Policy Issues Associated with Undertaking a New Large U.S. Population Cohort Study of Genes, Environment, and Disease

Report of the
Secretary's Advisory Committee on Genetics, Health, and Society

March 2007

A pdf of this report is available at
http://www4.od.nih.gov/oba/sacghs/reports/SACGHS_LPS_report.pdf.

Policy Issues Associated with Undertaking a New Large U.S. Population Cohort Study of Genes, Environment, and Disease

---

Report of the
Secretary's Advisory Committee on Genetics, Health, and Society

March 2007

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Secretary's Advisory Committee on
Genetics, Health, and Society
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892-7985
301-496-9838 (Telephone)
301-496-9839 (Facsimile)
http://www4.od.nih.gov/oba/sacghs.htm

March 1, 2007

The Honorable Michael O. Leavitt
Secretary of Health and Human Services
200 Independence Avenue, S.W.
Washington, DC 20201

Dear Secretary Leavitt:

On behalf of the Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) and in keeping with our mandate to serve as a public forum for deliberations on the broad range of human health and societal issues raised by the development and use of genetic technologies and, as warranted, to provide advice on these issues, I am pleased to submit the report *Policy Issues Associated with Undertaking A New Large U.S. Population Cohort Study of Genes, Environment, and Disease*. The report describes the foundational questions that must be addressed to help policymakers determine whether the U.S. Government should consider undertaking a new large population study to elucidate the influence of genetic variation and environmental factors on common, complex diseases.

In this report, SACGHS identifies a range of salient issues that warrant consideration and further analysis. These issues fall into five broad areas—research policy; research logistics; regulatory and ethical considerations; public health, social, and economic implications; and public engagement. The report sets forth recommended next steps for addressing policy gaps and for engaging the public. These are steps that we believe should be carried out expeditiously given the study's potential benefits, the need for broad public support for such a study, and questions about the feasibility of undertaking it from scientific, social, economic, and ethical perspectives. The results of these analytical efforts are likely to contribute insights to other important research and public health policy initiatives within the Department of Health and Human Services.

This report is the culmination of almost two years of fact-finding, public consultation, and deliberation by the Committee. SACGHS is grateful to the many individuals who shared their knowledge and perspectives with us in the development of the report.

We appreciate the opportunity to address this important topic and hope that our deliberations and recommendations will be helpful to you and the Department.

Sincerely,

Reed V. Tuckson, M.D.
Chair, SACGHS

About SACGHS

The Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) was first chartered in 2002 by the Secretary of Health and Human Services (HHS) as a public forum for deliberation on the broad range of human health and societal issues raised by the development and use of genetic tests and, as warranted, to provide advice on these issues. The charter sets out the following specific functions of the Committee:

- Assessing how genetic and genomic technologies are being integrated into health care and public health
- Studying the clinical, public health, ethical, economic, legal, and societal implications of genetic and genomic technologies and applications
- Identifying opportunities and gaps in research, and data collection and analysis efforts
- Examining the impact of current patent policy and licensing practices on access to genetic and genomic technologies
- Analyzing uses of genetic information in education, employment, insurance, and law
- Serving as a public forum for discussion of issues raised by genetic and genomic technologies

Structurally, SACGHS consists of up to 17 individuals from around the Nation who have expertise in disciplines relevant to genetics and genetic technologies. These disciplines include biomedical sciences, human genetics, health care delivery, evidence-based practice, public health, behavioral sciences, social sciences, health services research, health policy, health disparities, ethics, economics, law, health care financing, consumer issues, and other relevant fields. At least two of the members are specifically selected for their knowledge of consumer issues and concerns and the views and perspectives of the general public.

Representatives of at least 19 Federal departments or agencies also sit on SACGHS in an ex officio (nonvoting) capacity. The departments and agencies are the Department of Commerce, Department of Defense, Department of Education, Department of Energy, Administration for Children and Families (HHS), Agency for Healthcare Research and Quality (HHS), Centers for Disease Control and Prevention (HHS), Centers for Medicare & Medicaid Services (HHS), Food and Drug Administration (HHS), Health Resources and Services Administration (HHS), National Institutes of Health (HHS), Office for Civil Rights (HHS), Office for Human Research Protections (HHS), Office of Public Health and Science (HHS), Department of Justice, Department of Labor, Department of Veterans Affairs, Equal Employment Opportunity Commission, and Federal Trade Commission.

Committee Roster

Chair

Reed V. Tuckson, MD (2003-2007)
Senior Vice President
Consumer Health & Medical Care Advancement
UnitedHealth Group
Minnetonka, MN

Members

Sylvia Mann Au, MS, CGC (2005-2009)
Hawaii State Genetics Coordinator
Hawaii Department of Health
Honolulu, HI

Cynthia E. Berry, JD (2003-2007)
Partner
Powell Goldstein Frazer & Murphy
Washington, DC

Chira Chen (2005-2007)
Staff Research Associate
University of California-San Francisco Comprehensive Cancer Center
San Francisco, CA

James P. Evans, MD, PhD (2005-2009)
Associate Professor of Genetics and Medicine
Director of Clinical Cancer Genetics and the Bryson Program in Human Genetics
University of North Carolina at Chapel Hill
Chapel Hill, NC

Andrea Ferreira-Gonzalez, PhD (2006-2010)
Professor of Pathology and Director of the Molecular Diagnostics Laboratory
Virginia Commonwealth University
Richmond, VA

Kevin FitzGerald, SJ, PhD, PhD (2005-2008)
David P. Lauler Chair in Catholic Health Care Ethics
Research Associate Professor
Department of Oncology
Georgetown University Medical Center
Washington, DC

Debra G.B. Leonard, MD, PhD (2003-2006)*
Vice Chair of Laboratory Medicine
New York Presbyterian Hospital
New York, NY

Julio Licinio, MD (2007-2009)
Professor and Chairman
Leonard M. Miller School of Medicine
University of Miami
Department of Psychiatry and Behavioral Sciences
University of Miami/Jackson Memorial Hospital Mental Health Hospital Center
Miami, FL

Agnes Masny, RN, MPH, MSN (2003-2006)*
Adjunct Assistant Professor of Nursing
Temple University College of Allied Health Professionals
Research Assistant and Nurse Practitioner
Family Risk Assessment Program
Fox Chase Cancer Center
Philadelphia, PA

Barbara Burns McGrath, RN, PhD (2006-2010)
Research Associate Professor
University of Washington School of Nursing
Seattle, WA

Joseph Telfair, DrPH, MSW, MPH (2004-2008)
Professor
Public Health Research and Practice
Department of Public Health Education
University of North Carolina at Greensboro
Greensboro, NC

* Member's term expired prior to the Committee's acceptance of the final draft report.

Steven Teutsch, MD, MPH (2006-2010)
Executive Director
Outcomes Research and Management
Merck & Co., Inc.
West Point, PA

Huntington F. Willard, PhD (2003-2007)
Director
Institute of Genome Sciences & Policy
Vice Chancellor for Genome Sciences
Genome Sciences Research
Duke University Medical Center
Durham, NC

Emily S. Winn-Deen, PhD (2003-2006)*
Vice President
Strategic Planning & Business Development
Cepheid
Sunnyvale, CA

Ex Officio Members

Michael Amos, PhD
Department of Commerce

Lt. Col. Scott D. McLean, MC, USA
Department of Defense

Daniel Drell, PhD
Department of Energy

Department of Health and Human Services

Martin Dannenfelser
Administration for Children and Families

Gurvaneet Randhawa, MD, MPH
Agency for Healthcare Research and Quality

Muin J. Khoury, MD, PhD
Centers for Disease Control and Prevention

Barry M. Straube, MD
Centers for Medicare & Medicaid Services

Steven I. Gutman, MD, MBA
Food and Drug Administration

Denise Geolot, PhD, RN, FAAN
Health Resources and Services Administration

Francis S. Collins, MD, PhD
National Institutes of Health

Robinsue Frohboese, JD, PhD
Office for Civil Rights

Michael Carome, MD
Office for Human Research Protections

Anand Parekh, MD, MPH
Office of Public Health and Science

Thomas Alexander, JD
Department of Labor

Ellen Fox, MD
Department of Veterans Affairs

Matthew Daynard, JD
Federal Trade Commission

* Member's term expired prior to the Committee's acceptance of the final draft report.

SACGHS Task Force on Large Population Studies

Chair

Huntington F. Willard, PhD

Members

Sylvia Mann Au, MS, CGC
Chira Chen
Kevin FitzGerald, SJ, PhD, PhD
Debra G.B. Leonard, MD, PhD*
Julio Licinio, MD
Barbara Burns McGrath, RN, PhD
Joseph Telfair, DrPH, MSW, MPH
Steven Teutsch, MD, MPH

Ex Officio Members

Department of Health and Human Services

Centers for Disease Control and Prevention
Muin J. Khoury, MD, PhD

National Institutes of Health
Francis S. Collins, MD, PhD
Alan Guttmacher, MD
Phyllis Frosst, PhD

Department of Veterans Affairs

Ellen Fox, MD
Sherrie Hans, PhD

* Member's term expired prior to the Committee's acceptance of the final draft report.

Acknowledgments

The Committee wishes to thank the members of the SACGHS Task Force on Large Population Studies for their pivotal role in guiding the development of this report. The Task Force was chaired by Huntington Willard and was composed of SACGHS Members Sylvia Mann Au, Barbara Burns McGrath, Chira Chen, Kevin FitzGerald, Debra Leonard, Julio Licinio, Joseph Telfair, and Steven Teutsch and SACGHS Ex Officio Members Francis Collins (NIH), Ellen Fox (VA), Phyllis Frosst (NIH), Alan Guttmacher (NIH), Sherrie Hans (VA), and Muin Khoury (CDC).

The Committee is indebted to the following presenters for sharing their knowledge, expertise, and perspectives on issues related to large population cohort studies during the February-March and October 2005 and June 2006 SACGHS meetings: Ruth Brenner (NIH); Toby Citrin (University of Michigan); Mylene Deschenes (Public Population Project in Genomics); Troy Duster (New York University); Stephan Fihn (VA); Gerald Fink (Massachusetts Institute of Technology); David Goldstein (University College of London); Henry Greely (Stanford Law School); John Hewitt (University of Colorado); Sharon Kardia (University of Michigan); Yvonne Lewis (Faith Access to Community Economic Development); Teri Manolio (NIH); Richard Marchase (Federation of American Societies for Experimental Biology), John Newton (UK Biobank); Gil Omenn (University of Michigan); Pilar Ossorio (University of Wisconsin Law School); Charles Rotimi (Howard University); David Schwartz (NIH); Joan Scott (Genetics and Public Policy Center); and Mary Woolley (Research!America).

The Committee thanks all of the individuals and organizations that responded to the request for public comments during the development of this report (see Appendix C). The Committee gave careful consideration to each comment and incorporated the information as appropriate throughout the body of this report. The Committee benefited greatly from the public input, and we hope the public's insights are well reflected in this final report and in our recommendations.

The Committee also wishes to recognize the work of SACGHS staff members. Amanda Sarata served as lead staff member at the start, and Yvette Seger saw the project to completion. Ms. Sarata and Dr. Seger were responsible for organizing the deliberations of the LPS Task Force and the full Committee on this topic and for additional research and analysis required during the development of the report. Contract writers were also employed; Kathi Hanna prepared an initial draft of the report, and Betsy Earp brought the final document to fruition. Katherine Kolor, Amita Mehrotra, and Tony Tse contributed greatly to our review and analysis of the public comments. During a fellowship sponsored by the American Association for the Advancement of Science, Holly Campbell-Rosen prepared a helpful background report. Sarah Carr and Suzanne Goodwin provided guidance throughout the process.

SACGHS is managed by the NIH Office of Biotechnology Activities (OBA) within the NIH Office of Science Policy (OSP), NIH Office of the Director. The Committee wishes to thank NIH Director Elias Zerhouni, OSP Director Lana Skirboll, and OBA Director Amy Patterson for NIH's ongoing support of SACGHS.

Contents

Executive Summary

Characterizing human genetic variants and determining the ways in which they interact with environmental factors to influence health are critical issues being addressed by scientists who are trying to unravel the underlying causes of common diseases. By studying where variations among individuals occur within the human genome and how particular DNA variants interact in the context of environmental factors, major clinical and public health advances might be achieved. One approach for learning more about these relationships is through large longitudinal population studies that involve the collection of health and environmental data, as well as biological specimens from hundreds of thousands of people.

Such large national population projects are currently underway in Japan, Iceland, the United Kingdom, Estonia, and other countries, with the expectation that they will serve as essential resources for hundreds, if not thousands, of research studies (see Appendix A). A similar large-scale project in the United States is considered by some scientists to be a logical next step that would build on the complete sequencing of the human genome[1] to enhance the understanding of common diseases and improve treatments and therapies. Accordingly, the National Institutes of Health (NIH) has been investigating the initial questions concerning the possibility of mounting a new large U.S. population cohort study.

In 2004 the Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) also prioritized the need for an analysis of the opportunities and challenges associated with conducting a new large population study (LPS) aimed at understanding the relationships between genes, environmental factors, and their interactions and common complex diseases. SACGHS began with an analysis of the National Human Genome Research Institute (NHGRI) document *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel*. The report was developed in collaboration with several NIH Institutes and Centers and examines the scientific foundations and broad logistical outlines of a hypothetical U.S. cohort project for the study of genes, environment, and health.

In June 2005 NIH Director Dr. Elias Zerhouni asked the Committee to focus its inquiry on the preliminary questions that need to be addressed before considering whether the United States should undertake such a study. Dr. Zerhouni asked SACGHS to identify broad policy and process issues but not to provide answers to the questions raised. Specifically, Dr. Zerhouni requested that the Committee (1) delineate the questions that need to be addressed for policymakers to determine whether the U.S. Government should undertake a new LPS to elucidate the influence of genetic variation and environmental factors on common complex diseases; (2) explore the processes by which the identified questions and issues can be addressed, including any necessary intermediate research studies, pilot projects, or policy analysis efforts; and (3) determine the possible ways in which these questions could be addressed, taking into account the feasibility of those approaches that seem optimal, and recommend a specific course of action for moving forward.

SACGHS organized two separate, daylong factfinding and consultation sessions in 2005 that provided the Committee with a comprehensive understanding of science policy; ethical, legal, and social considerations; public health; and public engagement issues that are relevant to a new LPS in the United States.

[1] Collins FS. The case for a US prospective cohort study of genes and environment. *Nature* 2004. 429(6990):475-7.

The resulting report presents issues for consideration by the Secretary, U.S. Department of Health and Human Services (HHS), in separate chapters that address the areas of research policy; research logistics; regulations and ethics; public health, social, and economic implications; and public engagement. A comprehensive discussion of these areas concludes with specific recommendations for next steps by the HHS Secretary.

This report was submitted for public comment in the *Federal Register* on June 13, 2006. In addition, input was solicited through a posting on the SACGHS Web site, dissemination by SACGHS ex officio members, e-mail and listserv outreach to many stakeholder groups, and media outreach conducted through the NIH Office of Communications. All feedback received has been carefully considered and incorporated throughout this report. During its November 2006 meeting, the Committee reviewed these revisions and approved the report for transmittal to the HHS Secretary.

Overview of this Report

Chapter I. Introduction provides background information on the promise and challenges associated with an LPS and describes the efforts that have been undertaken to explore the feasibility of such a study in the United States.

Chapter II. State of the Science describes recent advances in molecular biology and genetics, including the research opportunities made possible by the recently completed Human Genome Project and the International HapMap Project, which have prompted a search for methods that will advance the understanding of the relationship between genetics and disease. The status of the field of environmental health science also is described. Detailed information is provided on the central topic of this report: a new prospective cohort study of genes and the environment, which enrolls individuals prior to disease onset and prospectively collects environmental and biobehavioral marker data, allowing for the analysis of nongenetic and genetic factors in disease. Chapter II also describes the population-based, gene-environment cohorts and related initiatives already taking place in the United States and abroad.

Chapter III. Policy Issues Associated with a New Large Population Study of Genes, Environment, and Common Diseases—the heart of the report—focuses on four key policy areas associated with a new LPS in the United States: (1) research policy, (2) research logistics, (3) regulations and ethics, and (4) public health, social, and economic implications. (A summary of these discussion topics is found on pp. 3-6.)

Chapter IV. Need for Public Engagement explores the critical function of public engagement. SACGHS concluded that extensive public engagement efforts are necessary to gauge public opinion on the feasibility of a new LPS in this country. If the initiative were to go forward, continued evaluation of public opinion would need to play a central role in the planning, implementation, and conduct of the study, as well as in decisionmaking that affects the application of research results.

A comprehensive discussion of the issues is contained in Chapters III and IV, and Chapter V includes recommendations for next steps by the HHS Secretary.

Appendixes A and B provide summaries of the factfinding and consultation sessions conducted by SACGHS in March 2005 and October 2005. Appendix C lists the individuals and organizations that provided public comment. Appendix D describes several international biobanking efforts. Appendix E details the public consultation approaches used by four international large cohort projects.

The sections below describe key discussion topics in Chapters III and IV and list related recommendations.

Research Policy

The topics discussed under Research Policy include the arguments that have been put forward for and against a new LPS, the possible effects on other areas of science, the capacity of the current U.S. research system to conduct interdisciplinary science, and the need for public and private and other creative partnerships. Other issues considered include access to the study data and materials collected by other researchers and the complexity of intellectual property concerns. This section presents one overarching recommendation and five additional recommendations on these issues for consideration by the HHS Secretary, as listed below.

Overarching Recommendation

As part of the process for determining whether to undertake such a large-scale research project, the HHS Secretary should initiate a thorough consideration of the full range of policy issues outlined in this report. The HHS Secretary should consult and engage the full range of potential partners for such a project during this decisionmaking process, including the public at large, the full scientific community, a wide spectrum of Government agencies and policymakers, and the private sector.

Recommendations: Research Policy

If a new large population cohort study is conducted:

1. **The HHS Secretary should continue to promote and facilitate ongoing consultation with the public, the private sector, and the international community to explore opportunities for collaboration on a large population study.**
2. **The HHS Secretary, in consultation with relevant HHS agencies and appropriate congressional committees, should assess support for sustaining a long-term and stable investment in a large population study.**
3. **Given the interdisciplinary nature of its scope, the HHS Secretary should establish a highly collaborative model of project leadership and management in multiple HHS and non-HHS agencies (e.g., NIH Institutes and Centers, CDC, EPA, and VA) and with other stakeholders. This includes the public and private sectors; biological, behavioral, social, public health, and population science disciplines; and basic biological scientists and epidemiologists.**
4. **The HHS Secretary, in consultation with relevant HHS agencies, should ensure that there are opportunities available to the general scientific community to (a) be informed about the potential for such a project; (b) present its views about the scientific validity and feasibility of such a project; (c) present its views on the commitment of resources to such an effort, including whether there are benefits to leveraging existing efforts; and (d) provide input on issues related to fair access by scientists to the project resources and the sharing of data and samples collected within it.**
5. **To ensure public benefits, the HHS Secretary should require that there are clear intellectual property policies in place for discoveries made using the data and samples collected.**

Research Logistics

The topics discussed under Research Logistics have important social and ethical consequences. They include a recruitment and enrollment strategy that includes individuals residing in all geographic regions of the United States, ethnically and socioeconomically diverse populations, and others who have not traditionally participated in research. The section explores the need for methods and technologies that can measure differences in health and risk factors in the population, as well as the need for methods to collect and analyze environmental exposure data. The requirement for interdisciplinary research teams and coordination across multiple institutions and health care systems is also discussed. There is an emphasis on the need for development of an information technology infrastructure that would support data collection, storage, sharing, and security and privacy protections. The section presents four recommendations on these issues for consideration by the HHS Secretary, as listed below.

Recommendations: Research Logistics

If a new large population cohort study is conducted:

1. **The HHS Secretary should encourage the project leadership and the scientific community to develop clear, consistent definitions and parameters for the stratification and classification of the projected sample population to ensure diversity and appropriate representation in the population to be studied.**
2. **The HHS Secretary should seek input from the public, as well as from researchers and clinicians, on the best approaches for identifying, recruiting, educating, and enrolling various subpopulations. Project organizers should be encouraged to consult with community-based organizations as part of their recruitment, assessment, and enrollment strategies.**
3. **The HHS Secretary, in consultation with both HHS and non-HHS agencies, should refine methods for collecting and analyzing environmental (i.e., physical, behavioral, and social) factors influencing health and ensure that resources are devoted to developing new tools to validate existing methods and improve assessments of the environment.**
4. **The HHS Secretary should encourage the project leadership to consult with health care providers and organizations to develop uniform and secure approaches for collecting, storing, tracking, and centralizing clinical information to be gathered over the course of the project, including the use of electronic health records.**

Regulations and Ethics

The topics discussed under Regulations and Ethics include institutional review board (IRB) oversight, informed consent, provision of care by the research team, privacy and confidentiality, control of the biological samples and the research data, the return of research results, and the need for an independent ethics review committee. The section presents four recommendations on these issues for consideration by the HHS Secretary, as listed below.

Recommendations: Regulations and Ethics

If a new large population cohort study is conducted:

1. **The HHS Secretary should convene a working group of representatives from the Office for Human Research Protections, Food and Drug Administration, Office for Civil Rights, and other relevant HHS and non-HHS agencies to address issues and questions raised by the public and to provide technical assistance and guidance to research sites on legal requirements regarding the protection of research subjects, health information privacy, and patient safety.**
2. **The HHS Secretary should establish an independent ethics committee to serve in an advisory capacity to the institutional review boards and the project leadership.**
3. **The project leadership should systematically and regularly seek the input of study subjects regarding their experiences, concerns, and recommendations for enhancing protections to ensure that the appropriate protections are in place and are being consistently implemented.**
4. **The project leadership should develop a policy regarding the use of data and samples to ensure the legal and ethical use of clinical and epidemiological data and specimens. This policy should be made available to study subjects.**

Public Health, Social, and Economic Implications

The topics discussed under Public Health, Social, and Economic Implications include the possibility of elucidating and/or exacerbating health disparities and the risks of misrepresenting genetic determinism by the lay public, professionals, and policymakers. Possible social and policy responses to research findings are considered, as are potential economic impacts. The section presents two recommendations on these issues for consideration by the HHS Secretary, as listed below.

Recommendations: Public Health, Social, and Economic Implications

If a new large population cohort study is conducted:

1. **The HHS Secretary and the project leadership should systematically and regularly integrate project findings with other emerging data from other types of studies and regularly disseminate the accumulated knowledge base in a manner to benefit the population's health. This information should be tailored to meet the information needs of the public, health care providers, and the public health community to use integrated information for the benefit of the population's health. Project resources should be sufficient for the integration, dissemination, and translation activities necessary to maximize the public health impact.**
2. **The HHS Secretary, in consultation with the project leadership, should establish an independent standing committee for the duration of the project to periodically assess the persistent and emerging social and economic implications of this initiative, with special attention to health disparities. The committee could consist of individuals with expertise in the relevant sciences, medicine, law, ethics, and patient and community advocacy. The committee would routinely seek input from the public on the implications of project results and report its findings.**

Need for Public Engagement

The topics discussed under Need for Public Engagement include a recent survey on the public's awareness of genomics and effective mechanisms for engaging the public. To illustrate specific approaches, examples of public engagement efforts used by the National Children's Study and an NHGRI public consultation initiative are described. The section presents two recommendations on these issues for consideration by the HHS Secretary, as listed below.

Recommendations: Need for Public Engagement

1. **Before embarking on such a large population study and in advance of any funding decision, the HHS Secretary should assess the public's willingness to participate in such an extensive endeavor.**
2. **If a decision is made to proceed with a large population study, it will be important for the HHS Secretary to ensure that public engagement occurs throughout all aspects and stages of the research process, from conceptualization through design, planning, implementation, conduct, and data analysis and reporting. Public engagement also will be important in applying the knowledge gained by the research and in addressing its implications. The HHS Secretary should ensure that sufficient project resources are dedicated to public consultation activities before and throughout the duration of the project.**

Conclusion

Throughout this report, SACGHS has identified the questions that must be addressed before a decision is made on whether to proceed with a new LPS in the United States. Because this issue has the potential to affect the health of millions of U.S. citizens, the Committee has comprehensively investigated it as a high priority area for more than 2 years. SACGHS has heard from numerous experts in relevant disciplines and carefully considered public comments submitted by academic researchers, industry representatives, Federal policymakers; members of policy research groups, IRBs, professional societies, and nonprofit advocacy groups; and individual consumers of health services.

On the basis of the knowledge gathered during this process, the Committee believes that a new LPS in the United States could lead to improved diagnostics, treatments, and preventive measures for common diseases such as cardiovascular disease, diabetes, and cancer. Because of the potential knowledge and health benefits to be gained, SACGHS urges the HHS Secretary to move forward expeditiously to address the policy gaps identified in this report. Until this analytical work is completed, it will not be clear whether such a study has the broad public support necessary for moving forward and whether the study is feasible from scientific, social, economic, and ethical standpoints.

I. Introduction

For scientists who are trying to unravel the underlying causes of common diseases, one of the most pressing goals is to characterize human genetic variants and how they interact with a variety of physical, physiological, behavioral, and other environmental factors to produce disease. Scientists hope that major public health advances will be realized by learning where variations among individuals lie within the human genome; how these variants differ among healthy, predisposed, and sick individuals; and how particular variants of deoxyribonucleic acid (DNA) interact with each other and with environmental factors.

Large, longitudinal population cohort studies have been proposed as one promising approach to learning more about these relationships. These studies involve the collection of data about and biological specimens from hundreds of thousands of people and the creation of a large database and biobank that could be mined by many investigators to enable the conduct of individual research studies. The cohort approach is complementary to the case-control method, which has been the backbone of genetic research for determining correlations between genotypic and phenotypic data.

A. Purpose and Organization of this Report

This report describes the efforts of the Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) to assess the need and readiness for a new large population study (LPS) in the United States and presents recommendations to the Secretary of the U.S. Department of Health and Human Services (HHS) so that this concept can be further explored. The HHS Secretary established SACGHS in 2002 as a public forum for deliberation on the broad range of human health and societal issues raised by advances in genetics and, as warranted, the development of advice on these issues. In a March 2004 priority-setting process, SACGHS identified 11 high-priority issues warranting its attention and analysis. One of those issues was the need for an analysis of the opportunities and challenges associated with conducting an LPS aimed at understanding the relationships between genes, environments,[1] and their interactions and common complex diseases. Among the considerations that led the Committee to this decision was the fact that discussions were underway at the National Institutes of Health (NIH) about whether the United States should mount a new large population-based study.

In June 2005, as SACGHS factfinding efforts were beginning, NIH Director Dr. Elias A. Zerhouni requested that the Committee develop a report on the preliminary questions, steps, and strategies that would need to be addressed before considering the larger question of whether the United States should undertake a new LPS. Specifically, the Committee was asked to (1) delineate the questions that need to be addressed for policymakers to determine whether the U.S. Government should undertake a new LPS to elucidate the influences of genetic variations and environmental factors on common complex diseases; (2) explore the ways in which, or processes by which, the questions identified in step 1 can be addressed, including any intermediate research studies, pilot projects, or policy analysis efforts needed; and (3) determine the possible ways in which these questions could be addressed, taking into account the feasibility of those approaches

[1] Defining "environment" is a key policy issue for consideration. For the purposes of the LPS under consideration, "environment" should be defined as relating to human health as broadly and as inclusively as possible. See the section on Research Logistics.

that seem optimal, and recommend a specific course of action for moving forward. Dr. Zerhouni did not expect the Committee to recommend solutions to the questions raised.

The next section summarizes exploratory work by the National Human Genome Research Institute (NHGRI) and factfinding and consultative efforts by SACGHS on this issue. Chapter II presents the scientific basis for an LPS. Chapter III outlines the key policy issues that SACGHS has identified as warranting further attention. Chapter IV discusses the critical role that public engagement must play in determining the willingness of U.S. citizens to support and participate in such an endeavor. In keeping with its agreed-upon charge, throughout this report the Committee explores the ways in which the identified policy issues could be addressed and describes possible approaches for the HHS Secretary's consideration.

B. Processes Used in the Development of this Report

To guide exploration of the LPS concept, the Committee created the SACGHS Task Force on Large Population Studies, consisting of a subgroup of Committee members and ex officio members.[2] The Task Force began its work by considering the publications and documents developed by NHGRI on this issue. In May 2004 NHGRI Director Dr. Francis S. Collins published an essay in *Nature* that asserted that drawing rigorous and unbiased conclusions about the causes of diseases and their population-wide impact would require the conduct of a new LPS over many years and that the time was right for the United States to consider undertaking such a project.[3] Later that year NHGRI, in collaboration with several other NIH Institutes, commissioned a panel of experts in genetics; epidemiology; biostatistics; and ethical, legal, and social issues in genetic research to examine the scientific foundations and broad logistical outlines of a hypothetical U.S. cohort project for the study of genes, environment, and health. The recommendations of that panel are summarized in the document *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel*.[4]

The Expert Panel concluded that identifying genetic and environmental factors that influence health, disease, and response to treatment is "essential to developing approaches to reduce disease burden." The Expert Panel also said that the goal of any large U.S. cohort study of genes, environment, and health should be ambitious and designed to ascertain and quantify all of the major environmental and genetic causes of common illnesses, setting the stage for a future of better preventive medicine and more effective therapy."[5]

The Expert Panel specified that the objectives of such a study would fall into the following four categories: (1) analysis of information that would be derivable very soon after enrollment of participants, such as the population prevalences of known genetic variants, common diseases, environmental exposures, and associations of genetic variants and environmental factors with traits and conditions present at baseline; (2) analysis of information that would be derived from incident cases as the cohort progresses, such as the determination of quantitative, unbiased risks of major genetic and environmental susceptibility factors for complex, common diseases, identification of major gene-environment and gene-gene interactions, and

---

[2] LPS Task Force members are listed in the Committee Roster on page v.
[3] Collins FS 2004. Op. cit.
[4] *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel.* See http://www.genome.gov/Pages/About/OD/ReportsPublications/PotentialUSCohort.pdf [accessed December 19, 2006].
[5] Ibid.

identification of biomarkers that represent early indicators of disease; (3) development of technology, such as sophisticated methodology for collection of large-scale phenotypic and environmental data, and data mining and statistical analysis appropriate for studies of genetic and environmental influences; and (4) broad access to this resource by the scientific community.

To achieve these objectives, the Expert Panel estimated that an unprecedented 500,000 to 1 million individuals would need to be enrolled and followed prospectively for at least 10 years. Enrollees would include representative samples of the U.S. population from defined census tracts. Subgroups that have not traditionally participated in research would be oversampled to ensure that these groups were sufficiently represented. The Expert Panel suggested that study subjects be contacted twice per year and examined every 4 years on average. Disease outcomes would be assessed using hospital records, outpatient records, and other data sources, such as Centers for Medicare & Medicaid Services data and registries. The Expert Panel also laid out other design and procedural elements that such a cohort study would entail.

In March 2005 the LPS Task Force organized a daylong factfinding session that provided a broad understanding of basic issues related to LPSs. Topics included the conceptual basis for these studies; public health perspectives; an overview of U.S. and international LPSs already underway or under consideration; and associated ethical, legal, and social issues. Federal representatives from the National Institute of Child Health and Human Development (NICHD), U.S. Department of Veterans Affairs (VA), NHGRI, and Centers for Disease Control and Prevention (CDC) made presentations on relevant activities in their agencies. (Appendix A provides a detailed description of the March 2005 presentations.)

At the October 2005 SACGHS meeting, another daylong session was held, this time to obtain expert consultation in the areas of science policy, bioethics, and public engagement.[6] (A summary of the October 2005 session is found in Appendix B.) The session crystallized the reasons why a new large cohort initiative in the United States raises so many policy issues, including the following: (1) It would involve an unprecedented number of participants and, thereby, would have a significant public profile and a direct impact on many U.S. citizens; (2) it would require a relatively large investment of public resources and, as such, would warrant close scrutiny of and deliberation about its relative value to science and society; and (3) the nature of the information that would be derived would raise ethical, legal, social, and public policy concerns that could be unique and/or significant. During the discussion that followed, most SACGHS members expressed support for a new LPS, balanced with the recognition that many areas required further exploration before a decision could be reached.

The first draft of this report, with recommendations for further action, was developed by the LPS Task Force and approved by SACGHS in March 2006. In May 2006 the Committee sought public comments on the report through the *Federal Register*,[7] including whether the policy issues identified were appropriately focused, whether any policy issues had been omitted, whether the issues were well organized, and whether the mechanisms for public engagement were adequately described. Public comments also were solicited through a variety of listservs and the efforts of SACGHS ex officio members, reaching as many as 48,000 subscribers, and a notice was posted on the SACGHS Web site. Selected media outlets were reached through the NIH Office of Communications.[8]

---

[6] See http://www4.od.nih.gov/oba/SACGHS/meetings/October2005/SACGHSOct2005postmeeting.htm [accessed December 19, 2006].

[7] *Fed Regist* June 13, 2006. 71:34134-5.

[8] For example, Kaiser J. Bank shot. *Science* 2006. 312(5781):1727.

SACGHS received a total of 69 responses during the public comment period (many of which provided numerous suggestions) from a wide variety of stakeholders, including academic researchers, industry representatives, Federal Government researchers, and members of policy research groups, institutional review boards (IRBs), professional societies, and nonprofit advocacy groups. The responses included comments submitted by several NIH Institutes. It should be noted that those comments reflected a lack of consensus within NIH regarding support for pursuing such a project.

Several themes emerged in the public comments. Some commenters recommended that the report describe more fully the current cohort studies in the United States and abroad and the ongoing interdisciplinary research that demonstrates the capacity of the research enterprise to collaborate. Some commenters stated that the socioeconomic and cost factors were not sufficiently addressed. Many thought that the study would need considerable ethical oversight because of the complex ethical, privacy, and confidentiality issues that might arise and that the Committee should strengthen its recommendations in this area. Another recurring comment emphasized the importance of the public engagement process. Many commenters also pointed out that the lack of comprehensive Federal protections against genetic discrimination in health insurance and employment was a major concern and would be an impediment to the conduct of the study. Some suggested that the study should not be undertaken unless and until a Federal genetic information nondiscrimination law is in place. Finally, several commenters were unclear about the purpose of the report, and some believed that the tone of the report should be more neutral. All of the comments were carefully considered by the LPS Task Force and were incorporated as appropriate throughout the body of this revised version of the report. This revised report was approved and recommended for transmittal to the HHS Secretary at the SACGHS November 2006 meeting.

II. State of the Science

The scientific possibilities resulting from the recently completed Human Genome Project[9] and the International HapMap Project[10] have prompted a search for methods that will advance the understanding of the relationship between genetics and disease. Now that the sequencing of the human genome is essentially complete, scientists hope to identify the genes and other functional parts of DNA that are correlated with health and disease, elucidate their functions, and gain an understanding of their interactions with each other and with the environment. Experts in a variety of scientific fields are debating the utility and feasibility of an LPS to achieve these goals. Due to the breadth of such initiatives, an interdisciplinary approach is needed that includes geneticists, epidemiologists, toxicologists, social and behavioral scientists, public health experts, dieticians, biostatisticians, information technologists, health providers, ethicists, community representatives, and others. To provide a context for the discussion of an LPS, this chapter begins with a brief review of two recent milestones in genetic research—the Human Genome Project and the International HapMap Project—and a discussion of environmental health and exposure science disciplines. The chapter also reviews methods for identifying the genetic basis of disease, the function of biobanks, and current LPSs and related initiatives that are underway in the United States and abroad.

A. Human Genome Project

Officially launched in 1990 as an international effort, the **Human Genome Project** (HGP) was coordinated in the United States by NHGRI and the U.S. Department of Energy. International HGP partners included the United Kingdom, France, Germany, Japan, and China. Its goal was the complete mapping and understanding of all the genes of a human being (i.e., the human genome).[11]

HGP researchers deciphered the human genome in three major ways: (1) determining the sequence of all the bases in the genome's DNA, (2) making maps that show the locations of genes for major sections of all chromosomes, and (3) producing linkage maps, through which inherited traits, such as those for genetic disease, can be tracked over generations.[12] A draft of the human genome sequence was published in February 2001, and the project was completed in April 2003. A final publication detailing the finished genome sequence appeared in *Nature* in October 2005.[13]

The HGP revealed that there are approximately 20,000 to 25,000 human genes. The completed sequence can identify their locations, providing a resource of detailed information about the structure and organization of the complete set of human genes.

---

[9] See www.ornl.gov/sci/techresources/Human_Genome/home.shtml [accessed February 12, 2007].
[10] International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005. 437(7063):1299-320.
[11] NHGRI Web site. An Overview of the Human Genome Project. See http://www.genome.gov/12011238 [accessed December 19, 2006].
[12] The hereditary material of all multicellular organisms is the double helix of DNA, which contains all genes. DNA, in turn, is made up of four chemical bases, pairs of which form the rungs of the twisted, ladder-shaped DNA molecules. All genes are made up of stretches of these four bases, arranged in different ways and in different lengths. NHGRI Web site. What was the Human Genome Project? See http://www.genome.gov/ [accessed December 19, 2006].
[13] Collins FS 2004. Op. cit.

B. International HapMap Project

The haplotype map (HapMap) is a tool that enables researchers to find genes and genetic variations that affect health and disease.[14] The elucidation of the entire human genome has made possible the development of the HapMap of the human genome.

The DNA sequence of any two people is 99.9 percent identical. The 0.1 percent that varies may greatly affect an individual's disease risk. Sites in the DNA sequence in which individuals differ at a single DNA base are called single nucleotide polymorphisms (SNPs). Sets of SNPs that are near one another on the same chromosome are inherited in blocks. The pattern of SNPs within a block is called a haplotype. Blocks may contain a large number of SNPs, but only a few SNPs may be sufficient to uniquely identify all possible haplotype patterns within a block. The HapMap is a map of these haplotype blocks, and the specific SNPs that identify the block haplotypes are called tag SNPs.

The HapMap will reduce the number of SNPs required to examine the entire genome for association with a phenotype from the 10 million SNPs that exist to roughly 500,000 tag SNPs. This will make genome scan approaches for identifying the regions that contain genes that affect diseases much more efficient and comprehensive, because there will be no need to waste effort by typing more SNPs than necessary, and all regions of the genome can be included.

In addition to its utility for studying genetic associations with disease, the HapMap can be a powerful resource for studying the genetic factors that contribute to variations in response to environmental factors, susceptibility to infection, and the effectiveness of and adverse responses to drugs and vaccines. All such studies will be based on the expectation that there will be higher frequencies of the contributing genetic components in a group of people with a disease or with a particular response to a drug, vaccine, pathogen, or environmental factor than in a group of similar people without the disease or response. Using just the tag SNPs, researchers should be able to find chromosome regions that have different haplotype distributions in the two groups—those with a disease or response and those without. Each region then could be studied in more detail to discover which variants in which genes in the region contribute to the disease or response, leading to more effective preventive, diagnostic, and therapeutic interventions.

C. Environmental Health and Exposure Sciences

A study aimed at advancing the knowledge of associations among genes, environment, and disease status will require the measurement of the effects of environmental factors and exposures on human health over time. Environmental health is "the field of science that studies how the environment influences human biology and the risk of developing disease."[15] Within this field, "environmental exposure" research covers a wide range of human exposures to chemical, physical, and biological agents,[16] as well as lifestyle factors, such as diet and physical activity.

---

[14] NHGRI Web site. International HapMap Project. See http://www.genome.gov/10001688 [accessed December 19, 2006].
[15] *New Technology For Detecting Biological Responses to Environmental Factors*. See http://www.niehs.nih.gov/oc/factsheets/pdf/gei.pdf [accessed December 19, 2006].
[16] Barr DB. Human exposure science: a field of growing importance. *J Expo Sci Environ Epidemiol* 2006. 16(6):473.

Environmental exposures can be quantified through three main mechanisms: (1) environmental or personal measurements, (2) biological sampling, and (3) indirect sampling.[17] Environmental and personal exposures are typically measured via sensors that detect the presence and concentration of specific chemicals or pollutants in an individual's environment (e.g., home, work, school). Biological samples such as blood, urine, hair, saliva, and nail clippings also can be assayed to detect the presence of potentially harmful chemicals, as well as to measure secondary side effects of exposure. Examples of indirect measurements include subject questionnaires and activity and diet diaries that are used to establish baseline behaviors and exposures in a qualitative manner.

The National Institute of Environmental and Health Sciences (NIEHS) is conducting several projects geared toward determining linkages between environment and disease status. The **Sister Study** is investigating the environmental and genetic causes of breast cancer by examining 50,000 sisters of women diagnosed with breast cancer.[18] Initiated in 2004, the Sister Study is focused on determining correlations among genes, environmental exposure (particularly, common exposures in the home, workplace, and community and certain personal care and household products), and susceptibility to breast cancer.[19] It involves questionnaires as well as the analysis of blood, urine, and other personal body samples and household dust samples. NIEHS also conducts and funds research that evaluates the role of the environment in Parkinson's, Alzheimer's, and other neurodegenerative diseases. Typically, these studies have focused on long-term exposure to environmental agents and disease status (e.g., aluminum and the onset of Alzheimer's). NIEHS also is promoting research in biomarker development for increased susceptibility to environmental agents, additional epidemiological studies to identify chemical combinations associated with increased disease risk, development of models for long-term exposure and neurodegenerative disease risk, and other related areas.[20] The goal of such research is to develop better treatments, as well as to explore potential preventive measures for these debilitating diseases. NIEHS also funds research and community-based programs to identify and resolve the environmental causes of health disparities. For example, because minority and low-income communities are more likely to be exposed to environmental pollution, they may have an increased susceptibility to diseases such as cancer, cardiovascular disease, diabetes, and asthma.[21] Therefore, much of the research in this area seeks to identify practical solutions to improve the overall environment (e.g., use of household air purifiers), while also preventing disease.

The NIH **Genes and Environment Initiative** (GEI) seeks to combine the rapidly emerging field of genome sciences with studies in exposure biology to determine gene variants and environmental factors that may result in increased risk for both acute and chronic diseases. Approximately 30 percent of the GEI's proposed fiscal year 2007 budget of $40 million will be specifically targeted to improving and expanding the tools available for determining environmental and chemical exposures.[22]

---

[17] Ozkaynak H, Whyatt RM, Needham LL, Akland G, Quackenboss, J. Exposure assessment implications for the design and implementation of the National Children's Study. *Environ Health Perspect* 2005. 113:1108-15.
[18] *What is Environmental Health?* See http://www.niehs.nih.gov/oc/factsheets/pdf/e-health.pdf [accessed December 19, 2006].
[19] Ibid.
[20] *Research Initiatives: Parkinson's, Alzheimer's, and Other Neurodegenerative Diseases*. See http://www.niehs.nih.gov/external/resinits/ri-17.htm [accessed December 19, 2006].
[21] *Health Disparities Research*. See http://www.niehs.nih.gov/oc/factsheets/disparity/home.htm [accessed December 19, 2006].
[22] *The NIH Genes and Environment Initiative (GEI)*. See http://www.gei.nih.gov/index.asp [accessed December 19, 2006].

D. Methods for Identifying the Genetic Basis of Disease

As noted earlier, any two individuals differ in their genetic makeup by only about 0.1 percent.[23] Characterizing this small fraction of variation is one of the most pressing goals for scientists who are trying to unravel the influence of genes on human health and disease. Personalized medicine and major advances in public health are expected to result from understanding the variation in DNA that makes humans different from one another in their susceptibility to disease; their physiological, mental, and emotional responses to physical, behavioral, and social environmental exposures; and their responses to medicines. Understanding gene-environment interactions is particularly important, because recent epidemics of chronic diseases have developed over the span of a few generations. Although this is far too short a period for the genome to change dramatically, it is sufficient time for substantial environmental changes to occur and to have adverse effects on individuals who are genetically predisposed to respond poorly to environmental challenges. Advances in public health and personalized medicine will be possible if it is known where in the genome the variations lie, how these variations differ between healthy and sick people, and how individuals with particular variants of DNA are affected by environmental factors.

Classic genetic research methods involving "family linkage analysis" have been used in many instances to identify the genetic basis of disorders that exhibit the rules of Mendelian inheritance (i.e., dominant v. recessive traits). Linkage analysis examines the patterns of cotransmission of genetic markers and diseases within families through a comparison of affected and nonaffected family members to identify chromosomal regions that may contain disease-related genes. Efforts then can be made to identify candidate genes and understand their roles in disease.

Many of the common diseases that affect the U.S. population, however, are complex and multifactorial—that is, they are caused by a complex interplay of multiple genes and environmental factors—factors in the physical environment as well as in the behavioral and social environments.[24] In other words, although the presence of one or more genetic variants contributes to the underlying cause of disease, it is the body's exposure to environmental factors, including behavioral and social influences, that helps determine whether and how genetic variants contribute to disease manifestation. These markers of variation in biobehavioral reactivity might produce cleaner intermediate phenotypic markers of early disease vulnerability.[25]

A key component of an LPS is the elucidation of genetic associations with disease, as well as gene-gene and gene-environment interactions. "Genetic association studies" assess correlations between previously identified gene variants and phenotypic differences (i.e., the presence or absence of disease) on a larger, population-based scale, rather than just within families.[26] These larger-scale studies are now more feasible because of recent advances in identifying genetic variants in humans, including more rapid and cost-effective gene sequencing technologies, the mapping of SNPs within the genome,[27,28] and the ability to compare groups of genetic loci simultaneously in a single test.

---

[23] Sachidanandam R et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001. 409(6822):928-33.
[24] Palmert MR, Hirschhorn JN. Genetic approaches to stature, pubertal timing, and other complex traits. *Mol Genet Metab* 2003. 80(1-2):1-10.
[25] Moffitt TE, Caspi A, Rutter M. Strategy for investigating interactions between measured genes and measured environments. *Arch Gen Psychiatry* 2005. 62(5):473-81.
[26] Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001. 2(2):91-9.
[27] Mullikin JC et al. An SNP map of human chromosome 22. *Nature* 2001. 407(6803):516-20.
[28] Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 2000. 407(6803):513-6.

The identification of genetic associations and interactions is only the first step on the continuum from gene discovery to clinical and public health applications. The evaluation of the validity and utility of genetic information in clinical trials, observational clinical settings, and population settings will rely on systematic integration of data across human genome epidemiology studies, such as the Human Genome Epidemiology Network (HuGENet), to assess the impact of human genome variation on population health.[29] Additionally, infrastructure is needed to translate the discoveries of genome-based research into effective interventions to improve population health. Global efforts are underway to address this challenge, including the Genome-based Research and Population Health International Network.[30] An important part of initiating a large population-based study is determining the number of patients that must be enrolled to obtain relevant information on the associations among genes, environment, and disease. The study population must be large enough to capture genetic variants that not only do not express complete dominance or recessiveness but also do associate with disease status, as well as detecting the multiple variable combinations between genes and environment that interact to contribute to disease status.

Population-based genetic association studies rely on samples obtained from both affected and unaffected individuals. For these studies, the frequency with which certain alleles are present in each of these groups is tested for association with a disease. A common approach is to use biological information about the molecular pathology of the disease to guide the selection of candidate genes for such testing.[31] Several sampling strategies can be used in association studies, including case-control studies and prospective cohorts. The case-control method typically has been used, in which genetic and environmental data are collected from persons with specific diseases or conditions and compared with the data of those free of disease. Although case-control studies are of great value in suggesting potential etiologic factors, they cannot provide information on predictive biobehavioral markers, are prone to biases related to case ascertainment, and often provide incomplete or biased assessment of risk modifiers or gene-environment interactions.[32]

In comparison, prospective cohort studies of genes and environment enroll individuals prior to disease onset and prospectively collect environmental and biobehavioral marker data, allowing for the examination of contributing nongenetic and genetic factors in disease. Prospective cohort studies must enroll more individuals than case-control studies to ensure that a sufficient number of affected persons within the study population eventually develop the disease of interest. By increasing the sample size, scientists increase the study's power to detect subtle differences among individuals. Prospective cohort LPSs are designed to find significant associations among genetic variants, traits, and environmental exposures. Collecting phenotypic and environmental information in a standardized and unbiased manner is crucial to such efforts. However, it is even more challenging to collect indices of variation in biobehavioral reactivity that might produce cleaner intermediate phenotypic markers of early disease vulnerability.

---

[29] Khoury MJ, Little J, Burke W. Human genome epidemiology: scope and strategies. In: Khoury MJ, Burke W, Little J, eds.. *Human Genome Epidemiology*. New York, NY: Oxford Press, 2003. p. 8.

[30] Burke W, Khoury MJ, Stewart A, Zimmern RL; Bellagio Group. The path from genome-based research to population health: development of an international public health genomics network. *Genet Med* 2006. 8(7):451-8.

[31] *Genetic Association Studies-Overview*. See http://slack.ser.man.ac.uk/theory/association_overview.html [accessed December 19, 2006].

[32] *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel.* Op. Cit.

E. Role of Biobanks in Large Population Studies

A biobank is "a stored collection of genetic samples in the form of blood or tissue that can be linked with medical and genealogical or lifestyle information from a specific population, gathered using a process of generalized consent."[33] Large-scale cohort studies of different diseases involve the enrollment of a large number of individuals who are willing to provide access to their specimens and medical information and the collection of information about their environmental exposures (including physical, social, and behavioral information). Data stored in databases and specimens stored in repositories (sometimes variously referred to as "biobanks," "biorepositories," or "genebanks") are accessed by qualified investigators for specific research purposes (e.g., studies of specific diseases or genes of interest).

In recent years, many biobanks have been initiated in parallel with LPSs to facilitate the simultaneous analysis of individual genetic material, disease status, and environmental exposures. In some cases, the biobank is directly linked to a study with a predetermined goal, such as identifying the genes causing a specific disease. In others, the biobank literally serves as a repository of genetic material, patient exposure data, and medical history information that is available as a resource to researchers who request samples for the study of a particular disease. The characteristics of biobanks, such as participant population, age, size, ethnicity, and environmental exposures, vary widely.

The types of analyses to be used and the hypotheses to be addressed determine what kinds of biological sample(s) will be collected in a biobank. At present, the types of analyses commonly used include the genotyping of specific biomarkers; transcript profiling, the measurement of how a gene or a set of genes is expressed in tissue samples; gene quantification, the analysis of how altered copy numbers of genes and chromosomes differ between normal and malignant tissues; and proteomic analysis, the analysis of protein expression and modification in response to genetic and environmental factors.[34]

For example, genotyping requires a supply of DNA that can be obtained from blood samples or a swab of cells from the lining of the mouth cavity, as well as other methods, provided that the samples contain cells with intact nuclei. The study of changes in genes that occur during life, such as those in tumor tissues, can utilize transcript profiling and gene quantification, both of which require samples of nucleated cells from the relevant tumor and tissues. Proteomic analysis, however, does not require nucleated cell samples; rather, any body fluid or specimen related to the disease process can be collected. In many cases, biobank samples are stored in a manner that preserves the DNA, gene transcripts such as ribonucleic acid, or proteins for study at a later date. However, some biobanks have used methods that preserve the living samples in a cell culture system, which provides a renewable and permanent source of the cells and materials for analysis.

---

[33] Austin MA, Harding S, McElroy C. Genebanks: a comparison of eight proposed international genetic databases. *Community Genet* 2003. 6(1):37-45.

[34] Jonsson L, Landegren U. Storing and using biobanks for research. In: Hansson MG, ed. *The Use of Human Biobanks. Ethical, Social, Economical and Legal Aspects. Report I: Ethical, Social, Economical and Legal Aspects.* Uppsala, Sweden: Center for Bioethics at Karolinska Institutet and Uppsala University, 2004. See www.bioethics.uu.se/biobanks-report.html [accessed December 19, 2006].

F. Current Population-Based Gene-Environment Cohorts and Other Initiatives in the United States

U.S. investigators already have conducted or are conducting many smaller-scale studies to detect associations among environmental factors, genetic and biobehavioral markers, and disease. Although these studies and initiatives are important and informative, it is not clear whether any individual study has the statistical power needed to definitively identify associations of the magnitude anticipated for many diseases. The following efforts are described below:

- Women's Health Initiative[35]
- Framingham Heart Study[36]
- National Health and Nutrition Examination Survey[37]
- National Cancer Institute Multiethnic/Minority Cohort Study of Diet and Cancer[38]
- National Children's Study[39]
- Marshfield Clinic Personalized Medicine Research Project[40]
- Veterans Administration Genomic Medicine Program[41]
- Genes and Environment Initiative[42]
- Genetic Association Information Network[43]
- Human Genome Epidemiology Network[44]

The **Women's Health Initiative** (WHI) is a long-term national health study that focuses on strategies for preventing heart disease, breast and colorectal cancer, and bone fractures in postmenopausal women. The 15-year project, sponsored by the National Heart, Lung, and Blood Institute (NHLBI), involves more than 161,000 women ages 50 to 79 years. The study is one of several federally funded research projects for which participants provided blood samples years ago, which creates a rich resource for matching genetic variations with disease on a large scale.

With support from NIH, and in collaboration with the biotechnology company Amgen Inc., Brigham and Women's Hospital is organizing a study of the DNA collected more than a decade ago from 28,000 WHI participants.[45] They will look for differences between those who have developed serious illness and those who have remained healthy. The results of the new project, called the Women's Genome Health Study, are expected to help physicians predict a woman's risk for disease and tailor effective treatments. Because of the large cost, the study involves corporate, Federal Government, and academic players who are working together to translate existing genetics research into practical medical knowledge. Results describing associations between genes and diseases will be posted in an NIH database and made available to any

---

[35] See http://www.nhlbi.nih.gov/whi/.
[36] See http://www.framingham.com/heart/.
[37] See http://www.cdc.gov/nchs/nhanes.htm.
[38] See http://epi.grants.cancer.gov/resport/multiethnic.html.
[39] See http://www.nationalchildrensstudy.gov/.
[40] See http://www.marshfieldclinic.org/chg/pages/default.aspx?page=chg_pers_med_res_pri.
[41] See http://www4.od.nih.gov/oba/sacghs/meetings/feb2005/fihn.pdf.
[42] See http://www.gei.nih.gov/.
[43] See http://www.fnih/org/gain/gain_home.shtml.
[44] See http://www.cdc.gov/genomics/hugenet/default.htm.
[45] Winslow R, Regalado A. Tying Diseases to DNA in Thousands of Women. *The Wall Street Journal*, October 24, 2006.

scientist interested in conducting additional research. Once in the public domain, the genetic associations cannot be patented, but scientists can use the data as a starting point to develop diagnostic tests and treatments that could be patented.

The **Framingham Heart Study** was launched in 1948 as a prospective cohort study of cardiovascular disease, the leading cause of death and illness in the United States. The study continues to be supported by NHLBI and now involves three generations of U.S. citizens. The original study prospectively examined the cardiovascular health of more than 5,000 adults in Framingham, Massachusetts, and was eventually extended to the children and grandchildren of the original participants. The study followed participants by providing extensive medical examinations, blood tests, and other measures of health status. The results of this longitudinal study established the role of risk factors such as elevated cholesterol, high blood pressure, and diabetes in cardiovascular disease risk. These findings dramatically altered the treatment of patients with cardiovascular disease and increased public education about the risk factors for cardiovascular disease. A new phase of the study involves examining the extent to which genetic factors relate to cardiovascular disease and its risk factors.

The **National Health and Nutrition Examination Survey** (NHANES), sponsored by CDC's National Center for Health Statistics, has collected more than 35 years of survey and laboratory data that have resulted in significant public health improvements. NHANES is unique as the only national survey that collects biologic specimens from a representative sample of the U.S. population,[46] although on a much smaller scale than the proposed LPS discussed in this report. "The goals of NHANES are to monitor the nation's health by estimating the number and percentage of persons in the U.S. population and in designated subgroups with selected diseases and risk factors; monitoring trends in the prevalence, awareness, treatment, and control of selected diseases; monitoring trends in risk behaviors and environmental exposures; analyzing risk factors for selected diseases; studying the relationship between diet, nutrition, and health; exploring emerging public health issues and new technologies; and establishing a national probability sample of genetic material for future genetic research."[47] As part of NHANES III (1991-94), blood specimens were collected from more than 7,000 participants age 12 years and older.

The use of these samples for genetic research was not specifically included in the informed consent process, and confidentiality concerns arose regarding linking genetic test results with other NHANES data. After careful deliberation, a plan was approved for using these specimens for genetic research, linked with other NHANES data, under closely controlled conditions to ensure that the confidentiality of the research participants' identities is maintained.

The ethnic diversity of the National Cancer Institute's (NCI) **Multiethnic/Minority Cohort Study of Diet and Cancer** (MMCS) provides a unique dimension to the search for the causes of cancer and the means for cancer prevention. This study is harnessing recent developments in cancer research to explore the relationships between lifestyle and genetic susceptibility to cancer risk in 215,000 people from 5 distinct ethnic groups.[48] The Japanese, African American, Latino, Native Hawaiian, and Caucasian men and women chosen for the study differ in their relative rates of cancer and in their dietary habits. By comparing these

---

[46] McQuillan GM, Porter KS, Agelli M, Kington R. Consent for genetic research in a general population: the NHANES experience**.** *Genet Med* 2003. 5(1):35-42.
[47] Ibid.
[48] Kolonel, LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat Rev Cancer* 2004. 4(7):519-27.

diverse populations, it may be possible to determine the relative contributions of inborn and environmental risks. Investigators are focusing on breast, prostate, and colorectal cancer. In addition to examining environmental factors (e.g., tobacco use, diet, obesity, sun exposure, occupational exposures), the study began a prospective collection of biological specimens from a substantial portion of the cohort, which should yield a biorepository of 100,000 members within 3 to 4 years. Investigators are exploring the interactions of variations in genes, diet, and lifestyle in causing cancer, with the hope of developing improved diagnostic, therapeutic, and preventive interventions for broad use across the world's population.

The **National Children's Study** (NCS) is designed to focus on the influence of environmental exposures on childhood disease and development. As part of the Children's Health Act of 2000 (Public Law 106-310), Congress authorized a longitudinal cohort study of 100,000 children. The Act assigned NICHD to lead a consortium of Federal agencies to plan, develop, and implement this prospective study from birth to adulthood to evaluate the effects of chronic and intermittent exposures on child health and human development and investigate basic mechanisms of developmental disorders and environmental factors that influence health and development. The NCS established seven study centers and a coordinating center and is preparing for implementation this year.

The NCS convened 21 working groups, including more than 2,000 representatives of the academic and lay communities from around the country, to facilitate the development of the study plan, specifically, defining research hypotheses, exposure assessment methods, and outcome measures. A wide array of mechanisms were used, including workshops, public meetings, postings for public comment, pilot studies, focus groups, and the establishment of the National Children's Study Advisory Committee. Data collection is planned to begin prior to participants' births and continue until they reach at least the age of 21 years. Subjects would take part in approximately 15 in-person visits with a local research team, parents/guardians would complete questionnaires every 3 months until age 5 years and annually thereafter, and biological samples would be collected for genetic analysis. Samples of the air, water, soil, and dust from the children's environments would be collected regularly. To capture U.S. ethnic, social, and geographic diversity, the NCS is enrolling a representative sample of families from 105 locations in the United States. The study is estimated to cost $2.7 billion over 25 years. Although funding for the project was not included in the President's fiscal year 2007 budget, the House and Senate appropriations committees expressed strong support for continuation of the study.[49]

The goal of the **Marshfield Clinic Personalized Medicine Research Project**[50] is to translate genetic data into clinically relevant knowledge about disease that will enhance patient care. It is based out of the Marshfield Clinic Research Foundation, a large group of private clinics, and uses the Marshfield Epidemiologic Study Area (MESA) in central Wisconsin, which has a longstanding electronic health records program. The project requires completion of a written informed consent document and a 30-minute questionnaire, as well as extraction of a blood sample that will be used for DNA analyses. The project expects to enroll 40,000 subjects age 18 years and older. In 2004 the objectives of Phase I were to educate, inform, and consult with the MESA population and communities concerning potential studies, establish the DNA component of the personalized medicine database, and build the bioinformatics tools to securely

---

[49] Report of the Committee on Appropriations, U.S. Senate on S. 3708. Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriation Bill, 2007. Report 109-287, p. 165.

[50] Marshfield Clinic Research Foundation. *Personalized Medicine Research Project.* See http://www.marshfieldclinic.org/chg/pages/default.aspx?page=chg_pers_med_res_prj [accessed December 14, 2006].

store and analyze genotypic and phenotypic data. The project created an ethics and security advisory board, a scientific advisory board, and a community advisory group. More than 18,000 residents participated. The objectives of Phase II are to create the phenotypic database, establish the scientific and administrative infrastructure to support genetic mapping of the DNA and the initial discovery projects, and genotype a sufficient portion of the genetic material to support these discovery projects. The objectives of Phase III are to expand the discovery projects, complete genotyping of the genetic database, and expand physician/health care provider education and community consultation. Physicians then could diagnose genetically influenced problems, prescribe personal preventive measures, and select the most effective medications that are least likely to cause adverse reactions.

VA is developing a new **Genomic Medicine Program** (GMP) using the Department's existing and extensive health care databases. VA has established the GMP Advisory Committee (GMPAC), which convened for the first time on October 16, 2006, to advise VA on emerging issues in genomic medicine. The GMPAC will help VA establish policies for this new large-scale project by using genetic information to optimize the medical care of veterans by enhancing the development of tests and treatments for diseases that are particularly relevant to veterans and developing genetic assessments that will enable pharmacogenomic customization. The GMPAC is expected to meet three times annually to help lay the groundwork for the future development of a comprehensive genomic medicine program for VA by recommending policies to govern the collection and using both genetic and other medical information for medical care and research.

In February 2006 HHS announced two large-scale efforts: (1) the **Genes and Environment Initiative** (described previously), a research effort at NIH to combine a type of genetic analysis and environmental technology development to increase understanding of the causes of common diseases, and (2) the **Genetic Association Information Network**, a public-private partnership among NIH, the Foundation for the National Institutes of Health, Pfizer Inc., and Affymetrix, Inc., to conduct laboratory studies to determine genetic contributions to five common diseases.[51] These projects will rely on existing cohorts using a case-control method; that is, they will study people who have a disease and compare them with those who do not.

NIH is also developing a policy to facilitate access to data resulting from the NIH-funded **Genome-Wide Association Studies**.[52] These initiatives use recently developed technologies to look for genetic differences between people with specific illnesses, such as diabetes or heart disease, and healthy individuals. It is hoped that these differences will point to genetic risk factors for the development or progression of disease and, ultimately, to better diagnostic tools and highly effective treatments. NIH is in the process of standardizing the policies by which the results of these studies will be made available to researchers.

As noted previously, the United States also participates in a global effort to integrate epidemiologic evidence on genes and population health from existing and planned epidemiologic studies of all types. The **Human Genome Epidemiology Network** (HuGENet) is committed to the development and integration of the knowledge base on human genetic variants and health. HuGENet maintains a database of published genetic association studies and is working with journals to encourage publication of negative results to

---

[51] U.S. Department of Health and Human Services. *NIH News* Press Release. Two NIH Initiatives Launch Intensive Efforts to Determine Genetic and Environmental Roots of Common Diseases. See http://www.niehs.nih.gov/oc/news/gei.htm [accessed December 2006].

[52] NIH Web site. Genome-Wide Association Studies. See http://grants.nih.gov/grants/gwas/index.htm [accessed December 19, 2006].

address the well-known publication bias in peer-reviewed scientific literature. HuGENet guidelines are available for conducting systematic reviews and meta-analyses of epidemiologic evidence for gene-disease association in collaboration with several participating journals. Networks of HuGENet investigators are sharing best practices and methods for the analysis of associations between genetic variation and common disease, and HuGENet is planning to develop guidelines for the consistent reporting of results from genetic epidemiologic studies.[53]

G. International Population-Based Gene-Environment Cohorts and Other Initiatives

In addition to the individual U.S. projects discussed above, some other countries are already conducting national population research projects. Large-scale cohort projects are under discussion or underway in the United Kingdom, Iceland, Estonia, Germany, Canada, Taiwan, Japan, and other countries (see Appendix D). These efforts capitalize on genome-wide scanning for SNPs and haplotypes that could provide population-based information about associations between common polymorphisms and common diseases.

These studies, which include biobanks for the long-term storage of biological samples, differ in their design and approach to LPSs, with population diversity being a major influence on design. In Iceland, for example, where the population is homogenous and extensive genealogies are maintained by the public, a large-scale version of linkage analysis is possible. In LPSs of more diverse populations, such as those found in the United Kingdom and Estonia, association studies will be used to examine the population distribution of genetic variants and their association with disease.

These projects also differ in their ultimate objectives. In the case of UK Biobank, the goal is an epidemiological analysis of risk factors that contribute to disease, whereas the goal of the Estonian Genome Project is to maintain genetic information in a database as a resource for public health and biomedical research, as well as for the clinical management of participants. Biobank Japan aims to develop tools for personalized medicine, choosing medical procedures and drugs based on patients' genetic profiles. Ultimately, the designers of biobanks anticipate that the biobanks and associated LPSs will become part of the research infrastructure from which future discoveries in medicine and public health can be derived. Because they are considered research tools and part of these countries' research infrastructures, these biobanks have been funded primarily by governments and nonprofit organizations. In some cases, the exclusive rights to the development of therapies and diagnostics have been assigned to private companies that are funding the research, such as deCODEgenetics, Inc., in Iceland.

International experiences with designing and implementing these efforts have demonstrated the need to proceed with careful deliberation and public input. In addition, some of these countries have infrastructures that more easily allow for uniform data collection of health information on large numbers of the population (e.g., universal health care records).

---

[53] Ioannidis JP et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006. 38(1):3-5.

III. Policy Issues Associated With a New Large Population Study of Genes, Environment, and Common Diseases

The SACGHS Task Force on Large Population Studies organized presentations by experts to the full Committee during its 2005 and 2006 meetings to gather information about the issues associated with undertaking a new LPS in this country and to facilitate the identification of approaches to address them. The issues that were identified fall into four broad policy areas: (1) research policy; (2) research logistics; (3) regulations and ethics; and (4) public health, social, and economic implications.

Each of these four areas raises a number of specific policy issues that are discussed in this report. Some are issues that must be addressed to assist in deciding whether such a proposed LPS can or should be undertaken in the United States at this time; other issues might be addressed only once such a decision has been made. Some decisions relate to downstream issues (e.g., the consequences of the knowledge that would be generated by this large-scale project and the impact of its findings on individuals, groups, and society). In this chapter, following the discussion of these issues within each of the four broad areas, the Committee sets out recommended approaches for addressing them.

In addition to the recommendations in each of the broad areas, the Committee identified a single overarching policy recommendation to guide the HHS Secretary as he considers the possibility of a new LPS in the United States:

As part of the process for determining whether to undertake such a large-scale research project, the HHS Secretary should initiate a thorough consideration of the full range of policy issues outlined in this report. The HHS Secretary should consult and engage the full range of potential partners for such a project during this decisionmaking process, including the public at large, the full scientific community, a wide spectrum of Government agencies and policymakers, and the private sector.

This recommendation is intended to highlight the fact that the decision to pursue a new LPS in the United States will require thorough consideration of the many and varied individual policy issues discussed in this report.

A. Research Policy

Based on the input provided to NIH by the scientific community[54] and on testimony to SACGHS during public meetings, it is clear that experts are divided about the wisdom of proceeding with a new LPS of the magnitude outlined in the *Design Considerations* report. For example, some experts assert that such a large study will be uniquely capable of producing data of sufficient power to elucidate the roles of genetic variation and environmental factors in common diseases. Other experts suggest that similar statistical power could be achieved with lower costs by combining data from existing cohort studies (e.g., through methods described in Chapter II). Some contend that the data collected through such a project may not lead to more understanding of common diseases or population health benefits unless they are hypothesis-driven. Other experts state that such a study should *not* be hypothesis driven but, rather, should serve as a data resource for researchers to mine. Other experts have focused their concerns on the significant costs of such a study and on the direct and indirect effects that such a project might have on funding in other research areas. There are also divided views about the opportunities and challenges associated with structuring the project as a collaboration with international efforts and with the involvement of private sector expertise and resources.

The fundamental question policymakers will need to address is whether the development of a new LPS is the only or most effective way for the United States to advance understanding of the interactions between genes, environments, behaviors, and their interactions and common diseases. In the next section, the views of the experts on the value of and need for a new LPS are presented in more detail.

1. Arguments Favoring a Large Population Study

Proponents assert that a new LPS would have a number of advantages. It would be prospectively designed and would include detailed plans for data and specimen collection. Genetic data and environmental data on exposure, lifestyle, and behaviors would be collected using a standardized protocol and the most advanced technologies for data and specimen collection and storage. A broad informed consent process could be developed to cover all research needs, and the participants would be recruited from diverse populations and age ranges.

Perhaps its major advantage, according to advocates, is the increased statistical power it would have over traditional linkage studies and small association studies,[55] which have had little success in identifying genes involved in common complex diseases. When associations have been found, most attempts to replicate them have failed.[56] According to this view, small association studies lack the power to identify the multifactorial bases of common diseases, often causing false-positive associations, and to confirm initial findings through replication studies. Proponents say that the scale of an LPS would allow for the confirmation or refutation of existing hypotheses that would otherwise remain uncertain due to the constraints of the data currently available in smaller association studies.

---

[54] National Institutes of Health (2005). *Summary of Public Responses to Request for Information: Design and Implementation of a Large-Scale Prospective Cohort Study of Genetic and Environmental Influences on Common Diseases.* See http://www.genome.gov/Pages/About/OD/ReportsPublications/ResponsestoRFINot-OD-04-041.pdf [accessed December 18, 2006].

[55] Rosand J, Altshuler D. Human genome sequence variation and the search for genes influencing stroke. *Stroke* 2003. 34(10):2512-7.

[56] Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002. 4(2):45-61.

One alternative to a completely new LPS would be to integrate existing prospective cohort studies, such as the Framingham Heart Study, into the design of the project. However, this approach could present several challenges.[57] Existing studies are limited in terms of the particular characteristics and phenotypes studied and the environmental exposures measured. Pooling of data could be difficult because of ascertainment biases, differing sampling strategies, missing data on critical variables, and survivor bias. In addition, existing studies have a greater number of older participants and would not represent the full age range. Also, combining existing cohorts would not necessarily reflect the ethnic heterogeneity of the U.S. population.

The strengths of the design of an LPS include the replication of associations and the estimation of their magnitude, consistency, and temporality. Another benefit is that, ultimately, the same population of participants, along with their genetic, medical, and exposure data, can be used to study the etiology of many common diseases. It may be difficult to obtain these benefits from existing cohort studies, particularly because most such studies focus on only one or a few common diseases. Conducting multiple, individual case-control studies ultimately may require the same number of participants and resources. In an LPS, the links between various common diseases, such as hypertension and obesity, could be studied.

A collaboratively planned and implemented LPS within the United States may increase the participation of populations that are currently underrepresented in research and allow for enhanced subsequent analysis of their genetic variations and disease risks. This includes populations that are not fully represented in current international studies, such as African Americans, Latinos, Asian Americans, and Native Americans. In addition, there would be a need to determine the environmental or exposure risks associated with disease among U.S. citizens requiring the collection of detailed behavioral, exposure, and sociocultural data (e.g., poverty, education, diet) on U.S. populations, as well as information about behavior and social conditions. Finally, although many existing biobanks intend to make their data available to researchers outside their countries, it is possible that access to international data and biological materials would be limited.

As discussed more fully later in this chapter, the potential benefits of mounting a new LPS would need to be balanced against the expected cost of the project, its implications for other areas and types of research, and the willingness of subjects to tolerate the questions and examinations required for participation.

2. Arguments Favoring the Pooling of Existing Cohort Studies and Biobanks

Some experts point out that increasing the number of individuals in a cohort may not provide sufficient power to detect the gene-environment interactions correlated with disease because of the large number of strata present in multivariable conditions, such as complex diseases. Thus, these experts question whether a large project will facilitate the discovery of population health benefits[58] or only suggest some associations that then would have to be pursued in smaller, more targeted studies. Some have suggested that pooling data and samples across multiple ongoing LPSs (e.g., those listed in Chapter II) could be a reasonable alternative or addition to mounting a new LPS. Furthermore, they argue that this approach might be necessary to achieve the statistical power to detect gene-environment interactions, thereby reducing false-negative findings. Additionally, validation of initial findings across cohort studies would help minimize false-positive associations.[59]

---

[57] National Institutes of Health (2005). *Summary of Public Responses to Request for Information: Design and Implementation of a Large-Scale Prospective Cohort Study of Genetic and Environmental Influences on Common Diseases*. Op. Cit.
[58] Khoury M. The case for a global human genome epidemiology initiative. *Nat Genet* 2004. 36(10):1027-8.
[59] Ibid.

As another alternative, existing cohorts could be mined more extensively or expanded to address the same experimental questions that a large project would address. The pooling of these existing cohorts might save time and/or money because researchers could build on existing DNA repositories, environmental exposure assessments, and infrastructure, rather than leaving these cohorts' genetic analyses relatively underfunded and their samples unanalyzed. Many in the research community who commented on this report believe there has been a failure to access and use information that could be readily tapped from existing cohorts and specimens and that there is a risk of failing to invest financially in these resources in the future. Given the methodological limitations of the proposed large cohort project and the lack of good hypotheses and instruments, many argue that existing cohort studies should be mined before attempting to embark on the proposed LPS. Existing cohorts could be used to refine and pilot the processes of standardizing exposure, phenotype determination, and subject recruitment methods prior to the full-scale launch of an LPS.

Another suggested alternative to a single LPS is the development of a worldwide collaboration of existing large population biobanks. Because genetic markers do not change over time, the use of existing DNA repositories could reduce costs and allow genetic discovery projects to begin almost immediately. By using existing collections, resources could be dedicated to genotyping, data set creation, and statistical analyses, rather than cohort assembly. This type of collaboration would increase the statistical power to detect gene-environment interactions and could reduce errors in which false-positive associations are found. Such collaborations and their extremely large samples would be particularly valuable for allowing the separation of cases into independent groups to allow for the replication and confirmation of findings.

Given current restricted budgets for biomedical research overall and the difficulty of ensuring adequate support for existing cohort studies, some in the scientific community question the wisdom of beginning a new LPS at this time.

3. Arguments Against a Large Population Study

Experts opposing the mounting of a new LPS put forward a number of arguments. First, they point out that an LPS may not clarify the interactions among genes, environments, and phenotypes.[60] Pembrey and colleagues argue that, "Quite simply, the proper study of multifactorial traits or disorders demands the analysis of all the likely multiple factors in the same subjects over time."[61] Even the widest possible study of social, behavioral, physiological, and physical exposure factors still would have limitations, because by the time it is completed and the environmental influences are understood, the environments would have "moved on" (i.e., the environmental factors that were studied may not have the same degree of social relevance or public health importance).

In addition, LPSs do not necessarily hypothesize a priori ideas of the underlying mechanisms of disease, and therefore, questionable results might be obtained by screening large numbers of potential etiologic factors for correlations with multiple diseases. In case-control studies, a particular disease is identified for study, and correlations with exposure or genetic variants are sought. Some in the scientific community worry that if the project does not involve a priori hypotheses about environment-gene-behavior, gene-disease-behavior, or environment-disease-behavior interactions, the results may be suspect.

---

[60] Foster MW, Sharp RR. Will investments in large-scale prospective cohorts and biobanks limit our ability to discover weaker, less common genetic and environmental contributors to complex diseases? *Environ Health Perspect* 2005. 113(2):119-22.
[61] Pembrey M; ALSPAC Study Team. The Avon Longitudinal Study of Parents and Children (ALSPAC): a resource for genetic epidemiology. *Eur J Endocrinol* 2004. 151 Suppl 3:U125-9.

It also is possible that too few clinically relevant events may accrue in an LPS, making it difficult to detect genetic or environmental factors of small effect. Because common diseases are thought to be genetically and environmentally heterogeneous, the multiple influences underlying disease will require sample sizes large enough, and analytic techniques sensitive enough, to detect the multiple combinations of contributing genetic and environmental factors to the same phenotypic disease. Some scientists are not convinced that the necessary experience, infrastructure, and scientific culture exist in which to responsibly carry out a large cohort study. They argue that the genetic science of common complex diseases is simply not yet mature enough to launch such a costly initiative.

4. Effects on Other Areas of Science

Since an LPS would involve as many as 1 million participants and be carried out over a decade or more, with additional time required for data analysis, the level of funding needed to develop and maintain the study would be enormous. The costs of the project would include identification of, recruitment of, and clinical followup with study participants; storage and security of biological samples and associated clinical data; sample processing and genotyping for all study participants; and dissemination and general management of samples and data.

Although there are no publicly available estimates of the project's cost, the figure could be as high as $3 billion or perhaps higher.[62] Annual costs could be significant and have implications for other critical research and training programs. Some members of the scientific community have expressed concern about the impact of such a large allocation during flat funding periods and argue that a large project should be undertaken only if it is funded through sources that do not compromise investigator-initiated projects. A number of public commenters also pointed to the potential cost of an LPS and the possible effects on the funding available for other biological, behavioral, and social sciences. They urge careful consideration of the cost of a new project versus other priorities within HHS or the biomedical research community. Depending on the funding mechanisms, broad public support for Federal Government funding would be needed. However, some have questioned whether it is appropriate for the Federal Government (and HHS in particular) to be the sole sponsor of such a project. Questions for consideration include: Does the Government have the necessary financial and technological infrastructure to carry out this type of effort? Should the private sector also be expected to invest funds and technical resources to complete some of the work? What are the implications of extensive public-private partnerships and collaboration? Who would control access to the knowledge generated, and would commercialization be a concern if the project were organized as a public-private partnership?

5. Importance of Collaborative Governance

Regardless of the sources of project financing, a number of Federal Government agencies would undoubtedly contribute expertise to such a study, perhaps in partnership with academia and private industry. Therefore, a cross-cutting team approach to governance would be needed at the Federal study organizational level. Given the interdisciplinary nature of the proposed study's scope, SACGHS believes the HHS Secretary should establish a highly collaborative model of project leadership and management in multiple HHS and

---

[62] The $3 billion figure is a very rough, and probably conservative, estimate based on the projected long-range cost of the smaller National Children's Study, which was reported to be an estimated $3.2 billion. Testimony before the House Appropriations Committee, Subcommittee on Labor, Health and Human Services, Education, and Related Agencies, on the National Institutes of Health, April 6, 2006.

non-HHS agencies (e.g., NIH Institutes and Centers, CDC, U.S. Environmental Protection Agency [EPA], and VA) and with other stakeholders. This includes the public and private sectors; biological, behavioral, social, public health, and population science disciplines; and basic biological scientists and epidemiologists. In addition, a working group could be formed that includes HHS and other Federal agency staff members, as well as representatives of academia, industry, and participant communities. This group could address scientific and methodologic issues and develop a process to engage other scientific disciplines that may not be represented on the core research team. The working group also could inform issues related to privacy, consent, intellectual property (IP) concerns, translation of findings into public health and health care, community advice and engagement, management of individual-level information of study participants, project logistics, information systems technology, and communications and dissemination.

Developing and managing such a complex and broad-based project will require many scientific and academic disciplines and the involvement of teams of experts in human genetics and genomics, environmental health sciences, clinical care, behavior, public health, epidemiology, sociology, nutrition, and behavioral psychology, among others. Some have suggested that organizing an interdisciplinary team of such broad range could be an insurmountable challenge. Although the need for interdisciplinary approaches has been recognized in science and efforts are underway at NIH to promote interdisciplinary efforts,[63] the disciplines that would be needed to plan or implement a new LPS go well beyond current practice.

Concerns also have been expressed about the need for additional work to link the perspectives and methods of multiple disciplines and develop an interdisciplinary outlook before embarking on a large and extensive project of this type. Experience with other large-scale population projects has shown that it takes tremendous effort to reach agreement on what should be measured, how to measure it, and how to interpret and translate the results.

Another issue that arises when contemplating this interdisciplinary study is the fact that the current promotion system in academia emphasizes independence rather than teamwork. Unless this system is redesigned, the contributions of individuals participating in an interdisciplinary team effort would not be recognized. New mechanisms of authorship recognition also might need to be developed, given that publications resulting from the study would in some cases be developed by teams of scientists.[64]

6. Models of Successful Interagency Collaboration

There are useful examples of successful collaborative models of project governance by Federal agencies that could be leveraged if such a study were to proceed. One example of a successful multiagency initiative is the National Toxicology Program, which involves NIEHS, EPA, and CDC. Productive collaborations also are taking place between NIH and CDC to address issues related to screening for genetic conditions, such as cystic fibrosis and hemochromatosis.

Other interagency efforts are exemplified by the WHI and the NCS, which are described in Chapter II. The WHI included a Community Prevention Study (CPS) that investigated strategies to enhance adoption of healthful behaviors.[65] For the CPS, NHLBI collaborated with CDC and worked with existing Community Prevention Centers to develop community-based public health programs and promote healthful behaviors

---

[63] *NIH Roadmap Initiatives. Theme: New Pathways to Discovery*. See http://nihroadmap.nih.gov/initiatives.asp [accessed December 19, 2006].
[64] Altshuler JS, Altshuler D. Organizational challenges in clinical genomic research. *Nature* 2004. 429:478-81.
[65] NHLBI Women's Health Initiative Web site. See http://www.nhlbi.nih.gov/whi/factsht.htm [accessed December 19, 2006].

in women. The NCS is led by NICHD in collaboration with NIEHS, CDC, and EPA and was designed in partnership with a Federal advisory committee and multiple working groups with more than 2,400 members from Federal agencies, private sector representatives, and members of nongovernmental organizations. The Federal agencies collaborating in the NCS reflect the expertise needed to conduct an epidemiologic study of environmental factors in the health of children. CDC's expertise in conducting epidemiologic studies was utilized for the development of the sampling design for the NCS.

7. Access and Intellectual Property Concerns

NIH has ensured that open and rapid access to research data from the HGP and the International HapMap Project were made rapidly available to the scientific community. However, sharing genotypic and phenotypic data on individual research participants raises a host of concerns about protecting the privacy of these subjects and the confidentiality of their data. In the past, access to large cohort population-based studies (e.g., Framingham Heart Study, Nurses' Health Study) has been limited for these reasons.

The "Bermuda Rules," drafted in 1996 in the context of the HGP, state that "all human genomic DNA sequence information, generated by centers funded for large-scale human sequencing, should be freely available in the public domain in order to encourage research and development and to maximize its benefit to society."[66] However, the Bermuda Rules do not address the issue of sharing biological specimens. A similar model of openness and access may need to be developed for use of and access to data and materials gathered and/or analyzed as a result of an LPS. It would be critical for investigators to have access to data and specimens to cross-validate markers and accelerate the clinical utility of the knowledge emerging from the project. Therefore, the system of governance would need to establish a method for review of requests for access. In devising a policy for sharing data and specimens, it also would be important to institute procedures for protecting subject confidentiality.

Some commenters pointed out a practical limitation in the sharing of specimens: Cell cultures may not be representative of the initially collected sample. In addition to the limited number of passages that will be successful, there is the problem of genetic and population drift, since cell samples are not completely homogeneous. There are also issues with storage conditions, since conditions suitable for one type of assay are not necessarily appropriate for another. Pilot projects may be needed to address these issues.

A fundamental tension exists between individual control and the value of broad public access to data to capitalize on the research results. Study policies on this issue would require periodic reexamination to allow for data uses that will arise and that were not foreseen in the original consent process. One idea that has been put forward is for varying fees depending on the type of user. For example, the fees could be cost-based for academic or governmental investigators who are willing to place their discoveries in the public domain. Other users who intend to patent their discoveries could be required to pay higher fees or a royalty on discoveries. However, the results of any additional testing and research should be rapidly placed in the public domain. Follow-on development may lead to patentable ideas, but the basic scientific information that derives from the study should be publicly accessible. A balance must be struck between individuals' rights to control their own information with the societal need to derive benefit from these data, even if it means that private gain results from additional private investment. If the latter could not be ensured, substantial scientific benefit would be sacrificed, resulting in a major decrease in the value of the study.

---

[66] *Policies on Release of Human Genomic Sequence Data: Bermuda-Quality Sequence ("Bermuda Rules*"). See http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml [accessed December 19, 2006].

Various well-known approaches (such as compulsory licensing or the use of collective rights organizations) would allow patenting while making it subject to public duties. Additional approaches might be developed by thinking of the national LPS database as a public resource and drawing on traditional concepts, such as those seen in electric utility regulation, that require nondiscriminatory access for all users but charge different access fees to different classes of users. The national database presents novel issues, and all options should be carefully explored. One commenter suggested that appropriate statutory authority should be sought from Congress to implement a framework that fairly apportions the value of discoveries made using the national database between the public and private interests and that creates incentives for commercial translational research but preserves access by noncommercial researchers.

Given the increasing rate at which genomic and proteomic discoveries are being patented, issues of ownership of IP associated with the proposed LPS would have to be thoroughly explored. The prospect of patent thickets or restrictive licensing of patents by institutions conducting the research could raise obstacles to the rapid development of public health measures associated with findings arising from the project.

8. NIH Genome-Wide Association Studies Policy Initiative

Data protection, access, IP, and authorship concerns are now being addressed in the development of an NIH policy to facilitate access to data resulting from genome-wide association studies.[67] These initiatives, several of which have been launched by NIH (described in Chapter II), use recently developed technologies to analyze the genetic differences between people with specific illnesses, such as diabetes or heart disease, and healthy individuals. It is hoped that these differences will point to genetic risk factors for the development or progression of disease and, ultimately, to better diagnostic tools and highly effective treatments. The proposed policy indicates that NIH-funded investigators will be expected to quickly submit de-identified, coded genotypic data and relevant health information (phenotypic data) about study participants to a centralized NIH data repository as soon as quality control procedures have been completed at the local institution. These detailed data will be made available through a controlled access process. Although NIH encourages the patenting of IP that addresses public health needs, such as creating new treatments that can be brought to the clinic, the proposed policy states that an effort will be made to prevent premature or inappropriate patents based on genome-wide association study results if they will impede future research. Because publication credit is critical to academic promotion, the proposed policy defines a grace period during which the data will be available for access, but the principal investigators submitting the data will be the only ones allowed to publish analyses in scientific journals. The policy asks that recipients of genome-wide association study data acknowledge the submitting investigator in any published work. NIH's proposed policy for whole-genome association studies might serve as a useful model to consider in an LPS.

9. Recommendations: Research Policy

The HHS Secretary should consider the complex research policy issues involved in undertaking a new large cohort population study. A decision to move forward with the study should be made only after widespread agreement is reached about its merit within the U.S. scientific community, HHS agency leadership, the international community, the public sector, and congressional policymakers.

---

[67] NIH Office of Extramural Research. Genome-Wide Association Studies. See http://grants.nih.gov/grants/gwas/index.htm [accessed December 19, 2006].

If a new large population cohort study is conducted:

1. **The HHS Secretary should continue to promote and facilitate ongoing consultation with the public, the private sector, and the international community to explore opportunities for collaboration on a large population study.**
2. **The HHS Secretary, in consultation with relevant HHS agencies and appropriate congressional committees, should assess support for sustaining a long-term and stable investment in a large population study.**
3. **Given the interdisciplinary nature of its scope, the HHS Secretary should establish a highly collaborative model of project leadership and management in multiple HHS and non-HHS agencies (e.g., NIH Institutes and Centers, CDC, EPA, and VA) and with other stakeholders. This includes the public and private sectors; biological, behavioral, social, public health, and population science disciplines; and basic biological scientists and epidemiologists.**
4. **The HHS Secretary, in consultation with relevant HHS agencies, should ensure that there are opportunities available to the general scientific community to (a) be informed about the potential for such a project; (b) present its views about the scientific validity and feasibility of such a project; (c) present its views on the commitment of resources to such an effort, including whether there are benefits to leveraging existing efforts; and (d) provide input on issues related to fair access by scientists to the project resources and the sharing of data and samples collected within it.**
5. **To ensure public benefits, the HHS Secretary should require that there are clear intellectual property policies in place for discoveries made using the data and samples collected.**

B. Research Logistics

A new LPS would face a number of logistical challenges. The design of the study is itself a major logistical issue, and details such as the study population; the variables to be collected (including thorough genetic and genomic tests and environmental monitors); and the range, frequency, and details of environmental and clinical testing would need to be carefully considered. Technologies that ensure fair and representative enrollment of individuals in the study, as well as the accurate collection of health information with minimal burden or inconvenience to study participants, would need to be developed. The difficulty of coordinating study enrollment and data collection among multiple health care systems and providers in the absence of interoperable electronic health recordkeeping systems would need to be addressed. Another relevant issue is the need for new databases for storing data and technologies to enable information about social and environmental exposures to be monitored and measured. Forming and sustaining interdisciplinary research teams, as discussed in the previous section, is a logistical issue as a well as a policy issue. These topics are discussed in the next section.

1. Operational Definition of "Environment"

The definition of "environment" must be carefully considered to determine how it would apply in an LPS of genes, environment, and common disease. In its broadest sense, the term "environment," as it relates to human health, includes the following factors:

- Physical factors (e.g., geographic location, climate)
- Chemical factors (e.g., toxic exposures, pesticides)
- Infectious agents (e.g., bacteria, viruses)
- Biological factors (e.g., biological sex, gestational factors, age)

- Behavioral factors (e.g., diet, physical activity)
- Social factors (e.g., education, socioeconomic status, gender, cultural factors)
- Lifespan events (e.g., stresses, quality of life)

Determining which environmental factors to study and measure and developing measurement tools are key logistical issues that would have to be resolved if the project were to be undertaken.

2. Recruitment and Enrollment

Determining the composition of the cohort to be studied is another important logistical challenge. The study population would need to include participants residing in all geographic regions of the United States, ethnically and socioeconomically diverse populations, and others who have not traditionally participated in research. A study cohort that is fully representative of the dynamic U.S. population would enhance the applicability of the research findings. Since the recruitment and retention of subjects from particular groups may be extraordinarily difficult, an LPS also would require a considerable amount of preparatory work—both to assess the barriers to recruitment and to develop effective incentives to encourage enrollment and sustain active participation. The following sections address some of the diverse population issues that would likely present barriers to recruitment, including race and ethnicity, gender and sex, socioeconomic and lifestyle factors, and children as study subjects.

a. Race and Ethnicity. Race and ethnicity are complex factors in human health and have been shown to play a role in disease risk. The complexity of these factors stems partially from the fact that race and ethnicity have been described in both biological and social terms.[68] Race is a particularly charged issue in the United States, and care would need to be taken when drawing correlations between health status and race and ethnicity.[69] A focus on ethnicity sometimes can discount the role of social and environmental factors in disease or overlook variations among persons of the same socially defined racial or ethnic group. Therefore, a plan for recruiting subjects into an LPS would need to consider race and ethnicity within social contexts, recognizing that biological differences may sometimes, but not always, be found.

In his presentation to SACGHS in March 2005, Dr. Charles Rotimi warned that the International HapMap Project, by including ethnic labels and concentrating on common genetic variants, could be used "to reinforce existing racial or ethnic categories."[70] Dr. Rotimi pointed out that ancestry is an important consideration but can be a confounding issue if it is not carefully assessed. For example, the ancestral history of African Americans is broad and represents a diverse genetic history from many countries. Yet scientists often group all African Americans together, as if their characteristics are uniform and homogeneous. The term "Hispanic" is used to refer to individuals whose ancestors are from such diverse places as Puerto Rico, Cuba, Mexico, South America, and Spain. An individual with an "Asian" background may have ancestors from one of at least seven countries. In grouping people together in these categories and studying them as if they have the same biological ancestry, valuable information is lost. At the same time, biological ancestry is a matter of gradation and, therefore, cannot be perfectly captured.

---

[68] Committee on Assessing Interactions Among Social, Behavioral, and Genetic Factors in Health. In: Hernandez LM, Blazer DG, eds. *Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate*. Washington, DC: The National Academies Press, 2006.

[69] Ioannidis JP, Ntzani EE, Trikalinos TA. Racial differences in genetic effects for complex diseases. *Nat Genet* 2004. 36(12):1312-8.

[70] Rotimi CN. Are medical and nonmedical uses of large-scale genomic markers conflating genetics and 'race'? *Nat Genet* 2004. 36(11 Suppl):S43-7.

While raising concerns about the use of racial and ethnic categories in research, Dr. Rotimi also acknowledged that there is much to be gained from generating and exploring hypotheses about disorders that occur more frequently within defined populations. The clustering of disorders among closely affiliated racial and ethnic groups is not unusual and reflects the more recent common ancestral origin, heritage, history, and environmental exposure of individuals within the group. Therefore, careful consideration must be given to all factors that could play a role in disease initiation and progression.

Thus, the mechanism for categorizing individuals for recruitment purposes has both scientific and social implications. Care must be taken to avoid confusing the social definitions of race or ethnicity with biologically informed definitions of populations. Such confusion could lead to inappropriate clinical interpretations or conclusions that could stigmatize or harm entire groups of people.

b. Gender and Sex. Because factors and traits such as genetics, biochemical interactions, physiology, and behavior characterize and differentiate males and females across the lifespan,[71] the distinctions between sex and gender are important factors that should be taken into consideration when recruiting participants into an LPS. As defined by the 2001 Institute of Medicine report *Understanding the Biology of Human Health: Does Sex Matter?*,[72] "sex" refers to the classification of an individual as male or female according to reproductive organs and sex chromosomes (in humans, XX for female and XY for male), whereas "gender" refers either to an individual's self-representation as male or female or to how that person is responded to by social institutions based on the individual's gender presentation. Whereas sex is purely defined by biology, gender is influenced by biology, self-definition, and an individual's environment and experiences. The Society for Women's Health Research has suggested that, given that biological sex has a significant impact on disease risk and response to environmental exposures, the need for efforts to recruit, enroll, and retain representative populations of males and females cannot be overstated. The Society also suggested that consideration also should be given to recruiting a representative population of participants whose biological sex is other than XX or XY.[73] These points and perspectives make clear that it would be crucial for such a study to recruit a study population that is representative in terms of sex and gender and to include questions related to sex differences and participant-defined gender.

c. Socioeconomic and Lifestyle Factors. A wide range of socioeconomic and lifestyle factors would need to be considered to ensure that the U.S. population is adequately represented in an LPS. Social determinants can influence health at multiple levels and throughout the life course[74] and would be important factors to assess in the study. However, because the benefits of the study would be realized in the future and may not be uniformly shared by all participants, it may be difficult to recruit individuals who, given the other circumstances of their lives (e.g., poverty, discrimination, low education or employment status, lack of access to health care, or lack of citizenship), perceive the potential benefits as outweighing the burdens and/or risks.

Furthermore, if health care facilities are the locus of recruitment, two major segments of the U.S. population, uninsured and unemployed individuals, are likely to be underrepresented. Special efforts would need to be

---

[71] Institute of Medicine, The National Academies. In: Wizemann TM, Pardue ML, eds. *Exploring the Biological Contributions to Human Health: Does Sex Matter?* Washington, DC: The National Academies Press, 2001.
[72] Ibid.
[73] Society for Women's Health Research. *Comments on the Draft Report "Policy Issues Associated with Undertaking a Large U.S. Population Cohort Project on Genes, Environment, and Disease."* July 13, 2006.
[74] Hernandez, LM, Blazer DG 2006. Op. cit.

made to include individuals who lack access to health care. In addition, steps would need to be taken to ensure the enrollment of a representative sample of individuals who rely on alternative methods of health care.

d. Children as Study Subjects. In addition to ensuring a diverse study population in terms of race, ethnicity, sex, gender, socioeconomic status, and lifestyle, an LPS would most likely require the inclusion of individuals at various biological ages, particularly given Federal policy encouraging the inclusion of children[75] in research.[76] Similarly, many researchers believe that the only way to truly account for the role of environmental exposures in disease is to enroll subjects prenatally and document prenatal exposures. However, recruitment of children as study subjects can present several ethical challenges, particularly with regard to autonomy and informed consent,[77] as well as logistical challenges, given that even the definition of "child" varies. In addition, regulations governing HHS-funded research with pediatric subjects (45 Code of Federal Regulations [CFR] Part 46, Subpart D) include specific protections afforded to children recruited as research subjects.[78] The work of the NCS planners may be helpful in thinking through the challenges of involving children in a longitudinal cohort study.

3. Measuring Differences in Health and Risk Factors in the Population

Currently, since no adequate methodologies and technologies exist for measuring the effects of the physical and social environments on health, we cannot accurately distinguish between the roles of environmental factors versus genetic factors in disease etiology and progression. The development of methods and technologies to collect data on factors such as diet and lifestyle, the initiation and progression of disease, and physiological and biobehavioral biomarkers would be a necessary component of any LPS.

The HGP required the dedication of financial resources specifically for the development of high-throughput data sequencing and computational tools to assemble, compare, and analyze high volumes of data. An LPS would involve the analysis of myriad questions related to the interplay among genes, environment, and health, necessitating the identification, measurement, and tabulation of numerous biological variables as well as the definition, measurement, and assessment of the effects of environmental changes (including behavioral and social factors) over time. To yield statistically significant results, data would have to be collected not only in the personal environments of the study subjects but also in the larger environments in which they live. Compared with the technology development required to carry out the HGP, an LPS most likely would require the development of far more complex and robust assessment and computational tools.

One of the goals of the GEI (described in Chapter II) is the development of new technologies to determine the contribution of environment, diet, and physical activity to illness. Investments will be made in emerging

---

[75] *NIH Policy and Guidelines on the Inclusion of Children as Participants in Research Involving Human Subjects* defines a "child" as an individual younger than 21 years of age, FDA regulations define a child as an individual from infancy to 16 years of age, and State law definitions also differ (many use younger than age 18 years). Federal Regulations (45 CFR 46, Subpart D, Sec. 401-409) rely on the State definitions of child for consent purposes.

[76] *NIH Policy and Guidelines on the Inclusion of Children as Participants in Research Involving Human Subjects.* See http://grants.nih.gov/grants/guide/notice-files/not98-024.html [accessed December 19, 2006].

[77] *Protections for Children in Research: A Report to Congress in Accord with Section 1003 of P.L. 106-310, Children's Health Act of 2000.* See http://www.hhs.gov/ohrp/reports/ohrp502.pdf [accessed December 19, 2006]. In addition to children, other subjects with vulnerabilities that affect their autonomy or ability to consent, such as pregnancy, incarceration, or incapacity due to illness, would be afforded additional protections, as provided under other Subparts of 45 CFR 46.

[78] Ibid.

technologies, such as small, wearable sensors capable of measuring environmental exposures. It is expected that such devices will provide researchers with the means to identify environmental factors that influence disease risk.

In addition to factors contributed by the physical environment (e.g., quality of air, water, and housing conditions), social, behavioral, cultural, and other environmental factors also must be measured as they interact to affect health and disease. The effects may vary depending on stage of life, extent and duration of exposure, and individual response. Data gathering and assessment of the relative influence of these factors will be complex and challenging, yet critically important.

4. Status of Environmental Health Sciences

During a presentation to SACGHS in June 2006, NIEHS Director Dr. David Schwartz stressed the importance of precise measurements of environmental exposures to elucidate environmental factors that may contribute to disease.[79] He pointed out that current measurements of environmental exposures rely heavily on questionnaire responses, which provide qualitative, rather than quantitative, data. Such measures lack specificity and precision and frequently fail to include lifestyle factors (e.g., diet, physical activity) that may also affect disease susceptibility.

Researchers also can measure exposures to environmental factors by evaluating blood and urine to determine the presence of biologically harmful chemicals (e.g., carcinogens) through the analysis of DNA and protein samples. Although these methods may determine associations between environmental exposures and disease states, they do not provide information on the mechanisms of such associations and typically provide data only after damage has occurred (i.e., reactive data). Therefore, determining the mechanisms of such associations and acquiring proactive data will require the development and implementation of new technologies and protocols, including biological sensors and remote sensing devices, and the adaptation of existing technologies to allow for more refined measurements of environmental exposures. In addition to improving data quality, these new technologies may determine individual real-time exposure levels and physiological responses. However, it is anticipated that these new technologies will not be available for at least 3 to 5 years.[80] Efforts to move forward in this field could be facilitated by collaborative efforts among relevant HHS and other Federal agencies (e.g., EPA, Occupational Safety and Health Administration, U.S. Department of Agriculture, National Institute of Standards and Technology).

5. Interdisciplinary Research Teams

Development and implementation of an LPS would require the input and involvement of investigators with wide-ranging expertise. Research teams would be responsible for guiding the selection of health conditions to be evaluated and for developing study endpoints. Given the complexity of study factors (e.g., public health impact, estimated genetic and environmental influences, feasibility of data collection, potential relationship to other disease endpoints) and overall relevance within the U.S. population and various subgroups, interdisciplinary research teams would need to include expertise in, among other areas,

---

[79] Secretary's Advisory Committee on Genetics, Health, and Society (2006). Summary of 10th Meeting, June 26-27, 2006. See http://www4.od.nih.gov/oba/SACGHS/meetings/June2006/transcripts/Pharmaco_Session-Schwartz.pdf [accessed December 21, 2006].

[80] Schwartz DA. The importance of gene-environment interactions and exposure assessment in understanding human diseases. *J Expo Sci Environ Epidemiol* 2006. 16(6):474-6.

genetics, human development, medicine, epidemiology, environmental science, behavior, bioinformatics, statistics, ethics, and community consultation. The expertise of investigators involved in existing cohort studies, such as the Framingham Heart Study, the MMCS, and the NCS, should be tapped as well.

6. Coordination Across Multiple Institutions and Health Care Systems

A new LPS would likely seek access to personal medical information from study participants. However, accessing such data would be difficult because of the absence of a national system of health care coverage and the lack of an interoperable, nationwide system of health records. These issues have raised questions about whether the study design would be valid. Since health insurance is primarily employer based and study subjects might change plans and/or providers over the course of the project, it may be difficult to ensure continuing access to their clinical data. In addition, as discussed earlier, enrolling uninsured and underinsured individuals may be difficult.

The HHS Office of the National Coordinator for Health Information Technology, which has an important role in promoting the nationwide adoption of electronic health records (EHRs)[81] and health information exchange among medical providers, could inform the development of an information technology infrastructure that would support data collection, storage, sharing, and security and privacy protections. Existing EHR systems, coupled with the use of effective health care delivery systems, such as those found in the Veterans Health Administration and the U.S. Department of Defense, could serve as prototypes for piloting an LPS.

7. Recommendations: Research Logistics

If a new large population cohort study is conducted:

1. **The HHS Secretary should encourage the project leadership and the scientific community to develop clear, consistent definitions and parameters for the stratification and classification of the projected sample population to ensure diversity and appropriate representation in the population to be studied.**
2. **The HHS Secretary should seek input from the public, as well as from researchers and clinicians, on the best approaches for identifying, recruiting, educating, and enrolling various subpopulations. Project organizers should be encouraged to consult with community-based organizations as part of their recruitment, assessment, and enrollment strategies.**
3. **The HHS Secretary, in consultation with both HHS and non-HHS agencies, should refine methods for collecting and analyzing environmental (i.e., physical, behavioral, and social) factors influencing health and ensure that resources are devoted to developing new tools to validate existing methods and improve assessments of the environment.**
4. **The HHS Secretary should encourage the project leadership to consult with health care providers and organizations to develop uniform and secure approaches for collecting, storing, tracking, and centralizing clinical information to be gathered over the course of the project, including the use of electronic health records.**

[81] U.S. Department of Health and Human Services Web site. See http://www.hhs.gov/healthit/ [accessed December 21, 2006].

C. Regulations and Ethics

Although most of the regulatory and ethical issues likely to be relevant to an LPS are not unique, the study's size and magnitude may amplify ethical concerns. Issues that would need to be addressed include study oversight through the IRB system, the applicability of existing Federal regulations related to privacy, and possible gaps in privacy protections. These concerns are further complicated by additional privacy, confidentiality, and IP concerns that are raised by the use of genetic information. Special protocols for informed consent may need to be developed because study data or samples might be used for purposes not identified at the outset of the study and because there are unresolved issues related to the ownership of biological samples and data. Decisions related to the provision of routine medical care would raise ethical challenges, and the length of the study would complicate the communication of research findings to subjects. Development of an independent ethics committee to work with the IRB and the project leadership might help address the complexity of these ethical issues, which are addressed in more detail below.

1. Institutional Review Board Oversight

Several Federal agencies play a role in defining and regulating the legal and ethical requirements for research involving human subjects, including banking human biological materials and clinical data. Regulations at 45 CFR 46, which are administered by the HHS Office for Human Research Protections, apply to research funded by HHS. Regulations at 21 CFR 50 and 56 apply when the research is regulated by the Food and Drug Administration (FDA), that is, when the protocols involve a clinical investigation regulated by FDA or support an application for research or marketing of an FDA-regulated product (i.e., most drug, biologic, and device studies). Both sets of regulations stipulate that human subjects research must be reviewed by an independent body known as an institutional review board and that subjects must provide informed consent to participate unless the requirements for informed consent are waived by the IRB according to the regulations. Typically, the IRB is situated locally to ensure that local community standards and norms are taken into consideration when the research is reviewed.

An extraordinary number of IRBs would be involved in the oversight of a large multisite study involving as many as 500,000 to 1,000,000 subjects, unless alternative oversight models were used. Several alternative models exist (e.g., central IRBs), but they have not been universally accepted among institutions because of concerns about relinquishing local control. Would a central IRB be the only feasible approach for an LPS? Would participating institutions be required to agree to review by a central IRB? If local review were allowed, how much modification would be possible to review and approval procedures?

In addition, if vulnerable subjects were enrolled, such as individuals who are cognitively impaired or children, additional regulatory requirements would apply, and mechanisms would need to be developed and implemented to ensure that such requirements are met.[82]

2. Informed Consent

Informed consent must be effective and prospectively obtained. The informed consent process involves three elements: (1) disclosing information to potential research subjects, (2) ascertaining that the subjects understand what has been disclosed, and (3) ensuring that subjects are voluntarily agreeing to participate

---

[82] *Protections for Children in Research: A Report to Congress in Accord with Section 1003 of P.L. 106-310, Children's Health Act of 2000.* Op. cit.

in the research.[83,84] The consent process for an LPS would need to be more comprehensive than such processes used in other research studies. Some public commenters suggested that prospective subjects should be provided with one-on-one counseling sessions to explain the basic science involved, overall study objectives, study questions, and subjects' rights and responsibilities.

HHS and FDA regulations permit IRBs to approve research when informed consent is sought and documented from each prospective subject (45 CFR 46.111[a][4])&[5]; 21 CFR 56.111[a] [4]&[5]). Although the informed consent issues that would arise in the context of an LPS are not unique, several deserve careful consideration. The study population and data would serve as resources for many additional research studies. Future research questions may not yet be conceptualized, or data and samples might be made available to additional investigators at other institutions for different studies. Therefore, it would be difficult at the time of recruitment and initial collection of clinical information, exposure data, or human biological materials to have an informed consent process that is broad enough to take into account future uses, yet specific enough to address immediate study objectives. There is an ongoing ethical debate about the adequacy of consent when subjects are not provided with details about how their data and specimens will be used.

IRBs and investigators will need to consider whether a one-time consent would be sufficient or whether periodic reconsent is needed. In addition, if participating investigators and institutions are not all identified at the outset of the study, it would be even more challenging to develop appropriate informed consent procedures. A variety of informed consent processes have been discussed to address the use of genetic samples.[85] One process is referred to as "enlarged consent," which allows the use of samples or data to be modified in the future. "Blanket consent" allows subjects a one-time choice concerning the future use of their data. There is also consent that allows subjects to opt in or out of specific studies. "Presumed consent" presumes that the researchers do not need to go through a reconsent process (this option sometimes includes an opt-out component).[86] Some surveys of research subjects suggest that many find consent for future unspecified use to be acceptable.[87] On the other hand, a study conducted in the United Kingdom by the Human Genetics Commission found that 82 percent of the respondents either strongly agreed (44 percent) or tended to agree (38 percent) that fresh consent should be sought from individuals before new research could be conducted on existing DNA samples held in medical genetic databases.[88]

3. Privacy and Confidentiality

Risks to privacy and confidentiality must be considered in any research involving the long-term collection and storage of clinical data and human biological materials. Concerns have increased as a result of advances

---

[83] U.S. Department of Energy Human Subjects Protection Resource Book. See http://www.orau.gov/communityirb/HS_book_6-22-06.pdf [accessed December 21, 2006].

[84] Faden RR, Beauchamp TL. *A History and Theory of Informed Consent*. New York: Oxford University Press, 1986.

[85] Cambon-Thomsen, A. The social and ethical issues of post-genomic human biobanks. *Nat Rev Genet* 2004. 5:866-73. In addition, previous advisory bodies have recommended approaches for crafting informed consent policies in human biological materials research. See, for example, National Bioethics Advisory Commission. *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance*. Rockville, MD: National Bioethics Advisory Commission, 1999. See www.georgetown.edu/research/nrcbl/nbac/hbm.pdf [accessed December 21, 2006].

[86] Caulfield T, Upshur R, Daar A. Debate: DNA databanks and consent: a suggested policy option involving an authorization model. *BMC Med Ethics* 2003. 3;4:E-1. Also see http://www.biomedcentral.com/1472-6939/4/1 [accessed December 21, 2006].

[87] Jack AL, Womack C. Why surgical patients do not donate tissue for commercial research: review of records. *BMJ* 2003. 327(7409):262. See also Wendler, D. One-time general consent for research on biological samples. *BMJ* 2006. 332(7540):544-7.

[88] Human Genetics Commission: Public Attitudes to Human Genetic Information. London: MORI Social Research, 2000. Also see www.hgc.gov.uk/UploadDocs/DocPub/Document/morigenticattitudes.pdf [accessed December 21, 2006].

in genetic, genomic, and other molecular technologies, in part because research can be conducted many years after specimen collection. Research results have the potential to identify genetic or other molecular alterations that may have implications for the current or future health of subjects and their family members, such as the risk of disease. The improper use or disclosure of such information could result in the loss of employment or insurability or in psychosocial harms (such as stigma). Concerns about the privacy and confidentiality of biological, demographic, or clinical data are not unique to LPSs. However, no researchers or organizations have previously sought to address the structural and legal problems that would likely arise from a new LPS.

For many research activities involving health information, neither researchers nor IRBs focus comprehensively on privacy. There is no law that establishes fair information practices for researchers, with the exception of the Privacy Act of 1974,[89] which applies only to Federal agency researchers. There are no "industry standards" for research privacy or any other set of research privacy policies that could be used as a model for an LPS. Laws pertaining to research records generally address only how researchers can obtain the records they need. Some laws provide limited protection for research records against compelled disclosure, (e.g., 42 U.S.C. 241[d]). Fair information practices (FIPs)[90] form the basis of many privacy laws in the United States; however, most elements of FIPs remain unaddressed in the research arena.

Donors of biological samples and identifiable clinical information generally expect their samples and data to be used to advance scientific knowledge or medical treatment without violating their privacy. However, maintaining the confidentiality of participants' genetic information would be challenging in a study of hundreds of thousands of people. In fact, if a sufficient fraction of an individual's genome is sequenced or genotyped, the genetic data become a unique identifier. If additional information about the source of the sample is also available, the data could be used to identify the individual. Over time, it may become easier to identify individuals from their DNA sequence data as further technological advances are made and genotyping becomes less expensive and routine.[91]

The protection of medical, exposure, and genetic information is critical for participants and groups who fear discrimination and stigmatization related to their genotypes. Varying levels of anonymity and coding schemes have been proposed as ways of protecting the identity of subjects, including the use of completely anonymous samples and data that cannot be traced to the identity of the subject. However, in a longitudinal cohort study, anonymity is not possible because it would prevent the collection of additional health status and exposure information in later phases. It would be important to ensure that study participants understand at enrollment that complete anonymity concerning their participation and data is not possible.

In addition to the regulatory requirements for protecting the rights and welfare of human subjects, privacy regulations provide additional requirements regarding medical information. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule,[92] which is enforced by the HHS Office for Civil Rights, governs the protection of individually identifiable health information. It was promulgated to increase the privacy protection of identifiable health information and to regulate known and unanticipated risks to privacy that may accompany the use and disclosure of such identified personal health information. HIPAA

---

[89] See http://www.usdoj.gov/oip/privstat.htm [accessed December 21, 2006].
[90] Organisation for Economic Co-operation and Development (1980). *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.* See http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html [accessed December 21, 2006].
[91] McGuire AL, Gibbs RA. No longer de-identified. *Science* 2006. 312(5772):370-1.
[92] See http://www.hhs.gov/ocr/hipaa/finalreg.html [accessed December 21, 2006].

covers individually identifiable health information that is held or maintained by "covered entities" (e.g. health plans, health care clearinghouses, or health care providers that transmit health information for certain transactions, as defined by HHS)[93] or by business associates acting for a covered entity. HIPAA does not apply to biological specimens per se, but it would apply to identifiable clinical information that may be associated with a specimen. In most cases, unless an exception applies, HIPAA will require authorization from individuals to use their protected health information in research. This authorization is distinct from informed consent.

An LPS would involve institutions that have different roles and legal requirements and are subject to different privacy laws. For example, the "Common Rule"[94] is the term used by 18 Federal agencies that have adopted the same regulations governing human subjects research. The policy is designed to ensure minimal standards for the ethical treatment of human research subjects. The Common Rule is also referred to as "Subpart A" of the HHS regulations of human research 45 CFR 46.[95]

However, the Common Rule standards and the HIPAA privacy standard are not the same. Research institutions involved in the study would have to comply with the Common Rule and, if they were covered entities, with HIPAA as well. In addition, the Privacy Act of 1974 mandates that each Federal agency have an administrative and physical security system in place to prevent the unauthorized release of personal records. Most Federal agencies likely to be involved in an LPS would be subject to both HIPAA and the Privacy Act of 1974. Most health care providers involved would be subject to HIPAA, and health care providers and researchers would also be subject to applicable State health privacy laws.

These differing and overlapping requirements could have many consequences. The privacy rights of a research subject might vary depending on which institution holds the subject's information and, possibly, where that institution is located. The privacy rights might also vary from data element to data element within a record maintained by a health care provider or a researcher. None of these circumstances is unique to the proposed LPS. What distinguishes this project from others is its large scope, the perceived sensitivity of the information, both in general and because of its connection to the Government, and the likely high visibility of the project. The extent to which these rules and laws overlap and/or conflict and the potential for problems that might result would have to be assessed early in the LPS planning stages. Pilot projects or studies to explore these issues in advance of any decision to proceed with the project may be helpful to inform the HHS Secretary, others in HHS, and the public at large about acceptable approaches and practices.

To address these concerns, an LPS privacy policy could be developed under the auspices of an independent project ethics committee. As stated, FIPs form the basis of many privacy laws in the United States and could serve as the basis for a privacy policy. HIPAA health privacy rules are based on common principles of FIPs.[96] FIPs do not, however, automatically translate into a privacy policy for any given set of data or activities. The establishment of a specifically applied policy requires the balancing of competing values, cost considerations, and practicality.

---

[93] U.S. Department of Health and Human Services. *Standards for Privacy of Individually Identifiable Health Information*, 45 CFR Parts 160 and 164. See http://www.hhs.gov/ocr/hipaa/guidelines/guidanceallsections.rtf [accessed December 21, 2006].
[94] See http://www.hhs.gov/ohr/humansubjects/guidance/45cfr46.htm [accessed December 21, 2006].
[95] Ibid.
[96] U.S. Department of Health and Human Services. *Standards for Privacy of Individually Identifiable Health Information; Final Rule*, 65 *Federal Register* 82461, 82487 [accessed December 28, 2000]. See also http://www.hhs.gov/combinedregtext.pdf [accessed December 28, 2000].

A privacy policy developed by an ethics committee could increase the trust of the public in the policy's provisions. A set of privacy protections established as a condition of participation by researchers could afford subjects sufficient rights without creating complex legal or practical problems. Since privacy laws generally permit higher standards to be applied, the privacy policy of the LPS could have a high standard and would not create conflicts in most instances. In some cases, additional protections would be needed for narrower classes of data (e.g., if they pertained to human immunodeficiency virus status, psychiatric information, or substance abuse).

a. Privacy Officer and Privacy Impact Assessment. Comments from the World Privacy Forum[97] suggested that the proposed LPS have a full-time privacy officer in place for the operational life of the project. The privacy officer would be independent of participating institutions and would have sufficient authority, staff, and resources to initiate and conduct audits and investigations of compliance with privacy and security obligations, receive and act on any complaints about the study, and propose changes in the privacy policy. The privacy officer would report directly to the project leadership and would issue public reports without the need for clearance from the project leadership.

The World Privacy Forum[98] also recommended that, prior to the start of such a study, a thorough privacy impact assessment (PIA) should be prepared by the privacy officer or a third party with suitable experience and independence from project planners and should be made publicly available. The E-Government Act of 2002 requires Federal agencies to prepare a PIA under specified conditions,[99] but the Forum recommended that the assessment conducted in this case should exceed the Act's requirements. The LPS planners and project leadership should be required to consider and respond publicly to the assessment's findings.

b. Third-Party Use of Project Records.[100] The LPS files and databases would have the potential to attract interested users who have no relationship to the project, such as law enforcement officials, national security agencies, private litigators, welfare agencies, and others. Existing laws establishing protection against compelled disclosure could be of assistance in protecting records from third parties, although it is not clear that they would apply in every case to a database held by the Government. The certificate of confidentiality laws available within HHS include 42 U.S.C. § 299c-3(c), 42 U.S.C. § 242m(d), and 42 U.S.C. § 241. These laws offer covered records some disclosure protection, but they have limits that would have to be explored and recognized. The mixing of research and clinical records that might occur as part of the project could weaken the already limited protections of certificates. Furthermore, most certificates protect against compulsory disclosures and do not expressly prohibit voluntary disclosures. As part of the PIA, an independent assessment could be conducted on the scope and gaps of existing certificate of confidentiality laws to determine whether additional protections are needed and what prospective subjects must be told about the legal risks of participation.

c. Identifiability.[101] As discussed earlier, technological advances and the proliferation of DNA databases for law enforcement, medical, and other purposes are expected to make genomic sequence data identifiable

---

[97] World Privacy Forum. Detailed public comments on the draft version of this report. This final report incorporates several ideas from these public comments.
[98] Ibid.
[99] 44 U.S.C §3501.
[100] World Privacy Forum. Op. cit.
[101] Ibid.

at some point in the future. In addition, it is increasingly difficult to de-identify health records.[102,103,104] As a result, the challenge of addressing the identifiability issues for genomic and clinical information collected and shared in an LPS is likely to grow.

Legal standards for identifiability are considered by some privacy experts to be incomplete and inconsistent. The HIPAA approach to de-identified data is more specific but may not solve all the problems that could arise. Study-specific standards for de-identification that are based on appropriate legal, policy, and technical measures will need to be developed, evaluated in the privacy impact assessment, and addressed in the published privacy policy.

4. Control of Samples and Data

Ownership of biological samples and research data and benefit sharing are also relevant issues. Biological samples traditionally have been viewed as belonging to the researchers or institutions to which they were donated, and recent court rulings support the concept that individuals do not own their biological samples, regardless of whether a commercial benefit is expected from the research.[105] Cases in which investigators have been sued for profiting from discoveries that were derived from study samples have highlighted the need for appropriate informed consent procedures and ownership agreements.[106]

It would be helpful to take stock of how similar projects have addressed these issues. For example, the Marshfield Clinical Personalized Medicine Research Project excludes private ownership of samples and data and requires the return of any profits from commercial applications to the Marshfield Clinic Research Foundation.[107] Some international biobanks, such as the Estonian Genome Project, include provisions for benefit sharing by allowing participants to access their genetic information. Participants have the right to access their personal genetic information for use in personalized medicine and the diagnosis of disease.

Efforts are underway to establish greater standardization in the technical aspects of data and specimen repositories. For example, the Public Population Project in Genomics (described in more detail in Appendix D) is a nonprofit organization that is building an international consortium to promote the types of discussions and collaborations needed to reach a consensus on optimal, standard procedures in the field of population genetics research. NCI has taken steps to standardize specimen collection, storage, and distribution among NCI-funded investigators by developing guidelines for common technical and operational standards, including ethical and legal issues. NIH also is developing an agency-wide framework that addresses the ethical and legal issues associated with research with human specimens and data and a trans-NIH policy for the sharing of genotypic and phenotypic data from whole-genome association studies.

5. Return of Research Results

There is ongoing debate about whether research findings should be communicated to subjects and, if so, at what point (e.g., during the study, when the study is completed, or only after study findings have been

---

[102] Malin B, Sweeney L. Determining the identifiability of DNA database entries. *Proc AMIA Symp* 2000. 537-41.
[103] See Dr. Latanya Sweeney's homepage at http://lab.privacy.cs.cmu.edu/people/sweeney [accessed December 21, 2006].
[104] Carnegie Mellon Data Privacy Lab: Genomic Privacy Project. See http://privacy.cs.cmu.edu/dataprivacy/projects/genetic/index.html [accessed December 21, 2006].
[105] Hakimian R, Korn D. Ownership and use of tissue specimens for research. *JAMA* 2004. 292(20):2500-5.
[106] Kaiser J. Biomedical research. Court decides tissue samples belong to university, not patients. *Science* 2006. 312(5772):346.
[107] See http://www.marshfieldclinic.org/chg/pages/default.aspx?page=chg_pmrp_faqs [accessed January 18, 2007].

validated by others). This issue is relevant to all research studies. However, a longitudinal LPS would continue for a decade or more, raising complex questions about investigators' responsibilities to report potentially useful information.

Opponents of sharing preliminary research data argue that the data may not be valid and may do more harm than good by causing physical, medical, psychological, social, or economic harms. Research results may affect life choices, such as whether to have children, or cause recipients to forgo screening or preventive care for specific diseases.[108] On the other hand, proponents of returning individual research results to research participants often cite the ethical principle of autonomy as a reason why study participants should have access to their own information, along with appropriate caveats if the results are of uncertain validity. In addition to establishing provisions for the use and disclosure of protected health information, the HIPAA Privacy Rule established a right of access by patients to their medical records.[109] Under the Privacy Rule, research subjects may access their health information records and must be notified of their right to do so. However, research that includes treatment, including clinical trials, permits a covered entity to suspend individual access rights until the end of the study, provided that the individual agreed to the suspension when consenting to participate in the research and was informed that the right of access would be reinstated upon completion of the research. The issue of how this aspect of the Privacy Rule applies in the context of a 10- or 20-year study has not yet been addressed.[110]

Another consideration with regard to returning research results relates to the applicability of the Clinical Laboratory Improvement Amendments of 1988 (CLIA). CLIA requires laboratories to be CLIA-certified for them to return test results used for the purposes of treating, diagnosing, or assessing an individual's health.[111]

6. Provision of Care

The provision of clinical care can be an important potential benefit to participating in clinical research. Critical ethical considerations for an LPS would include clarifying whether routine medical care would be provided as part of the protocol and, if so, whether it would create an undue influence on participation. Ethical issues can arise if subjects confuse the purpose of the study and believe that they are participating in the study to receive health care services rather than to be study subjects. This could unduly influence their willingness to participate and their ability to provide valid informed consent. In addition, the dual role of the physician-investigator can further confuse matters, given that investigators have competing obligations to funders, their institutions, and to science, all of which could affect the care they can offer participants.[112]

---

[108] Secretary's Advisory Committee on Genetic Testing (2000). *Enhancing the Oversight of Genetic Tests: Recommendations of the SACGT.* See http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf [accessed December 21, 2006].

[109] U.S. Department of Health and Human Services. 45 CFR Parts 160 and 164. *Standards for Privacy of Individually Identifiable Health Information; Final Rule,* August 14, 2002. See http://www.hhs.gov/ocr/hipaa/privrulepd.pdf [accessed December 21, 2006].

[110] HIPAA Privacy Rules. See http://privacyruleandresearch.nih.gov/clin_research.asp [accessed February 12, 2007].

[111] Secretary's Advisory Committee on Genetic Testing (2000). *Enhancing the Oversight of Genetic Tests: Recommendations of the SACGT.* See http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf [accessed December 21, 2006].

[112] National Bioethics Advisory Commission (2001). *Ethical and Policy Issues in International Research: Clinical Trials in Developing Countries. Volume I. Report and Recommendations of the National Bioethics Advisory Commission.* See http://www.georgetown.edu/research/nrcbl/nbac/clinical/Vol1.pdf [accessed December 21, 2006].

A number of questions would arise that the core LPS research team would have to address during the LPS planning phase. The study would necessarily include subjects who are underinsured or uninsured. What would be the ethical responsibility of researchers in obtaining consent from these subjects, especially if they have conditions that require immediate medical attention (e.g., angina)? What would be the responsibilities of investigators in ensuring that participants have access to necessary care? Environmental exposures to certain common substances might be linked to increased health risks in participants with specific genotypes or in the general population (e.g., exposure to lead by children). What would be the obligations to participants concerning resources for limiting their exposure and to others in the community who might be affected?

7. Independent Ethics Review Committee

Because the ethical issues raised by such a project are so diverse and complex, an independent ethics committee could be considered to supplement the efforts of the IRB. The ethics committee would have a broader mandate than is normally addressed by an IRB and would require a higher level of ethics expertise. A significant percentage of its members could be consumer representatives, some of whom should represent the various racial/ethnic groups participating in the LPS. The committee could act in an advisory capacity to the IRB and/or the project leadership but should be independent of both. The public should expect that the committee is in a position to freely provide ethics advice, recommendations, and discussions to decisionmakers. The functions of such a committee might include:

- Establishing a core informed consent document drafted in language understandable to the general public. It would explain patients' rights, including the policies on provision of care and the right to withdraw from the study. It would describe a process (to be developed by the committee) for determining whether reconsent is required for new study proposals.
- Addressing whether potential subjects should be provided with one-on-one genetic counseling sessions prior to enrollment to explain the basic science and overall study objectives and enumerate study questions.
- Recommending mechanisms to promote transparency about new research studies, even when no additional consent is required, and communicating this information to participants when they initially enroll.
- Developing mechanisms to ensure genetic privacy and data security to the extent possible.
- Periodically evaluating whether research participants' understanding of the project and their informed consent are consistent with project goals and activities.
- Making recommendations for new areas of research on ethical issues as they evolve over time.
- Addressing other ethical questions that arise and providing ethics consultation to the project leadership and/or the IRB on individual study or patient issues.

To afford protections against genetic discrimination, the independent ethics committee could also develop a policy mechanism to help ensure that the U.S. Government would not access or use the data from the study for any purpose other than research. Data use agreements could prohibit the recipient from using individually identifiable information for marketing purposes or service determinations for insurers or lenders.

8. Recommendations: Regulations and Ethics

If a new large population cohort study is conducted:

1. **The HHS Secretary should convene a working group of representatives from the Office for Human Research Protections, the Food and Drug Administration, the Office for Civil Rights, and other relevant HHS and non-HHS agencies to address issues and questions raised by the public and to provide technical assistance and guidance to research sites on legal requirements regarding the protection of research subjects, health information privacy, and patient safety.**
2. **The HHS Secretary should establish an independent ethics committee to serve in an advisory capacity to the institutional review boards and the project leadership.**
3. **The project leadership should systematically and regularly seek the input of study subjects regarding their experiences, concerns, and recommendations for enhancing protections to ensure that the appropriate protections are in place and are being consistently implemented.**
4. **The project leadership should develop a policy regarding the use of data and samples to ensure the legal and ethical use of clinical and epidemiological data and specimens. This policy should be made available to study subjects.**

D. Public Health, Social, and Economic Implications

The interdisciplinary nature of an LPS will require a strong partnership among basic scientists, clinical investigators, and, given its potential benefits at the population level, public health researchers. A 2006 report from the Institute of Medicine makes a strong case for the need for a coordinated approach among the various disciplines to improve health from a population perspective beyond the traditional bounds of State, Federal and local agencies.[113]

The knowledge gained from an LPS will have the greatest public health impact in identifying when individuals and groups would benefit from specific interventions based on their risk. Knowledge of risk contributes to improved health outcomes only if effective measures are available for early treatment or prevention.[114] Moreover, the appropriate use of risk information requires input from a broad range of experts in fields such as health outcomes research, economics, public policy, and health communication and education.[115]

There are questions, however, about whether the results from an LPS would be definitive enough to lead to clinical applications and whether the findings would be generalizable to the entire U.S. population. Although there is sufficient power to detect causative polymorphisms of small effect with as few as 500 subjects, greater power is achieved by increasing the sample size. Yet, even though the power of the sample size might suggest that the findings are generalizable to the U.S. population as a whole, they may reflect only subsets of the larger population due to genetic and/or environmental factors. In addition, unless gene function is understood, the certainty of the association between disease and a gene or SNP may not be clear. In other cases, the association will be clear, but no treatment will be available, a therapeutic gap that is all the more troubling because of its uncertain duration. Moreover, there are tremendous challenges in moving from a statistical genetic association to an understanding of gene function and the mechanisms of

[113] Committee on Assessing Interactions Among Social, Behavioral, and Genetic Factors in Health, 2006. Op cit.
[114] Burke W 2006. Op. cit.
[115] Khoury MJ et al. 2003. Op. cit.

action that could lead to new therapies or preventive measures or that would withstand regulatory standards of evidence.

Effectively translating the discoveries from an LPS into improvements in public health will require the integration of knowledge with other disciplines, increased focus on a population perspective, and enhanced partnerships and coordination among the areas of health care and public health, health services research, and population health monitoring.[116] Applied research conducted in population-based settings will be needed to identify effective ways of disseminating study findings, delivering interventions that work in practice, and evaluating health outcomes at the individual and population levels.

Collaboration with State and local public health officials could enhance data collection through coordination with State public health surveillance or monitoring programs, data collection instruments, and access to medical records. Study investigators could link genomic data from the study with public health data, such as vital records, hospital records, cancer registries, birth defects registries, and diabetes registries.

Collaborations with related studies that are underway within the United States and around the world would enable data harmonization and integration, replication, pooled analyses, meta-analyses, and online databases and information systems. An ongoing mechanism could be developed for evidence-based evaluation of results coming from the study for the purpose of developing health services genetic information to be used in clinical practice or for public policy purposes.

The LPS also would require extensive community engagement and the education of both the public and health care providers. Policymakers and legislators at the State level would need to be engaged and informed about the study's many ethical, legal, and social issues. Experts in health communications would be needed to develop an effective communication plan and employ effective mechanisms to disseminate information about the meaning and limitations of LPS study results. Engagement of the public would help make research findings significant to and understandable by the general population (see Chapter IV). Mechanisms for dissemination could include conferences, publications, a Web site, and public forums. Since health care providers would likely be involved in disseminating information to the public and addressing questions from their patients, communication strategies would need to target these stakeholders in particular.

1. Impacts on Health Disparities

The elimination of health disparities is an important public health goal. It is unclear the extent to which genetic differences account for health disparities, because most current genetic studies do not take into account physical and social environments. Health disparities research provides an important opportunity to integrate biological knowledge with social and behavioral knowledge to better understand the determinants of disease, which will help reduce the risk of disease and provide better treatment when it becomes available. With regard to an LPS, Dr. Charles Rotimi, who addressed SACGHS during it's February-March 2005 meeting,[117] noted that the prevalence of disease in certain populations can be related to the etiology of disease as well as to disparities in health care. He also stated that it is the responsibility of the scientist to define how etiology is being investigated. In this context, multiple risk factors would need to be

[116] Khoury MJ et al. Manuscript in preparation.

[117] Secretary's Advisory Committee on Genetics, Health, and Society (2005). Summary of Sixth Meeting, February 28-March 1, 2005. See http://www4.od.nih.gov/oba/SACGHS/meetings/Feb2005/SACGHSFeb2005postmeeting.htm [accessed January 19, 2007].

studied simultaneously within subgroups (e.g., race, ethnicity, behaviors, geography, genetic backgrounds, exposures, and social environments) to understand how environmental and genetic risk factors interact and lead to health differences. Thus, LPSs can help clarify or change our understanding of health disparities. For example, research may determine that a particular group of individuals has an increased risk of developing a disease. Although it may be clear in some cases that health care disparities are the primary cause of disease, an LPS would need to consider the ways in which health disparities play out in different ethnic groups. However, if the particular group also is socially or economically vulnerable, the findings could exacerbate the disparity.

At the same time, and as discussed earlier in the Research Logistics section, race and other socially derived notions of groups should not be used as a proxy for genotype when studying individual variation and diagnosing and treating genetic diseases.[118] Socially defined groups, such as African Americans, could experience stigmatization if race is used as a factor in characterizing genetic variation. For example, if a particular genetic variant that predisposes one to a particular disease is prevalent in one subpopulation, the medical community and public may incorrectly assume that all members of the race to which the given subpopulation belongs are predisposed to having the disease.

Consultation by decisionmakers with leaders in the groups likely to be affected will be critical, especially with regard to historically vulnerable groups. These populations include low-income families; racial and ethnic minority groups; women; lesbian, gay, bisexual, and transgender groups; and people with low educational attainment.

2. Risks of Genetic Determinism

Although genes play an important role in determining biological characteristics, the idea that they determine everything about an individual—called "genetic determinism"—discounts the fact that individual characteristics are the result of a complex interaction between genes and the environment and that human beings are influenced by the physical, psychological, social, and political environment as well as by biology.[119] Even when an association between a gene and a disease has been established, it is not always possible to predict with certainty that a given person will get the disease or determine the age of onset. (An exception to this general rule is seen in autosomal dominant diseases, such as Huntington disease, where a gene mutation will confer almost certain disease risk.) This is because other factors, such as other genetic, epigenetic, environmental, and behavioral variables, may be altered that mask or compensate for the disease gene. Although an LPS might be expected to enhance understanding of the multifactorial nature of common diseases, it is possible that discoveries about specific gene variations may promote the idea of genetic determinism and lead to continued misinterpretation and misuse of genetic explanations in public policy, the courts, health and life insurance policies, and medical practice.[120]

---

[118] Royal CD, Dunston GM. Changing the paradigm from 'race' to human genome variation. *Nat Genet* 2004. 36(11 Suppl):S5-7.
[119] National Bioethics Advisory Commission (1997). *Cloning Human Beings: Report and Recommendations of the National Bioethics Advisory Commission.* See http://www.georgetown.edu/research/nrcbl/nbac/pubs/cloning1/cloning.pdf [accessed December 21, 2006].
[120] Sharon Kardia, Ph.D., Associate Professor of Epidemiology and Director, Public Health Genetic Program, University of Michigan School of Public Health. Presentation to SACGHS on October 19, 2005. Transcript available at http://www4.od.nih.gov/oba/SACGHS/meetings/October2005/SACGHSOct2005postmeeting.htm [accessed December 22, 2006].

Exhibit 1
SACGHS and Genetic Nondiscrimination Legislation

Since its first meeting in June 2003, SACGHS has supported the enactment of Federal legislation to prohibit genetic discrimination in health insurance and employment. Since then, SACGHS has provided the HHS Secretary with extensive information to help make the case for enactment of this legislation. Public testimony gathered by the Committee demonstrates how deep-seated public fears are about the potential for misuse of genetic information in health insurance and employment and the lack of specific Federal legal protections against genetic discrimination. The testimony also documented how health care decisions can be affected by such fears when they lead patients to seek genetic testing outside the formal health care system, request that their test results be kept out of their medical records, opt for anonymous testing, or forgo testing altogether.

SACGHS also commissioned an analysis of the adequacy of current law in protecting against genetic discrimination, which has been a critical part of the debate about the need for new Federal legislation. The analysis concluded that there are no Federal laws that directly and comprehensively address the issues of discrimination raised by the use of genetic information; the laws and court cases that address parts of these issues leave substantial gaps in coverage and offer uncertain and inconsistent safeguards at best; and current avenues for relief are uncertain and, if pursued, are likely to lead to confusion for consumers, insurers, and employers, as well as to costly litigation.

More information on SACGHS activities regarding genetic nondiscrimination legislation is available at http://www4.od.nih.gov/oba/sacghs/reports/reports.html#discrimination.

Some of the potential harms of genetic determinism could be mitigated by the enactment of comprehensive and specific Federal legislation prohibiting genetic discrimination in health insurance and employment. The need for such legislation, which is SACGHS's highest priority (see Exhibit 1 above), has been under discussion for almost a decade. In the 109th Congress, the Genetic Information Nondiscrimination Act of 2005 unanimously passed the U.S. Senate in 2005,[121] but was not considered by the U.S. House of Representatives. The legislation would have prohibited health insurers and employers from making adverse decisions based on a person's genetic predisposition to disease.

During the first session of the 110th Congress, genetic information nondiscrimination legislation was introduced again in both the U.S. House of Representatives (H.R. 493) and the U.S. Senate (S. 358). The Administration continues to support Federal legislation in this area. If such protections are not universally available to all U.S. citizens during the conduct of such an LPS, great care would need to be taken to prevent research results from being inappropriately interpreted and misused.

[121] Executive Office of the President (2005). *Statement of Administration Policy: S.306-Genetic Information Non-Discrimination Act of 2005*. See http://www.whitehouse.gov/omb/legislative/sap/109-1/s306sap-s.pdf [accessed on December 21, 2006].

3. Impacts on Public and Social Policy

Although there has been some progress in advancing the understanding of gene-environment interactions, particularly in the areas of toxicogenomic and pharmacogenomic research, the findings have only confirmed how challenging it is to integrate such knowledge into existing policy standards and methods. As such, the public policy impacts of an LPS may not be as significant as expected. Genetic findings in complex disorders, especially gene-environment interactions, may not provide a sufficient impetus for regulatory bodies to create policies to protect individuals.

Traditionally, public health policy has focused on population-level solutions. A one-size-fits-all model has been used for most public health campaigns (e.g., antismoking campaigns). Extensive research will be needed to understand how individual genetic differences will affect public health policies and regulations. For example, if the study finds that certain people are especially susceptible to the adverse effects of pollution, will the onus for addressing the problem shift from society to the individual? Will the findings affect the current population-based risk assessment paradigm used by U.S. regulatory agencies (e.g., EPA, FDA)? Will standards and guidelines be based on susceptible genetic subgroups? Will additional resources be needed to incorporate new scientific findings?

4. Economic Impacts

As with many other policy issues described in this report, the economic consequences of an LPS may be significant. Although the issues do not differ in type from those associated with basic research in general, the scale of such a study requires that they be thoroughly considered before a decision is made about whether to move forward.

The LPS would likely have economic benefits as well as costs. It would stimulate a creative research enterprise, a vibrant private sector, job creation, and innovation, all of which have economic benefits. The study's scientific findings may lead to technological innovation and products, many of which would be integrated into the health care system and lead to better health outcomes. Although technologies that address the underlying causes of disease and facilitate timely public health action may have substantial cost offsets, their development may contribute, at least initially, to escalating health care costs that the United States may not be able to afford. The rising expenditures of health care might divert resources from other important public needs.

Currently, there is no resource allocation system in place for deciding how to pay for the many emerging health care advances. Since rising health care expenditures are largely driven by technological innovation, and technological innovation is driven by research and development, it is important to systematically examine the consequences of an investment in an LPS from an overall societal perspective. Consultation with health economists and science policy analysts would help clarify the gains of the initiative in light of the anticipated large costs of the project. Creating a forum for such discussions would be important, since these concerns raise difficult economic and political issues for the health sector, the biomedical and behavioral research communities, and society as a whole.

5. Recommendations: Public Health, Social, and Economic Implications

If a new large population cohort study is conducted:

1. **The HHS Secretary and the project leadership should systematically and regularly integrate project findings with other emerging data from other types of studies and regularly disseminate the accumulated knowledge base in a manner to benefit the population's health. This information should be tailored to meet the information needs of the public, health care providers, and the public health community to use integrated information for the benefit of the population's health. Project resources should be sufficient for the integration, dissemination, and translation activities necessary to maximize the public health impact.**
2. **The HHS Secretary, in consultation with the project leadership, should establish an independent standing committee for the duration of the project to periodically assess the persistent and emerging social and economic implications of this initiative, with special attention to health disparities. The committee could consist of individuals with expertise in the relevant sciences, medicine, law, ethics, and patient and community advocacy. The committee would routinely seek input from the public on the implications of project results and report its findings.**

IV. Need for Public Engagement

Public accountability is a hallmark of Federal stewardship of the biomedical, behavioral, and public health research enterprises. Accountability helps ensure the relevance of research programs, as well as congressional and public confidence in the research missions. Given the many policy implications of an LPS, an extensive public engagement effort would be needed to gauge public opinion about the study and whether it should be undertaken. If the initiative were to go forward, ongoing evaluation of public opinion would be important to further inform study planning and implementation. In addition, input from various geographic, racial, ethnic, and cultural groups would be essential.

In light of how difficult it is to engage the public on complex issues with scientific content,[122] the public's understanding of genetics and genomics research would need to be assessed before trying to determine public support for such a study. Various types of efforts would be needed to inform the public and then solicit feedback about genetic and environmental factors in common diseases.

A. Public Awareness and Attitudes Toward Genetics

Recent survey data suggest that while most of the public may be aware of genetics (75 percent of survey respondents indicated that they had heard of genetics), the depth of public understanding is limited. Public attitudes toward genetics and personalized medicine appear to be very favorable. About half of survey respondents were interested in using genetic information to understand and optimize health and make informed choices about prescription drugs.[123]

At the same time, the survey data suggest that the public may have significant concerns about the privacy of their genetic information, including its storage for research purposes and the potential for its misuse. In addition, respondents who had a favorable attitude about the use of genetic information for health purposes nonetheless expressed misgivings about a Government-sponsored DNA databank. One-quarter to one-third of respondents thought that the Government should create a national database of DNA information to advance health.[124]

Because the success of an LPS depends on favorable public attitudes, it will be important to thoroughly assess the public's willingness to participate in such a study in advance of any funding decision. This assessment might be carried out prior to or in parallel with a public education effort.

---

[122] *Information and Attitudes: Consulting the public about biomedical science*. A report published by the Wellcome Trust. August 2005. See http://www.welcome.ac.uk/assets/wtx026426.pdf.

[123] Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS). Summary of Ninth SACGHS Meeting, March 27-28, 2006, Bethesda, Maryland. A Cogent Research Web-based survey of a random sample of 1,000 Americans over the age of 18. See http://www4.od.nih.gov/oba/sacghs/meeting/march2006/white.pdf.

[124] Ibid.

B. Mechanisms for Engaging the Public

A range of methods exist for consulting with the public about scientific issues. At one end of the spectrum are public information or education efforts. At the other end of the spectrum are more consultative approaches that seek to encourage public input from and dialog with the public about the issues. There are many "publics" to be consulted, including the general public, disease advocacy groups, scientific and professional organizations, health care providers, and health care organizations.

Typical mechanisms for public engagement include polling, surveys, moderated focus groups, workshops, and scenario development. Issue identification and agenda-setting are typically carried out by the study organizers, but another model involves the community in identifying the issues, framing and prioritizing the discussions, and devising outreach strategies. A more deliberative approach to engaging the public involves providing indepth background information about the topic to better facilitate public formulation of the issues. In the case of an LPS, engagement efforts could be aimed at the communities from which participants would be recruited, or regional and/or national conversations could be solicited.

Participation would have to be broad and representative, with balanced and accurate information presented. It would be important for policymakers to be part of the process to demonstrate to the public that their time and effort are worthwhile. Communities would need an understanding of the purpose of the project, how it is designed, and who would benefit from it. Community-based organizations and churches could serve as intermediaries. The knowledge gained should be bidirectional, going from the community to the researchers and from the researchers to the community, and should be iterative, using public input to modify design of the project and using aspects of the project design to "test the waters" for acceptability by the public. If the project were to go forward, multiple data collection methods using quantitative and qualitative measurements would allow investigators to examine public attitudes, beliefs, and concerns from several vantage points throughout the life of the project, from planning through implementation.

According to public engagement experts consulted by SACGHS, an LPS could generate mistrust, especially among members of racial and ethnic communities who have been harmed or underrepresented in research.[125] The extent to which communities support a study might depend on whether they embrace the concept of co-ownership of the study, which would likely create powerful advocates. According to these experts, the process would need to directly address the issues of race and racism and involve in meaningful ways various racial and ethnic groups and representatives of relevant national organizations. In addition, input on civil liberties and privacy issues should be obtained from prospective subjects and patient advocacy groups through such mechanisms as face-to-face venues and town hall meetings.

C. Public Engagement Models

Public engagement efforts could employ various strategies since they are not mutually exclusive. A number of approaches have been developed and could serve as models. Regardless of the public engagement strategy(ies) employed, public input must be considered and factored into policy decisions about whether to undertake an LPS and, if such a study goes forward, into all elements of the study itself.

---

[125] One recent study suggests that the willingness of minorities to participate in health research may not differ as much when compared with nonminority populations. See Wendler D, Kington R, Madans J, et al. Are racial and ethnic minorities less willing to participate in health research? *PLoS Med* 2006. 3(2):e19. Epub December 6, 2005.

The consultation efforts of four LPSs in Iceland, Estonia, the United Kingdom, and Quebec serve as models. Some were formal and structured, relying on a legislative or regulatory process; others involved a planned program of outreach to the public that included opinion-seeking and requests for comments. These models are described in more detail in Appendix E. The process used in developing pandemic influenza vaccination priorities provides yet another model.[126]

1. Public Consultation in the National Children's Study

If a decision is made to undertake a new LPS, the public consultation approach employed by the NCS might inform this new effort. The NCS approach involved conducting focus groups with five key stakeholder groups in 10 sites across the United States to gain a better understanding of the issues that would affect study recruitment and retention and to explore methods for involvement and support.[127] The stakeholders included parents, parents of children with disabilities, expectant parents, health care providers, and community organization representatives who reflected the study population.

All of the focus group discussions explored basic issues, such as how to generate and sustain interest in study participation, including the appropriateness of using incentives, as well as procedural and logistical issues associated with study participation. Quotes, key themes, and a tally of responses to specific questions were captured from the discussions with each stakeholder group and were used to create a comprehensive summary for preliminary analysis. Review and coding of the summaries from all sessions provided the basis for the second level of analysis and an overall synthesis of results.

The NCS focus group results indicated that all stakeholders thought that the study was important and that informational campaigns with complete information about study goals and stakeholder roles should be conducted for each type of stakeholder. The convenience of the study procedures was seen as critical to participation, and the use of certain types of incentives was considered appropriate. African American and Hispanic participants confirmed that mistrust of the study and study investigators could be an issue in their communities. The focus groups gave the NCS leadership a better understanding of the approaches needed to gain and maintain stakeholder commitment.

2. NHGRI Public Consultation Initiative

In 2005 NHGRI announced the availability of funds for a pilot public consultation study to obtain societal input to inform the design and implementation of one or more possible large U.S. population-based studies, including a longitudinal cohort study, of the role of genes and the environment in health and disease.[128] The goal of the project is to solicit opinions on the design and implementation of the study from members of the public who represent the demographic makeup of the country. The issues to be addressed include the acceptability of goals of the initiative for the United States as a whole; concerns regarding the uses of data for individuals, communities, and the public at large; expectations about privacy protection; the acceptability of open-ended consent; the acceptability of a central IRB; optimal approaches to recruitment,

---

[126] See http://www.keystone.org/spp/health-pandemic.html [accessed December 21, 2006].

[127] Dimitropoulos L, Rench J. *Final Report: National Children's Study Focus Groups*. Prepared by RTI International for the U.S. Environmental Protection Agency under Contract 68-D-02-069. See http://www.nationalchildrensstudy.gov/research/methods_studies/upload/Final-Report-National-Children-s-Study-Focus-Groups.pdf [accessed December 21, 2006].

[128] Request for Applications: Public Consultation to Inform the Design of Possible Large-Scale Studies of Genes and Environment in Common Disease (RFA Number: RFA-HG-06-008). See http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-06-008.html [accessed December 21, 2006].

particularly regarding identifying and contacting family members; the need for tailoring the study to individuals or communities with special needs; expectations about the return of information to individuals, communities, and the public at large; the need for ongoing dialog with participants regarding study goals and processes; the advisability of including or excluding children; and IP concerns. Public input is to be obtained through several methods, including a Web-based survey of 4,000 individuals, focus groups in 6 U.S. cities, and town hall meetings in 5 States. Community leaders will be interviewed as well. Educational materials on large, population-based studies will be developed for all participants.

Funds totaling $2.1 million were awarded in September 2006 to the Genetics and Public Policy Institute (GPPI) at Johns Hopkins University using a cooperative agreement mechanism that will enable NHGRI to collaborate with GPPI in the design of survey instruments and in the development of focus group guides and final protocols for a series of public, town hall-type meetings.[129]

The findings of the pilot public consultation study are to be analyzed as they proceed, and an overall analysis of the findings is to be conducted. A preliminary analysis of the data is to be completed by September 2008. If determined to be feasible and if funding is forthcoming within the next few years, the findings will be incorporated into the design of this LPS, its full-scale public consultation component, and other population-based studies. If a large-scale study proceeds after the completion of the pilot project, additional consultation and more targeted community engagement would take place with the specific communities that are to be recruited.

D. Recommendations: Public Engagement

Given the scope, magnitude, cost, and timespan of developing an LPS, the significance of research findings that would result from analyses of the data/specimens, and the need for broad public support and participation, the public must be consulted about the project's value, design, implementation, and application of research results.

SACGHS suggests several options for engaging the public in discussions and decisions about undertaking an LPS, including consulting with the scientific and international communities, communities that might be involved in the research, health care providers and their institutions, and those who volunteer to participate in the project as research subjects.

1. **Before embarking on such a large population study and in advance of any funding decision, the HHS Secretary should assess the public's willingness to participate in such an extensive endeavor.**
2. **If a decision is made to proceed with a large population study, it will be important for the HHS Secretary to ensure that public engagement occurs throughout all aspects and stages of the research process, from conceptualization through design, planning, implementation, conduct, and data analysis and reporting. Public engagement also will be important in applying the knowledge gained by the research and in addressing its implications. The HHS Secretary should ensure that sufficient project resources are dedicated to public consultation activities before and throughout the duration of the project.**

[129] See http://www.dnapolicy.org/resources/NIHPressRelease9-28-06.pdf.

V. Conclusion

In 2004 SACGHS identified as a priority of the Committee the need for an analysis of the opportunities and challenges associated with conducting an LPS aimed at understanding the relationships of genes, environments, and their interactions and common complex diseases. In considering the prospect of undertaking such a study in the United States, SACGHS first analyzed the NHGRI document *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel*.[130] The document lays out specific scientific and logistical details for a 1 million-person cohort study and concludes that the time is right for the United States to consider such an endeavor.

The Committee created a Task Force to guide its exploration of the issue and subsequently held factfinding and consultation sessions on the topic with experts from a variety of relevant disciplines. In June 2005 SACGHS was asked by NIH Director Dr. Zerhouni to identify broad policy and process issues that need to be addressed before the United States considers the question of whether or not to undertake an LPS. Specifically, Dr. Zerhouni suggested that the Committee:

1. Delineate the questions that need to be addressed for policymakers to determine whether the U.S. Government should undertake an LPS to elucidate the influence of genetic variation and environmental factors on common complex diseases.
2. Explore the processes by which the questions that are identified in Step 1 can be addressed, including any intermediate research studies, pilot projects, or policy analysis efforts needed.
3. Determine the possible ways in which these questions could be addressed, taking into account the feasibility of those approaches that seem optimal, and recommend a specific course of action for moving forward.

In accordance with Dr. Zerhouni's guidance, this report addresses these three elements. Over the course of almost 2 years, SACGHS has investigated extensively the policy issues associated with a large cohort study in the United States, as well as their implications. The Committee consulted experts in the areas of science policy; genetics, genomics, and epidemiology; ethical, legal, and social issues; public health; and public engagement. SACGHS also examined the policies of numerous large cohort studies already underway or being considered in this country and abroad to inform their deliberations. Members of the public were invited to weigh in on the issue. In response, representatives of the public articulated their hopes and concerns related to a new study that would involve 500,000 to 1 million U.S. citizens over a period of many years. Nearly 70 sets of thoughtful comments representing diverse perspectives were received from academic researchers, industry representatives, Federal policymakers, policy research groups, IRBs, professional societies, nonprofit advocacy groups, and individual consumers of health services. SACGHS has carefully considered all of this information, with the goal of identifying for the HHS Secretary the preliminary issues that must be addressed before a decision can be made about whether to undertake an LPS in the United States.

[130] Ibid.

In this report, SACGHS presents salient issues for consideration by the HHS Secretary in the key areas of research policy; research logistics; regulations and ethics; public health, social, and economic implications; and public engagement. In each chapter, a comprehensive discussion of each area concludes with recommendations for next steps by the HHS Secretary.

In addition, the Committee identified a single overarching policy recommendation to guide the HHS Secretary as he considers the possibility of an LPS in the United States:

As part of the process for determining whether to undertake such a large-scale research project, the HHS Secretary should initiate a thorough consideration of the full range of policy issues outlined in this report. The HHS Secretary should consult and engage the full range of potential partners for such a project during this decisionmaking process, including the public at large, the full scientific community, a wide spectrum of Government agencies and policymakers, and the private sector.

There is little doubt that a well-designed LPS to identify the environmental and genetic determinants of common diseases, such as cardiovascular disease, diabetes, and cancer, has potential benefits. For example, such a study could lead to the development of new and improved technologies resulting in more effective diagnostics, treatments, and preventive measures. For these reasons, SACGHS urges the HHS Secretary to proceed expeditiously with the significant analytical and consultative efforts that need to be undertaken to fill the policy gaps identified in this report, particularly in the areas of public opinion, data sharing, IP rights, and human subjects protections. Until this analytical work is completed, which requires the broad and inclusive consultation called for in the Committee's overarching recommendation, it will not be clear whether the study is feasible from scientific, social, economic, and ethical perspectives or whether an LPS has the broad public support necessary to move forward. Irrespective of whether such a study is determined to be feasible at this time, evaluation of the issues presented in this report will serve to provide the foundational knowledge to aid in the consideration of the many other important research and public health policy initiatives that are a part of the overall mission of the Department of Health and Human Services.

Appendix A

March 2005 SACGHS Factfinding Session on a Large Population Study

Large Population Studies: Opportunities and Challenges

Huntington F. Willard, Ph.D.
Chair, SACGHS Task Force on Large Population Studies

Dr. Willard provided background about the purpose of the session and an overview of the presentations. The main goal of the session was to gather information about the opportunities and challenges associated with conducting large population studies (LPSs) that focus on determining genetic and environmental factors underlying common complex diseases. Six presentations were organized to provide the Committee with information about the scientific, logistical, ethical, legal, and social issues. Three presentations focused on the scientific aspects of large cohort studies, including their benefits and challenges, and three explored the logistical, ethical, legal, and social aspects of such studies. Federal representatives from four agencies provided information on some Federal programmatic efforts in this area, the importance of global collaboration among those who conduct population-based studies around the world, and the recommendations of a Federal working group that was commissioned to examine the goals, scientific foundation, design elements, and logistical outlines for a hypothetical new LPS in the United States.

Conceptual Basis for Large Population Studies of Human Genetic Variation and Common Disease

David Goldstein, Ph.D.
Wolfson Professor of Genetics, University College of London

Dr. Goldstein explained that common diseases are not caused by a one-to-one correlation between a genetic variation and a disease. There are more than 10 million places in the human genome where individuals differ, and these variations change physiology in subtle ways. Deoxyribonucleic acid (DNA) variations can interact with one another as well as with environmental factors to influence individual disease susceptibility. The aim of an LPS is to understand how these genetic differences influence health.

He said the field is very good at sequencing and genotyping but has not made much progress relating genetic variation (i.e., single nucleotide polymorphisms [SNPs]) to human disease. Researchers have not yet been able to compare in a comprehensive manner the genetic variation of very large numbers of people who have the same health condition with those who do not have it. A relatively new method called haplotype tagging is now being used to help address this gap. It uses knowledge of a small subset of SNPs to predict information about adjacent SNPs. Selecting an appropriate subset of variable sites in one individual can provide information about the genetic variation of others with the same trait. Using haplotype tagging,

scientists are attempting to create a framework for relating the presence of particular SNPs with clinical traits, such as disease.

Using a data set collected by the University College of London, Dr. Goldstein conducted a study using haplotype tagging that looked for patterns of association among 55 genes that encode major drug-metabolizing enzymes in individuals of European and Japanese ancestry. The study, carried out in collaboration with GlaxoSmithKline, showed that a subset of 200 SNPs within the 55 genes was representative of the other 4,000 SNPs estimated to be within those genes. These findings and others like it demonstrate that haplotype tagging is an efficient and economical method for representing common genetic variation. The technique is not as useful for representing rare genetic variations, and substantial work remains to be done in that area.

He pointed out that the main value of large longitudinal prospective cohort studies is that they allow investigators to gather a great deal of data over time before and after individuals develop common diseases, furthering the knowledge of the genetic and environmental factors that contribute to disease. One of the big challenges with such studies, however, is that investigators have not yet determined either the amount or types of information that should be collected at enrollment and during followup examinations or the areas that should be studied first. Dr. Goldstein suggested that it would be important for researchers who conduct such studies to interface carefully with health care providers to make sure they capitalize on the most important types of information. He concluded by suggesting that the study of genetic predisposition to disease has been overemphasized and that more research is needed on how patients respond to treatment.

Public Health Perspective on Large Population Studies of Human Genetic Variation, the Environment, and Common Disease

Gilbert S. Omenn, M.D., Ph.D.
Professor of Internal Medicine, Human Genetics, and Public Health, University of Michigan

Dr. Omenn began by noting the vast amount of genetic information that has been produced as a result of the sequencing of the human genome and the recently completed haplotype map project (International HapMap Project) and the growth in the understanding of genes and alleles associated with disease risk. He said that environmental and behavioral data sets must be improved and linked with genetic information and that privacy and confidentiality protections must be in place for these data sets. The new technologies and therapeutics developed from these data linkages have the potential to result in medical breakthroughs that reduce health risks and allow for cost-effective treatment.

Dr. Omenn described the National Health and Nutrition Examination Survey (NHANES), a survey sponsored by the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics, which has been collecting survey and laboratory data for 40 years. The analysis of NHANES data has resulted in significant public health improvements. Ongoing data analysis is providing information on various environmental exposures. Recently, a collection of DNA from approximately 7,000 people was added as a component of the survey.

He pointed out that other public health research tools and approaches are becoming available, such as geographic information systems, wireless devices to track exposure to pesticides, and spatial models for households at risk for lead poisoning. Biosensors and nanoscale devices also permit feasible measurements of an individual's environmental exposures.

Dr. Omenn suggested that a conceptual strategy for integrating exposure assessment technologies into research would need to identify priority diseases and plausible environmental factors; potential genetic determinants, pathways, and model systems for exploring gene-environment interactions; and target study populations for feasible measurement. Such a strategy should define the genetic determinants of susceptibility, conduct targeted exposure assessments, and identify and validate biomarkers. This information then would be brought together with descriptions of gene-environment interactions. Literature searches, computer-based pathway mapping, body burden assays, genomic screening, and biomarker studies are among the technologies and approaches that could be used.

Programs in measurement and modeling sponsored by the U.S. Environmental Protection Agency (EPA), such as the Multimedia Integrated Modeling System (MIMS), are directly relevant to the study of common diseases in large populations. The MIMS simulates ambient airborne substances in urban settings, which then can be utilized to assess individual exposures. EPA also is working on prototypes and advanced modeling tools to measure human exposure to air pollution.

From a public health perspective, he noted that the fields of genetic research and environmental science share many interests. Both are attempting to unite heritable information with environmental cues such as nutrition; metabolism; lifestyle behaviors; pharmaceuticals; and chemical, physical, and infectious exposures.

Dr. Omenn emphasized the need to include genetics in protocols for health promotion and disease prevention, including host-pathogen interactions and drug and vaccine development, and the importance of population-based studies that can make sense of genetic variation. In his view, it would be tragic if the extensive knowledge of genetic variation gathered through the Human Genome Project (HGP) and the International HapMap Project, as well as others, were not used to reduce health risks. Dr. Omenn concluded with two key findings from the Partnership for Prevention report *Harnessing Genetics To Prevent Disease & Improve Health: A State Policy Guide*: (1) The greatest opportunity of the genomic era is to use personalized medicine to prevent or better manage chronic diseases, and (2) genetics and genomics should be integrated into existing health, social, and environmental policies, rather than establishing stand-alone genetics programs.[131]

Overview of International and National Large Population Studies

Teri Manolio, M.D., Ph.D.
Director, Epidemiology and Biometry Program, National Heart, Lung, and Blood Institute

Dr. Manolio provided an overview of several national and international large cohort studies and discussed their designs. Biobank Japan, which began in 2003, is expected to include 300,000 people age 20 years and older. It is focused on 47 common complex diseases and is endeavoring to understand their causes, as well as examine medication side effects related to genetic variation. The hope is that this knowledge will support the development of new drugs and diagnostics, which also will provide a source of financial support for the study. A network of collaborating organizations and private universities is collecting samples and data; 90,000 samples have been collected to date. Access to the data is limited to Japanese researchers.

[131] See http://genes-r-us.uthscsa.edu/resources/genetics/geneticsguide.pdf [accessed February 20, 2007].

CARTaGENE™ is a Canadian study that focuses on common diseases and anticipates enrolling about 50,000 people ages 25 to 74 years. GenomeEUtwin is taking place in 7 European countries with 800,000 twin pairs. The investigators are focusing on 7 key outcomes.

The goal of the Estonian Genome Project is to improve health care by trying to understand the role of genes and environmental factors in common diseases. The study began in 2002 and has recruited 10,000 subjects, with plans to enroll 100,000. Subjects complete a 60- to 90-minute questionnaire that includes genealogical information going back three to four generations. Other data collected include height, weight, blood pressure, and heart rate, as well as a blood sample. With subject consent, individual test results can be provided to an individual's physician. Those who participate are called gene donors and have access to a Web site that answers questions about the project.

The Marshfield Clinic Personalized Medicine Research Project in Marshfield, Wisconsin, began in 2002 and anticipates enrolling 40,000 subjects age 18 years and older who are patients of the Marshfield Clinic Research Foundation, a large group of private clinics with a longstanding electronic health record program. The project requires completion of a written informed consent document and a 30-minute questionnaire, as well as DNA analysis and blood extraction. The study is intended to translate genetic data into knowledge that will enhance patient care, focusing on adverse drug reactions.

deCODE genetics, Inc., is a biopharmaceutical company in Iceland that applies discoveries in genetics to the development of drugs for common diseases. The Icelandic population has been geographically isolated on an island in the North Atlantic for hundreds of years and is homogeneous, resulting in less genetic variation. Because genealogy is an important part of the Icelandic culture, there is an extensive genealogic database extending back to 900 A.D. when the island was first settled. Because many individuals can trace their ancestry back six generations, scientists often can find a common ancestor between two people with the same disease. The investigators anticipate enrolling the island's entire population of approximately 200,000 people to study 50 common diseases. A proposal for an opt-out consent for access to medical records and inclusion in a health sector database that would be publicly accessible generated tremendous controversy around the project. The opt-out concept was ultimately dropped and replaced by an explicit written informed consent process.

Dr. Manolio described the pros and cons of large-scale cohort studies. Although they are expensive, take a long time to conduct, require access to large numbers of people, are broadly based, and are sometimes criticized as "fishing expeditions," they provide risk information that cannot be obtained any other way, are understandable to the public and the media, and can identify modifiable risk factors for potential preventive interventions. Ideal design characteristics of these studies include large numbers of subjects; representative samples that can be generalized and are diverse in terms of geographical location, socioeconomic status, and race/ethnicity; extensive characterization of environmental exposures, risk factors, and diseases at enrollment; repeated interim measurements to assess change in measurements taken at enrollment and to add new exposure measurements; and a comprehensive, standardized assessment of outcomes. Failure to ensure these characteristics can result in study results that are biased and lead to erroneous conclusions.

With regard to case-control studies, Dr. Manolio pointed out that these studies can be problematic when appropriate design strategies are not followed. Three assumptions must be made to ensure a bias-free case-control study: (1) the cases must represent all persons who develop the disease or condition; (2) the controls must be representative of the general "healthy" population members who do not develop the disease; and (3) the collection of risk factor and exposure information must be the same for cases and controls. Case-control studies have advantages in that they allow for the study of rare diseases; existing health records often

can be used; multiple etiologic factors can be studied simultaneously; they can be less time consuming and expensive; and if the assumptions are met, the inferences are reliable. The potential disadvantages of case-control studies include data obtained through recall or from records that may be incorrect or incomplete; the difficulty of selecting an appropriate comparison group; biases that may be created due to spurious evidence of associations between risk factors and disease; rare exposures that usually cannot be studied; and the difficulty of determining the temporal relationship between exposure and disease.

Dr. Manolio concluded by introducing the possibility of nesting smaller case-control studies within a larger prospective study as a potentially powerful study design. Rather than amassing separate data sets on a disease-by-disease or protocol-by-protocol basis, which could become prohibitively expensive, biological samples and health information could be collected, stored, and later accessed for a variety of research analyses.

Ethical, Legal, and Social Issues of Large Population Studies

Mylene Deschenes, L.L.M.
Executive Director, Public Population Project in Genomics

Ms. Deschenes stated that researchers are increasingly studying common complex diseases through national and international collaborations. The field has moved from relying on small specimen collections (e.g., samples in a freezer) to large repositories containing DNA and associated clinical data organized in a systematic way for research purposes. These large-scale repositories are often referred to as "biobanks" or "human genetic research databases" (HGRDs), and they generally contain data on at least 10,000 individuals.

She noted the proliferation of national and international legislation and declarations related to HGRDs, some of which have been adopted by various organizations, such as the World Health Organization and the Human Genome Organisation. Some countries have implemented legislation that specifically relates to HGRDs, whereas others apply preexisting legislation to all research. This variety of systems is confusing and demonstrates the need for harmonization of principles and terminology across jurisdictions. Several countries, such as Israel, Australia, France, Germany, and Canada, are actively addressing these discrepancies in rules and terminology.

The lack of internationally agreed-upon rules is detrimental to research collaborations, database compatibility, and data sharing. At the national level, there is a need to recognize that HGRDs are not time-limited research projects but rather resources that will be used in the future for multiple applications. Traditional consent and privacy legislation was not written with HGRDs in mind.

Ms. Deschenes stated that the field needs a coherent and comprehensive regulatory framework that will protect participants and structure the conduct of this large-scale research. A consensus is emerging on the areas that need attention. Traditional consent mechanisms should be changed to take into account obligations to protect participants' identities over many years. In addition, ethical oversight and monitoring mechanisms should be put in place at the inception of a new database. The professional and public dialog must be strengthened so that the public can become active participants in the research.

She discussed three key elements that should be considered when establishing an HGRD. The first is ensuring the legitimacy of the database by protecting it with appropriate checks and balances. The use of

tremendous financial and other resources must be justified, and the potential benefits must be communicated to stakeholders and the media. In countries such as Estonia and Iceland, legislation was passed to establish and legitimize their HGRDs. They have adapted the science to the community's needs based on open discussion. However, transnational enterprises, such as the International HapMap Project, GenomEUtwin, and the Public Population Project in Genomics (P3G), are more complex; the success of these projects depends on trust and communication among the participants and is based on a common understanding of the scientific, ethical, legal, and social issues.

The second key element in establishing an HGRD involves building trust on various levels. Establishing public trust depends primarily on communication with the community from the outset. Ideally, all groups should be represented in the sampled population; however, financial constraints affect the selection process. It is important for data collectors to be sensitive to the ethical, legal, and social issues involved. Ms. Deschenes noted that trust is enhanced through the use of an appropriate consent process. The research team must make decisions about providing individual feedback and about whether to share results with the public. The commercial aspects of the project also must be taken into consideration, because matters of free, public access versus intellectual property (IP) rights likely will arise. Industry involvement can provide the financial resources needed for these kinds of studies, but questions about effective ways to involve industry have not yet been answered.

The third element that must be considered is the need to ensure adequate checks and balances in the governing of HGRDs. A project's framework and protocol assessment need a "stamp of approval" from authorities and should have input from the public. Ongoing review procedures should be a component of the project.

Ms. Deschenes concluded by emphasizing the importance of the monitoring and oversight of research projects and the public resources used and that adjustments to the protocol may need to be made over time.

Dichotomy Between Social Identity and Ancestry in Large Population Studies

Charles Rotimi, Ph.D.
Acting Director, National Human Genome Center, Howard University

Dr. Rotimi addressed the representation of different groups in LPSs in terms of social identity and ancestry. He explained that investigators seek different "levels of resolution," depending on the purpose of the study. In the case of large cohort studies, the levels of resolution relate to the characteristics of the population being studied. Dr. Rotimi said that in the human population, groups of people are socially identified (and self-identify) in an ever-evolving way that reflects political, economic, and other nongenetic factors. It is important, he suggested, to take this idea into account in the design of any LPS.

Dr. Rotimi noted that the prevalence of disease in a certain population might be related to etiology of disease and/or disparities in health care. He suggested that scientists should be responsible for defining specifically how each factor will be studied. Although it sometimes may be clear that health care disparities are the primary cause of disease, the ways in which health disparities play out in different ethnic groups will be less obvious and, as such, may need to be considered when developing strategies for conducting large cohort studies.

Dr. Rotimi emphasized that a study's design strategy must clarify which variables are being examined and that ancestry should be considered. For example, in the United States, although the ancestral history of African Americans is very broad and represents a diverse genetic history from many countries, scientists often group all African Americans together as if their characteristics were uniform and homogeneous. In such cases, group identity is confused with group ancestry. The same thinking holds true for the "Hispanic population," a label the census applies to those with such varying origins/ancestries as Mexico, South America, Cuba, and Puerto Rico. He noted that the label "Asian" is used to refer to people from at least seven countries. Dr. Rotimi stated that much information is lost when people are grouped together and studied as if they have the same biological ancestry. He closed by emphasizing that because ancestry is not absolute, it is important for scientists to realize that some level of compromise will be needed in defining the population studied when designing large cohort studies.

UK Biobank

John Newton, Ph.D.
Chief Executive Officer, UK Biobank

Dr. Newton explained the importance of large prospective cohort studies of genes and the environment and their relationship to health and illness. This type of research is possible because of the transformation of biomedical science following the HGP. He noted that it is important to take this next step to see improvements in public health. Dr. Newton said that prospective studies are feasible, that the public responds well to them, and that they must begin soon because it may take at least a decade, if not two, to achieve meaningful results. A sense of urgency notwithstanding, it is important to pay close attention to issues of quality, value, comprehensiveness, and the scope of future results.

He reported that UK Biobank is designed to support long-term prospective studies and that it will enroll 500,000 participants ages 40 to 69 years. Although the specific study questions have not been formulated, the data collected at baseline are expected to include environmental exposures, physiological variables, neuropsychiatric evaluations, biochemical markers, and biological samples (DNA, blood, and urine). The subjects will be tracked through the United Kingdom's National Health Service and the Biobank's registration data. Investigators plan to study the subjects over a period of decades.

Dr. Newton said the overall goal of UK Biobank and other LPSs is to understand disease models in a way that has never been possible before. The three main scientific objectives of the UK Biobank project are to (1) determine the separate and combined effects of genes and the environment on the common causes of illness through nested case-control studies, prevalence studies, and exposure-based cohort studies; (2) genotype 500,000 people within 5 years for a limited number of SNPs; and (3) identify biomarkers as early risk factors for common diseases.

Most of the approximately $110 million in funds for the UK Biobank project (about 1 percent of the amount spent on biomedical research in the United Kingdom) comes from the Medical Research Council and the Wellcome Trust, a large biomedical research charity. Other funders include the U.K. Department of Health and the Scottish Executive. He stated that the value of the resources available in the Biobank would increase continually over time.

In terms of organization and governance, UK Biobank is an independent company with two separate advisory panels, the Science Committee and the Ethics and Governance Council. Implementation of UK

Biobank will be conducted by 6 regional collaborating centers that represent scientific groups around the country, including 22 universities. The project uses modern, efficient methodologies in patient recruitment, data collection, information technology, genotyping, and genomics. There is strong central coordination and scientific collaboration.

Dr. Newton addressed the recruitment process, stating that direct mail will be used. The project will start slowly in the first year to ensure that sound procedures are in place before it is implemented on a large scale. Subjects will register with UK Biobank through a dedicated clinic that will conduct data collection. Samples and data will be transported to a central resource and archived. Blood samples will be stored in a way that makes it possible to conduct genetic, proteomic, and metabolic studies. He noted that data management systems are in development by commercial suppliers that are being advised by experienced researchers.

Dr. Newton noted that UK Biobank participants will be asked to provide broad consent for future use of biological samples and health data and will be assured of data security and confidentiality. Participants will be able to withdraw from the study at any time. UK Biobank will retain control of the samples, but there will be full access to the data for appropriate purposes. Internal and external reviews will be carried out to ensure that all uses of the data are scientifically and ethically appropriate.

He emphasized the importance of collaboration and said that others should be encouraged to conduct similar studies. Dr. Newton cautioned that the systems used must have the ability to interact with one another and that large-scale studies are more easily conducted in countries with universal health care coverage, such as Canada and the countries of Scandinavia.

He also commented on the advantages of prospective cohort studies over case-control studies. Whereas, for example, cohort study data can provide a resource for investigating genetic risk factors for various diseases and other determinants of health and disease, case-control studies that provide information only on the disease being studied will not be able to predict unforeseen outcomes. Dr. Newton said that additional benefits of prospective studies include the promotion of high standards for ethics and governance in the field, broadened access to expensive research resources, opportunities to collaborate internationally, and an efficient and economic approach.

He concluded with a summary of the unique features of UK Biobank, which include its size, unprecedented biological resources, ability to recall subsets of individuals for intensive phenotyping, extensive use of written records, and an ethical approach that encourages public participation.

Federal Perspectives on Large Population Studies

Four Federal representatives provided information relevant to the topic of LPSs. Dr. Ruth Brenner of the National Institute of Child Health and Human Development (NICHD) presented background information on the National Children's Study, followed by a study update and timeline. Dr. Stephan D. Fihn of the U.S. Department of Veterans Affairs (VA) discussed the possible use of VA's extensive integrated health system for genomic research. Dr. Alan Guttmacher of the National Human Genome Research Institute (NHGRI) presented some design considerations for a new large population cohort study in the United States, including the recommendations of a working group that endorsed the study. Dr. Muin J. Khoury of CDC described the importance of global collaboration by those who conduct population-based cohort studies around the world.

Ruth Brenner, M.D., M.P.H.
National Children's Study, National Institute of Child Health and Human Development

Dr. Brenner presented an overview of the National Children's Study (NCS), a large-scale national prospective cohort study of environmental influences (e.g., physical, chemical, biological, and psychosocial) on children's health and development. The NCS was authorized by the Children's Health Act of 2000 (Public Law 106-310) and reflects growing concern about the numerous exposures, diseases, and conditions of children. A consortium of Federal agencies is involved, with NICHD assigned lead responsibility.

The study is hypothesis driven, with primary outcomes related to child health and development, and involves 100,000 children, starting from before birth and continuing through age 21 years. The focus is on children because they have an increased vulnerability to environmental exposures, critical windows of vulnerability during development, immature mechanisms for detoxification and protection, and differences in metabolism and behavior that may result in higher exposure than adults in the same environments. The longitudinal study design will allow inferences about causality and the study of multiple outcomes, multiple exposures, and mediating pathways between exposure and disease. Recall bias in relation to exposure will be minimized.

She described NCS milestones. The decision to use a national probability sampling method was made in 2004. In the first stage, 101 study locations were drawn from a list of all counties in the United States. In the second stage, investigators will select segments or groups of households within the study locations. They anticipate having a highly clustered sample to facilitate the study of community characteristics and increase logistical efficiency. Dr. Brenner stated that, from the initial list of study locations, eight were selected to serve as potential vanguard locations.

The study will enroll women and their partners prior to or early in pregnancy, with followup of children until 21 years of age. Enrollment will take place over a 4-year period, and data will be collected through face-to-face visits and remotely. Data will be collected via questionnaires and interviews, environmental samples and observations in the home and community, clinical and behavioral assessments, and biologic samples (blood, urine, cord blood, placenta, and breast milk). Initial centers were scheduled for selection in late 2005, and the initial protocol was to be completed and piloted in 2006. It was hoped that the first participants would be enrolled in early 2007.

The investigators planned to continue to hold meetings, peer reviews, workshops, and consultations. In September 2004 they held a workshop on the collection and use of genetic information that brought together experts in the Federal Government (National Institutes of Health [NIH], EPA, CDC, and the Food and Drug Administration) to explore opportunities and challenges and to provide recommendations to the NCS. The workshop report and recommendations are available on the NCS Web site.[132]

[132] See http://www.nationalchildrensstudy.gov [accessed February 20, 2007].

Stephan D. Fihn, M.D., M.P.H.
Proposed VA Genomic Medicine Program, U.S. Department of Veterans Affairs

Dr. Fihn reported that the proposed VA Genomic Medicine Program has been evolving for several years and is still in the early stages of development. If it moves forward, its goals will be to study the role of genetic factors in the cause and prevention of disease, develop clinical programs targeting therapeutic drug response and prevention of adverse reactions, and develop information systems to confidentially manage genetic data for patient care and research.

He pointed out that VA is interested in this project because the Department has the largest integrated health care system in the United States and possibly in the world. In addition, VA has a very stable patient population. The approximately 5 million users have electronic health records that contain copious amounts of clinical data. There is also a large intramural research program at VA, with many investigators working in the area of genomics on a small scale and utilizing existing resources, such as several sanctioned DNA repositories.

Since it is a Federal health care system, Dr. Fihn said that VA would insist on absolute control and ownership of all materials and information gathered. A stringent set of policies for human subjects protections, IP, conflict of interest, privacy, and scientific merit would govern the project. Additional protections that are being developed involve an independent oversight board composed of Federal and private representatives.

He said the protection of confidentiality, development of collaborations with other researchers, funding, IP issues, and decisions regarding the types of specimens to collect are ongoing challenges that must be addressed.

Alan Guttmacher, M.D.
American Gene-Environment Study, National Human Genome Research Institute

Dr. Guttmacher presented the case for undertaking a new LPS of genes, environment, and disease in the United States. He began by describing the different approaches to discovering and quantifying genetic and environmental contributions to disease risk, including case-control studies and prospective, population-based cohort studies. Dr. Guttmacher said that although case-control studies have benefits, they also have several shortcomings, including a bias toward the more severe end of the disease spectrum, a recall bias for environmental exposures and family history, and an inability to identify predictive biomarkers that signal the future onset of disease.

He noted that other countries are planning LPSs of genes, environment, and health; however, these studies would not adequately substitute for a major project based in the United States. Other countries do not reflect the population groups or the environmental factors found in the United States, and access to data from other countries' studies by U.S. researchers is limited.

Dr. Guttmacher described the American Gene-Environment Study as a working concept designed to create a large resource for the research community. He reviewed the recommendations of a working group organized to consider the design of such a study. The recommendations were that the cohort should be chosen to match the most recent U.S. census on the characteristics of age, sex, race/ethnicity, geographic region, education, and urban/rural residence; the household should be the primary sampling unit, with roughly 30 percent of cases consisting of biologically related individuals; the cohort should be of significant size to achieve adequate power for most common diseases and quantitative traits; participants should undergo a 4-hour

baseline clinical assessment, during which a core group of measures and some age-specific measures would be collected; and biological specimens should have core laboratory measurements taken and undergo DNA sequencing and genotyping. The working group also emphasized the need for extensive public consultation through town meetings and focus groups.

Followup telephone or e-mail contact with participants would occur every 6 months, with reexamination every 4 years. The investigators would seek open-ended informed consent and create an encrypted database to protect privacy and confidentiality. Data would be made immediately accessible to all investigators with institutional review board (IRB) approval. In terms of oversight of the research, Dr. Guttmacher indicated that a central IRB might be highly advantageous.

He stated that the time for this study is now, for several reasons: The study would provide opportunities to understand and address health care disparities, be a powerful stimulus for technology development (e.g., driving innovation in the measurement of environmental factors), and have the potential to reduce skyrocketing health care costs through a better understanding of disease etiology and treatment. He closed by posing the rhetorical question, "Can we afford NOT to do something like this?"

Muin J. Khoury, M.D., Ph.D.
Building the Knowledge Base on Genes and Population Health: Need for Global Collaboration, Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention

Dr. Khoury discussed the importance of building the knowledge base on genes and population health, emphasizing the need for global collaboration between biobanks and population-based cohort studies. He said that pooled analyses would increase the chance of finding true associations relevant to public health and noted the need for systematic integration of all human genome epidemiology studies, regardless of the type of study design, and the need for evidence-based processes that use human genome epidemiologic data to assess the value of genomic information in health care and disease prevention. Dr. Khoury stated that epidemiological data can be used to characterize the prevalence of gene variants in a population and help determine how these factors affect the relative and absolute risks for disease.

He challenged the idea that association studies are inferior to cohort studies. Because some association studies have not been well designed (e.g., the cases and controls came from populations that are not comparable), some genetic investigators have dismissed all association studies. Dr. Khoury suggested that cohort studies are not inherently superior to case-control studies, stating that a well-designed, population-based, case-control study is far superior to a poorly designed cohort study. Case-control studies can be very valuable, especially for rare outcomes. More than 15,000 association studies, mostly case-control studies, had been published in the previous 3 years. An increasing number of studies focus on gene-gene and gene-environment interactions, and some provide prevalence data on different genetic variants in populations. CDC is analyzing a 5-percent random sample of this database of literature to examine the quality of the association studies.

He described the Human Genome Epidemiology Network (HuGENet), developed by CDC and many partners in 1998. HuGENet represents a global collaboration of individuals and organizations and was formed to assess the population impact of genomics and determine how this emerging field could be used to improve health and prevent disease. HuGENet currently includes about 700 people from 40 countries and is open to anyone who wants to join. The Network provides technical assistance and training through workshops and has been developing a knowledge base so that information can be disseminated for both policy and practice. Dr. Khoury noted that HuGENet has been sponsoring systematic reviews of gene-

disease associations in collaboration with six journals and has a database of 200 meta-analyses of other gene-disease associations, is developing an information-sharing system among 14 other networks around the world, some of which study cancer or heart disease, and also collaborates with various biobanks.

He described a CDC meeting held in collaboration with P3G and NIH to discuss the harmonization of epidemiologic data through the development of standard criteria. The meeting resulted in a statement about the publication of studies that will be derived from biobanks. The statement relates to a worldwide movement for standards for epidemiologic studies outside of genetics.

Dr. Khoury described CDC's collaboration with NIH to estimate the prevalence of the top 50 genes of public health significance using a sample of approximately 8,000 individuals who participated in the NHANES. He also described a population-based, case-control study that uses surveillance systems to examine the association between genes and the environment and birth defects. Dr. Khoury stated that most single genes alone do not contribute to the etiology of the diseases being studied. Large sample sizes and replication across studies are needed to discover relative risks and odds ratios.

He said that the knowledge base on genes and population health can be built through different approaches, including a single large population cohort study; systematic synthesis of data from existing and planned cohort studies; systematic synthesis of data from all epidemiologic studies (cohorts, case control, and other); accelerated systematic synthesis of group and individual data using collaborative networks; and consortia of all types of studies (cohorts, case control, and other).

Committee Discussion

Dr. Reed Tuckson, SACGHS Chair, asked whether NIH, CDC, and other Government agencies represented at the meeting were working together on the issue of LPSs, including a plan to coordinate resources and approach the HHS Secretary. Dr. Guttmacher and Dr. Khoury replied that extensive discussions had taken place among the agencies and that their efforts would interrelate and be complementary, although no document had been prepared for the Secretary's review. They clarified that all the agencies are funded through the same appropriations committee, with the exception of VA.

Dr. Willard asked whether the Committee needed additional information before developing recommendations for the HHS Secretary. Some Committee members commented that, although it is critical to build on the knowledge gained through the HGP, they did not believe they had enough information to recommend an LPS to the HHS Secretary.

Committee members discussed the difficulties that would be inherent in a centralized IRB mechanism with responsibility for overseeing numerous local IRBs. Dr. Guttmacher acknowledged that this issue has created challenges for biomedical researchers, as multicenter studies have become more common. He stated that the nongenetics communities have looked at the question of centralizing large studies and have begun to develop some models that would provide guidance if a new LPS were undertaken. Dr. Michael Carome, SACGHS ex officio member from the HHS Office for Human Research Protections (OHRP), pointed out that the HHS regulations governing HHS-funded research with human subjects allow for cooperative or joint review arrangements for multicenter trials. In recent years, these joint arrangements have become increasingly useful, as the number of such trials and IRB burdens have increased, and that a central model could be implemented with no regulatory or policy changes within OHRP. Barriers to the use of the model, however, often exist at the institutional level. For a number of reasons, there is reluctance on the part of

some institutions to rely on another institution's IRB. Committee members acknowledged the issue and suggested that educational efforts among local IRBs would be needed and that this idea could be included in any recommendations made to the HHS Secretary.

The Committee discussed approaches to public engagement and decided that there was a knowledge gap in this area and that SACGHS needed more information from external experts. It was agreed that the public should be extensively involved in any plans to conduct a study that could enroll a million or more U.S. citizens from many different racial and ethnic groups.

The Committee acknowledged the complex questions that would be have to be addressed during the design stage, such as who would have access to the new treatments developed. Dr. Guttmacher suggested that the Committee should call the importance of an LPS to the HHS Secretary's attention but refrain from recommending specific study design approaches. He said that the emphasis should be on the potential value of these studies and on general design issues that would make a study effective.

Dr. Khoury suggested that the Committee was well situated to recommend an initiative to the HHS Secretary that would take the HGP to the next level. He said the Committee should encourage HHS to develop agency-wide collaborations to determine how the human genome influences health and address the number of studies that should be conducted.

Dr. Tuckson noted that a letter to the HHS Secretary would need to be specific concerning the study's aims, coordination of resources, and planned deliverables, and he asked the group to comment on the level of detail that should be included. Committee members suggested that the letter should address the idea of a public/private partnership, with an investment up front from the private sector. Dr. Emily Winn-Deen recommended a three-phase approach: (1) review the valuable work that has already been done and present it in a document, (2) identify knowledge gaps, and (3) fund studies to fill the gaps, either through the Federal Government or through a public/private partnership. Dr. Willard suggested a review of the NIH Working Group document[133] so that the Committee could take into account ongoing activities within NIH and CDC, which would inform their thoughts on an HHS-wide initiative.

In summing up the discussion and taking the pulse of the Committee, Dr. Tuckson suggested that the Committee prepare a letter to the HHS Secretary expressing the view that the initiative should move forward. The LPS Task Force was asked to develop a draft letter along the lines suggested by Dr. Winn-Deen emphasizing that public perspectives about an LPS remain an important knowledge gap and that specific efforts would be needed to gain the trust of U.S. citizens.

---

[133] *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships among Genes, Environment, and Health: Recommendations of an Expert Panel.* Op. Cit.

Appendix B

October 2005 SACGHS Consultation on a Large Population Study

Report from the SACGHS Task Force on Large Population Studies

Huntington F. Willard, Ph.D.
Chair*, *SACGHS Task Force on Large Population Studies (LPS Task Force)

Dr. Willard reported on the work of the LPS Task Force since the June 2005 meeting. The Task Force reviewed the National Institutes of Health (NIH) Work Group report *Design Considerations for a Potential United States Population-Based Cohort to Determine the Relationships Among Genes, Environment, and Health: Recommendations of an Expert Panel*. The National Human Genome Research Institute (NHGRI), in collaboration with several other NIH Institutes, commissioned a group of experts to consider some of the scientific and logistical questions raised by a hypothetical new large population study (LPS) in the United States. The Task Force also reviewed the article in *Nature* by NHGRI Director Dr. Francis S. Collins, "The Case for a U.S. Prospective Cohort Study of Genes and the Environment."

The NIH Work Group was established to examine the goals, scientific foundation, design elements, and logistical outlines related to a hypothetical U.S. LPS. The national experts involved represented fields such as genetics; epidemiology; biostatistics; and ethical, legal, and social issues in genetic research. The Work Group report recommended that the goals of such a study should be to ascertain and quantify all the major environmental and genetic causes of common illnesses in the United States to set the stage for better preventive medicine and more effective therapy. The report recommended that approximately 500,000 to 1 million participants would be needed. Ideally, participants would be sampled from different census tracts and would be recruited door-to-door over a 4-year period, and individuals from underrepresented minority groups would need to be overrecruited to provide a level of power sufficient to detect significant trends in minority populations.

The data to be collected at the time of a participant's enrollment into the study would include a wide range of phenotypic information and environmental factors. The largest scientific issue would be deciding on the specific list of factors to study. These decisions would be balanced against the expected cost of the project and the potential burden on individual participants. The conclusion of the Work Group was that core group baseline variables should be collected for all or nearly all of the participants. Disease outcomes over the course of the study could be assessed using hospital and outpatient records and data sources, including Centers for Medicare & Medicaid Services registries.

The Task Force strongly emphasized the importance of several policy issues described in the NIH report: (1) public engagement, since the success of an LPS would require a well-informed, fully supportive public; (2) determination of a representative cohort, because the U.S. population is extremely heterogeneous;

(3) investigation of possible collaborations with ongoing U.S. and international projects; (4) protection of data, which would be particularly complex for a large-scale longitudinal study; (5) decisions regarding a policy for notifying participants of their results and providing genetic or genomic counseling; (6) intellectual property concerns; (7) confidentiality and privacy; (8) informed consent; (9) institutional review board (IRB) management of such a large project; and (10) use of electronic health records. Based on these considerations, the Task Force identified four categories of issues and questions for discussion by the full Committee:

1. Broad Social Issues. Are there data to support the inherent value of a new LPS? Is a large cohort study the best way to obtain information about genetic and environmental influences on common disease? Given the existing cohort studies already underway within the U.S. Department of Health and Human Services (HHS), are there other ways to produce new knowledge? How much would a new study cost, and how does one balance this cost with other priorities within HHS and the biomedical research community? What tradeoffs would be necessary in terms of resource allocation? Would the benefits of such a study be distributed evenly to all groups in society? Would the study increase or decrease the stigmatization of individuals belonging to subgroups of the population on the basis of their genomes?

2. Engagement. How can public trust of science (i.e., genetics) and the Government be increased? How should preparation for such a study engage the public? How can input from the broader scientific community be gathered? Although the NIH Work Group involved a number of experts, a much broader section of the scientific community has not been engaged on these questions.

3. Access and Health Care System Issues. Would the study reduce or exacerbate health care disparities? Would the results benefit people with limited access to care? How would such a project deal with the ethical dilemmas associated with diagnosing conditions for which there are currently no treatments? What would the cost of participation be for study participants? How would the costs affect access to study participation across different strata of the population? How should minority communities and uninsured individuals be reached?

4. Research Issues. How would a new LPS leverage the existing HHS cohort studies that are already underway and are, at least in part, addressing many of the same questions? What does the term "environment" encompass in this context, and how should environmental, socioeconomic, and behavioral variables be measured? What steps must be taken to recruit and retain a cohort of between 500,000 and 1 million participants? How can protocols for study recruitment, enrollment, and withdrawal be kept free of incentives that are coercive? How would biological samples be secured, stored, and disposed of? To what extent would information be shared with family members, and what processes would be in place to address the familial implications of study participation? What guidelines should be developed related to the application of the research findings and anticipated technology developments? Special attention must be paid to avoid discrimination and stigmatization as research findings emerge over time.

Dr. Willard said the goal of the day's session was to obtain input on key policy and process issues from the scientific and bioethics communities and effective mechanisms for public engagement. Since assessing either the scientific need for a study or the specific scientific aspects of the study or its research design was beyond the SACGHS's charge, those issues were not part of the day's exploration. The panels representing science, public engagement, and bioethics would be followed by question-and-answer sessions. Dr. Willard introduced the first panel.

Policy Perspectives from the Scientific Community

Gerald R. Fink, Ph.D.
Professor of Genetics, Massachusetts Institute of Technology

Dr. Fink began his presentation by drawing an analogy between the proposed large-scale population study and the Human Genome Project (HGP). He described the early history of the HGP to demonstrate how this large science project eventually gained support and achieved success after an initial period of skepticism on the part of the scientific media and the public. Dr. Fink emphasized the importance of defined benchmarks, endpoints, and costs as key factors in gaining support for the project. The scientific community's confidence in the project grew as benchmarks were met, endpoints were reached, and actual costs came in lower than expected. As each phase was completed, basic scientists became the strongest supporters of the project because it added value to their endeavors.

He addressed areas in which skepticism about an LPS might be expected. For instance, an LPS would have to address policy issues concerning the kind of data to be collected and how to maximize the identification of key genes involved in a multigenic disease. Dr. Fink recommended conducting a pilot study first, as was done with the HGP. Choosing a heritable multigenic disease for the pilot, achieving defined benchmarks and endpoints, and staying within a predictable cost range could increase the confidence that researchers would be able to obtain statistically significant data from this type of study.

He posed several hypothetical questions: What would the Government do with information indicating that particular gene variants increase the risk of disease by several percentage points? Would the Government collaborate with pharmaceutical companies? Would this effort take away from investigator-initiated research?

Dr. Fink expressed the opinion that an LPS initially would be viewed with some alarm because of the complex scientific issues involved and the risk to funding for smaller-scale, investigator-initiated research studies, just as the HGP had been. He said the community would be much more supportive if there were proof of principle. Dr. Fink also recommended a pithy characterization of the overarching study question, stating that defining the goal of the study as "an understanding of the relationships of genes, health, and common complex diseases" is too general for the scientific community. He said that an LPS described as "seeking to identify risk factors for a specific disease" would generate more support. Dr. Fink also asked whether NIH was the appropriate organization to host the study, saying that others have suggested the Centers for Disease Control and Prevention (CDC) or that it be conducted as a public-private partnership with pharmaceutical companies as potential cosponsors. He closed by suggesting that a successful pilot study also could address some of the ethical concerns.

Questions and Answers

Dr. Francis Collins addressed Dr. Fink's proposal for a pilot project. He stated that the case-control study design is commonly used, particularly with the International HapMap Project, making it possible to conduct whole-genome association studies. Dr. Collins noted that there are already many pilot projects on case-control studies with good evidence of success. Rather than discovering gene variants of a specific disease, the population cohort study is designed to determine a linkage among genetic variation, environmental factors, and disease risk. He stated that case-control studies are somewhat biased in that regard, especially when assessing gene-environment interactions, because there is often recall bias. Dr. Collins added that a

U.S. population cohort study would take at least 1 to 2 years of planning and that sufficient pilot data would be in hand long before enrollment in the main study would begin.

Dr. Muin Khoury asked Dr. Fink for suggestions of benchmarks or parameters of early success in an LPS. He noted that public health agencies collect data through birth defect surveillance systems, cancer surveillance systems, and population surveys such as the National Health and Nutrition Examination Survey but that they are open-ended (i.e., they are not trying to test a scientific hypothesis). Dr. Fink suggested that advancing the understanding of spina bifida might provide an early benchmark. Although maternal folic acid deficiency is known to play a role in the development of spina bifida, all of the environmental and genetic components of the disorder have not yet been elucidated.

Sharon Kardia, Ph.D.
Associate Professor of Epidemiology, University of Michigan School of Public Health; Codirector, Michigan Center for Genomics and Public Health

Dr. Kardia noted that she is a human geneticist and statistician who studies cardiovascular disease, hypertension, diabetes, and the genetics of drug response and that she has a broad understanding of the social and policy issues surrounding genetics research and its implications. She stated that an LPS of genetic and environmental factors has advantages and disadvantages and suggested that the disadvantages currently outweigh the advantages.

Dr. Kardia stated that several critical social and regulatory policy issues make such a project premature. She cited a lack of genetic literacy by the public and by health professionals, a factor that will make it difficult to achieve valid informed consent for the study. Dr. Kardia pointed to the lack of a Federal genetic nondiscrimination law as a liability for the public to participate in an LPS. Also, since hundreds of investigators and clinicians would be involved in the project, the public also could question whether the privacy and confidentiality of personal medical information could be maintained. Based on experiences working at the Detroit Urban Research Center with community-based participatory research projects, she noted that she had become convinced that the public fears genetics research. Dr. Kardia noted substantial fear concerning inappropriate access to research data by doctors, insurance companies, employers, and Government agencies that might occur in an LPS. She stated that many public health practitioners are concerned that genetics research will increase health disparities and reduce access to care.

Dr. Kardia noted that health education and health behavior research has demonstrated that the public struggles with general comprehension of genetic concepts and the concept of genetic risk. Individuals are often unable to retain comprehension of genetic concepts following an educational session and frequently misinterpret information. Many do not know where genes are located or understand why a genetic test might be predictive of disease risk. This lack of basic understanding by the public presents a major barrier to meaningful communication of an individual's genetic risk. She stated that this lack of comprehension also is found among medical professionals and policymakers. Since most genetics research is focused on identifying single causative factors rather than more complex models of disease, Dr. Kardia said scientists continue to promote a naive interpretation of complex disorders. This leads to further misinterpretation and misuse of genetic explanations in public policy, courts, health and life insurance policies, and medical practices.

She suggested that the current research infrastructure and scientific culture are inadequate to carry out an LPS. Dr. Kardia stated that the genetic science of common complex diseases is not yet mature, and scientists are not always able to replicate findings from one study to the next. She noted that this caveat

is particularly relevant for an LPS, because researchers would want to use the power of the sample size to make definitive findings and statements. However, these findings would represent an overall average result for the study population and would not reflect the local heterogeneity of genetic and environmental factors, where clinical utility matters most.

Dr. Kardia suggested that most geneticists are not sufficiently well versed in the social, behavioral, and environmental causes of disease. True interdisciplinary research that integrates knowledge from the influence of the genome to the influence of human ecology is just beginning. She indicated her belief that geneticists and social behaviorists are currently pitted against one another within institutions and in competition for research funds. Representatives of these fields will have to learn one another's languages and methodologies. Dr. Kardia also stated that geneticists are criticized for their simplistic genocentric analyses, which lack key sociobehavioral measurements, reproducible results, and clear causative mechanisms for disease. She said that it is very difficult to move from a statistical genetic association to a more complete understanding of the genetic mechanism of a disease that would suggest new therapies and preventive measures capable of withstanding evidence-based regulatory decisionmaking. Genetic findings in complex disorders, especially gene-environment interactions, therefore would not help regulatory bodies create policies that protect people. Dr. Kardia stated that although there has been some progress in the field of gene-environment interactions, the results have exposed immense complexities in integrating this type of knowledge into existing policy standards and methods. Traditionally, public health policy has focused on population-level solutions (i.e., the one-size-fits-all model, as seen in antismoking campaigns). In contrast, genetic information is based on a variety of individual, familial, and ethnic factors. Genetics will require intense research on the implications of specialized policies and regulations before vulnerable populations can be adequately protected.

She questioned how regulatory agencies such as the U.S. Environmental Protection Agency and the Food and Drug Administration would set standards and guidelines for businesses and products based on complex susceptible genetic subgroups. For every disease, there is likely to be a different combination of genetic factors involved. Dr. Kardia stated that defining a vulnerable subgroup could be extremely challenging, especially when genetic definitions are overlaid with existing definitions of vulnerable populations. She suggested that the regulatory agencies do not have the resources or staff versed in genetic and genomic concepts to tackle an upheaval in their systems. Dr. Kardia also noted that genetic testing companies can market directly to consumers and doctors without any regulations. These companies are not required to disclose the real utility or makeup of their products. Direct-to-consumer gene testing kits are freely available, even though there is insufficient evidence to warrant their use.

Another factor affecting the timing for an LPS is that publicity surrounding the HGP has moved genetics into the public eye, but the research findings tend to be overstated. This has led to a cycle within society of aggrandizing simple genetic solutions to complex problems.

She expressed her belief that seeking input from the scientific community would lead to biased opinions from two groups: the outspoken antagonists from the social epidemiology field who are concerned about the "geneticization" of disease and the excessive use of resources by geneticists and the outspoken proponents of the project who are primarily concerned about access to the large funding source behind it.

Dr. Kardia also suggested that many genetic epidemiologists believe that the 500,000- or 1 million-person cohort is unrealistically large and too broad to ensure quality. In addition, a megascience model would fund only a few insiders, leaving little funding for the rest of the scientific community. There

is fear that the model would not build on the years of experience of genetic epidemiologists who have already accrued data from numerous cohort studies.

She described her work with the National Heart, Lung, and Blood Institute's Family Blood Pressure Program over the previous 10 years. Investigators collected data from 13,000 individuals and 5 racial and ethnic groups in more than a dozen field centers. Dr. Kardia noted that it takes tremendous effort to agree on what to measure, how to measure it, and how to package the results. Scientists have a long way to go in terms of learning to collaborate on and coordinate the use of existing resources. They are often competitive and have strong, conflicting opinions.

She suggested that although the project is too ambitious at this time, the infrastructure for such an LPS might be in place within 5 to 10 years. Dr. Kardia said that many intermediate steps could be taken along the way, such as encouraging genetics researchers to work together to mine existing cohorts. Genetic researchers also could begin to learn how to work with social and behavioral epidemiologists and researchers. She noted that genetics researchers have not typically utilized the resources available through State departments of health. Cancer registries, early death registries, and environmental health registries could be tapped as a first wave of research. State registries containing data on key environmental factors that influence the public's health should be used more widely by geneticists and policymakers; Dr. Kardia is currently engaged in such efforts in Michigan. State departments of health also would benefit from such collaborations, because staff members need more training to deal with issues surrounding genetic information on common disorders.

She summarized her remarks by stating that the regulatory agencies, State departments of health, and the public must be prepared for the use of genetics in the United States, lest they repeat the sickle cell screening debacle of the 1970s, when well-intentioned legislatures passed marriage laws to "protect" people. Given the right social and policy investments, Dr. Kardia said she would be greatly enthusiastic about the project in the future.

Questions and Answers

Dr. Collins challenged the idea that the project should wait, because of his belief that some of the current barriers are unlikely to improve without a stimulus. He stated that regulatory systems rarely change unless they perceive a need. A public project with great visibility would serve as an impetus for action. Dr. Collins suggested that the project would provide an opportunity to educate the public, the media, public policymakers, and the scientific community. He said it would bring together the communities that are now unable to work together, just as the HGP did. Dr. Collins said that since the data would be publicly accessible, they would provide a stimulus for researchers and would not deprive them of funding.

Richard B. Marchase, Ph.D.
Vice President for Science Policy, Federation of American Societies for Experimental Biology

Dr. Marchase described the Federation of American Societies for Experimental Biology (FASEB) as a coalition of 23 member societies representing more than 70,000 scientists in diverse areas of life science and medical research. In preparation for this presentation, discussions were held with FASEB's Clinical Research Subcommittee, the NIH Issues Subcommittee, and member societies, including the American Society of Human Genetics.

FASEB recognizes the potential of an LPS to improve health. However, he cited three key policy issues of concern to the scientific community: (1) the prioritization of this study relative to other large-scale studies; (2) study goals and study design that lead to useful data; and (3) the cost and possible effects on research grants, investigator-initiated studies, and other initiatives at NIH.

First, FASEB would like to see a dialog that would put a new LPS into perspective with existing studies (e.g., the National Children's Study and recent initiatives to increase NIH's presence in clinical and translational work). Dr. Marchase asked whether other long-term studies have been mined sufficiently to set the stage for a new LPS.

He also emphasized the need for optimal utility of study data and outcomes to the scientific community and stated that some questions, such as those raised by Dr. Kardia, should be considered in more detail. Dr. Marchase's questions included the following: How will the data be collected, stored, and made available to researchers? How will genetic and other personal information be protected? Does the current U.S. health care system have sufficient technology and infrastructure to support data collection and data sharing among multiple investigators and institutions? Is there a way to restrict or focus the study? He opined that pilot studies should be conducted first.

Of primary concern to FASEB is that the study would be very expensive and is being proposed at a time when NIH funding is not increasing. Dr. Marchase said the 2005 NIH budget was likely to have only a zero- to 1-percent increase above the 2004 NIH budget. Because small increases place a significant burden on investigators submitting their ideas for funding at NIH, FASEB believes that funding for investigator-initiated research projects should remain a high priority. He discussed a hypothetical example. Dr. Marchase stated that the estimated cost of the study could be as much as $3 billion or more. If one-tenth of that, or $350 million, were taken out of the budget for investigator-initiated grants (referred to as the "R01 budget"), it would result in 1,000 fewer grant awards. Based on 2004 data, the success rate for funding R01s would drop from 24.9 percent to 21.3 percent. Therefore, FASEB is concerned that the allocation of large funds to one project during a flat funding period would be highly detrimental to the current and next generation of biological scientists. FASEB does not want to see the study funded in a manner that deters the entry of new investigators into the research pool or takes a high toll on the funding available to established scientists. FASEB's longstanding principle has been that investigator-initiated, competitive, peer-reviewed grants should remain the core mechanism for distribution of research funding.

Dr. Marchase concluded by stating that FASEB recognizes the numerous potential benefits of such a study for public health. He said it is a visionary idea that could help break the flat-level funding cycle. However, the fact remains that discretionary spending is very limited.

Questions and Answers

Dr. Khoury commented that in other countries, work with biobanks and cohort studies is referred to as a "resource," because collecting information on a large number of people followed over time is not seen as an individual study. The biobank is considered a resource that will lead to thousands of studies in the future. Dr. Marchase agreed that in the long term, an LPS would be an important resource available to the full spectrum of biomedical scientists. His concern was that in the short term, it might jeopardize scientists who are currently working.

Dr. Collins stated that NIH is deeply concerned about the trends in terms of support for R01 investigators. He agreed that losing 1,000 new grants to fund the project is not tenable. Dr. Collins stated, however, that there is an opportunity for the biomedical research community to identify one or two compelling flagship initiatives that will benefit public health. This could generate enthusiasm and energy for a return to a more progressive course. He stated that it would be disadvantageous for scientists to try to get by within the current budget circumstances. Dr. Collins noted that no one is proposing that the project be initiated with existing funds. Dr. Marchase agreed that the project might move the community out of stagnation.

Dr. Debra Leonard commented that Dr. Kardia and Dr. Marchase seemed to be articulating a wish to hold on to the academic system as it currently exists. She said there is an impetus for change to the academic system in the NIH Roadmap Initiative, which values large collaborative efforts, some of which are not supported by the current academic tenure system. Dr. Leonard asked whether the academic system should therefore be reevaluated. Dr. Marchase agreed and said he recognized that "big science" will become important in moving research and scientists forward. However, FASEB wants to ensure that the thinking of individual scientists is not lost while moving in the direction of big science.

One participant cautioned Dr. Kardia about letting the perfect be the enemy of the good. He expressed concern that she was making the assumption that there must be perfect societal and infrastructure development prior to this study. He also noted that NIH and the Health Resources and Services Administration were engaged in a 5-year, $2.4 million contract to educate health care professionals about the use of genetics and genomics in their practices.

Dr. Kardia said that her concern lies with the large discrepancy between spending $2.4 million for education compared with the projected $3 billion pricetag of a new LPS. She stated that $100 million invested in the genetics education of the Nation, including health professionals, would accomplish something significant, but $2.4 million is not enough money for infrastructure building. Dr. Kardia noted that genetics policy research also should take place on a large scale. She agreed that academia needs to change to an interdisciplinary approach but questioned where the funding to accomplish this goal would come from.

Dr. Khoury stated that CDC acts locally through the States' infrastructures to build the capacity for the work that will come from the HGP. He asked Dr. Kardia what she would think about a two-pronged approach—building big science, while concurrently building infrastructure in smaller locales.

Dr. Kardia described her efforts working for 3 years with the Department of Health in Michigan, creating relationships to foster a bridge with academia. These efforts allow the University of Michigan to garner the resources of the Michigan Department of Health. She said it is critical to have enough people that are representative of the population being served so that replication studies can be done within that population. Dr. Kardia said scientists tend to group unlike people. For example, within the African American population, there are great differences in allele frequencies, genetic factors, and environmental factors. Therefore, the populations being served should be matched with the genetics research. She recommended focusing on big cities in which the largest public health burdens exist and working so that local and State departments of health and local clinicians can use the study data. Dr. Kardia added that the technologies of genomics, proteomics, and metabolomics should be used, because a pure genetic approach leaves a large gap between finding an association and moving it into treatment and prevention.

The Committee discussed some specifics of the study and asked questions of the panelists. All agreed that education would be a critical component. One SACGHS member suggested looking at the outcomes of

case-control studies and large cohort projects in other countries to see whether they are obtaining a good return on their investment, and it was agreed that financial issues would require close scrutiny.

One Committee member expressed ethical concerns about individuals in the proposed study who might not have health insurance. What would happen, he asked, when study participants in a longitudinal, observational study get sick? Would the investigators simply have to document the illness without offering help? Although it would not be feasible to provide health insurance to all enrollees who are uninsured, he suggested that there would be a social responsibility to participants. He questioned how the $3.5 billion cost of the project could be justified under these circumstances. Dr. Kardia suggested a model in which research also served as health care. She said that the University of Michigan is assisted by community-based participatory researchers who have money in their budgets to provide care for uninsured participants. Dr. Kardia suggested applying some of what has been learned from the social sciences about community support and participation to this project.

The Committee agreed that public health would be integral to an LPS and that participants' health care should be taken into consideration. However, one member noted that a substantial number of people who are uninsured would be enrolled. He said that, for these individuals, provision of care could become an inappropriate inducement to participate in the research. He proposed that ethicists and health economists be included in the project from the outset to address such issues. Dr. Collins stated that the obligation of research to provide medical benefits is critical but not new. There is a large body of ethical debate and literature on the topic. He suggested looking carefully at existing studies and documented discussions concerning uninsured research participants.

Another member said that funding is needed for the training of genetics professionals who would be working with the public. Currently, two-thirds of State genetics coordinators report having no formal training in genetics. She stated that primary care providers do not seek to enhance their training in genetics; they prefer to refer their patients to specialists from whom the patients can obtain these services. She added that geneticists and genetic counselors are leaving the field, and there is a paucity of minority health professionals to serve minority populations.

At the conclusion of the discussion, Dr. Marchase said the study should be done only if it does not disrupt ongoing research projects and if it does not discourage new investigators from entering the system. Dr. Kardia suggested that 25 percent of the money allocated for the project be invested in infrastructure development in public and regulatory systems because regulatory bodies, health professionals, departments of health, and the public lack an understanding of genetics. She said a strategic plan could identify the areas of greatest need so that these efforts could be prioritized.

Public Comments

Joann Boughman, Ph.D.
Vice President, American Society of Human Genetics

Dr. Boughman testified for the American Society of Human Genetics (ASHG). She stated that although there is widespread support for the concept of an LPS, there are diverse views on the manner in which the study should be implemented, the nature of the data to be collected, and the extent to which the data will translate into the promise of treatment or prevention. She said the design of an LPS would be an immense

challenge because the specific aims will necessarily evolve over time. The data gathered must be broad enough so that undefined or currently unrecognized questions could eventually be asked and answered.

ASHG identified a number of particular challenges to the study:

- Do existing data sets have sufficient breadth and depth to provide some of the information proposed in this study, and if not, are there ways that existing data can be further mined to limit study costs?
- Given the current fractious state of health care in the United States, can a coherent cohort study be designed, data collected and analyzed, and benefit returned to the participants and others in the United States at a reasonable cost? In the face of the patchwork health care system of the United States and the absence of systematic electronic health records or a realistic vision of uniform health care delivery, the direct applicability of the results to the broader community must be questioned.
- The cohort study demands the identification of a population that has sufficient breadth and depth to allow analysis of myriad relevant questions; identification of the numerous biological variables to be measured; tabulation of data; and creation of robust assessment and computational tools to define, measure, and assess the effects of environmental changes over time. These requirements are far more complex than for the HGP.
- The substantial costs of an LPS must come from outside the usual funding mechanisms. Otherwise, the effect on existing biomedical research funding could be highly deleterious or devastating.
- The choice of individuals and populations to be included is extremely complex in the highly heterogeneous society of the United States. The need for a diverse study population and the manner in which that diversity is handled must be carefully considered.

Exploring Mechanisms for Public Engagement

Dr. Willard introduced a panel of four experts in the area of public engagement who discussed effective methods that could be used for a new LPS.

Joan A. Scott, M.S., CGC
Deputy Director, Genetics and Public Policy Center, Johns Hopkins University

Ms. Scott described the spectrum of public engagement, stating that at one end, the public can be merely informed. Some would argue that public information stops well short of public engagement because it involves passive receipt of information. A more consultative approach to public engagement assumes that the public brings valuable experiences and values that will inform an issue. There are many different levels at which the public can be engaged, including surveys, focus groups, workshops, and scenario development. "Deliberative democracy" provides an opportunity for participants to hear from experts with different points of view and deliberate about the issues. Participants are asked to reach consensus about the best option.

A more collaborative approach to public engagement invites the community to participate early in the process with issue identification and prioritization. They help set the agenda for the engagements and devise outreach strategies. At the farthest end of the spectrum, not only are participants empowered to make a decision, but also the decision they arrive at is abided by. In the case of an LPS, engagement could be aimed toward the communities from which participants will be recruited or may require a more national or regional conversation.

Ms. Scott described the Genetics and Public Policy Center's approach to the genetic town hall meetings that were held in six cities around the United States during summer 2004. Fifteen concurrent discussion groups were held online. The topic was reproductive genetic technologies, and the Center had conducted extensive background work using surveys, focus groups, and interviews prior to the town hall meetings. She said that one of the criticisms of public engagement is that people sometimes are asked to comment about technologies with which they have little personal experience. The deliberative approach provides participants with indepth background information about the topic. Ms. Scott stated that a credible deliberative process has four requirements: (1) broad, representative participation, (2) balanced and accurate information, (3) an environment with ample opportunity for deliberation, and (4) involvement of policymakers from the outset. For the town hall meetings addressing reproductive genetic technologies, the Center prepared a DVD to ensure that the content was consistent and balanced in all locations around the country. The DVD provided an overview of reproductive genetic testing and comments from experts conveying different perspectives.

The Center partnered with the Public Forum Institute, and the recruitment for the town hall meetings was conducted through local coordinators with contacts in the communities. They used a variety of outreach strategies, including posting notices in libraries, hospitals, clinics, grocery stores, and community centers and targeting community organizations and leaders. The Center used the media by working with local reporters, talking on local radio shows, and placing ads in newspapers. They asked people to register for the town hall meetings so they could monitor recruitment. During the meetings, about 8 to 10 people participated in roundtables and were asked 36 questions, 8 of which were repeated at the end of the session to determine whether there was a shift in attitudes. The town hall meetings lasted 3.5 hours, during which content was presented and participants took part in small- and large-group discussions.

The online group met for three 1-hour sessions over the course of 3 weeks. Participants were recruited through a Knowledge Networks' Web-enabled panel, which is representative of the general population. Participants first took an 80-item survey and then were mailed headsets and instructions prior to the online sessions, which were moderated by genetic counselors. About 1 week after the last session, 76 of the questions were repeated to document changes in knowledge and attitudes. A control group of 400 individuals took the pretest and the posttest but did not participate in the discussions.

The results indicated that people who were willing to take nearly 4 hours out of their schedules to attend an in-person meeting were more likely to be stakeholders who already had a specific perspective. The Center was able to document clear shifts in opinion before and after the engagement, many toward disapproval of the technologies described, whereas the control group members did not change their opinions. Although numerous concerns were expressed about the use of genetic testing and equal access to the benefits of advances in genetics, the public was optimistic about the potential for health benefits.

The approach for an LPS would need to use methodologies appropriate for the entire engagement strategy. Ms. Scott suggested that televised town hall meetings and increased media involvement to reach a wider audience can have a broad ripple effect. She emphasized that tracking over time is important for monitoring the effects of engagement initiatives.

E. Yvonne Lewis
Executive Director, Faith Access to Community Economic Development

Toby Citrin, J.D.
Director, Michigan Center for Genomics and Public Health, University of Michigan School of Public Health

Ms. Lewis and Mr. Citrin described three of the engagement projects they have conducted that are relevant to the proposed new LPS. The Communities of Color and Genetics Policy Project engaged African American and Latino communities on genetics issues at the grassroots level to formulate policy recommendations that would enhance benefits and minimize harms to these populations. The University of Michigan, Michigan State University, and Tuskegee University partnered with 12 community-based organizations (CBOs) in Michigan and Alabama for this initiative. They taught a basic educational module on genetics research to a series of focus groups, followed by discussion. Each of the community organizations then sponsored a series of five 2-hour dialog sessions, typically attended by approximately 20 members of their organizations. The community organizations worked closely with the academic team to develop and implement the process and craft the summaries and recommendations. The community organizations and the academic partners met with policymakers in Michigan and Alabama to share the recommendations. In a 2-day visit to Washington, D.C., they also met with community partners, Members of Congress, congressional staffers, and domestic policy advisors to the President.

The Genetics Education Needs Evaluation project involved two communities in New York and Michigan. Building on previous relationships, they reached out to several CBOs, including churches, social organizations, fraternities and sororities, and a school system. To help determine educational needs, they provided community members with basic education about genetics and asked them how that education might be facilitated. The information that was gained was reported back to the community in a town hall meeting. This collaborative process led to a statewide initiative that looked at improving cancer outcomes for African Americans in Michigan. They worked with the Department of Community Health and a number of community leaders from five cities in Michigan to raise the level of awareness, reduce myths, and engage people in screening programs.

Ms. Lewis said that mistrust among certain minority communities stems from a history of research in which the benefits were not translated to the community. She said that to build trust, an honest and frank discussion must take place, and individuals must believe that they are an integral part of the project and understand the purpose of the project, how it is designed, and who will benefit from it. Ms. Lewis stated that recognition of the relevance of the project to the community results in engagement, and engagement raises the community's expectations. She warned that not fulfilling these expectations would lead to mistrust and opposition.

Mr. Citrin stated that the proposed LPS poses a major risk of generating distrust among communities of color. He remarked that the key lesson they learned from their projects is that a successful project must be a true engagement. The avoidance of distrust and the achievement of public support depend on the concept of co-ownership across the communities that are most at risk from the study. If a sense of co-ownership can be achieved, powerful advocates will support the necessary infrastructure-building. Ms. Lewis said that decisionmaking and planning must engage the community from the outset. The study process must be explicit in addressing the issues of race and racism, and individual representatives of racial and ethnic groups must be meaningfully involved in developing plans and methods. Mr. Citrin also noted that it is

important to pay attention to the use of the media to ensure that a lack of trust and the presence of fear do not become predominant messages.

He commented on the suggestion of an earlier presenter that an infrastructure must be in place prior to the start of such a project. If the project engages the community and is fully participatory, it can be a vehicle for community education as it moves forward. Ms. Lewis and Mr. Citrin said that to achieve full engagement, the community must be involved in all stages of the project, including study design, development of the instruments and materials, and reporting of outcomes to the public. Ms. Lewis added that the process for partnership-building must be continually evaluated, leading to a common language, understanding, and goals.

Mr. Citrin noted several national organizations that could help foster dialogs with scientists, professionals, practitioners, public health representatives, and grassroots community members. Ms. Lewis added that CBOs are valuable intermediaries that can maintain synergistic and consistent engagement and expand the level of involvement. These CBO connections can be made at local, State, and national levels. She noted that CDC's National Community Committee represents community-based individuals.

Ms. Lewis summarized by suggesting that research subjects be referred to as individual "study partners" and be kept informed through ongoing communication. The success of an LPS may depend on whether it is perceived as a project carried out *by* the public or conducted *on* the public.

Mary Woolley, M.A.
President, Research!America

Ms. Woolley described the key considerations of a possible LPS. She suggested that broad support from the scientific community and the public is critical. Ms. Woolley noted that the proposed study would compete with many other public agenda items and that the high overall costs of the study would affect public perception. She also said that to help address the current mistrust of Government, the LPS must be framed to underscore the fact that researchers work for the public good.

Another key consideration is the importance of identifying an urgent, compelling goal. People want to participate in an endeavor that is exciting and that they can understand immediately. Ms. Woolley explained that individuals and decisionmakers want to feel that they are part of history.

Public involvement is another consideration. She described the difference between public engagement and public relations, which is a different area of expertise that comes later in the process. Ms. Woolley also noted the importance of communication and terminology used in engaging the public. She stated that the fewer words used, the better, and researchers should be trained to describe their work in three or fewer sentences. An authentic messenger also can make a great deal of difference. Ms. Woolley remarked that it is also important not to try to engage the public before they are ready.

Ms. Woolley concluded by noting that project funding was the least significant of the considerations. If value and need for a study have been established and a project has the confidence of key people in the public and the scientific community, the money will follow.

Questions and Answers

During the discussion, Dr. Khoury discussed the history of the HGP and said the next initiative could lead to the translation of the first phase of gene sequencing to improvements in public health. He asked how, if the next project were to be considered a translational population-based effort to take genes from the "bench to the trench," money should be spent for the contextual elements that would allow the initiative to move forward. Mr. Citrin recommended that a substantial amount of money be spent on engagement and education and stated that a number of existing networks can play a role in this effort. Ms. Scott emphasized that starting early in the process is critical. Mr. Citrin suggested using the media to stimulate national attention, engage the public, and facilitate dialog between scientists and other stakeholders. Ms. Scott noted that it is not important that the public understand every scientific detail of the project, but how it applies to them. She said the public is capable of understanding complex technical, social, and ethical issues and putting them into the context of what they already know (i.e., complex diseases). Ms. Lewis observed that the science community needs to understand the language of those who are not in the science community.

Dr. Willard asked the presenters to suggest concrete steps for engaging the public. Ms. Woolley suggested surveying public opinion. Ms. Scott suggested clearly identifying the communities they wish to engage and going to them to obtain a sense of their concerns, then focusing on the long-term engagement process. Mr. Citrin recommended convening a group that would engage in a dialog at the national level. It would have representatives from the science community, public health leadership, and stakeholders with national prominence who have credibility with communities interested in reducing and eliminating health disparities. Even if it were an informal meeting, participants could engage in a dialog on how they might further the project and maximize the community engagement of their constituencies. Ms. Lewis added that representatives from the communities of concern should be identified up front so they can attend the first meeting. Resources should be allocated to ensure their long-term participation. Dr. Willard said that such a partnership concept would require everyone who comes to the table to be willing to listen to others' perspectives and to respect differing opinions.

Policy Perspectives from the Bioethics Community

Dr. Willard introduced a panel of three bioethicists to discuss a variety of perspectives on policy issues related to an LPS.

Henry T. Greely, J.D.
Deane F. and Kate Edelman Johnson Professor of Law, Stanford Law School

Mr. Greely recommended that SACGHS undertake a careful study of large-scale scientific projects conducted previously, such as the HGP, to help identify the "land mines" experienced and the solutions developed to resolve those problems. He spoke on several ethical issues concerning a large population resource, including access to and control of materials and data, return of information to participants, and confidentiality.

Mr. Greely discussed the sensitivities people have about how their data and biological specimens will be used and suggested that research participants must have the ability to control how their specimens will be used. They may not want information about themselves or their family members to be used for certain types of research. Not respecting this right can lead to loss of trust. This concern is particularly complex with the creation of gene libraries and databanks. It is impractical to obtain informed consent for the numerous

possible uses of such data over time. Even if resources were available to obtain consent for each use, participants do not necessarily want to be contacted each time their data might be used in research. Using the term "informed consent" for such studies is a misnomer, because no one knows what specific research will be conducted or what diseases, genes, or environmental effects will be examined. The very purpose of the resource is to make it available for scientists to do anything that seems useful over time. Mr. Greely said consent in these circumstances is not truly informed.

He said that none of the possible solutions to this problem is perfect, but some come close to mitigating the dilemma. Mr. Greely recommended asking research participants at the beginning of the process whether there are specific research areas for which they would not want their data or DNA used. He said that a short, targeted opt-out list of the most sensitive areas might be helpful. Another alternative used in a U.S. Department of Veterans Affairs project is to have continual monitoring of the research topics, both by an IRB and by a group drawn from the research participants. They could discuss the new protocols proposed, decide whether a significant number of participants might object, and determine whether individual reconsent is necessary. Mr. Greely noted that problematic issues are most likely to arise in the area of behavioral genetics. This monitoring approach would rule out the alternative mentioned by Dr. Collins, in which the material is open to anyone who has IRB approval. Mr. Greely stated that if that goal is a high priority, the informed consent must warn people up front, in plain English, that they have no control over the ways in which their data and materials will be used.

Regarding confidentiality, he said that U.S. citizens value health privacy, but the growing networks and the linking up of medical data undercut the possibility of complete confidentiality. In most research projects, participants are afforded a high degree of confidentiality, but even with the greatest security measures, it cannot be guaranteed. In a longitudinal study, privacy and confidentiality are harder to achieve because of the comprehensive data it contains and the repeated access to these data. This will become a bigger problem as more data are placed online. On the other hand, de-identification of data would result in an unquantifiable loss of potential scientific value. Mr. Greely recommended being completely straightforward with prospective participants and acknowledged that this approach could affect, and increase the cost and time involved in, recruitment. Some research participants will refuse to participate if absolute confidentiality cannot be guaranteed. This, however, is preferable to participants enrolling without a clear understanding and then feeling betrayed when they discover that full confidentiality is not possible. He closed by stating that treating research participants well is ethically important for science and for scientists. Research participants who feel betrayed and mistreated are unlikely to lobby for, vote for, or support biomedical research.

Pilar Ossorio, Ph.D., J.D.
Associate Professor of Law and Medical Ethics and Associate Director, Center for the Study of Race and Ethnicity in Medicine, University of Wisconsin Law School

Dr. Ossorio discussed the issues surrounding the return of research results to participants. She made a distinction between building a resource and conducting follow-on studies, saying that there are ethical and pragmatic differences between them involving the proximity of the researchers to participants in both space and time. In addition, follow-on studies may be more likely to generate information for which there are no clear processes for validation or regulatory requirements, which could potentially affect the level of difficulty in reporting information to participants.

Dr. Ossorio said that a number of policy committees have studied these issues and made recommendations but have not achieved consensus. At one end of the spectrum is the option not to return any individual results. Most genetic studies do not return individual results, partially because many of the data were not

validated or considered to be clinically useful. Current views about the permissibility of not returning individual results are changing. She noted that the proposed LPS would have the potential to result in clinically relevant information.

Dr. Ossorio said that there are good reasons not to return individual results, including lack of clinical validation. The costs of sharing results when they are ambiguous are high. By not returning results, the researcher also increases confidentiality and privacy protections. The relationship between the researcher and the participant is a factor because there may have been no direct personal contact, which lessens the feeling of mutual obligation. If time has passed, the information may be outdated. Finally, not returning results helps maintain the cognitive and legal distinction between research and the provision of medical care. A middle view is to provide a limited set of clinically relevant results.

She noted that the project under consideration almost certainly would require the return of at least some results. The following questions will therefore arise: Which results, to whom, what is the process of returning them, and when and how should it be done? Dr. Ossorio said there is wide agreement that researchers should not return results unless there is some analytic and clinical validity. The reasons in favor of returning results also are stronger when they have serious medical implications for the participant, would change medical management, or are unlikely to be discovered through routine medical care, or there is a robust relationship with the researcher so that the expectation of reciprocity is greater. Although little research has been done on whether participants want results back, the data thus far indicate that participants have a fairly high degree of interest in obtaining results and that there is a small risk of harm.

She discussed findings that are the focus of the research, those that are incidental but foreseeable, and those that are unforeseeable. Dr. Ossorio stated that these are different categories to which different degrees of permissibility or obligation to report might be attached. In the first category ("obliged to report back"), the researcher is obligated to report results, yet very few circumstances fall into this category. In these cases, it would not matter what the focus of the research was or who was doing it. The second category ("permissible to report back") is very broad and includes information that is permissible to be reported back at the researcher's discretion. This type of information and its reporting requirements should be delineated in the protocol, approved by the IRB, and included in the consent form. The third category ("impermissible to report back") includes information that is not permissible to be reported back. The method for reporting depends on the category. For all categories, if data will be reported back, this stipulation must be approved by an IRB, included in the consent, and reported by a person with relevant expertise. Genetic information should be validated in a lab approved according to the Clinical Laboratory Improvement Amendments of 1988.

She stated that the data access policy for nonmedical uses of the database, such as for apprehending criminals, must be delineated up front. However, law enforcement could, in some cases, obtain a court order to obtain access to the data. Dr. Ossorio noted that the Federal Bureau of Investigation is very interested in obtaining genetic information. Some State laws would protect a research database against uses for other purposes.

Participants should be given options regarding how data are reported back (i.e., either not getting information back or getting it back from a specific individual). She concluded by stating that there are tradeoffs concerning what is permissible versus what is obligatory, in part because reporting back to the participants adds costs to the study. Dr. Ossorio stated that community engagement could help researchers formulate the boundaries of "permissible to report back."

Almost all ethics guidelines state that there is a right not to know, but many clinical researchers disagree. The dividing line is clear for clinicians who are conducting a clinical exam in the course of their research and feel an obligation to disclose findings because their duties as physician overlap with their researcher role.

Troy Duster, Ph.D.
Professor of Sociology, New York University

Dr. Duster stated that it is important that a large research project accurately represents the population. In an LPS, people are likely to raise the question of whether race is being used as a taxonomic system, even though scientists know that race as a category is fluid biologically, socially, anthropologically, politically, and culturally.

He asked what it means to have a population study that is representative of the U.S. population. Most would assume the inclusion of whites, blacks, Hispanics, and Asians. Dr. Duster asked rhetorically whether a study should include participants who live near toxic waste dumps, given that cancer rates in this country are highest in those who live near toxic waste sites. By framing the study as a study of genes and the environment, there is an assumption that the interaction between them is more or less equal. Although there are good empirical data showing that, for many diseases, environmental factors play a more important role than genetic factors, science tends to focus more on the genetic contributions. The assumption is that if an ethnic group has a higher rate of a particular disease, it must be due to their racial or ethnic background. However, some epidemiological studies suggest that the differences disappear when cross-cultural studies are conducted. For example, hypertension is high among African Americans, but a study of eight countries on three continents shows no differences in hypertension between blacks and whites. To arrive at the conclusion that the different hypertension rates in the United States are a function of genetic differences is a leap, unless confirmed by functional genomics. If racial categories are used, the method of results reporting becomes increasingly important.

Full Committee Discussion

The Committee agreed that it would be appropriate to draw on the expertise of panelists and other experts in drafting a report to the HHS Secretary. Dr. Willard asked each of the Committee members to express their opinions on the feasibility of the study and to identify key issues or recommendations for consideration. The Committee consensus was one of support for the project, balanced with recognition of its complexity. It was agreed that a community consultation process should serve as the starting point. Some of the issues suggested for discussion in the report were as follows:

- The critical importance of public engagement and the question of how early these efforts should begin
- The possibility of developing a new paradigm for the way research is conducted using community engagement principles
- The difficulties of conducting the project in the context of the current U.S. health care system
- The need to integrate knowledge from similar studies taking place in the United States and around the world
- The importance of the project as a means of maximizing the utility of the HGP
- The need to define key elements of the study and commit to timelines for their completion
- The importance of broad Federal Government agency and private sector participation

- The potential for non-health-related outcomes as well as individual and public health benefits
- The need for an environmental taxonomy, which is not currently described for this study
- The need for new funding
- The need for protections against discrimination
- The need for smaller studies to lay the groundwork for the project as it moves forward

It was agreed that the staff would develop an outline for a draft report to the HHS Secretary for the Committee to review at the March 2006 meeting. The LPS Task Force would be involved in guiding the development of the draft report and would call on other experts as needed.

Appendix C

List of Public Commenters

Adelsman, Heidi

Affymetrix, Inc.

American Dietetic Association

American Medical Association

American Nurses Association

American Tinnitus Association

Anonymous mental health consumer

Association for Molecular Pathology

Association of American Medical Colleges

Austin, Finley M.J., PhD

Bowers, Beth

Center for Health Workforce Studies

Chak, Amitabh, MD, MS

Colditz, Graham A., MD, DrPH

Coltman, Charles A., Jr., MD

Community-Campus Partnerships for Health

Condit, Celeste M.

Coronary Artery Risk Development in Young Adults (CARDIA) Study principal investigators

Dashner, Ralph

DNA Fingerprinting and Civil Liberties project

Doyle, Debra Lochner, MS, CGC

Dressler, Lynn, DrPH

Eastman, Dale

Findlay, Steve

Galewski, Ralph, MD, MPH

Genetics and Public Policy Center

Grandbois, Ray

Greenway, Kirk

Harrison, Victoria

Health Resources and Services Administration

Heart Rhythm Society

Helzlsouer, Kathy J., MD, MHS

Institute of Electrical and Electronics Engineers, Inc.

International Society of Nurses in Genetics, Inc.

Ketterer, Mark W., PhD

Kittner, Steven, MD, MPH

Koenig, Barbara A., PhD

Lahey, Benjamin B., PhD

Marts, Sherry A., PhD

Maternal Outreach Management System

Modell, Stephen M., MD

Motulsky, Arno G., MD, ScD

National Cancer Institute

National Consumers League

National Institute of Allergy and Infectious Diseases

National Institute of Child Health and Human Development

National Institute of Dental and Craniofacial Research

National Institute of Diabetes and Digestive and Kidney Diseases

National Institute of Environmental Health Sciences

National Institute on Alcohol Abuse and Alcoholism

National Society of Genetic Counselors

Northwestern University Gene Bank (NUgene) Project

Ouellette, Eileen M., MD, JD, FAAP

Pickering, Rexann G., PhD, RN

Policy Working Group of the Community-Based Public Health Caucus

Pulley, Jill

Raizenne, Mark, MD

Rathjen, Alice

RTI International

Rupp, Mike

Society of Critical Care Medicine

Swanson, James, PhD

Voss, Joachim, PhD, RN

Wade, Deborly, PhD

Wadhwa, Pathik, MD, PhD

Witz, Samuel, PhD

World Privacy Forum

Appendix D

International Biobanking Efforts

UK Biobank

UK Biobank's main aim is to elucidate the effects of genetic and environmental factors on the risk of the common multifactorial diseases of adult life. The Biobank aims to enroll 500,000 male and female middle-age participants and is the largest and most ambitious biobank established to date. UK Biobank directors believe that a large sample is needed to provide the statistical power necessary to conclusively detect meaningful correlations of phenotype, genotype, and environmental exposure and to identify the multiple factors of often modest effect that contribute to disease. UK Biobank's goal is to establish a prospectively gathered collection of samples, in conjunction with comprehensive measures of exposure and phenotype, so that a wide range of gene/exposure/phenotype relationships can be studied. UK Biobank directors intend for the Biobank to serve as a resource for the biomedical research community for decades to come. Each participant will contribute a blood sample, complete a questionnaire on lifestyle, provide a medical history, and receive an examination by a nurse. Participants also will be followed regularly through their physicians for the reporting of morbidity and disease diagnosis and will be resurveyed periodically to update their exposure data. Blood samples will be stored for future retrieval for nested case-control studies. This information will be made available to researchers investigating the complicated interactions among genes, environment, and lifestyle that are believed to cause many complex disorders such as cancer, heart disease, diabetes, and Alzheimer's disease. Researchers will apply through a peer-review mechanism for access to data and for approval of their research proposal to investigate a specific disease. To ensure that the identity of gene donors is protected, scientists or medical doctors who request data will be given de-identified, coded data.

Biobank Japan

Biobank Japan is a fully funded national project designed to collect blood samples from 300,000 residents of Japan, with the goal of developing personalized medicine for a set of 40 diseases, including cancer, diabetes, rheumatoid arthritis, and other common disorders.[134] The samples will be genotyped by the Human Genome Center, Institute of Medical Science, University of Tokyo, with production-scale, single nucleotide polymorphism genotyping by BeadLab (Illumina, Inc.), a state-of-the-art tool for the analysis of genetic variation and function. A new facility to house the specimens will be built, and specimens will be kept separate from medical and genetic data. To ensure that the identity of gene donors is protected, samples will be coded. An explicit informed consent process will be utilized.

[134] Triendl R. Japan launches controversial Biobank project. *Nat Med* 2003. 9(8):982.

Estonian Genome Project

The Estonian Genome Project (EGP) is a biobank of information on diseases, lifestyle, demographics, genealogy, and deoxyribonucleic acid (DNA) in a database that is accessible by researchers. These data might be used to inform the clinical treatment of sample donors and for public health research and statistical applications. The data are free of charge to academic researchers. Foreign researchers can use the data in collaboration with Estonian scientists. The EGP plans to enroll 100,000 of Estonia's 1.4 million people. In addition to advancing the development of diagnostics and therapies, the EGP is designed to be of direct benefit to participants and allows study participants to access their own genetic information. For example, should scientists discover that a genetic variant is an indicator of a particular disease or of an adverse reaction to a medication, a donor can request information on his or her genotype at the particular variant. To ensure that the identity of gene donors is protected, scientists or medical doctors requesting data are given de-identified, coded data.

Icelandic Genetics and deCODE genetics, Inc.

In 2000 the Icelandic Parliament granted deCODE genetics, Inc. (deCODE), a pharmaceutical company, exclusive rights to the country's medical records. These records exist in the Icelandic Health Sector Database. At the same time, Iceland's government authorized deCODE to begin the construction of a biobank of the Icelandic population. Iceland's population is a unique resource because it comprises a homogenous population for which an extensive genealogical database, including information on previous generations dating back hundreds of years, is available. Iceland's population of 275,000 has been geographically isolated over time. Its genetic homogeneity, limited population, and extensive genealogical information are frequently cited by deCODE as the major reasons for its success in discovering new genes and genetic material related to several common diseases.

Because this extensive genealogy is available for the entire Icelandic population, individuals from extended families who have the same disease can be grouped and studied using linkage analysis to identify biomarkers and segments of particular chromosomes. These segments of chromosomes are likely to contain genes related to the disease that can serve as targets for pharmaceutical development. The approach taken by deCODE is augmented by the country's extensive and well-developed system of medical records, which has been well maintained since 1915. The identities of the participants in the deCODE studies are kept secret through use of an encrypted identification code. To date, deCODE has identified genes involved in several common complex diseases, including myocardial infarction, stroke, osteoporosis, and asthma.

GenomEUtwin Project

GenomEUtwin is studying population genetics in Europe through the collaboration of twin researchers, genetic epidemiologists, molecular geneticists, and mathematicians.[135] Their goal is to identify critical genetic and lifestyle risk factors for common diseases using genetics, epidemiology, and biocomputing. The project is applying new molecular and statistical strategies to analyze unique European twin and other population cohorts to define and characterize the genetic, environmental, and lifestyle components for health problems such as obesity, migraine, coronary heart disease, and stroke. The cohorts include Danish, Finnish, Italian, Dutch, and Swedish twins and the many European members of the MORGAM population

[135] GenomEUtwin Web site. See http://www.genomeutwin.org/index.htm [accessed December 21, 2006].

cohort. The MORGAM Project is a multinational collaborative study that is exploring the relationships between cardiovascular diseases and their classic and genetic risk factors.

Centre for Integrated Genomic Medical Research

The Centre for Integrated Genomic Medical Research (CIGMR) in the United Kingdom focuses on the investigation of common complex disease phenotypes that can be readily identified through epidemiological, statistical, and genetic approaches.[136] The Centre is initially concentrating on arthritis, neuropsychiatry, gerontology, cancer, infectious disease, inflammatory conditions, degenerative disease, and tissue regeneration. CIGMR has commissioned a large-scale DNA archiving facility that is designed for long-term storage of DNA and other biological samples for use in academic projects. CIGMR offers advice on genetic study design, ethics committee applications, bioinformatics, and data analysis. Clinicians and research scientists wishing to pursue genetic approaches to understanding complex traits are encouraged to work in collaborative partnership with the Centre.

Public Population Project in Genomics

The Public Population Project in Genomics (P3G) is an international consortium to create an interdisciplinary infrastructure for comparing and merging results from population genomic studies.[137] P3G plans to help the international research community deliver more effective health care strategies aimed at preventing disease and tailoring medicines and other treatment regimens to individuals, families, and communities.

P3G is dedicated to fostering collaboration among researchers and projects in the field of population genomics in an open and transparent manner. The Project develops research tools for effective collaboration among biobanks so that the international research community can share expertise and resources and facilitate knowledge transfer. P3G members are leading public organizations conducting large-scale genetic epidemiology projects and biobanks, with their own independent governance structures and objectives. In addition to the three founding regular members—CARTaGENE™, the EGP, and GenomEUtwin (involving eight countries)—the P3G regular membership includes CIGMR (United Kingdom); the Western Australian Genetic Health Project (Australia); the Danubian Biobank Consortium (involving six countries in Central Europe); the National Heart, Lung, and Blood Institute (National Institutes of Health [NIH], United States); KORA-Gen (Germany); and the Swedish LifeGene Project (Sweden). P3G's associate members include the Centers for Disease Control and Prevention's National Office of Public Health Genomics (formerly Office of Genomics & Disease Prevention) (United States); the Spanish Genome Foundation (Spain); the Research Center for Public Law (Canada); the McGill University and Genome Quebec Innovation Centre (Canada); the National Human Genome Research Institute (NIH, United States); and individual scientists, ethicists, and others.

There are three main P3G International Working Groups (IWGs), each composed of a multidisciplinary team of investigators, to promote knowledge exchange on P3G. They work in the areas of social, environmental, and biochemical investigations; information curation and technology; ethics; governance; and public engagement. The IWGs focus on core research activities, milestones, and deliverables for fulfilling P3G's

---

[136] Centre for Integrated Genomic Medical Research Web site. See http://www.postgenomeconsortium.com/cigmr/index.html [accessed December 21, 2006].
[137] Public Population Project in Genomics Web site. See http://www.p3gconsortium.org/ [accessed December 21, 2006].

strategic objectives. They oversee the content to be included in the P3G "knowledgebase" (P3Gdb). P3G "Cores" are the key work units of P3G, focusing on specific issues related to biobanks, using full-time personnel with expertise on specific issues related to biobanks, such as methodologies for integrating data, validating new technologies for biochemical analyses or genotyping, research on public engagement, relevant governance and ethics questions, and analysis methods in population genetics related to merging data from biobanks using different study designs.

Appendix E

Public Consultation Approaches by International Biobank Projects

Background

Within the four large population genetic studies that are underway in Iceland, Estonia, the United Kingdom, and Quebec, two major approaches are being taken to consult with the public and gather opinions. Godard et al. classify the type of consultation being conducted in Estonia and Iceland as focused more on quantitative rather than qualitative data.[138] In embarking on these consultations, deCODE genetics, Inc., which has exclusive rights to Iceland's medical records, and the Estonian Government did not specifically "reach out" to the public, but the launching of these biobanks and the rules governing them were established through the legislative process, which includes discussion.[139] The United Kingdom and Quebec, on the other hand, are engaging the public in a "participation or partnership approach," with a focus on both quantitative and qualitative measures of public opinion.[140]

deCODE genetics, Inc., and the Icelandic Health Sector Database

deCODE genetics, Inc., (deCODE) looked at community consent as a necessary prerequisite to performing a large cohort deoxyribonucleic acid (DNA) study. The Iceland Parliament passed a law allowing the development of the Icelandic Health Sector Database (IHSD), which can be viewed as proof of community consent. The debate took place through hundreds of newspaper articles and television programs and several town hall meetings across the country and informed the passage of the law, affecting the database license that was granted by the Parliament. According to deCODE, "debate is one of the most important mechanisms by which complex ideas are processed by democratic societies."[141] Following the debate and the media coverage, a survey indicated that 75 percent of the Icelandic population supported passage of the bill to allow the IHSD. A 2000 survey taken after the law was passed indicated that public support had grown to 90 percent (although this second survey may have used misleading wording/content and may not be accurate).[142,143] Organized and vocal dissent to the legislation still occurs, and more than 10 percent of the Icelandic population has chosen to opt out of the study, indicating that support for the project may be

---

[138] Godard B, Marshall J, Laberge C, Knoppers BM. Strategies for consulting with the community: the cases of four large-scale genetic databases. *Sci Eng Ethics*. 2004 10(3):457-77.
[139] Working Party on Biotechnology. *Tokyo Workshop Report: Human Genetic Research Databases—Issues of Privacy and Security*. Organisation for Economic Co-operation and Development, DSTI/STP/BIO 2005. 14.
[140] Godard B et al. 2004. Op. cit.
[141] Gulcher J, Stefánsson K. The Icelandic Healthcare Database and informed consent. *N Engl J Med*. 2000 342(24):1827-30.
[142] Ibid.
[143] Godard B et al. 2004. Op. cit.

weaker than surveys suggest and that the public consultation did not successfully address the concerns of a significant portion of the population. One opposing group, Mannvernd (Association of Icelanders for Ethics in Science and Medicine), has brought attention to some public concerns over human rights and the private control of medical and genetic information that are not addressed in the legislation allowing the IHSD.

Estonian Genome Project

Estonia used an approach similar to that of Iceland for consultation. The test of public opinion has been limited to Gallup poll results.[144] The Estonian Genome Project (EGP) informed the public about the pertinent scientific information regarding the Project, and public support/approval was based on this educational process. The EGP Web site provides information, definitions, and news stories related to the Project.[145] There appears to be less opposition to the EGP than to the deCODE project in Iceland, but this is likely because of some aspects of the projects themselves (i.e., the EGP carefully addresses concerns in the areas of consent, confidentiality, trust, and discrimination) rather than because of public education and consultation efforts.[146] In addition, the EGP received more media attention than did the deCODE project during the stages before the legislation was enacted. Since the inception of the EGP, Estonia has commissioned a marketing and consulting company to conduct polls to assess the public's knowledge about and opinion on the Project.[147]

UK Biobank

The funders of UK Biobank acknowledge the importance of consulting with the public as instrumental to the project's success and as valuable in shaping policies and practices.[148] The Wellcome Trust and the British Parliamentary Office of Science and Technology believe that the engagement model" of dialog between scientists and the general public is a better form of public consultation and communication than the "deficit model," which merely provides information about science to the public.[149] The idea for UK Biobank first appeared in 1999, and by 2000 the first public consultation was undertaken. UK Biobank consultations were preceded by reports from the Wellcome Trust on public views of science and consultations and the role of scientists in public consultation and debate.[150] The first Biobank-specific consultation focused on public perceptions of human biological sample collection.[151] In establishing principles to govern this collection, in the context of the large population cohort, the Wellcome Trust/Medical Research Council framework for collection was discussed with spokespersons for certain public groups and with scientists (those with an interest in medical research). Sixteen focus groups, composed of a diverse range of members of the general public, were formed to address policy concerns surrounding biological sample collection. Factors such as ethnic group, age, socioeconomic group, and geographic location were taken into account

---

[144] Ibid.
[145] Ibid.
[146] Ibid.
[147] Working Party on Biotechnology. Op. cit.
[148] The Office of Science and Technology and the Wellcome Trust. *Science and the Public: A Review of Science Communication and Public Attitudes to Science in Britain.* A Joint Report by the Office of Science and Technology and the Wellcome Trust, October 2000.
[149] Ibid.
[150] Ibid.
[151] UK Biobank. See http://www.ukbiobank.ac.uk/ethics/consultations.php [accessed December 21, 2006].

to ensure adequate representation of the entire population.[152] The topics looked into by the focus groups included awareness of, understanding of, and attitudes about topics such as medical and genetic research and human biological samples. Questions were addressed about the decision to donate, anonymity and confidentiality, consent, and ownership of these samples.[153] In addition to these focus groups, indepth interviews were conducted with specific stakeholder groups (including medical professionals, individuals or family members of individuals with a disease or disability, and community and religious leaders).[154] These initial consultations were followed by several subsequent consultations, including consultations with primary care health professionals on the recruitment of patients; consultation with social groups that were underrepresented in the initial consultation; a workshop with medical professionals, social scientists, patient advocates, lawyers, ethicists, and members of civil society groups to discuss ethics; consultation with industry representatives; a public panel (of previously consulted individuals without a stake in the project) on governance and framework; and a workshop with stakeholders on governance and framework.[155] The second round of consultations, with health care professionals, looked at developing the protocol for the project. The consultations that followed focused on oversight and ethical concerns in the areas of feedback, access to the database, and withdrawal from the study.[156] In total, UK Biobank sponsored 12 different consultations with various public groups.

Reports from each of these consultations, detailing the objectives, methods, findings, and more, can be found on the UK Biobank Web site at www.ukbiobank.ac.uk/ethics/consultations.php.

CARTaGENE™

In establishing its biobank, CARTaGENE™ in Quebec hopes to engage the public in a partnership decisionmaking process.[157] The first stage of communication with the public—focus groups that included members of Quebec's population (randomly selected from the phonebook[158]) representing Quebec's diverse linguistic, cultural, and regional groups—were held to look at the social and ethical implications and the social perceptions of the CARTaGENE™ project.[159] The primary goal of this stage of the consultation was to identify the concerns of the public regarding the establishment of the biobank.[160] In November 2001 four preliminary sets of these focus groups were held to gauge the popular opinion of this project.[161] In fall 2003, 19 of these focus groups (7 to 8 people each) were held to obtain a larger-scale view of the social and ethical concerns of the public. This initial set of focus groups was followed by a large-scale survey developed to assess how true the results of the focus groups were to the general public. This survey was conducted in all regions of Quebec, with more than 1,300 people agreeing to participate.[162] The second

---

[152] The Wellcome Trust and the Medical Research Council (2000). *Public Perceptions of the Collection of Human Biological Samples*. London. See http://www.phgu.org.uk/ecard?reference_ID=3870 [accessed December 21, 2006].
[153] Ibid.
[154] Ibid.
[155] UK Biobank Web site. See http://www.ukbiobank.ac.uk/ethics/consultations.php [accessed December 21, 2006].
[156] Working Party on Biotechnology. Op. cit.
[157] Godard B. *CartaGene (Abstract). Genome Canada GE³LS Project Presentations*. 2004. See http://www.genomecanada.ca/ge3ls2005/proceedings/08_04.asp [accessed August 2, 2005].
[158] Godard B et al. 2004. Op. cit.
[159] Godard B. *Consulting Communities: A Matter of Trust and Communication* (Presentation). 2003. See http://www.humgen.umontreal.ca/genconsult/docs/9.pdf [accessed August 2, 2005].
[160] Godard B et al. 2004. Op. cit.
[161] Godard B 2003. Op. cit.
[162] Godard B et al. 2004. Op. cit.

stage of public consultation consisted of developing a plan for communicating with the public before and during the project. To this end, workshops with ethics, law, and policy experts were held to look at the sorts of communication that were needed with the public before embarking on the project. During summer 2001 information was shared with the public through the CARTaGENE™ Web site, newsletters, and ongoing press releases and interaction with the media. In June 2003 a second workshop of professionals was held, and starting 6 months before beginning recruitment through the project's initiation, a telephone hotline was set up to respond to the questions and concerns of the public, and information about the project was dispersed through fliers, posters, and the Web site.[163] The final stage of the CARTaGENE™ consultation involved the establishment of a "deliberative electronic forum" by which the public can discuss concerns and share opinions with researchers, allowing for a dialog between the two groups that could continue throughout the project.[164]

[163] Godard B 2003. Op. cit.
[164] Godard B et al. 2004. Op. cit.