximizing the information available on as representative a sample as possible, a platform is created on which many studies of different kinds can be based, including currently unspecified studies.

Greg Duncan (University of California, Irvine) asked O'Muircheartaigh and Kalton for clarification about subsequent births. He referred to the comments in the first panel (see Chapter 2) about the importance of exposures very early in pregnancy and the potential importance of exposures preconception. Unless subsequent births are included in the design, Duncan said there would not be representative samples of births with preconception and very early pregnancy exposure information.

Kalton replied he had been describing the basic design. Within a two-year enrollment period, there may be good grounds for including any subsequent births in a selected family with certainty. Then to retain an equal probability sample, it will be important to ensure that prenatal care providers and hospitals do not independently allow the sample to include subsequent pregnancies or births to mothers with a previous birth within the NCS enrollment period.

He noted that there may be some potential advantages to having sibling data. The design has some operational efficiency, and the data can be used for sibling comparisons. He views the inclusion of siblings over a more extended time period to be an adjunct study that warrants careful assessment of its operational feasibility and associated costs, as well as a full examination of how it can be applied to yield data for the very early period of pregnancy. O'Muircheartaigh followed up saying that an earlier, widely advocated design was household-based probability sampling of women to be interviewed if they were of childbearing age. Clearly, a preconception sample is possible, and the previous approach would be a good solution if there were no costs or practical considerations. However, some evidence shows that the approach is impractical. Any

design that does not involve recruitment of women in those age ranges regardless of pregnancy status is not going to collect a representative preconception sample. However, recruiting subsequent births to a mother recruited into the NCS has strong advantages in terms of providing some preconception and prenatal data.

Duncan said he agreed that a two-year recruitment period would capture some second births, but it would be an unusual sample with a fairly short time interval between births. A longer interval, perhaps five years rather than two, would include more births and be more representative. He suggested amending Kalton's proposal to include a longer interval with oversampling of births early in the period, which would include some subsequent births and some first births later on. Although difficult for samplers, it may be another way of potentially providing a single integrated sample over a five-year period that would include both initial and subsequent births.

Kalton responded that Duncan's suggested design might be very expensive. He suggested another way to describe it might be to extend the two-year enrollment period to five years and then follow on the model that he had put forward. He said he would argue in the other direction, a one-year enrollment period, for a variety of reasons related to efficiency of data collection. It would avoid such problems as field workers going to one household to conduct a fifth interview while going to another to conduct a second interview. These types of mixtures, he said, make data collection much more difficult. Additionally, providers change over time, coming in and out of business. He said Duncan's suggestion is possible, but it is not clear whether or not it is feasible to tie that design into the basic child-related data collections. If the design were to provide preconception information, all the women would have to be followed and go through questionnaires, although only some of them (perhaps 20 percent) would become pregnant.

Kalton noted the unified design will collect child data on a schedule of every three months initially and then every six months, and he asked how that would work with the desire to know, almost immediately, when a woman becomes pregnant. He said he had heard the suggestion of pregnancy tests by mail. To operationalize the sibling data collection, a method for capturing data from a woman at point of pregnancy should fit in with the ongoing collection of child information. O'Muircheartaigh agreed that it is a complex and difficult question. To address a simpler question, he suggested the number of initial recruitments might be reduced if the sample were supplemented with siblings over a two-year or five-year period, which, he said, would not affect the principle of the design.

Duan followed up on Duncan's point about the duration of the recruitment window, noting a longer duration would enhance the repre-

sentativeness of the sibling cohort and might have its own merits. If the sample were entirely within one year, it would reflect the idiosyncrasies of that year, while a longer duration gives a better representation over time. The population of interest is not just children born in 2014, so a longer duration could help capture variations in economic and environmental events and in the weather. He said he appreciated Kalton's point that it might be more costly for the same sample size, but potentially it might yield more useful scientific results.

Duan agreed with Kalton's proposal to look at the likely missing early pregnancy data in this unified approach. Many of the women who can be recruited through providers will not be in the sampling frame until after the first trimester. He said Kalton's point about using statistical methods like weighting and imputation and applying them backward is an interesting idea, but there is a difference between time forward and time backward. Looking at time forward, a study with a strong field operation usually can maintain the sample over time, and the conditional response rate after recruitment is usually very good. Going backward for imputation, the missing data are not under the study's control because they are missing from a time before participants were recruited. With the missing data rate going backward much higher than going forward, this sort of missing data methodology can be very sensitive to underlying assumptions. For that reason, Duan said it is important to supplement the data with either a sibling cohort or with alternative ways to get to early pregnancy data.

Kalton noted one of the key issues is the proportion of women who can be recruited during pregnancy and in the first trimester. If the study could pick up 70 percent of participants in the first trimester, it would be similar to the PSID experience of 75 percent response to the first wave.

Garfinkel asked O'Muircheartaigh about the possibility of high early attrition. If very expensive prenatal data are collected and attrition is high after that time, then expensive data have been wasted. Because of this, he said he disagreed with the statement that it does not matter when the money is spent. If siblings are important, then it matters greatly when the money is spent. As an example, if it costs $18,000 to collect data on a prenatal birth, $2,000 to enroll them, and $2,000 to enroll the mothers at the hospital, 10 times more women could be enrolled in a birth cohort than would be possible in a prenatal cohort. Thus, he said, if many women in the prenatal cohort are lost to attrition, it matters greatly when the money is spent.

O'Muircheartaigh countered that if it cost $18,000 to recruit one way and $2,000 the other, it does not mean the sample can have nine times as many one way as the other, because these people will be maintained in the sample throughout the 21 years. He explained that is why consideration of both long-term costs and short-term costs are important in determin-

ing the optimum allocation. If it costs $200,000 to cover each child for 21 years, then the comparison is between $218,000 and $202,000 in terms of the cost of a case in the NCS. He said it is only if the decision has to be made based on the money spent this year that one would make that decision, but, in his opinion, that is entirely the wrong decision. He reminded the audience that it is critical to remember that short-term recruitment costs are only a small fraction of the costs of a case in the NCS. The appropriate basis of comparison is the total cost of the case under each of the scenarios. He said there is no reason to believe the later costs are any different depending on the method of recruitment, and therefore the imbalance is not 9 to 1 but perhaps 1.05 to 1.

Kalton agreed with O'Muircheartaigh and Duan, who made the same point in the previous session, noting the cost of investing in a good sample is paid over the life of the study. He noted he did not fully understand Garfinkel's point about costing of the cohorts.

Jennifer Madans (National Center for Health Statistics) asked for clarification about how women who have had no prenatal care or had a provider not on the frame would be identified at the hospital. Kalton responded that, as currently conceived, the data collectors at the hospital have a list of all the prenatal care providers on the sampling frame, and they are instructed to exclude from hospital recruitment all the women who attended any of these providers. The exclusion can be determined either prior to data collection (based on hospital records) or as part of the screening interview.

Kalton said a woman has different potential routes of getting into the sample, but each woman is eligible for the sample through only one of these routes. In one route, they come in for their first prenatal visit to a sampled provider who has agreed to participate. Only those women who have had no prior visits for that pregnancy at that provider location are potentially eligible. When a woman is interviewed, she is asked if she has had any earlier prenatal care visits for that pregnancy with this or another provider. If she has visited this provider before, she is ineligible for the sample. If she has visited another provider, a check is made to see whether that provider was on the provider sampling frame. If so, the woman is again ineligible. Since eligibility is based on whether it is the first visit, there is only one route for sample selection. The same approach applies for the hospital cases: if the woman has had a prenatal care visit at a provider that was on the sampling frame, she is ineligible. The eligibility screener also includes questions on age, and whether the woman lives in the sampled county or not. Women can often be prescreened by the hospital as not being eligible based on hospital records.

5

Moving Forward

This chapter begins with the question that panelists were asked to address in the final wrap-up session on moving forward. These were provided in advance to workshop participants via Kwan et al. (2013, p. 9). The chapter continues with key points made by the panelists, followed by a summary by the panel moderator and highlights of the open discussion.

QUESTION ABOUT MOVING FORWARD

> Panelists were asked to synthesize the tradeoffs among factors, issues, and values that need to be balanced and considered by NCS leadership.

KEY POINTS OF THE DISCUSSION

The moderator for the final session of the workshop was Greg Duncan (School of Education, University of California, Irvine). The panelists were Ana Diez Roux (Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor), Roderick Little (Department of Biostatistics, University of Michigan, Ann Arbor), and Edward Sondik (National Center for Health Statistics, Centers for Disease Control and Prevention).

Remarks by Edward Sondik

Sondik said he would try to identify recurring themes, issues where there does not seem to be agreement, tradeoffs, and finally offer a comment and suggestion. He said he tried to look at his remarks from the standpoint of the study and what kinds of information might move it forward.

He listed five recurring themes. First, the population of interest is children born during a fixed time period, although there was not agreement on the exact time period. Second, no one raised the issue that the proposed design includes only a very small preconception sample, although collection of preconception data in the NCS has been part of the dialogue from the beginning. He said prioritization of the collection of preconception data might rely on examination of the hypotheses that drive collection, which he called a significant decision. Third, examination of total costs over the lifetime of the study in addition to up-front costs might lead to different conclusions, referring to the point that although recruitment costs are high, they can be viewed in the context of study results and the total cost of the study. Fourth, mobility of sampled families is an important issue and may result in potential loss to the sample if families cannot be followed; further, it will be important to be able to link environmental measurements to (possibly changing) family residence. Fifth, he noted discussion about the tradeoffs in the number of primary sampling units (PSUs), which he said is a complex and important issue that has not been discussed completely. Allied with that is the question of how to handle the geographic environmental variables, and whether or not they are clustered. If environmental exposures are uniformly distributed geographically, a good probability sample is sufficient. If exposures are not uniform, then it is not clear how to include geography in the sample design. He said he called this an area of agreement because he said he thought that people would agree that, in general, more is better.

Under important issues with little or no agreement, he noted very little discussion during the workshop about study objectives. In his opinion, a clear articulation of the ability of the design to produce information important to public health would be invaluable to justify the study. He observed that the first panel (see Chapter 2) specifically addressed the proposed measurements in light of their ability to learn about asthma, endocrine disruptors, and neurological problems, but this was the only discussion about public health impacts and how well the design supports understanding them. He said this raises the question of whether or not the design will support learning about the impact of specific environmental exposures on health.

As tradeoffs, Sondik noted one of the open questions is the extent to which the design will include prenatal measures. Referring to Graham Kalton's earlier estimate that 70 percent of women see a doctor during the first trimester, he pointed out that another participant had said a measurement at six weeks is more important. Sondik asked about the importance of collecting information at six weeks and how to do so.

The second tradeoff has to do with the sample composition, noting the agenda started off posing that the design will include a 50-50 split between the birth and the prenatal cohorts, and invited the panel to consider other compositions. Sondik proposed the optimal composition of the sample is related to the science, noting that he was concerned about the challenges associated with recruiting women in labor or just post-labor. The question of coverage of first-born versus siblings is important and brings up many other tradeoffs to consider. He praised the elegance of Kalton's sample design.

Sondik's final suggestion also concerned the relationship of the science to the design. He expressed concern about the power of the study to determine relationships. He stated that he understands the reluctance to identify a single specific set of hypotheses, but one way to evaluate a sample design might be in terms of an exemplar set of hypotheses and consideration of the power that exists within the design to evaluate them. As an example, he asked about the power of the study to determine relationships among the variables of first-born, the income level or poverty level of the family, race, and asthma. Given the short time before the Main Study is to begin, he suggested commissioning a panel to look at and prioritize the set of reasonable science questions the study might address. Given those priorities, perhaps a second panel (or NCS staff) could look at the set of priorities in terms of the design and assess the power of the design to identify relationships. He said this would provide NCS with a very powerful argument to give to decision makers about the study's focus.

He noted in closing that, in the past, the Framingham Heart Study was raised as an example of data collection that did not start with a set of hypotheses. Clearly, a broad database could provide possibilities for exploring relationships, but it is important to be able to give examples of the kinds of analytic power the study is expected to have. Sondik said keeping track of those estimates early on, as the study progresses and as the sample develops, would be a very important management tool.

Remarks by Roderick Little

Little suggested the NCS needs to articulate specifically what it is adding over existing studies, pointing to other new studies, many in

other countries. He observed that when he was on the NCS Advisory Committee, a huge effort was made to develop hundreds of hypotheses. While he said he was somewhat critical of the scope of that effort, it was laudable in some ways. He said now with no hypotheses, the NCS is proposed to be a data platform that somehow will address many different things and suggested searching for a happy medium between these two extremes. He said he does not see an obvious way to make decisions about optimal design without some specific objectives articulated through hypotheses. He suggested seeking a relatively small number of "sentinel" hypotheses that could be viewed as being the current burning issues and providing the power calculations for those hypotheses.

Little noted the workshop focused on the role of prenatal exposures, particularly environmental exposures. Though clearly important, they are not the only component of the NCS and pointed to work to be done with exposures after birth.

He praised the supplemental sample of 10,000 in the draft sample design and agreed with O'Muircheartaigh and Kalton that, for most of the sample, an equal probability sample design makes sense, particularly without clearly articulated hypotheses. On the other hand, he said getting good variability in exposures is important and suggested developing an index of environmental risk and oversampling areas that have high risk. This might increase the power of the sample for looking at some of these associations.

He observed the progress made in sample design, with the debate now about which particular kind of probability sample to use. He said he supports probability sampling, or something as close as possible to a probability sample with the possibility of some missing data. He said it may not be possible to collect early trimester information on everyone. Even if there is only partial information, it is still a probability sample. He called the NCS a very useful study because there are many things that will be analyzed that do not necessarily use first-trimester information.

He noted three overarching issues in terms of specific choices of a design: (1) choice of the frame and whether to use a provider frame or a hospital frame, (2) point of contact, and (3) timing of the initial visit. He said very detailed specifications for these alternatives are important. He suggested that costs will illuminate the argument about the birth cohort versus siblings versus prenatal provider cohort, but there is divergence of opinion and confusion about the relative costs. Given the utility of the prenatal information for some hypotheses, he said he favors trying to get direct information for as many people as possible during early pregnancy to help ensure the representativeness and utility of the sample.

He noted even though there has been conflicting information about cost and practicality, he is somewhat more inclined toward the provider

approach if it can be operationalized satisfactorily. He said he defers to those who have been in the field doing work on these issues, but the provider sampling approach looks promising, as does the unified way of thinking about the design described by Kalton and O'Muircheartaigh.

Little reminded the audience about his comments earlier in the workshop about the imputation of missing early pregnancy data. Good auxiliary data are needed for imputing early pregnancy exposure (or other) data. Auxiliary data may come from proxy interviews or some other source (such as dust measurements or linkages to EPA exposure databases, as discussed in Chapter 2). The validity of an imputation method relies on its ability to estimate the relationship between exposure variables and outcomes.

Little's final comment concerned the original 105 PSUs versus a smaller number of PSUs. He suggested that there would be added variance in going to a more highly clustered design and doubts that it would be worth the cost savings if amortized over the life of the study.

Remarks by Ana Diez Roux

Diez Roux began her comments by acknowledging that the NCS is trying to address a very complex and broad-ranging set of issues that require engagement of many different disciplines. She addressed several big-picture items that she said could be done relatively quickly. She commented on the criteria for making design decisions, the process, and other points that came up during the workshop.

Diez Roux said the study will continue to grapple with prioritizing study objectives in order to make design decisions, and there are tradeoffs in that the Main Study will not be able to properly address certain items. She said this is acceptable, but it is important to acknowledge what the study will and will not do. When people are collecting data, having a sense that specific objectives are being targeted helps to keep the group focused on a common goal.

She noted two big sets of objectives. One potentially important objective, not discussed during the workshop, is estimating incidents and prevalence of different conditions among U.S. children. She was part of a National Research Council/Institute of Medicine panel (National Research Council and Institute of Medicine, 2013) that looked at the U.S. health disadvantage compared to other high-income countries. She reported that health under age 50 and specifically among children and adolescents features prominently as an area in which the United States does substantially worse than other high-income countries. One of the panel's findings was insufficiency of data on the prevalence and incidence of many health conditions among U.S. children that can be compared

to other high-income countries. Filling this information void may be an objective for the study to consider in order to demonstrate its value.

She said the other big study objective has to do with etiological investigation, which can be driven in two ways. First, it can be driven by very specific research questions, although the disadvantage, as the NCS experienced in its early years, is that too many hypotheses make the process unmanageable. Conversely, the other extreme approach is to be completely agnostic and say data will be collected to support unknown analyses in the future. She expressed her belief that a purely agnostic approach is virtually impossible because decisions have to be made, and in making those decisions one needs to know what questions are to be answered and what their priorities are.

She noted, however, that some aspects of the design may be applicable to many different questions, so the hypothesis-driven and agnostic approaches are not totally incompatible. However, a number of other decisions may require thinking about priority objectives. She agreed with Little that a middle ground between these two extremes exists and suggested thinking about a typology of the priority questions that the study might answer. For example, she asked whether there is one set of important questions about prenatal exposures. If so, that would indicate the importance of collecting prenatal information. She asked whether there are sets of environmental factors that are important, and if they are common factors, important factors, factors that potentially have very adverse impacts, or factors that may vary geographically. The latter, she noted, would indicate a geographically distributed sample is more appropriate. Deciding what kinds of environmental factors are of primary interest will help the NCS make decisions.

She argued less for a specific list of questions than a typology. She noted the typology might include rare outcomes, outcomes that have a public health impact, or common outcomes that are causing the United States to have much worse health than other high-income countries for unclear reasons. She suggested this kind of thinking might also help in some of the design decisions. She said a third kind of question has to do with investigation of disparities and asked whether disparities are key to this study. This kind of typology has implications for the core design, sample size, and other issues, as well as for the core measures and key exposures of interest.

She said she thinks that developing this typology would not take much time. A group would have to develop consensus and not everyone would agree. She observed the differences in opinion expressed during the workshop's discussion of design and measurement reflect underlying differences of views about what the study should address. Making these things explicit, she said, will assist in decision making.

Diez Roux noted a first step is to establish a set of core typology questions that prioritize the study without being overly specific or detailed. Collecting as much additional data as possible is a good idea because so much is unknown. Many new questions will emerge in the future, including exposures and outcomes of predisease markers and support for epigenetics. She said some criteria have to do with the expected utility based on what is known. While incomplete because there are many things that are unknown, it is one starting point. She noted some data are easy to collect, such as global positioning system locations of residences to allow linkage to a wealth of information down the line.

She noted storage is a second criterion. If samples can be stored and may be useful, collecting and storing them (within logistical constraints) is a good idea.

Another aspect of the study that needs balance is complexity versus simplicity in design data collection decisions. She said she weighs toward simplicity, which has advantages in terms of running a study on site and of analyzing the data later. She reminded the audience more complicated designs make it more difficult to use these data later. She noted simpler is better, but not so simple that it defeats the purpose. To her, the unified design approach proposed by the previous panel (see Chapter 4) is very appealing because it is a simpler approach than having multiple cohorts that have to be weighted differently and combined.

Drawing on her experience in multisite studies, she said it is important for the study to find the right balance between centralized and decentralized activities. The NCS can learn from other multi-site studies and the Vanguard sites, including the investigators involved in these studies. She said when she served on the NCS Advisory Committee she had the sense that a wealth of information was not being used as much as it could be.

She concluded with two comments on ideas that came up during the workshop. Given differences in the biology of first pregnancies versus subsequent pregnancies, and also birth order effects on a number of social and health outcomes, she stated that it is very plausible that prenatal factors interact with birth order.

She stated that the tradeoffs concerning the sibling sample are important to consider. The advantages might be the potential ability to get preconception information, with perhaps some cost implications, and within family sibling comparisons that could be informative. There are also logistical issues to work out. However, depending on what the priority questions are, adding siblings to the sample will reduce sample variation in exposures that are invariant within families and reduce statistical power if outcomes are clustered within families.

Closing Comments by Greg Duncan

Duncan stated that he detected agreement among Sondik, Little, and Diez Roux on some issues. They all strongly endorsed probability samples, and the two presenters who discussed PSUs (Sondik and Little) endorsed more versus fewer of them. All three panelists expressed the opinion that there ought to be some version of hypotheses or, as Diez Roux proposed, typologies.

Duncan said he would push to first consider the kind of objectives that have important bearing on the design. He said if it is true that preconception exposures and exposures very early in pregnancy are not very important, then a sibling sample may not be worth it. The question that remains is whether preconception exposures and very early exposures in pregnancy are important to address with this study.

His second point concerned geographically varying environmental exposures. If environmental measures are either not geographically varying or are not important to the study, then geography is not important to the sample design. However, if measures vary by geography and are important, considering geography in the design may be important.

Open Discussion about Moving Forward

Michael Bracken (Yale University) noted he found it encouraging that the moderator and three panelists agreed that hypotheses are important, at least selected sentinel hypotheses. He also agreed that hypotheses that reflect what this study could do and that demonstrate its value are important, noting a study's targeted health effects help communities decide to support a project, rather than just serve as a data platform. He agreed with Diez Roux in urging some speed in developing sentinel hypotheses and the value of reviewing previous work to determine whether hypotheses of current interest are included. He noted experts in various disciplines spent countless hours developing these hypotheses, and the documentation is available.

Nigel Paneth (Michigan State University) thanked the panel for emphasizing the need for prioritization, for systematically developing a schema that would allow NCS to prioritize properly and to find the balance between enormous numbers of hypotheses and the current state of no hypotheses. He went on to support Diez Roux's suggestion for developing a typology of hypotheses that would help the study to make important decisions.

He noted the NCS has already specified a sample size of 100,000, which he termed the wrong approach. Because the sample size has been specified to be 100,000, many hypotheses cannot be considered while

others are easily accommodated. He urged NCS leadership to pay attention to the views expressed during the workshop.

Naihua Duan (Columbia University) asked about the plan not to proceed with the household screening sample to collect preconception measures. He noted his purpose in raising the issue is for brainstorming. One message he heard from the first panel was the importance of early pregnancy or maybe even preconception data. Some of those data could be captured in the unified sampling approach but might not be captured. He said he wondered what could be obtained from the prenatal sample relative to what potentially could be accomplished with a household sample.

Even though the household sample has been found to be too expensive, Duan noted that many at the workshop commented on the importance of early pregnancy and preconception measures. This suggests that recruitment costs are not the only costs to consider; total lifetime costs for the study are important as well. He noted several other panelists commented that the difference in recruitment costs might not be that large when combined with follow-up costs. Part of the high cost of recruitment of preconception women is following women who never become pregnant. To reduce these costs, one solution might be to limit the preconception sample to women who are actively trying to get pregnant. He said there may be a benefit to retaining some household sample (perhaps in hot spots, or other geographic areas of interest) in order to answer important scientific questions.

Duan agreed with Irwin Garfinkel and the other panelists that the sibling sample has much merit but also some limitations. Questions about the first-born and what might not be covered in the sibling sample are not trivial. One advantage of the household sample, if it could be retained in some affordable way, would be to fill that gap.

Sondik agreed, saying a comparison of the characteristics of household sampling versus provider-based sampling in terms of data that could be collected might provide a useful basis for evaluating each method and understanding in what circumstances it would be useful.

Kalton commented that he perceived that the participants understood the value or potential value of the NCS. In terms of costs, he said there is also an issue of political will. He said political will helps with this study because both political parties understand the study well enough to support it. He urged making sure the NCS is a good strong study that is properly implemented.

Steven Hirschfeld (National Children's Study director) thanked participants for a very informative and stimulating discussion, saying that he and others will continue the evolution of the design of the National

Children's Study. He said its potential is well appreciated, and the goal of the NCS is to have that potential not only realized to meet current expectations but also to have a platform that would exceed expectations in the future so it becomes an ongoing resource for informing about the health and development and growth of children.

References

Annett, R., Baker, D., Bancroft J., Belanger, K., Bracken M., Buka, S., Cauley, J., Chambers, C., Cherry, D., Culhane, J., Dabelea, D., Dearborn, D., Dole, N., Drews-Botsch, C., Dudley, D., Durkin, M., Flick, L., Forman, M., Hale, D., Hertz-Picciotto, I., Hobbs, C., Hovell, M., Hudak, M., Kozinetz, C., Leuthner, S., Lipshultz, S., Lister, G., McLaughlin, T., McGovern, P., Murray J., Paneth, N., Robbins, J., Miller Sauvage, L., Siega-Riz, A.M., Specker, B., Swanson, J., Sweet, R., Tollerud, D., White, L., Wilhelm, M., and Wyatt, S. (2013). *A Cost-Effective and Feasible Design for the National Children's Study: Recommendations from the Field*. Available: http://www.epi.msu.edu/faculty/paneth/costeffective_ncs.pdf [June 2013].

Bracken, M.B. (1984). *Perinatal Epidemiology*. New York: Oxford University Press.

Environmental Protection Agency. (2010). *Workshop on Optimizing Exposure Metrics for the NCS,* Available: http://www.epa.gov/nerl/download_files/documents/national_childrens_study_june_2010.pdf [June 2013].

Herbst, A.L., Ulfelder, H., and Poskanzer, D.C. (1971). Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *New England Journal of Medicine* 284(15):878-881.

Kiefer, J. (1959). Optimum experimental designs (with discussions). *Journal of the Royal Statistical Society, Series B (Methodology)* 21:272-319.

Kwan, J., Dellarco, M., Park, C., Hirschfeld, S., and Kirkendall, N. (2013). *Background for Discussion at the Workshop on the Design of the National Children's Study* [background paper provided to workshop participants]. Working draft dated January 3. Available: http://www.nationalchildrensstudy.gov/research/workshops/Pages/nationalacademyofsciencesworkshop.aspx [June 2013].

National Children's Study. (2012). *The National Children's Study—Institute of Medicine Workshop, Steering Committee Briefing Document.* October 16. Available: http://www.nationalchildrensstudy.gov/research/workshops/Pages/nationalacademyofsciencesworkshop.aspx [June 2013].

National Research Council and Institute of Medicine. (2008). *The National Children's Study Research Plan: A Review.* Panel to Review the National Children's Study Research Plan. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Board on Children, Youth, and Families and Board on Population Health and Public Health Practice, Institute of Medicine. Washington, DC: The National Academies Press.

National Research Council and Institute of Medicine. (2013). *U.S. Health in International Perspective: Shorter Lives, Poorer Health*. Panel on Understanding Cross-National Health Differences Among High-Income Countries, S.H. Woolf and L. Aron, Eds. Committee on Population, Division of Behavioral and Social Sciences and Education, and Board on Population Health and Public Health Practice, Institute of Medicine. Washington, DC: The National Academies Press.

Appendix A

Workshop Agenda

Workshop on the Design of the National Children's Study
Committee on National Statistics
Board on Children, Youth, and Families
National Academy of Sciences Main Building
2101 Constitution Ave NW Lecture Room

Friday, January 11, 2013

8:30 am **Welcome and Introductions**
Sara McLanahan, Chair, Workshop Steering Committee, Center for Research on Child Wellbeing, Princeton University

8:35 **Welcome to the National Academies**
Constance Citro, Director, Committee on National Statistics, National Research Council

8:45 **Welcome to the Workshop and Statement of Its Purpose**
Steven Hirschfeld, Director, National Children's Study, National Institute of Child Health and Human Development

9:00 **Panel Discussion: Decisions about Environmental Measures**

Moderator: Marie McCormick (Steering Committee Member), Harvard School of Public Health and Harvard Medical School

Panelist 1: Melissa Perry, Department of Environmental and Occupational Health, School of Public Health and Health Services, George Washington University

Panelist 2: Nicole Cardello Deziel, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute

Panelist 3: Linda Sheldon, National Exposure Research Laboratory, Office of Research and Development, Environmental Protection Agency

Panelist 4: Antonia Calafat, Personal Care Products Laboratory, Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention

10:10 Floor Discussion, Moderator

10:30 Break

10:45 **Panel Discussion: Composition of Sample: Alternatives for Cohorts of Women**

Moderator: Barbara Carlson (Steering Committee Member), Mathematica Policy Research

Panelist 1: Irwin Garfinkel (Steering Committee Member), Columbia University School of Social Work and Population Research Center

Panelist 2: Naihua Duan (Steering Committee Member), Columbia University Medical Center

Panelist 3: Nancy Reichman, Robert Wood Johnson Medical School and Center for Health and Wellbeing, Princeton University

Panelist 4: Michael Bracken, Yale University School of Public Health

11:55 Floor Discussion, Moderator

12:15 pm Break for Lunch

1:15 **Panel Discussion: Weighting, Imputation, and Estimation in Proposed Design**
Moderator: Steven Cohen (Steering Committee Member), Agency for Healthcare Research and Quality
Panelist 1: Graham Kalton, Westat, Rockville, MD
Panelist 2: Colm O'Muircheartaigh, Harris School of Public Policy and NORC at the University of Chicago
Panelist 3: Richard Valliant, Joint Program in Survey Methodology at the University of Maryland and University of Michigan, Ann Arbor

2:25 Floor Discussion, Moderator

2:45 Break

3:00 **Panel Discussion: Factors, Issues, and Values to Balance and Consider in Reaching Decisions About the NCS Design**
Moderator: Greg Duncan (Steering Committee Member), School of Education, University of California, Irvine
Panelist 1: Ed Sondik (Steering Committee Member), Director, National Center for Health Statistics, Centers for Disease Control and Prevention
Panelist 2: Roderick Little, Department of Statistics, University of Michigan, Ann Arbor
Panelist 3: Ana Diez Roux, Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor

4:10 Floor Discussion, Moderator

5:00 Adjourn, Sara McLanahan, Chair

Appendix B

Registered Participants

WORKSHOP ON THE DESIGN OF THE NATIONAL CHILDREN'S STUDY MAIN STUDY

Deborah Baca, National Institute of Child Health and Human Development (NICHD) Office of Acquisitions
Marion Balsam, NICHD/National Children's Study (NCS)
Arthur Bennett, National Institutes of Health (NIH)/NICHD
Valerie Betley, Circle Solutions, Inc.
Lizbet Boroughs, American Psychiatric Association
Andrew Bowman, Drinker Biddle and Reath, LLP
Elizabeth Boyle, Westat
Michael Bracken, Yale University School of Public Health
Amy Branum, Centers for Disease and Control Prevention(CDC)/ National Center for Health Statistics (NCHS)
Ruth Brenner, NICHD/NCS
Andrew Briggs, Booz Allen Hamilton
Margot Brown, U.S. Environmental Protection Agency (EPA) Office of Children's Health Protection
Antonia Calafat, CDC
Nicole Cardello Deziel, National Cancer Institute
Barbara Carlson, Mathematica Policy Research
Kirti Chadha, Booz Allen Hamilton
Richard Chestek, Booz Allen Hamilton
Kendall Cislo, Michigan Alliance of the National Children's Study
Marguerite Clarkson, PricewaterhouseCoopers

Steven Cohen, Agency for Healthcare Research and Quality
Leslie Cooke, NICHD/NCS
Dorr Dearborn, Case Western Reserve University
Mark del Monte, American Academy of Pediatrics
Michael Dellarco, NICHD/NCS
Ana Diez Roux, University of Michigan, Ann Arbor
Naihua Duan, Columbia University
Greg Duncan, University of California, Irvine
Eric Fatemi, Labor, HHS Subcommittee, Senate Appropriations Committee
Alexa Fraser, Westat Health Sector
Mischka Garel, Johns Hopkins University
Irwin Garfinkel, Columbia University
John Gohagan, NIH
Jessica Graber, NICHD/NCS
Alan Guttmacher, NICHD
Alycia Halladay, Autism Speaks
Brian Harris-Kojetin, Office of Management and Budget
Paymon Hashemi, Booz Allen Hamilton
Maire Heikkinen, Westat
Steven Hirschfeld, NICHD/NCS
Charlotte Hobbs, University of Arkansas Medical Sciences and Arkansas Children's Hospital
Mary Jo Hoeksema, Population Association of America
Valerie Hsu, Westat
David Hubble, Westat
Grace Ji, Avar Consulting, Inc.
Lisa Kaeser, NICHD/NIH
Graham Kalton, Westat
Carol Kasten, NICHD/NCS
Jean Kerver, Michigan State University
Jennifer Kwan, NICHD/NCS
Colleen Lee, NICHD/NCS
Kate Lefauve, NORC at the University of Chicago
Roderick Little, University of Michigan, Ann Arbor
Maria Lopez-Class, NICHD/NCS
R. Eric Lorenzo, NICHD/NCS
John Lumpkin, NICHD/NCS
Cora MacPherson, Social and Scientific Systems
Jennifer Madans, NCHS
Nina Markovic, University of Pittsburgh
Dave Mason, Mason Consulting, LLC
Marie McCormick, Harvard University

Sara McLanahan, Princeton University
Thomas McLaughlin, University of Massachusetts Medical School
Lori Merrill, Westat
Mark Meschter, Circle Solutions, Inc.
Aubrey Miller, National Institute of Environmental Health Sciences (NIEHS)
Susan Mitchell, RTI International
Mary Mortensen, CDC
Nolan Morton, Circle Solutions, Inc.
Jack Moye, NICHD/NCS
Sheila Newton, NIEHS
Cindy Nowinski, National Children's Study South Regional Operations Center
Colm O'Muircheartaigh, University of Chicago
Nigel Paneth, Michigan State University
Nancy Parfit Hondros, NICHD/NCS
Christina Park, NICHD/NCS
Albert Parker, Avar Consulting, Inc.
Melissa Perry, George Washington University
Nancy Reichman, Princeton University
Katie Rush, NICHD
Rajni Samavedam, Booz Allen Hamilton
Christian Sauter, PwS
Susan Schechter, NORC at the University of Chicago
Margo Schwab, Office of Management and Budget
Angela Sharpe, Consortium of Social Science Associations
Linda Sheldon, EPA
Ed Sondik, NCHS
Mike Spittel, NIH Office of Behavioral and Social Sciences Research
Aditi Srivastgav, American Academy of Pediatrics
Gerald Sroufe, American Educational Research Association
Karen Studwell, American Psychological Association
Gitanjali Taneja, NICHD/NCS
Richard Valliant, Joint Program in Survey Methodology at the University of Maryland and University of Michigan, Ann Arbor
Sandra Wadlinger, The Children's Hospital of Philadelphia
Joan Wang, Avar Consulting, Inc.
Emil Wigode, March of Dimes
Anne Zajicek, NICHD
Tia Zeno, Population Association of America fellow

COMMITTEE ON NATIONAL STATISTICS

The Committee on National Statistics (CNSTAT) was established in 1972 at the National Academies to improve the statistical methods and information on which public policy decisions are based. The committee carries out studies, workshops, and other activities to foster better measures and fuller understanding of the economy, the environment, public health, crime, education, immigration, poverty, welfare, and other public policy issues. It also evaluates ongoing statistical programs and tracks the statistical policy and coordinating activities of the federal government, serving a unique role at the intersection of statistics and public policy. The committee's work is supported by a consortium of federal agencies through a National Science Foundation grant.