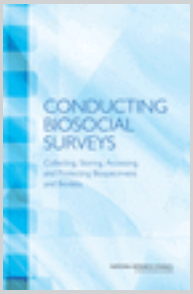

Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata

ISBN
978-0-309-15706-3

124 pages
6 x 9
PAPERBACK (2010)

Robert M. Hauser, Maxine Weinstein, Robert Pool, and Barney Cohen, Editors; Panel on Collecting, Storing, Accessing, and Protecting Biological Specimens and Biodata in Social Surveys; National Research Council

CONDUCTING BIOSOCIAL SURVEYS

Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata

Robert M. Hauser, Maxine Weinstein,
Robert Pool, and Barney Cohen, *Editors*

Panel on Collecting, Storing, Accessing, and Protecting Biological Specimens and Biodata in Social Surveys

Committee on National Statistics

Committee on Population

Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

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

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

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

This study was supported by Contract No. N01-OD-4-2139 between the National Academy of Sciences and the National Institutes of Health. Support for the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (award number SES-0453930). Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organization or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-15706-3
International Standard Book Number-10: 0-309-15706-4

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Research Council. (2010). *Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata.* Robert M. Hauser, Maxine Weinstein, Robert Pool, and Barney Cohen, Eds. Panel on Collecting, Storing, Accessing, and Protecting Biological Specimens and Biodata in Social Surveys. Committee on National Statistics and Committee on Population, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

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

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

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

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

www.national-academies.org

PANEL ON COLLECTING, STORING, ACCESSING, AND PROTECTING BIOLOGICAL SPECIMENS AND BIODATA IN SOCIAL SURVEYS

ROBERT M. HAUSER (*Chair*), Department of Sociology, University of Wisconsin, Madison
GEORGE M. CHURCH, Department of Genetics, Harvard University
GEORGE T. DUNCAN, H. John Heinz III School of Public Policy and Management, Carnegie Mellon
HENRY T. GREELY, Stanford Law School, Stanford University
MYRON P. GUTMANN, Directorate for Social, Behavioral and Economic Sciences, National Science Foundation
ROBERT J. LEVINE, Interdisciplinary Center for Bioethics Institution for Social and Policy Studies, Yale University
JOHN QUACKENBUSH, Department of Biostatistics, Harvard University
JEROME P. REITER, Department of Statistical Science, Duke University
ROBERT B. WALLACE, Department of Epidemiology, University of Iowa
MAXINE WEINSTEIN, Center for Population and Health, Georgetown University

BARNEY COHEN, *Study Director*
ROBERT POOL, *Consultant*
ULYANA VJUGINA DESIDERIO, *Christine Mirzayan Fellow*
JACQUELINE R. SOVDE, *Program Associate*

COMMITTEE ON NATIONAL STATISTICS

COMMITTEE ON POPULATION

Acknowledgments

This report reflects the efforts of many people, each of whom has contributed their time and expertise. In November 2008, the committee organized a public workshop and benefited greatly from the assistance and insight of many colleagues including: John Abowd, Cornell University; Paul S. Appelbaum, Columbia University; Ellen Wright Clayton, Vanderbilt University; Jennifer Harris, The Norwegian Institute of Public Health, Oslo; Kathie Mullan Harris, University of North Carolina; Murat Kantarcioglu, University of Texas; Alan F. Karr, National Institute of Statistical Sciences; Bartha M. Knoppers, University of Montreal; Barbara A. Koenig, Mayo College of Medicine; Karen J. Maschke, Hastings Center; Leslie Shaw, University of Pennsylvania; Kathi Shea, SeraCare, Inc.; Mary Fran Sowers, University of Michigan; Barbara Stanley, Columbia University; Holly Taylor, The Johns Hopkins University; Alan Westin, Columbia University (emeritus).

The project was undertaken at the request of the Division of Behavioral and Social Research at the National Institute on Aging (NIA) and funding from the NIA has made this report possible. Particular thanks go to Dr. Richard Suzman who was a catalyst for this report, both intellectually and financially, and we are grateful to him and the NIA for their support.

Several members of the staff of the National Academies made significant contributions to the report. The committee was established under the auspices of the Committee on National Statistics, directed by Connie Citro, who was instrumental in developing the study and provided guidance and support to the staff throughout the project. Particular thanks are due to Barney Cohen,

who served as the study director, Robert Pool for research and writing assistance, Ulyana Vjugina Desiderio for research assistance, and Jacqui Sovde for logistical support, Kirsten Sampson Snyder for help guiding the report through review, Rona Briere and Eugenia Grohman for skilful editing, and Yvonne Wise for managing the production process.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's (NRC's) Report Review Committee. The purpose of this independent review is to provide candid and critical comments that assist the institution in making its report as sound as possible, and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

The committee wishes to thank that following individuals for their review of this report: Kathleen Mullan Harris, National Longitudinal Study of Adolescent Health (Add Health), University of North Carolina, Chapel Hill; Meena Kumari, Department of Epidemiology and Public Health, University College London; Nancy A. Mathiowetz, Public Opinion Quarterly, University of Wisconsin, Milwaukee; Thomas McDade, The Center on Social Disparities and Health, Institute for Policy Research, Northwestern University; Eleanor Singer, Institute for Social Research, University of Michigan; Richard L. Sprott, Office of the Executive Director, The Ellison Medical Foundation, Bethesda, Maryland; James W. Vaupel, Office of the Executive Director, Max Planck Institute for Demographic Research, Rostock, Germany; and Kenneth M. Weiss, Department of Anthropology, Pennsylvania State University.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the report's conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Richard A. Kulka, Survey Research, Abt Associates Inc., Durham, North Carolina. Appointed by the NRC, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

I close by expressing my great appreciation to my fellow committee members. This report results from the exceptional efforts of the members of the committee, all of whom had many other responsibilities but who nonetheless generously gave much of their time and their expertise to the project.

Robert M. Hauser, *Chair*
Panel on Collecting, Storing, Accessing, and Protecting
Biological Specimens and Biodata in Social Surveys

Contents

Summary

Recent years have seen a growing tendency for social scientists to collect biological specimens, such as blood, urine, and saliva as part of large-scale household surveys. By combining biological and social data, scientists are opening up new fields of inquiry and are able for the first time to address many new questions and connections. But including biospecimens in social surveys also adds a great deal of complexity and cost to the investigator's task. Along with the usual concerns about informed consent, privacy issues, and the best ways to collect, store, and share data, researchers now face a variety of issues that are much less familiar or that appear in a new light.

In particular, collecting and storing human biological materials for use in social science research raises additional legal, ethical, and social issues, as well as practical issues related to the storage, retrieval, and sharing of data. For example, acquiring biological data and linking them to social science databases requires a more complex informed consent process, the development of a biorepository, the establishment of data sharing policies, and the creation of a process for deciding how the data are going to be shared and used for secondary analysis—all of which add cost to a survey and require additional time and attention from the investigators. These issues also are likely to be unfamiliar to social scientists who have not worked with biological specimens in the past. Adding to the attraction of collecting biospecimens but also to the complexity of sharing and protecting the data is the fact that this is an era of incredibly rapid gains in our understanding of complex biological and physiological phenomena. Thus the trade-offs between the risks and opportunities of expanding access to research data are constantly changing.

This report, which was funded by the National Institute on Aging (NIA), offers findings and recommendations concerning the best approaches to the collection, storage, use, and sharing of biospecimens gathered in social science surveys and the digital representations of biological data derived therefrom. It is aimed at researchers interested in carrying out such surveys, their institutions, and their funding agencies.

COLLECTING, STORING, USING, AND DISTRIBUTING BIOSPECIMENS

This report's initial message to social scientists undertaking the collection of biospecimens is that there is no need to reinvent the wheel. Although working in this emerging area may be new and unfamiliar, they will find available a number of existing documents from the biomedical field that offer advice and describe recommended procedures and laboratory practices for dealing with biospecimens. **The panel recommends that social scientists who are planning to add biological specimens to their survey research familiarize themselves with existing best practices for the collection, storage, use, and distribution of biospecimens. First and foremost, the design of the protocol for collection must ensure the safety of both participants and survey staff.** At the same time, many issues arise when biospecimens are collected as part of a social science survey that are not encountered in biomedical research. Thus it is often necessary to move beyond the biomedical model to find answers and best approaches for the social science context.

The panel notes that there is a growing tendency among social scientists to propose the collection of biospecimens in surveys regardless of whether they are needed to test a specific hypothesis. Yet many social scientists who decide to add biospecimens to their surveys are not fully prepared to provide for the storage and distribution of the specimens they collect. Indeed, the panel concluded that the issues involved in the storage and distribution of biospecimens are too complex and involve too many hidden costs to assume that social scientists without suitable experience can deal with them unassisted. Therefore, **the panel recommends that NIA and other relevant funding agencies support at least one central facility for the storage and distribution of biospecimens collected as part of the research they support.**

The collection of biological specimens along with the traditional social and behavioral data promises a number of benefits that are likely to extend beyond the original research team. However, advances are continually being made in genetic analysis and the ability to identify individuals through their social and biological data, and the sharing of biospecimens implies the depletion of a nonrenewable scientific resource. For these reasons, sharing biospecimens with other investigators is highly complex, and best practices in this area have yet to be established. Thus **the panel recommends that early in the planning process,**

principal investigators who will be collecting biospecimens as part of a social science survey develop a complete data sharing plan. In general, there is no one best plan for the use and reuse of specimens, but the plan should include a discussion of the adequacy of the storage and retrieval protocols. It should spell out criteria for allowing other researchers to use (and therefore deplete) the available stock of specimens, as well as to gain access to any derived data. The plan should also specify the procedures for accessing the specimens and data. It should include provision for the storage and retrieval of specimens and clarify how the succession of responsibility for and control of the specimens will be managed at the conclusion of the project. Finally, the plan should contain information on how specimens and data derived from them are to be documented and provide for public access to that documentation. To ensure the inclusion of all essential information, **the panel recommends that NIA (or preferably the National Institutes of Health [NIH]) publish guidelines for principal investigators containing a list of points that need to be considered for an acceptable data sharing plan. In addition to staff review, Scientific Review Panels should read and comment on all proposed data sharing plans. In much the same way as an unacceptable human subjects plan, an inadequate data sharing plan should hold up an otherwise acceptable proposal.**

SHARING DIGITAL REPRESENTATIONS OF BIOLOGICAL AND SOCIAL DATA

Once a survey has been conducted and biospecimens have been collected and analyzed, the survey team is left with a large amount of valuable social and biological data in digital form. Yet given the above-noted advances in genetic analysis and the ability to identify individuals through their social and biological data, a difficult issue facing the field is how to share the digital representations of these data as widely as possible while ensuring the protection of confidentiality. This issue is especially acute when detailed genetic information is generated from survey participants' biological samples and linked to social science data, which may be as sensitive or even more sensitive in their own right. At present, no data restriction strategy has been demonstrated to protect confidentiality while preserving the usefulness of the data for drawing inferences involving multidimensional interactions among genomic and social variables, which are increasingly the target of research.

For these reasons, **the panel recommends that both rich genomic data acquired for research and sensitive and potentially identifiable social science data that do not change (or change very little) with time be shared only under restricted circumstances, such as licensing and (actual or virtual) data enclaves.** Making confidential genomic data available for unrestricted public use would require such intense data masking to protect confidentiality that it would distort genomic analyses and sharply limit their usefulness. **As a security**

measure, the panel recommends that genomic data and other individual-level data containing uniquely identifying variables that are stored or in active use by investigators on their institutional or personal computers be encrypted at all times.

At the same time, some digital biosocial data can be shared if first subjected to procedures that alter the original data; restricted access should not be the only mode of data protection. Yet evaluating the specific risks of sharing data and devising ways to protect data from breaches are complex and specialized tasks requiring an expertise in disclosure protection methods not possessed by most principal investigators and their institutions. Currently, not enough is known to be able to represent these risks either fully or accurately. Determining the best protection schemes for the sharing of sensitive social and biological datasets also requires a significant investment of resources, and it would be wasteful for individual investigators to expend their resources on such efforts rather than on collecting and analyzing the data. Instead, **the panel recommends that NIA (or preferably NIH) develop new standards and procedures for licensing confidential data in ways that will maximize timely access while maintaining security and that can be used by data repositories and by projects that distribute data. The panel also recommends that NIA and other funding agencies assess the strength of confidentiality protections through periodic expert audits of confidentiality and computer security. Willingness to participate in such audits should be a condition for receipt of NIA support. Beyond enforcement, the purpose of such audits would be to identify challenges and solutions.**

Further, NIH should consider funding Centers of Excellence to explore new ways of protecting digital representations of data and to assist principal investigators wishing to share data with others. NIH should also support research on disclosure risks and limitations.

OBTAINING INFORMED CONSENT

If participants are to provide truly informed consent to taking part in any study, they must be given a certain minimum amount of information. They should be told, for example, what the purpose of the study is, how it is to be carried out, and what participants' roles are. In addition, because of the unique risks associated with providing biospecimens, participants in a social science survey that involves the collection of such specimens should be provided with other types of information as well. In particular, they should be given detail on the storage and use of the specimens that relates to those risks and can assist them in determining whether to take part in the study. To this end, **the panel recommends that, in designing a consent form for the collection of biospecimens, in addition to those elements that are common to social and biomedical**

research, investigators ensure that certain other information is provided to participants:

- **how long researchers intend to retain their biospecimens and the genomic and other biodata that may be derived from them;**
- **both the risks associated with genomic data and the limits of what they can reveal;**
- **which other researchers will have access to their specimens, to the data derived therefrom, and to information collected in a survey questionnaire;**
- **the limits on researchers' ability to maintain confidentiality;**
- **any potential limits on their ability to withdraw their specimens or data from the research;**
- **the penalties (such as the elimination of research support) that may be imposed on researchers for various types of breaches of confidentiality; and**
- **what plans have been put in place to return to them any medically relevant findings.**

Additionally, **the panel recommends that NIA locate and publicize positive examples of the documentation of consent processes for the collection of biospecimens. In particular, these examples should take into account the special needs of certain individuals, such as those with sensory problems, the cognitively impaired, or children.**

The panel also notes that participants in biosocial surveys are likely to have different levels of comfort with how their biospecimens and data are used. Some may be willing to provide only answers to questions and others to provide specimens as well. Some may be willing for their specimens and data to be used only for the current study, while others may consent to their use in future studies. One effective way to deal with these different comfort levels is to offer a tiered approach to consent, allowing the participant to determine just how his or her specimens and data may be used. Accordingly, **the panel recommends that researchers consider adopting a tiered approach to obtaining consent.** Tiers might include participating in the survey, providing specimens for genetic and/or nongenetic analysis in a particular study, and allowing the specimens and data (genetic and/or nongenetic) to be stored for future use. Additionally, **participants who are willing to have their specimens and data used in future studies should be informed about the process that will be used to obtain approval for such uses.**

As part of the informed consent process, **the panel also recommends that NIA direct investigators to formulate a plan in advance concerning the return of any medically relevant findings to survey participants and to implement**

that plan in the design and conduct of their informed consent procedures. In addition, **the panel recommends that NIA, the Office of Human Resource Protections (OHRP), and other appropriate organizations sponsor training programs, create training modules, and hold informational workshops on informed consent for investigators, staff of survey organizations, including field staff, administrators, and members of Institutional Review Boards (IRBs) who oversee surveys that collect social science data and biospecimens.**

A final issue facing social science researchers who include biospecimens in their surveys is obtaining approval from IRBs. A number of challenges exist, including the fact that few IRBs are familiar with both social and biological science; thus investigators may find themselves trying to justify standard social science protocols to a biologically savvy IRB or explaining standard biological protocols to an IRB that is used to dealing with social science. Another issue is that institutional IRBs are increasingly busy, and they are particularly demanding whenever potential risk to human subjects is at issue. Therefore, **the panel recommends that investigators considering collecting biomarkers consult with their IRBs early and often.**

CONCLUSION

The panel believes that, by following the above recommendations and several others offered in the full report, it should be possible to overcome many of the practical issues related to the collection, storage, retrieval, and sharing of biospecimens and derived biodata. The result should be improved access to research data without compromise to appropriate protection for research participants.

1

Introduction

Rapidly developing technology has made it increasingly feasible and attractive for researchers to collect blood and other biological specimens in nonclinical settings. As a result, those who conduct multipurpose household surveys have become increasingly interested in collecting various types of biospecimens along with responses to the more familiar social and behavioral questions (see, for example, National Research Council, 2008). Doing so enables researchers to extend their standard analyses of social and behavioral measures by integrating various biomarkers into their theoretical frameworks and empirical models. This practice of collecting biological specimens along with the traditional social and behavioral data promises a variety of benefits with respect to the sorts of questions that can be answered and the types of connections that can be explored, but it also adds a great deal of complexity—and cost—to the investigator's task. Although social scientists have long had to be concerned about such things as informed consent, privacy, collection and storage issues, and data sharing, the addition of biospecimens to their studies creates new issues and casts old issues in a new light.

Social science researchers wishing to collect biospecimens must address a wide variety of additional legal, ethical, and social issues, as well as a number of practical issues related to the storage, retrieval, and sharing of data. For example, deriving biological data from biospecimens and linking them to social science databases adds considerable effort and costs associated with developing a biorepository, establishing data sharing policies, implementing an increasingly complex informed consent process, establishing an additional process for reviewing how the biodata are going to be shared and used for secondary

analysis, executing material transfer agreements, dealing with intellectual property issues, and navigating a more complex process for obtaining Institutional Review Board (IRB) approval that encompasses both human subjects protection and biosafety compliance (Box 1-1 presents the panel's definitions of some key terms used in this report that need to be clearly distinguished in the context of this study). Researchers also must consider what steps are necessary to protect the confidentiality of participants, especially when data obtained from biospecimens are uniquely identifying. Finally, a number of questions must be answered about what happens to the biospecimens beyond the life of the particular investigation: Will the biospecimens be stored? If so, who will be allowed to use them? What permissions will be necessary? Who owns the biospecimens? Who can discard them? How long will they be retained? Can subjects demand their destruction? Does this include destruction of any biodata derived from them? Can the specimens and data be used for purposes other than those specified in the subjects' original consent? Will the investigator report back to subjects on

BOX 1-1
A Note on Terminology Used in This Report

The terms "biospecimens," "biomarkers," and "biodata" are sometimes used interchangeably, and researchers should be aware that these terms can have different meanings in different fields of study. In this report, the term "biological specimens" or "biospecimens" refers to the actual biological material that is collected from a study participant, such as blood, urine, or saliva. A "biomarker," often derived from a biospecimen, is a measurable factor that is associated with a particular medical condition. In population-based research, biomarkers are used to identify such things as cardiovascular risk factors, metabolic process measures, immune system activity, and nervous system activity. Examples include levels of cortisol (a stress hormone), C-reactive protein (a marker of acute inflammation), Epstein-Barr virus antibodies (a marker for immune function), total and HDL cholesterol (indicators of cardiovascular risk), and hemoglobin A1c (a marker of glucose intolerance). Biomarkers can also consist of gene alleles that are associated with a higher probability of a particular medical condition (e.g., ApoE-ε4 and Alzheimer's disease). In population-based research, biomarkers are often obtained by collecting biological specimens in nonclinical settings, but they can also be derived from, for example, anthropometric measures; accelerometry-based activity monitors; spirometry; or other measures of functional capacity, heart rate, blood pressure, and grip. "Biodata" refers to the digital data derived from biospecimens.

Two additional terms used in this report need to be distinguished. The term "biorepository" refers to a facility used to store human specimens for research purposes, while "biobank" denotes a facility used to store biodata. The contents of biorepositories range from large multinational collections of thousands of specimens to several dozen specimens in an individual researcher's freezer.

findings with health implications for them or their family? Will the investigator contact other family members? What are the limits on the investigator's ability to maintain confidentiality? Are there circumstances in which biological specimens may be acquired for research purposes without consent?

THE VALUE OF BIOSPECIMENS

The ability to collect biospecimens along with social survey data opens up a wide range of research opportunities. It becomes possible, for example, to estimate the distribution of a particular genetic variant within a representative sample of the general population and to correlate genetic variations with differences in human phenotypes. It also becomes possible to use the biodata derived from biospecimens to verify certain responses to survey questions, such as influenza exposure or infection with a sexually transmitted disease. But the potentially most far-reaching applications result from combining genetic and other biological data with data on social and environmental factors. The collection of biological specimens in population surveys that also collect data on socioeconomic, demographic, behavioral, physical health, and psychosocial factors opens up new avenues for research and may allow researchers to build integrated biosocial models of various biological and social phenomena.

For example, by combining biological and social survey data, it may be possible to document the linkages among social, behavioral, and biological processes that affect health and various other measures of well-being. To the extent that biomarkers reflect health, one can examine the effects of social factors on health or look at how health affects social status and social inequality. The ability to examine genetic data in conjunction with environmental and phenotypic data offers an important opportunity to study gene–environment interactions. It is now widely recognized that phenotypes are generally the product of an interplay between genetic and environmental factors; the availability of individual-level genetic and environmental information should make it possible to study this interplay in much greater detail than has previously been possible. Researchers could use such surveys to study the genetic determinants of longevity, for example, or to examine the association between genetically determined low monoamine oxidase levels and violent behavior and to learn whether that association is affected by whether the subjects were abused as children (Huizinga et al., 2006; Widom and Brzustowicz, 2006). Researchers could also examine the relationship between measures of life stress and the length of telomeres at the ends of chromosomes that serve as a biomarker of a cell's biological (versus chronological) age (Epel et al., 2004).[1]

[1]Telomeres are repetitive DNA-protein complexes at the ends of chromosomes that protect the chromosomes from deterioration. Recent research points to the crucial role of telomeres in cellular aging. See Aubert and Lansdorp (2008) for a recent review.

At the same time, not every social science survey will benefit from collecting biospecimens, and the significant costs involved must be weighed against the benefits in deciding whether to do so. The collection of biospecimens should be integral to the study design and the hypotheses being tested, rather than being tacked on to the study just because it can be done. Indeed, the collection of biospecimens may even detract from the principal mission of a survey. It might, for example, be so expensive and time-consuming that it would lessen the survey's effectiveness. Certainly, as noted above, the collection of biological samples will necessitate the expenditure of resources for storage, data sharing, and other purposes. Furthermore, the collection of biospecimens imposes a burden on participants as well as investigators and could conceivably affect contemporaneous and subsequent response rates. It is important to recognize, moreover, that in many cases, the potential benefits of biodata—particularly genetic data—are not altogether clear. This point is illustrated by genome-wide association studies (GWASs) focused on the linkages between single nucleotide polymorphisms (SNPs)[2] and common diseases such as diabetes and cancer. The high level of enthusiasm for such studies has been tempered by the finding that SNPs account for only a small percentage of the genetic risk for these diseases (Dickson et al., 2010; Wade, 2010). On the other hand, biobanking—the collection, storage, processing, and distribution of biological specimens—makes it more likely that biospecimens collected as part of a survey will have a valuable payoff, even if it is one that cannot be predicted when the specimens are collected. And while the specimens themselves are depletable, the biodata derived from them have potentially limitless uses. The trade-offs remain complex here as well, however, since the use of biospecimens and biodata for purposes other than the original research raises issues related to informed consent (see Chapter 4).

CURRENT STATUS

Today many surveys sponsored by the National Institute on Aging (NIA) and other federal agencies either collect biological specimens such as blood, saliva, urine, and buccal swabs or plan to do so in the near future. As discussed above, these data provide population-representative data from nonclinical samples that can be used for a variety of purposes including the calibration of self-reports of health and as a way to explore new pathways and causal linkages between biological and social variables. Some of this data is also being banked with the intention of using it in the future, for as yet unspecified purposes. Surveys that have collected or that currently collect biospecimens include the Dynamics of Health, Aging, and Body Composition Study (blood and saliva),

[2]A single nucleotide polymorphism (SNP) is a single-base variation in the genetic code, the most common form of polymorphism.

the Framingham Heart Study (blood), the Health and Retirement Study (blood specimens and buccal swabs), the National Health and Nutrition Examination Survey (NHANES) (blood, urine, hair, and buccal swabs, although there is some variation year to year), the National Longitudinal Survey of Adolescent Health (blood, saliva, and urine specimens), the National Long-Term Care Study (blood and buccal specimens), the NIA Alzheimer's Initiative (blood specimens and autopsy tissues), the Study of Women's Health Across the Nation (SWAN) (blood and urine specimens), and the Wisconsin Longitudinal Study (DNA). DNA amplification has been carried out on the biological specimens collected in some of these surveys, such as NHANES and SWAN, and could be performed on specimens collected in some of the other surveys as well. As a result of these efforts, a great deal of work has been done to develop policies and guidelines for the acquisition, collection, storage, and use of biological specimens, and this report draws on that experience. At the same time, these are complex subjects, and many questions remain unanswered. Even when the challenges are similar to those familiar to biomedical researchers, they will be new to most social scientists engaging in biosocial research.

PUBLIC ATTITUDES, PERCEPTIONS, AND RATES OF PARTICIPATION

Public Attitudes Toward the Collection of Biospecimens

Because the collection of biospecimens as part of social surveys depends on the willingness of individuals to contribute them, public attitudes and perceptions play an important role in the success of such efforts, and researchers must take these attitudes into account. Some authors claim that the public perceives biospecimens and the data derived therefrom to be significantly different from the traditional demographic, social, and economic data collected in surveys (see, for example, Greely, 2009). The former can sometimes be seen as more "objective" or "real" and thus potentially more powerful. They can also be perceived as being more hidden or secret because they can reveal things that cannot be known in any other way—things that those contributing the specimens may themselves not know, such as whether they possess a biomarker that is related to the probability of developing a certain disease. Thus they can be seen as more worthy of being protected or kept secret, in line with the strong tradition of keeping health information private (Greely, 2009).

Whether any of these public perceptions are well grounded in reality is another matter. A great deal depends on the types of biospecimens being collected. For example, survey participants may be more sensitive about sharing their earnings history than about providing a saliva sample. It is clear, however, that survey participants have less control over what is revealed through biospecimens than through traditional survey responses. A participant can refuse

to answer—or lie about—inquiries concerning, say, sexual history or income, but cannot prevent a blood sample from revealing the presence of a sexually transmitted disease or a DNA sample from indicating the existence of a genetic condition or predisposition.

In discussing public attitudes and perceptions toward the collection of biospecimens, it may be useful to talk about a gradient of sensitivity with respect to confidentiality. Some biological measures derived from biospecimens vary sufficiently across time that they do not raise the risk of reidentification. Some biological measures derived from biospecimens, such as cholesterol level, pose no more (or less) of a problem for confidentiality protection than many socioeconomic measures, while others, such as indications of illicit drug use or HIV or other disease status or genetic measures (such as a DNA sequence), may raise far more difficult issues of confidentiality and privacy protection. The potential harms from a confidentiality breach are significant because such measures, once associated with a specific person, may not only stigmatize the individual but also be used against him or her with regard to employment or in some other way. Moreover, genetic measures on one individual in a family may reveal characteristics of other family members. Additionally, the development of databases of genetic specimens that are stored for long periods increases not only the research potential but also the potential risks as new knowledge is discovered about genetic associations with health and behavior. Further increasing both the potential for innovative research and the potential for breach of confidentiality is the growing practice of linking survey records with administrative records, such as Social Security benefits and Medicare claim files.

It is also worth noting that people's attitudes and behaviors with regard to biospecimens are often inconsistent. People frequently report, for example, that they worry about the release of their genetic information because insurance companies might use it to discriminate, even though this has been expressly prohibited by law since the passage of the 2008 Genetic Information Nondiscrimination Act (GINA). At the same time, some measures that people do not consider sensitive and share readily, such as cholesterol levels, are at least as determinative of future serious disease as are genes given the present state of knowledge.

Findings from Opinion Surveys

Americans differ in their attitudes toward the collection of biospecimens and their willingness to participate in surveys that collect them (Westin, 2008). Greater knowledge generally leads to more favorable attitudes, and people are more likely to participate when they understand the importance of the research. Westin (2008) argues that the general public can be categorized as falling into one of three groups—(1) "privacy intense," (2) "privacy unconcerned," or (3) "privacy pragmatists." He argues that approximately 25 to 35 percent of the

population can be characterized as "privacy intense": they are skeptical about the motives and interests of government and business, they consider privacy to be extraordinarily important, they believe that the risks of their information being disclosed are very high, and they tend to be skeptical of the benefits.

In surveys, Americans consistently express the view that medical and health information is the most sensitive personal information. Furthermore, although Americans tend to trust doctors and health care providers with this information, they worry about third parties, such as insurance companies or employers, obtaining it. Not surprisingly, people with health problems are the most sensitive in this regard.

A large majority of the public—78 percent in a survey Westin performed for the Institute of Medicine (Westin, 2007)—say they are interested in health research, and three-quarters of the public believe health research is very important for society. As part of the survey, respondents were asked about participating in research that would require access to their medical records and other health information: How willing would they be to participate in such a study, and would they demand full disclosure of the study before they gave consent? Thirteen percent of respondents said they would not want to be contacted under any circumstances, and they would not even want to talk with somebody about participating. One percent said that they would always be willing to take part in any such study and that they did not even have to be asked for their consent. Eight percent said they would be willing to give general consent in advance if they were asked by the institution that held their medical records or health information. Another 19 percent said they would agree if they were given assurance that their identity would not be revealed and that an IRB would administer the study. And 38 percent—the largest single group—said they would have to have the research described to them each time so they could decide whether to participate. Thus a total of 57 percent would agree to having their information used if certain privacy-oriented conditions were met.

Observed Participation Rates in Social Surveys Collecting Biospecimens

Evidence from a number of different social surveys provides a sounder basis than opinion surveys for assessing people's willingness to participate in social science research that includes the collection of biospecimens. Most social surveys report higher rates of willingness to participate in such research than are perhaps suggested by opinion surveys, although these rates vary depending on the method of data collection and the health, age, and social characteristics of the subjects. Marmot and Steptoe (2008), for example, report on the experience of the Whitehall II and the English Longitudinal Study of Aging (ELSA) in the United Kingdom. Data collection in these surveys involved both face-to-face contact with participants and the assessment of physical measures, including blood sampling, and the authors were concerned that participants might find

the test sessions too long or burdensome. On the other hand, participants could benefit from the periodic medical screening sessions, which might reveal health problems that would otherwise have gone undetected. The authors report only a 16 percent loss of participants between the baseline survey and the first clinical follow-up. Involving participants in more intensive investigations did not deter them from taking part or lead to widespread sample attrition. Rather, participants in the more intensively studied group were more likely to remain involved in the study. The authors hypothesize that many participants found these more intensive studies to be intrinsically interesting and that they derived from the study detailed clinical information that would help them monitor their health status (Marmot and Steptoe, 2008).

Lindau and colleagues (2009) report participation rates in a nationally representative probability survey of 1,550 community-residing women aged 57–85 conducted in 2005 and 2006. All 1,550 female respondents in the study were asked to provide a self-administered vaginal swab specimen midway through the interview; 1,028 agreed to do so. Hauser and Weir (in press) report a similar response rate (approximately 65 percent) for saliva collection by mail for DNA analysis in the Wisconsin Longitudinal Study.

ISSUES SURROUNDING THE COLLECTION OF BIOSPECIMENS IN SOCIAL SURVEYS

Survey researchers intending to collect biospecimens must grapple with a number of issues, many of which will be unfamiliar to them. The majority of these issues can be grouped into three broad areas: (1) the collection and storage of biospecimens, (2) sharing of biospecimens and the data collected therefrom, and (3) informed consent. In each of these areas there are concerns that must be addressed, questions that must be answered, and policies that must be devised if the benefits of collecting the biospecimens are to be fully realized while the interests of research participants are protected.

Concerning the collection and storage of biospecimens, for example, what precautions should be taken in collecting the specimens from survey participants? What considerations should factor into an investigator's choice of a storage facility in which to maintain the specimens from a survey? What should the policies be for sharing those specimens?

With respect to the sharing of biospecimens, what are the risks to confidentiality in sharing specimens or the data derived from them? How can those risks be minimized while the usefulness of the data is maximized? What are the advantages and disadvantages of restricting access to the data versus restricting the data themselves?

Concerning informed consent, there are a great many uncertainties and controversies: How does one arrange for informed consent for specimens and data to be used in some future unspecified research project? How should

researchers handle situations in which an analysis of data has revealed significant health information about a participant—information that may not be known to the participant? What happens when a study participant withdraws his or her consent? What should be included in an informed consent form for a social science survey that will include the collection of biospecimens? How should one deal with IRBs when submitting a proposal to conduct this sort of study?

The following chapters address each of these issues in detail.

STUDY PURPOSE, SCOPE, AND APPROACH

To address the issues outlined above, in 2008 NIA's Behavioral and Social Research (BSR) Program asked the National Academies to convene an ad hoc panel of experts for the purpose of identifying best practices with respect to collecting, storing, protecting, and accessing biospecimens collected in social science surveys and the biodata derived therefrom. It is worth stating at the outset that these issues are not new: the research community is familiar with the challenge of reconciling the benefits of providing wider access to research data and the resulting increased risk of a breach of confidentiality. Several previous National Research Council (NRC) reports have addressed aspects of the subject (see, for example, National Research Council, 1993, 2005). However, these issues have not been sufficiently examined in the context of biosocial surveys that collect both biospecimens and typical social science data, a discussion that is becoming increasingly salient. BSR is continuing to develop a portfolio of new research directions linking social and behavioral research with data on genetics and genomics. Several large longitudinal data collection efforts funded by BSR (e.g., the Health and Retirement Survey and the Wisconsin Longitudinal Study) are now collecting various types of biospecimens. In other cases, plans for collecting new biospecimens are currently under way. For many BSR-supported researchers, the procedures and protocols surrounding the collection, storage, and sharing of biospecimens are new. Furthermore, ongoing advances in bioinformatics (see, for example, Homer et al., 2008) have raised issues of confidentiality and security that have prompted BSR to review its procedures with respect to data sharing.

The 10-member panel that conducted this study was appointed under the auspices of the Committee on National Statistics and the Committee on Population of the National Academies. Its members included leading experts in social, behavioral, genetic, ethics, and genomic studies who were familiar with the wide range of issues involved. The panel was charged with preparing a report that would address these issues and provide recommendations for best practices, procedures, and guidance for funding agencies, IRBs, and researchers.

To accomplish its task, the panel organized a public workshop as a means of interacting with other leading scientists engaged in (or considering) the col-

lection of biospecimens. (The workshop agenda is presented in Appendix A, while the participants are listed in Appendix B.) The workshop discussions were designed to explore issues related to informed consent, data collection, confidentiality protection, data archiving, and data access for multipurpose population surveys that collect biological specimens and measures in addition to socioeconomic/demographic, behavioral/lifestyle, and physical and mental health measures. Specifically, the panel was tasked to review the following issues, with particular reference to surveys sponsored by NIA:

- information that should be provided to survey respondents for informed consent and how the language of consent forms affects people's willingness to participate in surveys;
- methods for collecting and processing genetic and biological specimens and measures to minimize the burden on respondents, maximize research potential, and protect confidentiality and privacy;
- relevant laws, regulations, and policies, including the Common Rule for Protection of Human Subjects, the Confidential Information Protection and Statistical Efficiency Act of 2002, the 2002 regulations issued under the Health Insurance Portability and Accountability Act of 1996, and relevant National Institutes of Health policies on data sharing, certificates of confidentiality, and related topics, including the repository for genomewide association studies;
- factors for IRBs to consider in reviewing requests for the collection of biological specimens and measures in surveys;
- the risks of and evidence for actual misuse of biological specimens and measures in surveys;
- whether and which statistical techniques can be used to make genetic and other biological measures anonymous in microdata files while preserving their utility for research;
- the costs and benefits of alternative systems for archiving genetic and other biological specimens and measures derived from population surveys to permit later research use while protecting confidentiality; and
- the costs and benefits of alternative forms of access to microdata containing genetic and biological measures, such as secure research data centers and licensing.

A word about this statement of task and the scope of this study is in order. Although the statement of task mentions both biological specimens and measures, the panel chose to focus this report on the former, for two reasons. First, as will be clear from the discussion in the following chapters, unique issues of collection, storage, and sharing and of informed consent and confidentiality are associated with biospecimens that do not arise with respect to biological measures such as the taking of height and weight or blood pressure. Second,

whereas the data derived from measures are well defined and finite in scope, a wide-ranging and potentially limitless set of data can be derived from biospecimens, further complicating the issues that must be addressed.

The panel also wishes to emphasize that the starting point for this study is a decision that the benefits of collecting biospecimens as part of a social science survey outweigh the costs noted earlier. This study does not address the calculus that factors into the decision about whether to collect biospecimens as part of a social science survey. The panel emphasizes, however, that the trade-offs involved are complex, not least because, as noted above, the potential benefits are as yet not fully understood.

To carry out its charge, the full panel met four times. In March 2008 the panel met to discuss its statement of task with the sponsor; to review prior work; and to identify critical themes and speakers for the public workshop, which was held in November 2008. Following the workshop, the panel met three more times (November 2008, June 2009, and August 2009) to discuss the presentations and the rich interactive discussions that had occurred at the workshop, to deliberate, and to outline this report. The report is based on the deliberations of the panel as informed by the workshop but is the product of the panel, not merely an account of the workshop.

REPORT ORGANIZATION

The remainder of this report presents the panel's findings, conclusions, and recommendations. Chapter 2 deals with issues concerning the collection, storage, use, and distribution of biological data, including issues of custodianship and ownership. Chapter 3 reviews issues related to confidentiality and data sharing, including deidentification and other approaches to preserving the privacy of participants. Chapter 4 addresses issues related to informed consent, including biobanking, the use of blanket consent, and the role of IRBs. Finally, Chapter 5 offers the panel's recommendations for practices and procedures that can best facilitate research and protect participants as the collection of biospecimens in social science surveys moves forward over the next 5 to 10 years.

2

Collecting, Storing, Using, and Distributing Biospecimens

Social scientists have long experience with population surveys, but the collection of biospecimens as part of a survey protocol requires different technical and logistical skills and introduces complex legal and ethical issues. Additional, and often unforeseen, costs in terms of both money and time frequently must be borne as well. To be sure, epidemiologists have a long tradition of collecting biological data (often mimicking a clinical setting in the field) along with social and demographic data (that are often less rich than social surveys) but typically they are not designed to be representative of a larger population universe. Furthermore, social scientists generally have expectations regarding data sharing and access that differ from those of epidemiologists, clinicians, and medical researchers.

Adding the collection of biospecimens to social surveys provides the opportunity to identify or test biomarkers that can address questions in many areas of interest, from susceptibility to disease to measures of environmental exposure to a wide range of compounds. It is important, however, for proposed biomarkers to be aligned properly with the goals, hypotheses, and concepts of the research. Each team of investigators will need to consider carefully numerous questions that will guide their decisions regarding the collection, use, and storage of biospecimens: What biospecimens will best advance the research questions or hypotheses being investigated? What kinds of specimens will provide the most reliable measures (and be practicable in the field)? What constraints are involved in transporting specimens to a laboratory or tissue repository (in time)? What kind of preparation needs to be done in the field (and is it feasible)? What safety issues need to be addressed? Which assays should be done

first? In addition, a host of logistical questions must be considered by the participating scientists and their Institutional Review Boards (IRBs). These include how the specimens will be collected, managed, and stored; how access to them will be regulated and monitored; how they will be used to extract relevant biomarkers; and whether and how the laboratory data and survey information will be made available to scientific collaborators, the broader scientific community, and the public.

There is no single formula for adding the collection of biospecimens to a survey; each investigation will have different requirements that will have to be weighed and balanced against constraints of budget, time, field conditions, and participant burden. Significant advance planning, piloting, and revision will be required, and even then it is wise to expect the unexpected.

These cautions are particularly important given the rapid pace of technology development both with respect to the collection of specimens (e.g., blood, urine, saliva, or hair) as well as the ability to analyze data derived from such specimens (ranging from blood glucose levels to C-reactive protein to mercury levels to DNA). The prime example here is the analysis of genetic markers. Just a few years ago, analysis of large numbers of biological specimens was limited to examining a small number of candidate genes or, at best, a few thousand genetic markers. But the advent of microarrays that allow the measurement of expression levels of unprecedented numbers of human genes in a single experiment or the profiling of a million single nucleotide polymorphisms (SNPs) across the genome began to change the way scientists and the IRBs that oversee their projects conduct their work. For example, pooled data from genome-wide association studies (GWAS), representing the genomes of multiple individuals, were viewed for some time as acceptable for public release. But when recent work on forensic analysis of DNA samples showed that the presence of a single individual could be detected in a large pool of such samples (Homer et al., 2008), researchers and policy makers at the National Institutes of Health (NIH) reconsidered and changed data release policies.

Concerns about data release and protected health information have been compounded by the rapid pace of development of next-generation DNA sequencing technologies. Sequencing the first human genome was a 15-year project that cost billions of dollars; new technologies, however, allow sequencing of human genomes in times that are on the order of 1 month at a cost of less than $100,000, and technology advances may reduce the cost to $1,000 or less. This capability raises the prospect of having to deal with unprecedented amounts of personal information—the entire genome sequence from large numbers of individuals. While such data may have great potential for the discovery of biomarkers and functional studies, dealing with the data and their implications for privacy and protection of human subjects will require addressing many as yet unanswered questions.

The focus in this chapter is on the specimens themselves, not the data

derived from them (discussed in Chapter 3), although clearly the kinds of data that ultimately can be obtained are constrained by the specimens that are collected, the quality of the collection and storage procedures, the timing of the assays, and the integrity of the procedures. The chapter begins by describing some documents that provide guidance regarding best practices for investigators. It then reviews cross-cutting considerations for studies that include biospecimens. The subsequent sections offer a detailed discussion of issues related to collection, use, storage, custodial responsibility and ownership, and access and distribution.

BEST-PRACTICE REFERENCE DOCUMENTS

Fortunately, researchers undertaking the collection and analysis of biospecimens for the first time do not have to reinvent the wheel. First, an increasing number of social science investigators have hands-on experience with collecting, storing, using, and distributing biospecimens. A simple piece of practical advice is to talk with them. Second, documents that describe recommended procedures and laboratory practices are already available. Although they were not developed with social surveys specifically in mind, these protocols have been field tested and approved by numerous IRBs and ethical oversight committees. These best-practice documents are updated frequently to reflect the growing knowledge and changing opinions about the best ways to collect, store, handle, and distribute biological specimens. They are useful places to begin.

The panel identified three documents that it believes are most relevant to the concerns of newcomers to biospecimen collection. Although these documents focus on guidelines for biospecimen repositories, they also provide more general guidance and important reminders for consideration during the planning process. They will be most useful for identifying factors that need to be considered in choosing a repository or in establishing even a small biospecimen archive.

First is *2008 Best Practices for Repositories: Collection, Storage, Retrieval, and Distribution of Biological Materials for Research*, prepared by the International Society for Biological and Environmental Repositories (ISBER) (International Society for Biological and Environmental Repositories, 2008). ISBER was formed in part to develop effective strategies for the long-term storage of biological specimens; it "fosters education and research and promotes quality and safety in all activities relating to specimen collection, storage and dissemination" (International Society for Biological and Environmental Repositories, 2005, p. 5). The document covers repository organization, management, and facilities; storage equipment and environment; quality assurance and quality control; safety; training; tracking of biological materials; packaging and shipping; specimen collection, processing, and retrieval; and legal and ethical issues for human specimens.

Second, in 2002 the National Cancer Institute (NCI) set out to understand the quality and characteristics of biospecimens used in cancer research and to develop a set of principles for their collection and handling. A 5-year process of research, workshops, and feedback led to the publication in 2007 of *National Cancer Institute Best Practices for Biospecimen Resources* (National Cancer Institute, 2007). Less detailed than the ISBER publication, the NCI guidelines do not offer specifics, such as the temperature at which tissue samples should be stored or the precise sorts of labels that should be used. Rather, they provide "salient guiding principles that define state-of-the-science biospecimen resource practices, promote biospecimen and data quality, and support adherence to ethical and legal requirements" (National Cancer Institute, 2007, p. 1).

Third, in 2007 the Organisation for Economic Co-operation and Development (OECD) released *OECD Best Practice Guidelines for Biological Resource Centres* (Organisation for Economic Co-operation and Development, 2007). These guidelines are broader than those of NCI in that they address more types of biological specimens, including viruses, bacteria, and other microorganisms, as well as samples taken from human subjects. They are also broader in the sense that they represent best practices from 30 countries rather than being drawn primarily from the best practices of U.S. institutions. However, they take the same general approach as the NCI guidelines in that they provide guiding principles rather than specific suggestions.[1]

Collectively, these documents provide excellent advice on how to store biospecimens in ways that preserve them,[2] ensure adequate documentation, and protect confidentiality. However, they do not address questions that are most closely related to the design of the research itself, such as choice of biospecimen (e.g., blood, urine, saliva), choice of biomarker, and choice of assay.[3] Another important distinction for biosocial researchers is that the data archive (i.e., the collection of data derived from the specimens, as well as the data from the survey) is likely to be maintained separately from the specimens themselves, while documentation about the specimen collection and survey protocols may be archived in yet another location. Unlike the researchers addressed in the best-practice documents, therefore, social scientists are likely to use a biorepository as a facility for the storage and distribution of specimens, not as a partner in the substantive or intellectual concerns of the study or as an entity with a claim to ownership of the specimens. These differences translate to a different set of

[1]Two organizations in the United Kingdom (UK) have also produced valuable best-practice guides. The Wellcome Trust, the UK-based charity that funds innovative biomedical research, published *Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility* (Wellcome Trust, 2003), while the UK Biobank, a long-term biobank study investigating the genetic and environmental components of disease, produced *UK Biobank Ethics and Governance Framework* (UK Biobank, 2007).

[2]Blood spots are not addressed.

[3]For discussion of some of these issues, see National Research Council (2008).

challenges for data security and the protection of confidentiality from those covered in the best-practice documents. That said, the documents provide useful guidelines for the choice and evaluation of a repository for biospecimens: even if the biorepository is not an intellectual partner, it must track and protect information in its custody.

CROSS-CUTTING CONSIDERATIONS

Ethical, Legal, and Policy Issues

Safeguarding the rights and safety of study participants is of paramount importance. Honoring promises regarding the privacy and confidentiality of participants and ensuring that all procedures conform to what has been established during the informed consent process are fundamental to ethical research. The damage that can be caused by breaches of confidentiality—possible at many stages of research using biospecimens—can be devastating. Policies and protocols for using biospecimens properly in scientifically sound research and for ensuring adherence to federal, local, state, and international laws governing the collection, storage, and use of biospecimens by all members of the research team and its collaborators and contractors are fundamental. Chapter 3 examines issues related to confidentiality and the sharing of biological and social data, while Chapter 4 contains a more detailed discussion of issues related to informed consent.

Informed consent, privacy, confidentiality, and identifiability in genomic research are all closely interrelated (Lowrance, 2006a). All are relevant to decisions about deidentification and coding of data (and thus to protection, which in turn allows for dissemination). *Privacy* refers to the protection of an individual from unwanted intrusions, including the acquisition of information about that individual. It is a general protection against other individuals, groups, and society as a whole. By contrast, *confidentiality* refers to the protection of information about an individual that has already been provided willingly to one party; if the information is confidential, the recipient is responsible for ensuring that it is not released to others without the donor's permission. *Identifiability* is the ability to associate data with a particular person. The identifiability of data can be thought of in terms of a spectrum from data that are impossible to identify, to those that can be identified as possibly being linked to a given individual, to those for which that linkage is known with certainty.

A primary consideration is that the privacy of study participants should be respected and protected. Sound scientific research involving biospecimens depends on protecting the privacy of the participants who contribute them and the confidentiality of their data. Yet many of the very scientific and technological advances that make the collection of biological specimens so valuable—progress in genomic and proteomic science, the sequencing of the human

genome, the culture of rapid release of genomic data, and the increasing use of Internet-based searchable databases containing those data—increase the risk of potential breaches of confidentiality and make it increasingly important to anticipate and prevent such breaches (National Cancer Institute, 2007). Advances in digital computing, communication, and storage technologies also are linked with a new world of research characterized by immense data sets, new forms of interdisciplinary collaboration, and unprecedented levels of data sharing among researchers, all of which raise the risk of breaches throughout the data collection, analysis, publication, and distribution process (see National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2009).

While summary-level data are of great value, biosocial research often requires that individual-level survey data be linked to each biological specimen to facilitate the analysis of biodata derived from those specimens. However, the unrestricted release of both the individual survey data and the biodata poses potential risks for breaches of confidentiality. Although the public release of nominally deidentified data used to be widely accepted, the concept of deidentification needs to be reconsidered when the data can represent an individual's entire genome sequence or the state of a million or more variant positions along the genome, possibly together with various social, economic, psychological, or physiological characteristics. There is as yet no way of knowing what the frequency and consequences of breaches in data security might be, but it appears clear that they have the potential to lead to individual or group discrimination or stigmatization, as well as to diminished public trust and reduced participation in and support for research. It is crucial to have policies and protocols in place to prevent the unauthorized release of sensitive information and to respond effectively if it occurs.

At the federal level, constitutional, legislative, and administration protections of privacy apply to situations in which biospecimens are collected, stored, analyzed, or disseminated. In at least one case, the U.S. Supreme Court implied that the Constitution might guarantee individuals some level of "informational privacy" (*Whalen v. Roe*, 429 U.S. 589 [1977]). The Court has not developed that right further, however, and in any event, it would apply only to actions by governments and not private organizations, such as most health plans (Jones and Sarata, 2008, p. 17). The federal court finding that is perhaps most relevant to the issue of genetic privacy is *Norman-Bloodsaw v. Lawrence Berkeley Laboratory*. A group of workers at Lawrence Berkeley Laboratory (LBL) complained that their privacy rights had been violated when the laboratory used their stored blood samples to test for pregnancy, sickle-cell anemia, and sexually transmitted diseases without their consent. The district court sided with LBL, but the U.S. Court of Appeals for the Ninth Circuit disagreed and sent the case (later settled out of court) back to the district court. The Court of Appeals found that, although the employees had originally agreed to provide blood samples,

"the ensuing chemical analysis of the samples to obtain physiological data is a further intrusion of the tested employees' privacy interests." The court added, "One can think of few subject areas more personal and more likely to implicate privacy interests than that of one's health or genetic make-up," and suggested that the tests "may also be viewed as searches in violation of Fourth Amendment rights."[4]

Among the federal statutes that may apply to genetic privacy are the Privacy Act of 1974 (5 U.S.C. § 552a) and the Freedom of Information Act (FOIA) (5 U.S.C. §§ 552 et seq.). The Privacy Act makes it illegal for federal agencies to disclose information (including medical information) from records maintained on individuals except under certain conditions. FOIA was intended to make much of the information maintained by federal agencies available to the public, but the act specifically exempts "personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy."[5]

Additional federal privacy protections flow from the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and the Genetic Information Nondiscrimination Act (GINA) of 2008. Pursuant to HIPAA, the Secretary of Health and Human Services (HHS) put forth a set of regulations designed to protect the privacy of health information. The regulations cover all types of health information in any form—electronic, on paper, or oral—and they define health information very broadly so that it encompasses genetic information as well as such details as disease status or family medical history; however, they apply only to "covered entities," that is, institutions and providers that are subject to HIPAA. According to HHS regulations, the entities covered by HIPAA are health care providers that carry out certain types of electronic transactions, health plans, and health care clearinghouses (U.S. Department of Health and Human Services, 2005a). Many biosocial surveys will not be conducted in HIPAA-covered entities, but following the guidelines is a good step toward protecting privacy, so the spirit of that law can perhaps serve as a useful guiding principle. In cases where HIPAA does apply, researchers "are required to have in place reasonable safeguards to protect the privacy of patient information and limit the information used or disclosed to the minimum amount necessary to accomplish the intended purpose of the disclosure" (Jones and Sarata, 2008, p. 19; but see also Institute of Medicine, 2009). Serious violations may be punished by up to 10 years in prison and fines of up to $250,000. The consequences of breaches of confidentiality can be so dire that the penalties should be proportionately large; some would argue that $250,000 for an institution might not be enough. In addition to HIPAA, GINA protects people from being discriminated against by health insurers or employers as a consequence

[4]*Norman-Bloodsaw v. Lawrence Berkeley Laboratory,* 135 F.3d 1260, 1269 (N.D. Cal. 1998).
[5]5 U.S.C. § 552(b)(6).

of differences in their DNA that could potentially affect their health, say, by increasing their chances of getting a particular disease. The law enables people to participate in research studies or obtain medical tests without fear of having their DNA information subsequently used against them by health insurers or in the workplace.

An alternative to promising confidentiality is obtaining consent from participants in advance with the understanding of full disclosure (Church, 2005; Lunshof et al., 2008) and acknowledging that biological samples are convertible into identifiable genomic and trait data. Large security gaps are often social in nature; for example, even though high-security defense access requires psychosocial security checks (well beyond any employed by health research organizations), classified materials still slip outside of secure environments (e.g., via theft, mislabeling, cross-contamination, willful sharing of data despite consequences, or unanticipated reidentification algorithms). To some, it may appear disingenuous to continue to promise confidentiality of samples in light of numerous recent examples to the contrary. To help ensure informed consent (rather than merely obtaining legal signatures on consent forms), some groups require 100 percent scores on tests of comprehension of the potential risks and benefits for the subjects, families, and society (Church, 2005). This check has the additional benefit of educating participants before the study, rather than after they need to be informed of some alarming result.

The NCI recommendations regarding privacy and confidentiality (National Cancer Institute, 2007) include a detailed section designed to help guide biorepositories in their role as honest brokers of sensitive information. Although the recommendations refer to biorepositories, they are applicable to multiple stages of research, analysis, storage, and distribution processes:

> Biospecimen resources should establish clear policies for protecting the privacy of identifiable information. These policies may include data encryption, coding, establishing limited access or varying levels of access to data by biospecimen resource employees, use of nondisclosure agreements, and use of an honest broker system. (National Cancer Institute, 2007, p. 22)

Researchers (and biorepositories) should consider obtaining a certificate of confidentiality (National Cancer Institute, 2007). NIH is authorized to issue certificates to researchers involved in clinical, biomedical, or other research that would allow them to refuse to disclose identifiable information in federal, state, local, civil, criminal, administrative, or other proceedings (see also National Institutes of Health, 2009). If such a certificate is obtained, the participants should be informed about it during the informed consent process. The informed consent process should also include explaining what such certification means and the limits to the protection it can afford (National Cancer Institute, 2007).

Technical and Logistical Issues

Most biospecimens are classified as biohazardous materials. Personnel who handle them at any stage of a project should be trained in their safe use and handling and take adequate precautions. Protection of those who collect and handle the specimens, including laboratory personnel, should be ensured. Potential threats include not only exposure to infectious agents, but also, for example, exposure to toxic chemicals used in processing specimens, cuts from broken glass or shards, exposure to dry ice during the shipping process, and irritation from adhesives. Protocols for collection need to be established that will ensure the safety of both research participants and the study staff, and protocols for processing, shipping, storage, and use must take account of the safety of each person who may be exposed. Research institutions may have specific requirements to ensure the safety of biological materials.

Documentation is needed at all stages of a study. This requirement may be especially complex for social scientists who, in general, will be archiving not only specimens but also the more familiar kinds of self-reported or interviewer-assessed data that are typically included in social science surveys. As noted earlier, biosocial survey data and specimens usually will be archived in more than one location or type of facility. The data requirements associated with the biospecimens themselves include information about the administrative and operational "trail" of the specimens—collection site, date and time of collection, freeze/thaw occurrences, assay procedures, and laboratory quality control (QC) and quality assurance (QA), to name but a few that will be new to most social survey investigators. Researchers must also establish protocols for linking the different archives in a manner that protects individual confidentiality while having a minimal effect on research. Excessively restrictive rules for linking the data and specimens may hinder the originally envisioned research, the kind of research to which the participant consented in the first place.

The specimens themselves will generally be stored in a biorepository. Detailed information on all specimen collection, processing, and storage procedures should be recorded and tracked by the biorepository's information and database systems. The NCI guidance discussed earlier includes detail on the necessary systems: "An informatics system should support all aspects of biospecimen resource operations, including (but not limited to) research participant enrollment and consent; biospecimen collection, processing, storage, and dissemination; QA/QC; collection of research participant data; data security; validation documentation; and management reporting functions. In addition, the system should manage clinical annotations to the biospecimens" (National Cancer Institute, 2007, pp. 11–12). The NCI guidelines cover such areas as identification of biospecimens, integration with local systems, interoperability, and ethical and legal issues pertaining to informatics systems.

Pretesting is important at all stages of a study. Those new to working with biospecimens will find it necessary to test processes and equipment that social

scientists usually do not have to consider. The adhesive on specimen labels, for example, should be checked to ensure that the labels will not come off in the freezer or if they get wet; in a similar vein, the ink must be checked to make sure it will not smudge or run. It is wise to use ice water or "scrap" specimens to test shipment protocols for frozen specimens and ensure that the packaging includes adequate dry ice. It is wise as well to anticipate possible delays at the shipping docks, at customs inspection stations, in traffic, and at airports; establishing fall-back procedures for misadventures is advisable. In short, every step of the protocol should be pretested with the equipment and processes to be used at that step. Sufficient training of survey staff (data and specimen collectors and handlers) is also critical to eliminate the human error that may occur in the collection and processing of biospecimens.

COLLECTION: DESIGN AND OPERATION

As discussed above, the design of the protocol for collecting biospecimens must ensure the safety of both participants and staff. Interviewers or other survey staff who collect biological specimens from human subjects should take precautions to avoid infection from any pathogens that may be present in the specimens, especially when blood samples are being collected (Twitchell, 2003). Again, these issues should be addressed during the training of survey staff. Well-trained interviewers also can facilitate the process of obtaining participants' informed consent, decrease data collection errors by ensuring that the correct protocols are followed, and minimize the risk to study participants. In addition, there has been a movement recently to develop measures and methods that can be administered by nonclinicians, in some cases yielding immediate information that can be imparted to participants by an interviewer who is not competent to offer any clinical or diagnostic interpretation of the information. Thus it is important to emphasize in interviewer training and supervision that, while interviewers may tell respondents what they have just measured (e.g., blood pressure), they should not offer any interpretation of such information.

Consideration must also be given to the different kinds of preparation that will need to occur in the field. For example, urine may need to be aliquoted or acidified, or blood may need to be centrifuged. Some assays must be performed on freshly collected specimens, and in these cases, it is important to plan for rapid transit or on-site laboratory work.

Once a laboratory has been selected to analyze the biospecimens for the biomarker(s) of interest, the laboratory itself should conduct tests to evaluate assay reliability, and the investigator may want to test for both intra- and interlaboratory reliability (see Box 2-1 for a discussion on selecting a laboratory). Thus, the protocol may need to include the preparation of duplicate aliquots (for intralaboratory tests) or even triplicate aliquots (to test interlaboratory agreement of assay results). Ideally, the specimens used for these tests should

be collected from volunteers outside the target sample so as not to overburden the participants. The specimens to be tested should be indistinguishable in form from those of the participants so the laboratory cannot tell which specimens are intended for reliability testing. Duplicates should be sent to the laboratory at regular intervals to check for consistency of results. If possible, a third specimen should be sent to a different laboratory to validate results. For genomic assays, however, such replication may not be possible, and assay replication, in particular interlaboratory assay replication, may not be possible for assays that require unusually fresh specimens.

The kits a laboratory uses to perform assays may change over time because improvements occur in assay technology or because manufacturers discontinue making them. Investigators may want to purchase in advance all the assay kits needed for all the specimens to ensure consistency throughout the project. More generally, care should be taken to ensure that protocols remain constant in the field as well as in the laboratory and do not drift over the duration of the work. Activities that need to be monitored and checked for reliability or drift include the calibration of measurement instruments, the timing of specimen collections, specimen treatment, shipping methods, transit duration, and field processing (e.g., aliquoting and reagents), as well as practices within the laboratories that perform the assays. In addition to monitoring for consistency over time, the checks should ensure that activities of all staff remain in conformity with the approved informed consent and safety procedures.

USE (AND REUSE)

Important issues with respect to use and reuse of biospecimens include the question of who uses them and when, costs associated with sharing the specimens, informed consent for reuse, and shipping protocols.

Who Uses Biospecimens and When?

As noted in Chapter 1, biospecimens in a repository are a nonrenewable resource: the amount of blood, tissue, or other material that has been collected from each participant is limited, whereas the data derived from those materials are limitless and can be used as often as desired. Thus an important question becomes: Who will use the specimens and for what purpose? Those who have ownership of the biospecimens will need to develop procedures for choosing among competing demands.

A related question is when analysis of biospecimens and biodata should be undertaken. Costs of many analyses are steadily falling, so waiting to perform analyses can save money or make it possible to perform more analyses for the same cost. But there are also costs to waiting—time wasted by investigators, for example, and delays in uncovering what might be valuable information.

BOX 2-1
Selecting a Laboratory for Analysis of Biospecimens in Population Studies

Several issues must be addressed in selecting a laboratory to analyze biospecimens for biomarker(s) of interest in population studies.* Once the decision has been made on which specific biomarkers are of interest, a place to start in laboratory selection is to understand whether the markers are to be available for clinical purposes or are purely research measures. Clinical measures, such as conventional blood counts, clinical biochemistries, many antibodies to infectious agents, and many hormone levels, can be obtained in a number of places and are likely to be performed with standardized methods and rigorous quality control, often with procedures dictated by the Clinical Laboratory Improvement Amendments (CLIA). Determinations that are made only in research laboratories raise many additional issues:

- The investigator should check to see whether all of the determinations of interest can be performed in one laboratory.
- Quality control procedures, including those for long-term validity and reproducibility of results, should be established. Examples of prior data and tolerances, including coefficients of variation, should be requested. The survey protocol should allow for a sufficient number or quality of specimens so some can be divided and sent as "blind" duplicates and triplicates to check accuracy.
- It is important to check the usual funding source for the laboratory and the biomarker determinations. If the source is a grant of finite duration, it may not be possible to continue the analysis over a longer period of time. It may be useful to ask whether the laboratory equipment and reagents will continue to be available.
- Many research laboratories are good at what they do, but are not equipped to perform the large number of determinations that may be required by a population study. This constraint should be considered in advance. The investigator should inquire how quickly the necessary procedures can be performed. A related issue is whether one site can perform all aspects of a single assay or determination; if not, additional quality control measures will be needed.
- Pricing of determinations may be difficult for some laboratory personnel, who may lack experience with large quantities of materials over a long time frame. Thus, not only is obtaining the data at a reasonable price an issue, but it is also important that the price be estimated accurately for the longer term; renegotiations may be difficult and costly.
- Several important determinations, such as telomere length, may be performed by more than one method, and selecting the best method can be challenging. This is more of a scientific than a logistical question, but it must be discussed with technical and scientific experts so that whatever is done will be scientifically credible when the studies are published.

- The investigator should inquire what kind of publishing credits will be required. Will an acknowledgment suffice, or is coauthorship mandated? Does the laboratory analyst require sign-off on the relevant manuscript?

In addition, a number of administrative and logistical issues apply generally to biomarker determinations, although they may or may not affect clinical and research uses differently:

- The same principles that apply to general purchases of commodities apply here as well. Reputation, history of service, timely delivery of information, and longevity are all important in laboratory selection.
- Transportation and handling of specimens from the field to the laboratory are critical, and may depend on the type of determinations being made. An example of such considerations is whether specimens need preliminary processing, such centrifugation of blood into constituent parts, which may have to be done quickly after acquisition. Another example is whether specimens need to be frozen or refrigerated during transport. Analytes differ substantially in terms of rate of deterioration. Some may require highly controlled environments and rapid transit, while others may remain intact under usual shipping conditions. All such questions need to be answered in advance. In many cases, the portability of the specimens may not be known, necessitating preliminary test runs to ensure that they are preserved in transit. The same issues apply to long-term storage of the specimens, whether from prior collection or prospectively.
- There may or may not be differences in the ethical obligation to report specific biomarker findings. Almost all biomarkers determined in CLIA-certified laboratories are by definition clinically relevant, and thus a decision needs to be made as to whether the findings should be reported to participants or their physicians. In general, there is an ethical obligation to report findings from clinical laboratories, particularly if clinically abnormal findings surpass alert levels. The issue of reporting findings from research laboratories is more complex. The direct clinical import of the findings may not be understood, nor would most clinicians be able to interpret them. Also, as noted above, the level of standardization and quality control may not be as high as in clinical laboratories, further impeding clinical interpretation. Decisions about reporting on biomarkers determined in research laboratories need to be made on a case-by-case basis, with medical input. (See also the discussion of this issue in Chapter 4.)

---

*This discussion assumes that the markers of interest have already been selected, but such is not always the case. At times, discussion with investigators reveals that newer markers or those with greater specificity for the research question may be available. Identifying new and exploratory biomarkers is an iterative process and should involve reviews of the literature and discussions with relevant colleagues.

Costs Associated with Sharing Specimens

In addition to the costs of collecting and storing data, there are costs associated with sharing specimens once they are in a biorepository. The process of sharing specimens can be complex and time-consuming and involves a variety of steps, such as internal review and reutilization procedures, with nontrivial costs. These costs must be accounted for and should be reflected in the budget. As discussed later in this chapter, the panel agreed that the data sharing plan submitted with a research proposal should address these issues.

Informed Consent for Reuse

Reuse of specimens must be approved on the original consent form or through a separate IRB approval process. The latter process may require obtaining a new consent from the participants; under certain circumstances, however, some IRBs will consider waiving the reconsent process. Each IRB has its own criteria for deciding on future use, but those criteria generally include the possibility that it may not be feasible to obtain updated consent for every new use, so that any new use should lie within the general scope of the original research. Blanket permission for any and all future (unspecified) uses is generally considered unacceptable by today's increasingly conservative IRBs. Chapter 4 presents a more detailed discussion of this issue, including some proposed approaches for addressing it, while Chapter 3 reviews salient federal regulations.

Shipping

All biospecimens should be retrieved and shipped in a way that safeguards their integrity. Concerning retrieval, the NCI best-practice guidelines simply stipulate: "Samples are retrieved from storage according to biospecimen resource SOPs [standard operating procedures] that safeguard sample quality" (National Cancer Institute, 2007, p. 6). For shipping of specimens, however, the guidelines go into much greater detail. They specify, for example, exactly how samples are to be shipped to maintain them at various temperatures, ranging from 8°C to −150°C. They describe the packaging to be used for paraffin blocks and slides, and they suggest that for particularly valuable samples, test packages, such as frozen water samples, be sent out first. They also cover how to maintain the proper paperwork when shipping biospecimens (National Cancer Institute, 2007). It is worth repeating the earlier caution about testing shipment protocols with ice water or "scrap" specimens to check each potential vulnerable point in the protocol.

STORAGE

Most social scientists will lack the expertise and resources to establish or maintain their own facilities for the storage of biospecimens. The panel's best advice is to use experts in this area but to exercise careful oversight. A reasonable starting point might be a cancer or Alzheimer's disease center if the investigator is at a university that has one. Alternatively, many commercial facilities are available. The information in the best-practice documents described earlier can serve as an important guide for assessing the adequacy of potential sites.

To ensure the maximum value from biospecimens, it is necessary to store them so they do not degrade, to keep careful records on them, and to ship them in a way that preserves their quality and their identity. Given the complexity of the storage and distribution of biospecimens, investigators should have the option of delegating storage and distribution to a specialized institution—a biorepository—that is available to accept the biospecimens collected from social science surveys (see Chapter 5). Biospecimens need to be kept stabilized at all stages of the process: collection, storage, shipment, aliquoting, in the field, and in the repository. A biorepository should be chosen that can ensure storage of all biospecimens in an appropriate stabilized state to preserve their integrity and allow the maximum number of analyses to be performed. For example, the Australasian Biospecimen Network describes a number of specific recommended practices in its *Biorepository Protocols*:

> Audits should be conducted to check that biospecimen storage locations concur with database records; storage vessels (tubes, cassettes, etc.) should be checked to ensure they have remained intact; storage conditions should be monitored by a central alarm system and/or local alarms . . . ; back-up systems and enough empty freezer space should be allowed in case quick transfer of specimens from malfunctioning freezers is required . . . ; having multiple storage sites (on or off-site) ensures that all specimens will not be destroyed in the case of freezer malfunction or other emergency situations. (Australasian Biospecimen Network, 2007, p. 60)

Automated security systems should monitor all storage equipment, and backup systems should be in place in case of an emergency such as a power failure. For example, the ISBER best-practice guidelines call for using an uninterruptible power supply (UPS) for critical equipment: "Computer systems and electronic systems, such as environmental monitoring systems, safety systems (e.g., oxygen sensors, ventilations systems, etc.) or controllers for liquid nitrogen freezers, should be protected by a UPS. UPSs used in repositories should be tested on an annual basis to ensure their proper backup capabilities" (International Society for Biological and Environmental Repositories, 2008, p. 14).

For quality assurance and quality control purposes, training of personnel and harmonization of protocols throughout different repositories and the laboratories of individual researchers are essential to ensure reproducibility.

There are a variety of international standards for laboratories and related facilities, such as ISO9001:2000, developed by the International Organization for Standardization (ISO) for quality management systems (International Society for Biological and Environmental Repositories, 2005, 2008).

Perhaps the most important issue in storing biospecimens is whether to keep them in a central repository or use an alternative approach, such as maintaining them in separate repositories associated with the various institutions involved in their collection or keeping different types of specimens in different repositories. A major argument for using a central repository is that it should be more cost-effective than distributing specimens and data among several repositories, but this argument is based on general principles, not experience, as little is known about the cost of one option versus the other. Other issues to consider include which specimens (and which biodata) to keep and for how long. Storage is not free—there is always an incremental cost for additional specimens and data—so at some point one must decide which specimens or data it makes sense (economically) to maintain and which can be discarded without significant loss. In the case of redundant specimens or data, erroneous data that cannot be corrected, or specimens that have insufficient or no identifiers, it may make sense at some point to destroy them. On the other hand, the costs of data storage are declining, whereas the same cannot be said of the costs of biospecimen storage. Therefore, the issue of what to keep and what to dispose of concerns mainly specimens rather than data. No clear guidelines exist on this issue, but it is one that researchers should be aware of and be prepared to address, preferably at the start of a study.

Fortunately, new technologies are changing approaches to the storage of biological specimens and the associated costs. For example, while fresh-frozen tissue used to be essential for gene expression profiling (because of RNA stability issues), new methods such as Illumina's DASL[6] assay have made it possible to use formalin-fixed, paraffin-embedded (FFPE) tissue. This advance has reduced the need for expensive low-temperature storage and opened up the possibility of using large collections of FFPE samples that have been routinely archived in hospital pathology departments.

Finally, the informed consent process should include anticipating and requesting permission for the storage of specimens, as well as their future use (see above), and the specimen archive should include information that both links use of the specimens back to the informed consent documents and links the specimens to the data derived from them.[7] Specimens need to be stored with information that will allow determination of consent approvals. Did the participant agree to use of the specimen only at the time of collection? Was the

[6]DASL = cDNA-mediated Annealing, Selection, Extension, and Ligation Assay.

[7]To protect research participants, IDs for the biorepository and the archive of the derived data should be different, and access to the link between them should be limited.

agreement only for nongenetic tests? Has the participant withdrawn consent (and what does that mean with respect to the disposition of his/her specimens)? (See also Chapter 4.)

CUSTODIAL RESPONSIBILITY AND OWNERSHIP

Storage of biospecimens raises issues concerning their custodianship and ownership. A number of groups and organizations have examined these issues, and a variety of publications suggest policies and best practices for addressing them.

Responsible custodianship of biospecimens implies certain basic steps, including the development of transparent policies to ensure proper use and storage. It is important, for example, that researchers develop guidelines early on regarding who should receive biospecimens and biodata for use in other studies. These guidelines are particularly important for biospecimens because, as noted earlier, they, unlike data, can be exhausted. Thus researchers must ask themselves such questions as: Should remaining biospecimens be distributed on a first-come, first-served basis? Should there be a formal application process? In studies within the United States, should foreign investigators have access? And in general, what factors will be taken into account in deciding who will be allowed to share the data? These are complex questions. If the proposed analyses lie beyond the competence of the original investigators, for example, projects will need to ensure that they have adequate support for the purpose.

The panel agreed that the data sharing plan of a research proposal should specify policies and implementation plans for archiving and sharing both specimens and the data derived therefrom. In general, there is no one best plan for the use and reuse of specimens, but the plan should include a discussion of the adequacy of the storage and retrieval protocols. It should spell out criteria for allowing other researchers to use (and therefore deplete) the available stock of specimens, as well as to gain access to any derived data. The plan should also specify the procedures for accessing the specimens and data. It should include provision for the storage and retrieval of specimens and clarify how the succession of responsibility for and control of the specimens will be managed at the conclusion of the project. Finally, the plan should contain information on how specimens and data derived from them are to be documented and provide for public access to that documentation.

The NCI best-practices document describes a number of custodianship practices that should be followed to ensure that biospecimens are maintained and used as effectively as possible (National Cancer Institute, 2007):

- The biorepository should ensure the proper storage of the specimens to maintain their physical integrity and the integrity of study participants' data linked to the specimens. It is the responsibility of the

principal investigator to choose a biorepository that can meet these requirements.
- Clear and transparent protocols for the distribution of the specimens and the data to other investigators should be in place.
- Plans for handling, storing, and disposing of the biospecimens and associated data should be in place for such contingencies as the end of a grant, the end of a particular study, biospecimen depletion, and a study participant's request for discontinuation of participation.

Policies on Custodianship

Although the NCI best-practices document provides extensive guidance on many practices for biospecimen collection, storage, use, and maintenance, it does not specify the custodial roles and rights of biorepositories and their responsibilities to their host institutions or study participants, nor does it provide a functional definition of custodianship. To clarify these responsibilities and to help develop a transparent policy for biorepository governance, the NCI Office of Biorepositories and Biospecimen Research (OBBR) held a workshop in October 2007 (Custodianship and Ownership Issues in Biospecimen Research Symposium-Workshop). The workshop summary is an informative resource that makes a useful distinction between the concepts of "custodianship" and "ownership" (Office of Biorepositories and Biospecimen Research, 2008). It also provides recommendations for dealing with issues of financial and other types of conflict of interest (COI), intellectual property (IP), and access to products and benefits. Another complexity is the relationship between the principal investigator and the biorepository. Most social scientists will use a biorepository to store and distribute specimens, but they will not wish to delegate scientific decisions to the biorepository. Thus, it is important to specify at the outset the roles and responsibilities of the principal investigator and the repository. The OBBR workshop summary offers the following guidance on these issues:

- One OBBR recommendation is that "the custodian of biospecimens should be someone other than the investigator . . . investigators with small biospecimen collections should be encouraged to establish or join an existing IRB-approved regulated biospecimen resource" (p. 8).
- Decisions regarding the distribution and reuse of biospecimens have the potential to create a conflict of interest. An expert panel might be appointed to provide technical/scientific advice and to make recommendations to the investigators regarding the disposition and use of the specimens in accordance with all ethical and IRB guidelines.

- Conflicts of interest should be clearly identified. A starting point would be the Code of Federal Regulations (CFR) Part 50, Subpart F,[8] along with other NIH and university guidelines. All existing and possible institutional, NIH, and other COIs regarding biospecimens should be reviewed to determine whether they have been sufficiently addressed. All individuals responsible for biospecimen distribution should report COIs, and financial COIs should be disclosed to the public whenever possible.
- Biorepository personnel and staff, as custodians of biospecimens, are not considered inventors under patent law for any inventions resulting from the research on specimens housed in the repository. Biorepositories have no inherent rights to future IP associated with inventions developed by researchers using the biorepository's collection (see also National Institutes of Health, 2007).
- Educational materials on IP issues related to biospecimen research should be developed and made available to study participants. Possible financial and nonfinancial benefits and commercial products resulting from such research should be discussed in the informed consent document, including possible supplemental material.
- The existence of biospecimens should be made public when research data resulting from the use of those specimens are published, even if the biospecimens themselves are not available to the research community.

Court Decisions Regarding Ownership

There are few legal precedents related to the ownership of biospecimens, and they tend to be fact- and jurisdiction-specific (Office of Biorepositories and Biospecimen Research, 2008). In general, the courts have denied claims to ownership by research participants. In one recent prominent case, *Washington University v. Catalona*, the Federal District Court for the Eastern District of Missouri ruled that the research participants (human subjects) retained no rights to control their biospecimens after donation.[9] The court stated that in that particular case, the donation of biospecimens was an *inter vivos* gift. The informed consent document and the accompanying brochure were the key documents in this case, and the court used them as evidence that research subjects retain no rights to repossess or transfer their biospecimens after donation or to direct any future use of the specimens for research.

On appeal, the U.S. Court of Appeals for the Eighth District stated that

[8]Code of Federal Regulations Index, see http://www.gpoaccess.gov/CFR/INDEX.HTML (accessed January 26, 2010).

[9]*Washington University v. Catalona,* Case No. 4:03CV01065-SNL, Document 152 (March 31, 2006).

research participants do retain the right to discontinue participation in research by doing one of the following: not answering additional questions, ceasing donating additional tissue, or disallowing the use of their tissue in future research. The ability not to allow the use of donated specimens in future research is a key right of participants and is closely related to each participant's right to withdraw from a study at any time.

The decisions of the district court and the appeals court reflect a public policy that medical research can advance only with continuing access to biospecimens. Nevertheless, it is possible that in another set of circumstances, the informed consent document could be interpreted as giving study participants the right to withdraw and physically repossess their biospecimens.[10]

ACCESS AND DISTRIBUTION

Experience has led many IRBs to insist that biospecimens be destroyed when a clearly stated, focused use of the specimens has been completed, thus avoiding even the possibility of new and unapproved uses of the specimens. However, such stipulations impede the use of specimens to reproduce, challenge, or extend scientific findings. The panel finds that the latter goal outweighs the former. Consequently, the panel recommends the creation of at least one central facility for the storage and distribution of biospecimens, provided adequate procedural safeguards are in place (see Chapter 5 for a full discussion of the panel's recommendations).

Expanding access to biospecimens promotes further research and makes it possible to explore new questions without having to collect new specimens. It creates economies of scale, allowing different researchers to pursue a variety of investigations with specimens originally collected for a single purpose. Sharing biospecimens "fosters an open research community and reinforces transparent scientific inquiry" (National Research Council, 2005, p. 39). Thus ensuring that researchers have broad access to biospecimens is essential to maintaining and improving the quality of the research enterprise. Access to biospecimens by a wide variety of researchers also enables a feedback system that can reveal problems with the specimens or their collection and suggest methods for improvement (National Research Council, 2005). Eventually, moreover, people die, at which point they are no longer "human subjects," and the associated restrictions on the use of their biospecimens no longer apply (see the discussion of this issue in Chapter 3).

For all of the above reasons, it is important to provide the widest possible access to biospecimens. Decisions about who obtains access can be based on various considerations, including "the scientific merit and potential impact of the proposed research, whether the research use is appropriate to the nature

[10]See also the discussion in Chapter 4 on withdrawing consent.

and purpose of the repository, adequacy of the research design and funding, public health benefits and risks of the proposed research, legal and ethical considerations, and the qualifications of the research team and research environment" (International Society for Biological and Environmental Repositories, 2008, p. 49). But these considerations should not be used as a way of artificially limiting access.

To best serve the various research communities that can benefit from their biospecimens, principal investigators should develop clear and ethical access policies to facilitate sharing in ways that minimize disclosure risks and comply with all federal and state regulations. Policies should be flexible enough to respond to new technologies when they appear and general enough that they can be adapted to different kinds of biorepositories. The data sharing plan should include detailed information regarding these policies and their proposed implementation. To avoid any appearance of self-interest, a project might empower an external advisory board to make decisions about access to its data.

When one institution transfers materials to another for research purposes—for example, a biorepository making biospecimens available to an investigator at an institution not affiliated with the repository—a material transfer agreement (MTA) sets forth the conditions under which the transfer is to be made and delineates the rights of the two parties. If a discovery made using the material leads to a commercial application, for example, the MTA describes how any receipts or profits from the application will be allocated.

In the case of MTAs governing biospecimens from repositories and the data derived therefrom, investigators need to ensure that the agreement is specific about what can and cannot be analyzed. If a clear understanding is not reached, and an investigator analyzes and publishes data without the full informed consent of those who have the rights to those data, serious problems can arise. One example is the experience of researchers at Arizona State University who analyzed genetic data from the Havasupai tribe in an effort to understand its high prevalence of diabetes. Through a series of misunderstandings, the tribe came to mistrust the researchers performing the analysis, and demanded that all analysis stop and its biological samples be returned. Because the agreement had not been specific enough, the Arizona State researchers ended up with no data and no analysis after several years of work (Dalton, 2004).

3

Protecting Privacy and Confidentiality: Sharing Digital Representations of Biological and Social Data

One of the advantages of collecting biological specimens as part of social surveys is that digital representations of the data derived from the specimens—such as measurements of lipid levels or indicators of the presence or absence of genes or diseases—can be appended to the survey data and shared with other researchers. Wide dissemination of data facilitates advances in research and public policy. Indeed, the benefits of wide access to data have led the National Institutes of Health (NIH) to require data sharing as a criterion for funded proposals. However, biological and social data cannot be widely shared without consideration for the rights and interests of study participants from whom the data were derived, including their interest in confidentiality. As has been noted in a number of previous reports by the National Research Council (1993, 2000, 2005, 2007), there is an inherent conflict between confidentiality and data access: the obligation to protect confidentiality pushes data disseminators to restrict access, whereas researchers' demands push them to share highly detailed data. Balancing these conflicting demands can be complicated, especially for large surveys that combine biological and social data.

This chapter examines methods of sharing digital representations of biological and social data. It begins with a discussion of the risks inherent in sharing such data. The second section reviews existing data sharing approaches and considers their potential usefulness in the context of this report. The final section summarizes some of the federal regulations related to the confidentiality of combined biological and social data.

RISKS TO CONFIDENTIALITY IN LINKED BIOLOGICAL AND SOCIAL DATA

It is well understood that, before sharing data, organizations must remove all direct identifiers, such as names and addresses, from the files. *Deidentification* refers to a reversible process whereby the data are key-coded, encrypted, or pseudonymized to remove personal information, but a key is generated that allows the data to be reassociated with the personal information.[1] By contrast, *anonymization* is an irreversible process in which the data are completely stripped of all identifying information that can be linked to the study participants (Elger and Caplan, 2006; Shostak, 2006).

Deidentification and anonymization of biodata should decrease the risk that unauthorized individuals can identify the participants who were the source of the data. However, deidentification and even anonymization may not be sufficient to protect participants' confidentiality when the released data contain other variables that, in combination, might enable a malicious data user (hereafter called an intruder) to identify individuals in the file. For example, given precise values of such demographic variables as age, race, sex, education, and occupation, an intruder might be able to link records in a released file to records in other databases that include data subjects' names. Or, given a unique medical profile, such as a diagnosis of a rare disease, an intruder might be able to use public knowledge, health records, or research data sets to link to identifiers or other data. This determination of an individual's identity from data that have been deidentified or anonymized is referred to as *reidentification*, and it is a risk that is growing as more and more data become readily available electronically. For example, electronic health records may enable an intruder to access individuals' medical data and possibly link those data to social surveys with overlapping medical variables.

Broadly speaking, confidentiality risks are of two types: identification disclosure risk and attribute disclosure risk. Identification disclosure occurs when an intruder learns that information on a targeted individual is in a particular shared file; if this happens, it may be possible for the intruder to determine which of the records in the file belongs to the targeted individual by examining the demographic or other variables. Attribute disclosure occurs when an intruder learns the value of a sensitive variable for a targeted individual, which may make it possible to identify records belonging to that individual. (There are other types of disclosure, such as perceived identification disclosure and inferential disclosure, but they are not discussed here.)

[1]Note that some biological measures vary sufficiently across time that they do not raise a risk of reidentification that is any greater than that for most social, economic, or psychological characteristics. On the other hand, some of the latter characteristics are unchanging and thus pose a greater risk. In this chapter, the focus is on characteristics that are unchanging or sufficiently stable that they raise a risk of reidentification.

A great deal of research has been done to assess the risks of disclosure for biological and social data (e.g., National Research Council, 2005, 2007). Conceptually, the combination of biological and social data presents no new issues related to disclosure risk. What is important in the context of this report is that disclosure risk increases as more variables are added, and the risk of identification disclosure is greater with combined data than with any single type of data. Suppose, for example, that an individual is nearly identifiable from a combination of demographic variables that are publicly available; this would be the case, for instance, if only a small number of people matched those particular characteristics. Then adding a set of biomarkers that was available to an intruder could enable the intruder to identify that person. Similarly, an individual might not be identifiable from a small number of genes provided in a released file, but adding demographic data could provide enough information to result in identification.

Additionally, the degree of potential harm from attribute disclosure is greater in the case of combined data. For example, an intruder's learning that someone has a particular set of genes might not be especially damaging, but the potential damage would be much greater if the intruder also learned phenotypic information from social data, such as criminal histories or sexual habits. Likewise, learning someone's identity from social science data might be innocuous, but if the person's disease status were also uncovered, harm could result—for example, from discrimination.

Assessing the risks of sharing combined data is complex. First, the data steward must determine to the extent possible which variables would be available to intruders. The answer depends on the nature of the variables and may change over time. Furthermore, some biological and social variables are effectively permanent characteristics, so that sharing them poses risks for both the present and the future. A key example is genetic data. While there is currently no way to search for a person's identity using his or her DNA, ongoing work of researchers, government, and companies such as 23AndMe may well make individuals' genomes, in whole or in part, available to intruders in the near future. Hence, because of the permanence of genetic data, any sharing of such data now could lead to disclosure risks in the future. Indeed, research by Malin and Sweeney (2004) indicates that genetic data already pose risks: the authors show that unlabeled DNA sequences stripped of demographic data and identifiers, when interpreted for common disease genes and screened against publicly available data, could be narrowed down to several individuals.

Anyone who disseminates such data also must take into account the variation that may exist across databases in the variables available to intruders. High variation provides an additional layer of confidentiality protection since the same record may have different values in the shared file and in the intruder's external database, making correct matching difficult. While such variations can easily be found in social science data because of inconsistencies in self-

reporting, some clinically derived biological data are measured with very little error, so this additional layer of protection does not exist. Genetic data are again a good example, as the error rates for genotyping are known to be very small (Hao et al., 2004; Saunders, Brohede, and Hannan, 2007).

A third consideration is future technologies that might be used by intruders. Data that appear safe today may not be safe in the future. As an example, a recent report by researchers at the Translational Genomics Research Institute in Phoenix, Arizona, and the University of California, Los Angeles, shows that there is a much greater risk of identifying participants in a genetic survey than was previously thought (Homer et al., 2008). The report relates to genome-wide association studies (GWASs), which use the complete genomes from a number of individuals with a particular disease to scan for markers that may be associated with that disease. The surprising finding was that if one knows the DNA of a particular person, it is possible to tell whether that person is included in the GWAS database simply by looking at the allele frequencies. This finding led NIH to put some of its data behind a firewall, a step that was taken by the Wellcome Trust as well (Clayton, 2008). Once an intruder knows that someone is in a particular data set, that knowledge may provide information about the person's health or disease status, since GWASs typically focus on a particular phenotype—dementia, for example, or breast cancer.

Finally, while much of the research on identifiability has involved genomic data, similar risks can be expected for other highly detailed and high-dimensional data, such as proteomic and metabolomic data.[2] Given ongoing rapid technology advances and the likelihood that electronic health records will become commonplace in the next few years, it is possible that these data will be available to intruders in the near future.

APPROACHES TO SHARING BIOLOGICAL AND SOCIAL DATA

The literature on data sharing describes two broad approaches to protecting the confidentiality of individuals whose records appear in data collections: restricting access and restricting data. The restricted access approach involves controlling who can access the data for analysis and under what conditions. Specific strategies include licensing agreements and data enclaves. The restricted data approach involves providing a redacted version of the data to those who wish to use it. Redaction strategies are often termed statistical disclosure limi-

[2]Proteomics refers to the branch of molecular biology that deals with the full set of proteins encoded by a genome; proteomic data are data associated with proteins expressed by a genome, cell, tissue, or organism. Metabolomics refers to the study of the chemical fingerprints that specific cellular processes leave behind; metabolomic data are the small-molecular metabolite profiles present in cells or tissue.

tation (SDL); they include such approaches as recoding variables, suppressing data, and perturbing data.

Restricted Access

There are four primary strategies for restricted access: (1) licensing, (2) remote execution systems, (3) data enclaves, and (4) virtual data enclaves.

Licensing

To obtain data with little or no redaction other than removal of direct identifiers such as names and addresses or geocodes, researchers sign a licensing agreement not to use the data for malicious purposes, such as identifying individuals and subsequently taking injurious actions based on those identifications. A number of statistical agencies, including the National Center for Education Statistics, use this approach, as do public data archives such as the Inter-university Consortium for Political and Social Research (ICPSR) and the principal investigators of many social science and biosocial surveys. Licensing allows researchers to obtain highly detailed data and thus facilitates secondary analyses of the data. However, this approach relies on researchers not violating the terms of the license. Enforcement is generally the responsibility of the funding agency. For certain types of violations, substantial penalties can be levied, such as revoking funding from the responsible investigator and institution. One disadvantage of licensing is the difficulty that often characterizes the process; Box 3-1 describes some ways to make the process less burdensome for both researchers and data stewards.

Remote Execution Systems

In this approach, confidential data are maintained in a computer system owned by the data disseminator, and a secondary researcher who wishes to perform a study with the data submits a query to the system. As long as the information is not confidential, the system provides the results of the query (which may be computed by the system or by staff at the data-disseminating organization) to the researcher without revealing the individual data. The National Center for Health Statistics and the U.S. Bureau of the Census maintain such systems. These systems are not foolproof, however, as intruders can use judicious queries to glean sensitive information about data subjects. To minimize this risk, the system must limit the types of analyses that can be performed, which reduces the utility of the data to outside researchers. Furthermore, without full access to individual-level data, researchers find it difficult to perform exploratory data analyses or to check the fit of models.

BOX 3-1
Ways to Facilitate the Licensing Process

A number of authors have noted that the licensing process can be difficult, time-consuming, and even costly and thus may limit the number of researchers who use the specimens and data in a storage facility (see, for example, Nolte and Keller, 2004; National Research Council, 2005). The most successful researchers, who have many options from which to choose, may eschew projects for which the licensing process is too demanding; indeed, concern has been expressed that some investigators may not use data unless they have easy access. At the same time, managing licenses is difficult for data stewards. Those who produce data for research often lack the capacity to manage numerous contracts. Even professional data archives can be overwhelmed by the requirements of dealing with file cabinets full of paper contracts that require constant monitoring.

Thus it is important to find ways to make licensing easier. One recommendation to that end is offered in the National Research Council report *Expanding Access to Research Data: Reconciling Risks and Opportunities*: "Statistical and other agencies that provide data for research should work with data users to develop flexible, consistent standards for licensing agreements and implementation procedures for access to confidential data" (National Research Council, 2005, p. 79). A variety of options could be explored along these lines. For example, it might be possible to have a central licensing agency that would license individual researchers and laboratories in such a way that they would not have to seek a new license each time they sought access to a new data set. Or there could be a tiered licensing system with differing licensing requirements depending on such things as the sensitivity of the data or the amount of access requested, so that investigators whose projects presented a lesser risk to the confidentiality of the data might be subject to a less onerous licensing process. In one effort to contribute to solving problems with licensing, the NIH-sponsored Data Sharing for Demographic Research Project at the Inter-university Consortium for Political and Social Research (ICPSR) published guidance for developing and implementing a restricted-use data contract or license (see http://www.icpsr.umich.edu/DSDR/rduc/ [accessed May 27, 2010]). In the next step of this project, plans are to develop a largely automated system for managing licenses and for validating the security of computers with which licensed data will be analyzed.

Data Enclaves

With this approach, an investigator works in a room dedicated to accessing the data. Only approved researchers are allowed in the enclave. Computers in the room are not connected to the Internet or to other external resources. Researchers cannot take individual-level data from the enclave, and all results they obtain are checked by the data disseminator for potential confidentiality breaches before they can be taken out of the enclave. Because of these features, this approach offers the highest level of security. It also carries a variety of costs,

however. Data enclaves are inconvenient for investigators who must travel to them. Moreover, most enclaves require that data users pay a substantial fee to cover the costs of maintaining the facility and its staff; they also require potentially time-consuming background checks and proposal approvals. Hence, the amount of analysis likely to be carried out on the data in an enclave is limited (National Institutes of Health, 2006).

Virtual Data Enclaves

Virtual data enclaves combine features of licensing, remote execution systems, and data enclaves. The data are housed in a system owned by the data disseminator. Licensed secondary researchers access the data remotely; that is, the researcher's computer serves as a dummy terminal. The National Opinion Research Center (NORC) at the University of Chicago maintains such a system, called the NORC Data Enclave. Users of a virtual data enclave cannot store the data on their computers, and certain functions on their computers, such as printing and the ability to copy data to removable media (including disks and micro-vault storage media), are disabled. Virtual data enclaves thus allow researchers to access the data without traveling to a secure data enclave. They also avoid some of the disclosure risks posed by licensing researchers to store data on their own machines, such as the accidental loss of CD-ROMs or sharing of data with unapproved investigators or students. As with licensing, however, confidentiality protection depends on researchers not violating terms of the data use agreement; thus, virtual data enclaves are less secure than physical ones.

Restricted Data

Many data sets containing biological or social data are shared after identifying or sensitive values have been altered. Alterations can be made in a variety of ways, including coarsening the variables by, for example, releasing ages in 5-year categories rather than as exact values; top-coding, as in reporting incomes that exceed a threshold T simply as "above T"; swapping, or exchanging small amounts of data between records, with the goal of introducing uncertainty about identities; adding noise to numerical variables; and replacing sensitive values with synthetic data derived from a probability model (see the detailed discussion in National Research Council, 2007). One could characterize these methods as protecting confidentiality by obscuring relatively high-order features of the data. Top-coding, for example, destroys analyses of the tails of the distributions; swapping attenuates the correlations among swapped and non-swapped attributes; adding random noise distorts distributions and attenuates relationships; and synthetic data are guaranteed to preserve only the relationships in the synthesis models.

Generally speaking, greater modifications of the data lead to greater protection but also to greater reduction in the utility of the data. Thus, determining the amount and type of data alteration involves trade-offs. Weighing the type and extent of protection against the reduction in the utility of the data requires developing a way of measuring both the risk of disclosure for data altered in various ways and the utility of the data. Although some metrics for data utility exist (Karr et al., 2006), the assessment is a nontrivial task. In general, it is difficult for researchers not trained in methods of confidentiality protection to apply these methods in ways that optimize the utility of data while adequately protecting confidentiality.

It is difficult to imagine that standard protection methods will be effective for multidimensional biological and social data. Consider, for example, using these methods to protect 660,000 single nucleotide polymorphisms (SNPs) that are planned for release. It is not clear how many and which SNPs are needed to identify individuals, particularly if the data also contain psychological, social, or economic variables. If the data disseminator decides to protect a large proportion of the SNP data by, for example, swapping or synthesis, it is inevitable that many interaction effects among the SNPs will be nearly destroyed. This outcome is problematic since analyses of gene–gene or gene–environment interactions are inherently high-dimensional. One way to get around the limitations of data restriction is to combine this type of approach with restricted access. An example of such a combined approach is provided in Box 3-2.

Choosing a Data Sharing Strategy

The restricted access and restricted data strategies described above provide varying degrees of confidentiality protection, and each has its limitations with respect to preserving the utility of the data for research purposes. In general, there is a trade-off between the level of protection that is afforded and the level of utility that is preserved.

Of the restricted access strategies, remote execution systems and data enclaves offer the highest level of protection, but they also impose the greatest limitations on the utility of the data: the former because the types of analyses that can be performed are significantly limited, and the latter because researchers bear a heavy burden in having to travel to the enclave, pay substantial fees, and undergo background checks and proposal approvals. Licensing and virtual data enclaves impose fewer limitations, but the protection they provide depends on researchers abiding by the terms of the license or data use agreement. Nonetheless, the panel believes that, given rigorous enforcement (see the discussion later in this chapter), these two strategies hold the greatest promise for sharing combined biological and social data in ways that support and sustain promises of confidentiality while preserving the utility of the data. As discussed in Chapter 5, the panel also believes that an effort must be made to improve the

BOX 3-2
Data Sharing for the Health and Retirement Study

The Health and Retirement Study (HRS) at the University of Michigan uses a combination of approaches to share data while protecting confidentiality (Nolte and Keller, 2004). Investigators who wish to use sensitive or potentially identifiable data from the HRS can apply for a license for full access to the data if they agree to certain conditions, such as developing and implementing a data protection plan to protect participant confidentiality; allowing yearly inspections; providing annual reports; and in some cases, submitting to a prepublication review of the analysis results. The major disadvantages of this approach are the length of time required to obtain approval for a nonstandard data protection plan and the fact that only principal investigators of a federally funded project who are also affiliated with an institution with an NIH-certified human subjects review process can apply for a license. Thus, junior faculty members and students, for example, have difficulty accessing the data through such agreements (Nolte and Keller, 2004).

For this reason, the HRS also makes its data available through the Michigan Center on the Demography of Aging (MiCDA) Data Enclave in Ann Arbor. Investigators who do not meet the requirements for a data license can come to the data enclave to perform data analyses. There are few restrictions other than a review of the analyses by the data enclave staff to ensure that they include no information that could compromise the confidentiality of the participants whose data are in the database. The main disadvantage is that an investigator must go to the data enclave to perform an analysis (Nolte and Keller, 2004). To avoid this disadvantage, the HRS began a virtual data enclave program, through which researchers can access data on the HRS server remotely from their own institutions. To maintain confidentiality, all the data are kept in the HRS computers. Remote users send instructions to perform analyses that are carried out on the HRS computers and then receive the results. Various security systems ensure that no confidential data are released, and the data enclave staff still review the results of the analyses to ensure that no breaches of confidentiality occur. The main disadvantage of this approach is the cost of setting up secure computer systems at the investigators' home institutions (Nolte and Keller, 2004).

The HRS approaches to data access illustrate two access models available for dealing with biospecimens and associated data.

licensing process so as to make it less complex and time-consuming, employing options such as those outlined in Box 3-1.

Data restriction strategies are both imperfect in protecting confidentiality and likely to significantly reduce the utility of the data. Combining these strategies with some form of restricted access, as is done for the Health and Retirement Study (Box 3-2), can address these limitations by enhancing security and necessitating less extreme alterations of the data.

Regardless of which strategy for data sharing is chosen, it is essential to

BOX 3-3
The Importance of Planning for Data Sharing: Two Case Examples

Because placing biosocial survey data into data archives poses so many risks to participants' confidentiality, it is important to take the obstacles to sharing data into account when planning a survey, preparing consent forms, and working with the relevant Institutional Review Boards (IRBs). If the data sharing strategy is not properly planned in advance, it can be difficult to arrange later. A brief description of the types of issues encountered by two projects based in Italy is sufficient to illustrate some of those difficulties.

The SardiNIA Study of Aging is an international collaboration funded primarily by the National Institute on Aging (NIA) to study the effects of various genetic variants on aging in a group of people in Sardinia, the second-largest island in the Mediterranean Sea. The research has helped identify genetic variants linked to lipid levels and risk for coronary artery disease. But difficulties arose when the principal investigators were asked to provide their data to the Database of Genotypes and Phenotypes (dbGaP), a database operated by the National Center for Biotechnology Information to archive and distribute data from genome-wide association studies (GWASs).

Because the SardiNIA study began before dbGaP had been fully established, no provisions were made during the planning for the study to place the data in dbGaP. The SardiNIA investigators were able to provide dbGaP a listing of the traits and diseases that had been studied, a data dictionary, and the genotyping platform that had been used, along with p-values for every single nucleotide polymorphism (SNP) for every trait that had been studied. The investigators were able to supply that information because it did not include any personal data that could lead to breaches of confidentiality. On the other hand, the investigators were unable to provide dbGaP with information on family relationships and raw data connecting each participant to corresponding trait values and genotypes.

Placing the SardiNIA data into dbGaP would require approval from the IRB overseeing the study. In their original informed consent forms, the investigators had

make the decision during the process of planning for the study and to ensure that consent forms contain the necessary detail to enable the strategy's implementation. Box 3-3 provides two case examples of the importance of planning in advance for data sharing.

SALIENT FEDERAL REGULATIONS

A variety of federal regulations deal with issues related to data sharing and privacy. This section provides an overview of those that are most important to researchers dealing with biodata in social science surveys.

not discussed the possibility of depositing data into dbGaP or any other archive. Instead, the consent forms had specified only that investigators and staff with the National Research Council of Italy would have access to the data, and that NIA would have access only to data that were anonymously coded. Depositing all of the SardiNIA data in dbGaP would require participants' reconsenting, and only data from surviving participants who reconsented would be able to be deposited. Obtaining reconsents would require a great deal of additional expense, effort, and time.

One option the investigators suggested was providing a limited amount of data to dbGaP—only general information about the study and statistics regarding the prevalence of various genetic variants in the population—and then accepting requests for more complete data from individual researchers. The raw data would remain with the SardiNIA group rather than at dbGaP, and the SardiNIA IRB would decide on a case-by-case basis who would receive access to the more complete data.

Similar issues arose with the proposed archiving of data from the InCHIANTI (Invecchiare in Chianti, or "aging in the Chianti area") Study, which examined aging among the populations of two small towns in the Tuscany region of Italy. In this case, the principal investigator of the study spoke with the study's steering committee, explaining the position of the National Institutes of Health that data from research paid for with public funds should be shared. The steering committee agreed that at least part of the data should be archived, but its members were concerned that doing so might not be possible given the language in the consent forms, which said the data could be shared with the collaborators of the InCHIANTI group but did not mention making the data available in a public database. To archive the InCHIANTI data would require obtaining new consent forms from the participants that specifically allowed for placing the data in a public archive. In the case of participants who had since died, it would be possible to place some of their data in an archive because of the wording of the original consent form, but none of their genetic data could be included.

One clear lesson to be drawn from these two examples is the importance of including a discussion of data sharing in consent forms. Without such explicit discussion, it may be impossible to make data publicly available.

Federal Privacy Regulations

Among the federal regulations pertaining to privacy are the Common Rule (45 Code of Federal Regulations [CFR] 46 Part 160 and Part 164, Subparts A and E) (U.S. Department of Health and Human Services, 2000, 2005b) and the Standards for Privacy of Individually Identifiable Health Information, issued by the Department of Health and Human Services and commonly referred as the Health Information Portability and Accountability Act (HIPAA) Privacy Rule. These regulations are intended to protect the human subject's personal health information and identity while allowing society to benefit from the use of that information, including for research purposes (U.S. Department of Health and Human Services, 2003).

There is some complexity to the application of these federal regulations. The HIPAA Privacy Rule, for example, applies only to health data obtained from covered entities, for example, physicians, hospitals, and Medicare or Medicaid records, but not to reports or biodata obtained from individuals in the course of a social science survey. Moreover, both rules may be interpreted differently depending on the kind of research in question—for example, direct human subjects research versus research using leftover biospecimens in a biorepository or biodata stored in a biobank. Generally speaking, biorepositories and biobanks are not considered covered entities under the HIPAA Privacy Rule unless their contents were obtained in research requested and approved by a covered health provider.

The Common Rule and the HIPAA Privacy Rule need to be clarified to alleviate the scientific community's fears that they could be misinterpreted in a way that could be damaging to research on biospecimens. In particular, it is important to (1) develop a standard for what constitutes "minimal risk," especially in the context of the use of existing biospecimens and biodata, and (2) define "human subject" more clearly in light of the fact that research can be carried out on specimens and data that were collected from people who later died (Meslin and Quaid, 2004, p. 230). Under the Common Rule, someone who has died is not considered a "human subject," but in the case of biospecimens, it can be argued that the surviving interests of the deceased person, such as his or her reputation, should be protected. Another point of confusion is whether, if the research carried out on deceased persons can generate information about living people, those individuals should also be considered "subjects" (DeRenzo, Biesecker, and Meltzer, 1997; Meslin and Quaid, 2004, p. 230).

In some instances, the Common Rule and HIPAA Privacy Rule actually contradict each other, such as in informed consent for future, unspecified research: the Common Rule allows authorization of such unspecified uses, but the HIPAA Privacy Rule requires a specific research purpose for each authorization of release of protected health information (Vaught et al., 2007). Furthermore, the HIPAA Privacy Rule distinguishes between the creation of a biospecimen repository or a database containing protected health data and the release of data from such resources for research purposes. The rule requires a participant's authorization for each instance of data release unless a waiver is granted by the Institutional Review Board (IRB). This inconsistency between existing federal guidelines, along with differing interpretations of the guidelines, has prompted some experts to lament the burden on researchers that results, as well as to question the rules' effectiveness in protecting the privacy of research participants (Bankhead, 2004; Nosowsky and Giordano, 2006).

Since the privacy, confidentiality, and identifiability of human research subjects are important concerns, many institutions devote a good deal of time and resources to the discussion of best practices in these areas. The National Human Genome Research Institute (NHGRI) held a workshop in October

2006 to address these issues and how they pertain to genomic research. The discussion document developed for this workshop and the workshop report provide useful points to consider, especially with respect to the definition of "identifiable" data, strategies for data deidentification, and the risks that deidentified data may be reidentified (Lowrance, 2006a, 2006b). According to the workshop report, some of the "themes that were accepted as granted" are that (1) at this stage of advanced genomic science, the scientific community must do everything possible to respect and protect the privacy of human subjects whose genomic information is being used for research, and (2) everyone in the chain of data collection shares the responsibility for privacy and identity protection.

On this subject, it is also worth noting two other points concerning privacy. First, the Common Rule does not apply to research done by private companies, organizations, or individuals that do not receive support from the U.S. government, except in the case of institutions (e.g., universities) that have accepted blanket responsibility for enforcement of the rule, regardless of the source of research support. Second, neither the Common Rule nor the HIPAA Privacy Rule will protect an individual's privacy and identity in the case of a court-ordered subpoena requesting personal information, but that level of protection can be provided by a federal Certificate of Confidentiality.

Genetic Discrimination

On March 21, 2008, after 13 years of debate, Congress passed and President Bush signed into law the Genetic Information Nondiscrimination Act. This act protects Americans from being discriminated against in either employment or health insurance based on their genetic information (U.S. House of Representatives, 2008). Although this legislation provides a certain amount of protection from genetic discrimination, it is limited to asymptomatic individuals and thus does not protect individuals already presenting readily detectable symptoms of a genetic condition (Rothstein, 2008). At this point, there is relatively little in the judicial record to indicate how courts will deal with employers, insurers, or others discriminating against an individual on the basis of information about that individual's genetic makeup. No cases of genetic discrimination have yet been brought before federal or state courts, for example.

One case that offers some insight into how the courts may deal with the issue is a 2001 suit filed by the Equal Employment Opportunity Commission (EEOC) against Burlington Northern Santa Fe (BNSF) Railroad for testing its employees without their consent for a rare genetic condition that causes carpal tunnel syndrome. The doctors hired by BNSF also screened the employees for diabetes and alcoholism without their knowledge, and at least one BNSF employee was threatened with termination for refusing testing (National Human Genome Research Institute, 2008). The EEOC argued on behalf of BNSF employees that these tests were unlawful based on the Americans with

Disabilities Act (Public Law 101-336) as they were not job related, and any change in employment due to the results would constitute discrimination based on a disability. This lawsuit was quickly settled, and BNSF agreed with all the EEOC's requests.

Legal Sanctions and Enforcement

At present, NIH relies on its ability to restrict grant funding to ensure that best practices are upheld, data are shared, confidentiality is respected, and so forth. If, for example, an investigator signs a licensing agreement to obtain data from a repository and then fails to abide by the terms of the agreement, NIH can cut off all or part of that researcher's funding. The general perception is that enforcement of these agreements is weak, but there has been no evidence of breaches of confidentiality (National Institutes of Health, 2006).

As more and more biospecimens and biodata are collected by and made available through biorepositories and biobanks, it will become increasingly important to ensure effective enforcement of the rules governing the use of these materials and data, coupled with strong legal sanctions when the rules are broken. To this end, a number of questions need to be answered: What sort of enforcement scheme should be used? Should digital representations of biological data be turned over to an archive such as the Inter-university Consortium for Political and Social Research? Given that most researchers do not want to do their own policing, who should police the use of biospecimens and biodata—the Office of Research Integrity or some other agency?

4

Informed Consent

Although the informed consent process necessarily takes place before the collection, storage, and analysis of biological specimens and biodata, it requires giving potential participants an honest explanation of exactly what the research entails and what risks they face by participating. The consent process, therefore, hinges on the information and observations offered in earlier chapters of this report.

While clinicians and biomedical researchers have contended with issues of informed consent related to biological data for many decades, consensus on how to address these issues remains elusive. Active debate continues among clinicians, researchers, and ethicists about exactly what forms of consent are necessary in various situations. The debate has been particularly heated with respect to large-scale repositories of human biological specimens that are linked to clinical or socioeconomic data and are being used to assess interactions between genes and the environment similar to those discussed in this report. Thus, social scientists considering adding the collection of biospecimens to their surveys will find that doing so makes informed consent a much more complex process than is the case for survey research that involves collecting only self-reported psychological, social, or economic data. This is true in particular if the biospecimens and the data derived therefrom will form a large collection to be used by numerous researchers for an indeterminate period of time for research purposes that may all not be well defined at the time consent is obtained, especially if genomic data are involved. The controversies do not make this kind of research impossible, but they do add substantial complications that social science researchers need to anticipate. This chapter begins with an overview of the

informed consent process. It then examines how collecting biospecimens alters the informed consent process with which social scientists are familiar. The third section reviews a number of unresolved issues surrounding informed consent in social science survey research that includes the collection of biospecimens. The chapter ends with a discussion of the role of Institutional Review Boards (IRBs) in the informed consent process for such research.

OVERVIEW OF THE INFORMED CONSENT PROCESS

This section summarizes the purpose and the basic elements of the informed consent process.

The Purpose of Informed Consent

Informed consent as it is understood today did not exist until about midway through the twentieth century (see, for example, Faden and Beauchamp, 1986). In the medical arena, physicians historically were obligated to act in the best interests of their patients as they understood those interests, but they felt no obligation to inform a patient completely about a situation and then follow the patient's decision. A scattering of early cases established that physicians had to provide patients with information about proposed procedures and that the terms of the patient's consent could place legal limits on the procedures that could be performed. It was not until a series of cases in the 1950s and 1960s that clinical informed consent became well established as a legal right with full legal redress, equivalent to the penalties for battery, if it was not obtained. Battery theory was gradually replaced by negligence theory in the 1950s and 1960s (Levine, 1988). Similarly, there was historically no generally recognized obligation of those performing medical research to inform their subjects about the conduct, purpose, and risks of the research and to obtain their consent for participation, although there were some rudimentary versions of informed consent as early as 1900 (Vollman and Winau, 1996). A famous case in which informed consent was obtained was the Walter Reed yellow fever studies in the late nineteenth century.

The first internationally recognized code of research ethics, the Nuremberg Code of 1947, was developed in response to revelations concerning abhorrent Nazi human experimentation before and during World War II. It emphasized the importance of informed consent and the centrality of subjects' self-determination to participate in medical research. Some years later, in 1964, at the 18th World Medical Assembly in Helsinki, Finland, the World Medical Association adopted the Declaration of Helsinki, which provided a set of guiding principles for ethical considerations related to biomedical research. Until the 1960s, informed consent to research generally was obtained only from vol-

unteers. Then, as a result of U.S. Food and Drug Administration (FDA) statute and the Declaration of Helsinki, the practice was extended to patients.

In the United States, the importance of obtaining informed consent for medical research achieved national prominence in 1972 following revelations of unethical practices by the U.S. Public Health Service in the Tuskegee syphilis study. In 1974, the U.S. Congress passed the National Research Act, which created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. In 1979, the commission issued the Belmont Report, which set forth three guiding principles underlying the ethical conduct of research: (1) respect for persons, (2) beneficence, and (3) justice.

The purpose of informed consent and the reasons behind it have changed little since the Declaration of Helsinki was issued: no one should be subjected to experimentation—or other research activities—without giving permission, and that permission is meaningful only if the person understands what he or she is giving permission for. Thus the process of informed consent is designed to provide human subjects with sufficient information about the nature and purpose of the research, the procedures to be employed, and the potential risks and benefits of their participation so they can make an informed decision about whether to participate. As the types of studies have become more complex, the issues surrounding informed consent have become more complex and difficult to resolve, but the basic underlying principles have not changed.

Basic Elements of Informed Consent

The basic elements of informed consent are set forth in the U.S. Code of Federal Regulations (45 CFR 46 Part 116). First, the language of the informed consent document should be reasonably clear and specific. Some commentators have recommended that the document be written at the fifth-grade or some other low level of reading capability, but the precise appropriate reading level is somewhat controversial (see, for example, Levine, 1982). Among other topics, the document should address the following:

- the fact that research is being performed, the reason why it is being performed, and what the participant will be asked to do;
- the risks and potential benefits (including financial and nonfinancial) to the participant;
- the level of confidentiality that will be maintained, which can include such information as whether the data resulting from collected biospecimens will be deidentified or coded and whether the data or study results will be released to the participant, his/her family, or his/her health care provider; and
- the participant's right to withdraw consent without penalty.

Adding biological specimens to social science surveys changes the nature of the informed consent required. Because of the unique risks involved, social science surveys that collect biospecimens should include several additional elements in the informed consent document:[1]

- a process for dealing with blanket consent (see below), including what process will be used when the participant's specimens or data are to be used in future studies;
- a process for dealing with any significant medical information that is uncovered by the analysis (in general, any information that is not released to the subject will not be released to treating physicians. On the other hand, the subject may choose to have some data released to him/her without also being released to treating physicians);
- the process governing withdrawal and what it means with respect to what can be done with the specimens and data after withdrawal; and
- how third-party issues (see below) will be handled, including questions about third-party privacy and the furnishing of medically relevant information that may affect third parties.

Generally, although informed consent should be obtained before biospecimens are collected, in some instances postcollection consent is appropriate. This is the case, for instance, for the use of leftover specimens from earlier surgery or cases in which it was not initially possible to obtain consent because of the participant's illness, undue stress, or inability to comprehend the consent procedure.[2]

DIFFERENCES IN CONSENT PRACTICES BETWEEN SOCIAL SCIENCE SURVEY AND BIOMEDICAL RESEARCH

When social science surveys include the collection of biospecimens, two hitherto separate informed consent processes coincide. Social scientists must deal with requirements and issues different from those to which they are accustomed.

[1]An alternative approach, used in the Personal Genome Project (PGP) is to make volunteers score 100 percent on a consent exam. This is intended to ensure that participants understand that there is no real way to predict what can be done with genomic data in the future, nor any way to protect confidentiality or protect against possible misuses of the data in the future with 100 percent certainty. The panel does not believe that this is an appropriate or practical strategy in this context.

[2]Note that even the most sincere investigator cannot anticipate all of the ways in which a biological specimen may be used in the future. This is a major issue, but without a clear solution.

Traditional Consent Practices for Social Science Surveys

Traditionally, the informed consent process in social science survey research has been relatively straightforward. Because the risk of a breach is understood to be fairly small, the informed consent forms tend to be minimal as well; obtaining consent over the telephone at the beginning of a survey call is a common practice. At many institutions, the informed consent process is overseen by a nonmedical IRB, which generally has a very different idea of appropriate informed consent from that of a biomedical IRB (Martin and Marker, 2007).

Traditional Consent Practices for Biomedical Research

Biomedical informed consent has typically been a much more rigorous process than the traditional process in the social sciences. This is the case in part because in certain clinical trials and in other potentially invasive biomedical research, there can be a very real risk of physical harm. The informed consent process is generally seen as the responsibility of principal investigators and is usually overseen by medical IRBs. It involves extensive written consent forms containing detailed accounts of the potential risks, which are typically explained in face-to-face meetings with informed-consent personnel (Doyal and Tobais, 2001). This closer scrutiny of biomedical research has also influenced the treatment of informed consent for medical research that poses not physical risks but financial or psychological risks to participants.

Informed Consent for Social Science Surveys That Include the Collection of Biospecimens

Social scientists who wish to add the collection of biospecimens to their surveys will likely need to adopt standards of informed consent similar to those appropriate for minimal-risk biomedical research. (Minimal risk, as in the Common Rule [see Chapter 3], is taken to mean that the probability and magnitude of possible harm or discomfort resulting from participating in the study is no greater than that encountered ordinarily in daily life or during the performance of routine physical or psychological examinations or tests.) In some cases, as discussed in Chapter 3, the research may involve more than minimal concerns and risks. The addition of data derived from biospecimens to survey data may increase the risk of a breach of confidentiality and the amount of potentially damaging information that could be revealed in such a breach. To the extent that social science surveys that include the collection of biospecimens pose risks commensurate with those posed by biomedical research—including the risk, real or perceived, of disclosure—the informed consent process should be of parallel rigor. IRBs dealing with biosocial surveys should benefit from the standards and practices of both social and behavioral science IRBs and medi-

cal IRBs in light of a careful reading of the Common Rule (see also, National Research Council, 2003).

UNRESOLVED ISSUES

A number of issues remain unresolved with respect to informed consent for social science survey research that involves the collection of biospecimens. They include consent for potential future uses of the specimens and the data derived therefrom, return of significant findings, withdrawal of consent, third-party issues, and the nature and extent of communication with participants.

Consent for Potential Future Uses

The concept of informed consent assumes that participants can be truly informed about the various aspects of a study, including its benefits and risks. Yet with today's increasingly rapid changes in science and technology, obtaining (and retaining) informed consent becomes a moving target. This is particularly true with respect to whole-genome sequencing, genomics, and information technology, in which changes occur so rapidly that it is difficult to imagine what might be possible in just a few years' time. As a result, it may be impossible to guarantee a participant's privacy and confidentiality or to discuss comprehensively what risks the participant might face by taking part in a survey in which biospecimens will be collected and stored for future use.

Researchers who are collecting biological specimens (or establishing collection protocols for such specimens) as part of a social science survey therefore face a difficult issue concerning consent for potential future research uses of the specimens they collect and the data derived therefrom. Because it is impossible to know what types of research might be performed on the specimens and data in the future, researchers cannot fully inform study participants about those potential future uses when the original consent is sought. There are two basic approaches to dealing with this issue. One is to return to the participants to obtain their consent for each new kind of study, explaining its purpose, conduct, benefits, and risks. The other is to ask the participants at the time of the original survey to provide a blanket consent that permits use of the specimens and data for various studies in the future without the need to recontact them to request consent. There are also various combinations of these two approaches that entail varying degrees of contact with and commitment from participants, but the focus here is only on the two basic options. In addition, in some cases deidentification or anonymization of the data may obviate the need to obtain informed consent for future research using the data (see Chapter 3), although this means of avoiding informed consent poses ethical issues.

Recontact

Perhaps the most straightforward approach to ensuring that participants are comfortable with how their contributed specimens will be used is to recontact them to obtain informed consent whenever a new study is proposed. This approach guarantees that no specimens will be used without participants' explicit consent. However, it is a process many investigators find too costly, labor-intensive, and burdensome, and the consensus in the field is that it is economically and logistically infeasible. Moreover, it is impossible to reach every person involved in the original study after a significant amount of time has passed, as many of participants will have moved, and some will have died.[3]

Recontact procedures raise ethical questions as well, since it can be argued that the act of seeking to contact participants may violate privacy, especially if they asked not to be contacted again in their original consent document. Recontact also could be annoying to participants, especially if done frequently, and could reduce their willingness to continue to take part in the research.

As a result, a variety of organizations and researchers have concluded that recontact is not a workable option and that insisting on it would prevent potentially valuable research from being carried out. A working group of the UNESCO International Bioethics Committee found, for example, that a "system which required fresh consent would be extremely cumbersome and could seriously inhibit research" (United Nations Educational, Scientific and Cultural Organization, 2002, p. 16). Similarly, Strouse (2005, pp. 142–143) suggests that "going back to subjects to seek specific consent for each later study undertaken with the initial sample may seriously inhibit potentially important research."

There is one report of a new approach that may make recontact easier and more feasible (Shickle, 2006). First Genetic Trust, a U.S. biotech company, has developed "dynamic consent procedures in which subjects are contacted by e-mail each time their DNA is used." Study participants offer informed consent for the initial study, and later, if the same or a different group of researchers wishes to use the biospecimens or the data generated therefrom in a new study, the participants can say yes or no to the new research (Shickle, 2006, p. 507). One limitation of this approach is that it depends on the study participants having regular access to the Internet, which may be problematic, particularly for older individuals.

[3]The experience of the competition for the follow-up to the National Long-Term Care Survey may be instructive here. During the rebidding process, no applicant organization proposed retaining the existing longitudinal sample, partly because doing so would have required that the Census Bureau recontact each participant and obtain an additional consent before any new data could be collected.

Blanket Consent

Because of the difficulties involved in recontact, many organizations and researchers have used or have called for the use of blanket consent to the general use of specimens or data in future, unspecified projects. This type of consent can be advantageous both for participants, as less is demanded of them, and for researchers, who avoid costly and time-consuming recontact procedures. "If you are creating a biobank, the idea is that you want to build a resource that is going to have all kinds of uses going forward, many yet undefined. So blanket consent becomes something that, presumably, is more appealing" (Malinowski, 2005, p. 8).

Among the organizations that have recommended the use of blanket consent for data in biorepositories are UNESCO and the World Health Organization groups working on these issues (World Health Organization, 1998; United Nations Educational, Scientific and Cultural Organization, 2002). The UK Biobank, one of the largest biobanks in the world, is not planning to seek consent prior to each new research study using stored biospecimens once the initial consent has been obtained (Shickle, 2006). The biobank does plan, however, to send out frequent newsletters to remind subjects of their participation and to inform them of what studies are being conducted and planned for the future. The biobank also uses e-mail as a form of contact, which opens up the possibility of using this means of communication each time participants' DNA is used, giving them sufficient information and time to make an informed decision about whether to extend their consent to the newer projects. Again, however, any approach that relies on e-mail and access to the Internet may leave out some participants, particularly the elderly.

Approaches to Mitigating Problems Raised by Blanket Consent

One major problem with blanket consent is that it is inherently vague, and thus study participants can never know exactly what their biospecimens and other information will be used for in the future. This lack of knowledge presents a serious ethical issue: blanket consent arguably fails to meet the basic requirement for informed consent that it describe the possible risks of a study to potential participants. If the future studies are unknown, the risks of those studies cannot be described. For example, study participants may agree to enroll in a study that includes collecting biospecimens and to allow the sharing of their samples as long as their confidentiality is protected. But what happens if a future study results in diagnosing a particular participant with a disease? Should the participant be contacted with that information? What if the participant would prefer not to know? (See the discussion of this issue below.) In short, it is always possible to envision scenarios that neither the participant nor the researcher could have anticipated at the time the informed consent was

given. Thus, it can be argued that blanket consent cannot truly be "informed" consent (Deschênes et al., 2001).

Some researchers have argued that such issues are not serious concerns and that a simple blanket consent is sufficient to safeguard the rights of participants. For example, Rothstein argues that "as long as the potential research subjects are clearly apprised of the range of possible future uses of their sample, they should be permitted to give one-time blanket consent to such uses" (Rothstein, 2005, pp. 92–93). Others, however, have suggested that these issues need to be addressed. Moreover, as noted in Chapter 2, IRBs are increasingly conservative and therefore unlikely to grant approval for the use of blanket consent.

Various approaches to addressing this issue have been suggested. For instance, Greely (1999) proposes a variation on blanket consent that adds extra protections. His approach involves telling participants that their biological samples will be stored and explaining the possibility that they might be used for future research projects. At that point, participants could choose whether to consent to such future uses of their samples. Should researchers wish to undertake a future project using these samples, that project would have to be approved by an IRB or other appropriate body, which in certain circumstances could insist that the researchers return to the participants to obtain additional consent (Greely, 1999).

Weir and Olick (2004) suggest a hybrid approach: participants would offer their consent for the specific study for which their biospecimens were collected and would also offer consent for future, unspecified studies that fell within certain categories or satisfied certain conditions (see also Kapp, 2008). In effect they would be offered a menu of research options from which to choose, and they could consent to as many as they liked, depending on their values and beliefs. They might, for instance, consent to the use of their specimens for any research on diabetes or cancer but not for studies aimed at understanding the genetic components of behavior.

Another approach is to use a system of tiered consent, whereby participants are allowed to decide to what degree their specimens and data can be used. For example, participants might be given the option of having their specimens and data used only for the original study and not for any future research; in this case, the informed consent document should explain what will be done with the data and, especially, the specimens, when the original study is complete.[4] A second option might be to allow participants to specify the types of research for which their specimens could be used in the future. For example, a participant could agree for the specimens to be used only for cancer

[4]It is possible that a tiered consent procedure could introduce biases in research data, but there is a growing array of methods for dealing with missing data, including selectively missing data. A discussion of such methods is outside the scope of this report, but see Little and Rubin (2002) for more discussion on this point.

research or for research into Alzheimer's disease and diabetes, or perhaps for any sort of medical research. Care must be taken when considering such a tiered system of informed consent, however, since it may be inappropriate for a broad range of inter- and multidisciplinary research. A tiered approach should be used only if the biorepository (or other appropriate agent) has sufficient experience and a tracking system sophisticated enough to monitor the various levels of informed consent.

Deidentification or Anonymization and Consent

In certain cases when data have been deidentified—that is, meeting the requirements of the Health Insurance Portability and Accountability Act (HIPAA)[5] but not rendered completely and irreversibly anonymous—participants are not considered "human subjects" under the Common Rule (45 CFR Part 46, Subpart A), so informed consent is not required (see Chapter 3 for a detailed discussion of the Common Rule and the HIPAA Privacy Rule). This would be the case, for example, if a researcher were using coded data or specimens from a repository that had been collected by other researchers and if the researcher could not readily determine the identity of the individuals involved. (In this context "coded data" refers to records in which identifying information about individuals has been replaced by a number, letter, symbol, or combination thereof.) For example, the researcher may have signed an agreement prohibiting him or her from receiving the key to the code—the procedure under which data contributed to dbGaP have been declared not to be human subjects data (see Box 3-3 in Chapter 3) (Office for Human Research Protections, 2008). In short, if the data and specimens have been deidentified or anonymized, if they were collected by someone else for a different study, and if the researcher agrees to use the data without the possibility of learning the identity of the subjects, then in general it is not necessary to obtain informed consent because the

[5]The HIPAA Privacy Rule makes two methods available for deidentifying health information:

1. Remove 18 specific identifiers listed in the Privacy Rule and determine that there is no other information that may identify the individual. The identifiers are names; geographic subdivisions smaller than a state; all elements of dates (except year) related to an individual (including dates of admission, discharge, birth, death and, for individuals over 89 years old, the year of birth must not be used); telephone numbers; FAX numbers; e-mail addresses; Social Security numbers; medical record numbers; health plan beneficiary numbers; account numbers; certificate/license numbers; vehicle identifiers and serial numbers including license plates; device identifiers and serial numbers; web URLs; Internet protocol addresses; biometric identifiers (including finger and voice prints); full face photos and comparable images; and, any unique identifying number, characteristic or code.
2. Obtain an opinion from a qualified statistical expert that the risk of identifying an individual is very small under the circumstances; the methods and justification for the opinion should be documented. (See 45 CFR Subpart E 164.514 (b).)

participants who supplied the data and specimens are not considered human subjects. Aside from such special cases, however, for someone to be considered no longer a human subject for research purposes, complete and irreversible anonymization is generally required; deidentification is generally not enough.

The difficulty with this approach is that, as discussed in Chapter 3, using deidentification or even anonymization is no guarantee that participants cannot be identified at some point. This problem is even more pressing in databases that combine social data with biological, and particularly genomic, data because the combination can sharply increase the risk of reidentification. Thus avoiding informed consent by using deidentified or anonymized specimens or data raises a serious ethical issue: although the participants are still exposed to risks from breaches of privacy, they are no longer afforded the usual protection given to human subjects under federal regulations (Greely, 2007). Moreover, this procedure explicitly creates a double standard for what constitutes ethically and legally appropriate practice. The original investigators (and anyone who obtains access to the specimens or data through them or their storage facility) must play by the usual rules governing research with human subjects, while those using the deidentified or anonymized data are deemed no longer subject to the Common Rule. Finally, as long as the original specimens or data exist in identifiable form, it is relatively easy for anyone who has the deidentified version of the data to reidentify them through a merger with the identifiable versions.

Return of Significant Findings

Investigators need to consider what they will do if their research yields information with implications for a participant's health. In the course of analyzing the biospecimens collected for a study, for example, researchers may learn of a significantly increased risk of a life-threatening disease linked to a characteristic present in some of the samples, frequently, though not always, a genetic characteristic. Participants often expect that researchers have a moral obligation to share such information with them. Although in many cases the accuracy of genetic testing and its application in disease prediction are subject to debate (Zimmern and Kroese, 2007; Offit, 2008; see also Chapter 1), in other cases the connections are quite clear. It can be argued that if this information could affect a reasonable individual's health care options, it should be disclosed, at least under certain conditions (Greely, 2007)—for example, if the information clearly indicates the presence of an illness, and the illness is treatable. On the other hand, preliminary diagnostic data that reveal little about specificity or sensitivity should not be disclosed. Generally speaking, there should be different thresholds for disclosure depending on the severity of the illness in question and on whether an effective treatment exists. A further complication is that only laboratories covered by the Clinical Laboratory Improvement Amendments (CLIA) are allowed to disclose diagnostic information. Moreover, a participant

may not wish to be contacted with any information about the study or its results, in which case any kind of disclosure, even that deemed to be a moral obligation, could be considered an invasion of the participant's privacy. It is now customary to advise participants to contact a CLIA-approved laboratory independently themselves if they are interested in obtaining information from biodata that might have significant implications for their health.

Currently, IRBs and ethics oversight committees are encouraged to review disclosure statements on a case-by-case basis. One of the factors that can influence their decision is the expected validity of the prediction based on the presence of the risk factor, which can change over time as a result of new scientific discovery (McGuire et al., 2008). In any case, the protocols for sharing study results (if any) with the participants should be clearly explained in the informed consent document, subject to review by the IRB. The issue can be expected to arise with increasing frequency as more becomes known about the human genome, leading to a growing number of incidental findings that reveal genetic information with known implications for health that are not part of the research hypotheses of the study. Moreover, the issue extends beyond genetic and other data derived from biospecimens. A social science study that includes measuring blood pressure will, for example, reveal the occasional participant with dangerous hypertension. One possibility is to use the consent process to inform participants that no diagnostic information will be disclosed, but that they should have regular medical care and consider asking a physician to obtain genetic or other potential diagnostic indicators such as those collected in the study. This may be a tempting option for studies in which the investigators cannot claim competence to deal with diagnostic issues.

Withdrawal of Consent

There is widespread agreement that a basic component of any system of informed consent must be the ability to withdraw consent at any time. Initially, this concept was applied to human subjects research, as can be seen in the Declaration of Helsinki: "In any research on human beings, each potential subject . . . should be informed that he or she is at liberty to abstain from participation in the study and that he or she is free to withdraw his or her consent to participation at any time" (World Medical Association, 1964). This concept actually originated much earlier, in the Nuremburg Code. Later, with the development of biobanks, the principle of withdrawal of consent was generally included in ethical guidelines for dealing with subjects who provide specimens (Eriksson and Helgesson, 2005). The Common Rule specifies that "participation is voluntary . . . and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled" (45 CFR 46 Part 116(a)(8)).

The reasons for withdrawing consent to use stored specimens and data are

usually different from the reasons for withdrawing consent for participating in medical research.[6] In the latter case, participants are generally concerned about some sort of direct physical harm, but this is not an issue for stored specimens and data. Instead, participants are concerned about such things as sensitive information ending up in the wrong hands, leading in turn to discrimination, stigma, or personal distress. For example, prior to the passage of the Genetic Information Nondiscrimination Act of 2008, some participants might have worried that their genetic information would end up in the possession of an insurance company and possibly make it more difficult for them to obtain insurance.

It is difficult to estimate the risks of such occurrences, although some researchers believe them to be relatively small (e.g., Eriksson and Helgesson, 2005). Nonetheless, in most cases a study participant may withdraw from the study at any time, and this right should be clearly stated in the informed consent document.

When consent is withdrawn, the question arises of whether and to what extent the data gleaned from specimens can be used in research. As Shickle (2006, p. 515) notes:

> withdrawal raises some difficult questions. It is clear that new data collection should stop. However, the participant may still be content for their DNA and previously collected data to be used. Their decision may depend on whether withdrawal has been triggered by a desire not to share information about a new disease, perhaps that is associated with stigma, or by fundamental concerns about the way that the biobank is being maintained and used.

In the case of previous disclosure or publication of the data, it may, in fact, be impossible not to use the data. As Shickle (2006, p. 515) comments, "It would be unreasonable and impractical for a participant to require that data previously collected be removed from any analyses previously conducted and potentially published."

The more difficult question is whether data derived from a participant's biospecimens can be used in studies after consent is withdrawn. One possible solution to the continuing use of such data is anonymization. As with the use of anonymization to avoid the need for informed consent to uses of specimens and data beyond the original research, however, there are a variety of problems with using the procedure to deal with withdrawal of consent. As noted earlier, for example, anonymization renders the data less useful scientifically, especially when it comes to social survey data that are linked to administrative records and other health data (see Chapter 3). And as discussed earlier, anonymization may not completely cut the link to a given individual; it may still be possible to

[6]The regulations do not require that patients wishing to withdraw consent give a reason for doing so.

identify the individual through various means, for example, by using identifying characteristics such as age, sex, and location. Stripping the specimen of all such information may render it useless for research purposes. Moreover, anonymization may not truly satisfy a participant's wishes. If the participant desires, for whatever reasons, that his or her data not be used for any purposes at all, stripping the data of identifying information but still using them does not comply with this desire (Eriksson and Helgesson, 2005).

Arguing that biobank research produces a public good, Eriksson and Helgesson (2005, p. 1075) suggest that withdrawal of consent should require more than simply a change of heart: "During the course of the research, a participant should be at liberty to withdraw his consent if he can present sufficient reasons why it is no longer reasonable to ask for his continued participation." If the participant has sufficient reasons—something left to the researchers, biobank managers, or an ethics committee to decide—the data will be anonymized or destroyed, as the participant chooses. Likewise, samples may be destroyed, anonymized, or returned.

The panel is not comfortable with Eriksson and Helgesson's approach but does believe that there must be some practical limitations on the right to withdraw consent. In particular, the panel agrees that once data have already been published or relied upon in publications, it is not practical to delete them. (For a discussion of practical limitations on withdrawal of consent, see Levine [1988]). The larger message is that researchers should be aware of the issue, make plans for what they will do in cases of withdrawal of consent, and spell out those plans in the consent form.

Third-Party Issues

Because relatives of the participants in a study share some of the participants' DNA, many of the same risks participants face exist for their relatives as well, albeit in attenuated form. Any sensitive information, such as a genetic predisposition to a disease that is obtained from the genetic sample of a study participant may also apply to his or her biological relatives. Thus important questions arise concerning whether the researchers have any obligations to the participant's family.

Under federal regulations, third-party relatives are not generally considered to be research subjects, and researchers are not required to obtain their consent. However, a case involving Virginia Commonwealth University (VCU) researchers about a decade ago indicates the difficulties that can arise when a study accumulates information about third parties. A VCU study questionnaire had included questions about the psychiatric history of the participants' parents, and the father of one of the participants objected and complained to federal officials that the questions invaded his privacy (Amber, 2000). When the Office of Protection from Research Risks (OPRR, now the Office of Human

Research Protections, or OHRP) reviewed the minutes pertaining to the VCU IRB's review of the study, it found that the IRB had not considered whether the father or other family members should be regarded as research subjects whose consent to participate would be required. OPRR agreed that the father was a subject and faulted the IRB for its failure to consider whether, in this case, family members were or were not research subjects.

In the case of genome-wide association studies, it is impossible to avoid accumulating some information about third parties, since sequencing the genome of a study participant inevitably provides genetic information about that individual's biological relatives. Although it appears inappropriate to extend autonomy-based consent procedures to third-party relatives, it is important to be aware that these risks exist, and that researchers' ethical obligations to these individuals increase with the increasing risk of breaches of confidentiality and autonomy (Greely, 1997). Thus, it can be argued that study participants should be fully informed about these types of risks during the initial consent stage, even to the point of encouraging them to discuss their participation and possible consequences with their families before enrolling in a study (McGuire et al., 2008).

If information from a sample reveals potential health issues for a participant, those health issues could affect relatives as well. As noted above, there are no clear answers about what to do if analysis of data in a study indicates a link between a particular characteristic and the likelihood of developing a disease. One can argue that a moral obligation exists to inform the participants in a study of such a risk, but one can also make the opposing argument that at least some participants may not want to know and that recontacting them with the information is an invasion of privacy. The ethical issues are even muddier for third parties, and the question of whether researchers are morally obligated to inform third-party relatives of the results of a biomarker association or similar study remains unanswered.

Effective Communication with Participants

In obtaining informed consent, it is not sufficient simply to decide what information should be provided to participants; one must also take into account the participants' ability to understand and process that information. In other words, one must communicate the information effectively.

One of the most obvious factors to take into account is language needs. The necessary information should be provided in a language in which the participant is fluent. Those whose English skills are questionable should be given the information in their native language.

Even among native English speakers, it is important to consider exactly how the information is presented. Communications should be at an appropriate level. Furthermore, those preparing the consent form should be familiar with

research on health literacy. Studies among those with different levels of educational attainment imply that a good deal of variation exists in health literacy even among populations with more than a high school education.

To the extent possible, consent forms should be tailored to fit different populations. One population of special concern is the elderly, for whom cognitive impairment may make informed consent problematic. This is a particular issue among the oldest adults—those aged 85 and above.

Another issue to consider is how much information to provide. A balance must be achieved between presenting too little information, so participants cannot truly understand what they are consenting to, and presenting too much information, which may interfere with obtaining meaningful consent. A good approach is to consider the materiality of the information—a legal term of art—and provide only that information which is material to the participant's understanding of the benefits and risks of participation and therefore to his or her capability to make a decision.[7]

The Role of Institutional Review Boards

The Code of Federal Regulations (45 CFR Part 46) requires the use of IRBs for all research funded by the U.S. Department of Health and Human Services (HHS) that involves human subjects, and the Common Rule extends this requirement to most other federal agencies that fund research involving human subjects (see Chapter 3). Thus almost all federally funded research projects with human subjects must use IRBs, and they must follow the requirements concerning IRB use that are set forth in the federal statutes.[8]

Review and prior approval by research IRBs serves as the major protective oversight mechanism for all human subjects research in the United States, but the role of IRBs in protecting the rights of research participants—and in particular, their role in the process of informed consent—has been the subject of debate. Among other things, IRBs weigh in on the possible risks and benefits of participation in a study, the proper selection of participants, and whether informed consent should be obtained. If informed consent is required, an IRB must approve the informed consent document before the beginning of the study and then review it periodically. IRBs are also permitted to waive or alter the informed consent requirement if a study meets the conditions described in the Common Rule, the basic federal rule governing the role of IRBs in overseeing human subjects research. However, this kind of waiver is not always fitting and can sometimes be difficult for researchers to obtain (Littenberg and MacLean, 2006), although appropriate requests are rarely denied.

[7]With medical studies, patients are sometimes encouraged to take the document home and give consent only after "sleeping" on it.

[8]See 45 CFR 46, 101 for exemptions.

The Role of IRBs in Reviewing Informed Consent Procedures

There is growing concern about the ability of IRBs to be effective in their role of reviewing the informed consent process. Over the past decade, a number of calls have been made to examine closely how IRBs carry out this role. One editorial in a medical journal commented, for example, that the "tissue issue" has become increasingly vexatious for IRBs, which struggle with the question of how to phrase informed consent forms to explain the genetic studies that might be performed on the tissue in the future (Levine, 2001).

Furthermore, IRBs are becoming increasingly overburdened, with detrimental effects on their productivity and usefulness. In addition, persisting confusion about the federal guidelines that regulate human subjects research and the sometimes conservative interpretations of these guidelines by IRBs have raised concern in the research community as it strives for excellence that is both scientific and ethical (Levine, 2006; Fost and Levine, 2007).

The Role of IRBs in Biorepositories and Biobanks

The role IRBs should play in research involving biospecimens and biodata (especially genetic data) from a repository or biobank is unclear, and this subject is not considered in the federal regulations. The current IRB structure was designed primarily for direct human research, and it works especially well in such cases as clinical drug trials because the IRB paradigm assumes that researchers will be working with live human subjects to whom they disclose the probability and magnitude of potential harms that may be associated with the research (Meslin and Quaid, 2004). The current IRB structure is not as well suited to research involving biorepositories or biobanks in which the tissues for the studies are not obtained directly from subjects; the principal harms are nonphysical; the studies may involve someone other than the person from whom the sample was obtained, such as a family member; and the studies do not require the person to be present or even alive to participate (Meslin and Quaid, 2004). This is one good reason why biosocial research should not be delegated completely to medical IRBs.

The problem is especially difficult for social surveys that collect and store biospecimens and biodata; there are no clear and immediate benefits of participation for those who donate the specimens, yet risks are still present. The issues that arise from the storage of specimens and data from social science research were not anticipated when the federal regulations concerning IRBs were developed, and it may be necessary at some point to revisit these regulations and update them for use with biorepositories and biobanks. Such issues have long been a concern of social scientists (and their IRBs) because they have a history of sharing data, and for the most part no such tradition exists in biomedical research.

Issues Related to the Involvement of Multiple IRBs

When researchers from different institutions are involved in a single project or when researchers from one institution share biospecimens or biodata with researchers from another, multiple IRBs are likely to become involved. This involvement of multiple IRBs can cause a variety of problems that have no easy solutions (Burman et al., 2001). As Meslin (2006, p. 279) notes: "All researchers have experienced the frustration of submitting a protocol for a multi-center study only to have one IRB approve it, another decline to approve it, and a third require extensive modifications that the other two did not." This is a particularly difficult problem for multicenter biomedical research, and in November 2005 a national conference was held to discuss solutions. The conference was sponsored by the National Institutes of Health, OHRP, the Association of American Medical Colleges, and the American Society of Clinical Oncology. The workshop report details a number of the issues involved and some approaches to addressing them (National Institutes of Health, Office for Human Research Protections, Association of American Medical Colleges, and American Society of Clinical Oncology, 2005).

The issues to which Meslin refers are perhaps the most obvious ones that stem from the presence of multiple IRBs: How does an investigator proceed if one IRB approves the research plan and a second does not, or if one requires modifications that another does not? And what happens if two IRBs impose contradictory sets of conditions that both cannot be met? There are also concerns related to the IRBs' continuing oversight of a project, as disagreements between them can be even more problematic once a project is under way, participants have been recruited, and specimens and data are being analyzed.

As biosocial surveys and studies become more common, it is likely that the number of projects requiring multiple IRBs will increase. To be sure, this issue is not limited to biosocial surveys, but it will become increasingly important to find new ways to provide the oversight that IRBs offer without the complications entailed in having to answer to independent and uncoordinated entities at multiple institutions. Two possible approaches are to develop procedures for centralized review or to find ways for IRBs at different institutions to cooperate in their reviews of multi-institution research (Wolf, Croughan, and Lo, 2002).

CONCLUSION

Social scientists planning to add the collection of biospecimens to their surveys are likely to find that the informed consent process is quite different from their past practice. The best way for them to deal with these new challenges is to consult with investigators at their institutions or others with prior experience with biosocial or biomedical research, who can provide practical information about developing an informed consent document and about what to expect from their IRBs.

5

Findings, Conclusions, and Recommendations

As the preceding chapters have made clear, incorporating biological specimens into social science surveys holds great scientific potential, but also adds a variety of complications to the tasks of both individual researchers and institutions. These complications arise in a number of areas, including collecting, storing, using, and distributing biospecimens; sharing data while protecting privacy; obtaining informed consent from participants; and engaging with Institutional Review Boards (IRBs). Any effort to make such research easier and more effective will need to address the issues in these areas.

In considering its recommendations, the panel found it useful to think of two categories: (1) recommendations that apply to individual investigators, and (2) recommendations that are addressed to the National Institute on Aging (NIA) or other institutions, particularly funding agencies. Researchers who wish to collect biological specimens with social science data will need to develop new skills in a variety of areas, such as the logistics of specimen storage and management, the development of more diverse informed consent forms, and ways of dealing with the disclosure risks associated with sharing biogenetic data. At the same time, NIA and other funding agencies must provide researchers the tools they need to succeed. These tools include such things as biorepositories for maintaining and distributing specimens, better guidance on informed consent policies, and better ways to share data without risking confidentiality.

TAKING ADVANTAGE OF EXISTING EXPERTISE

Although working with biological specimens will be new and unfamiliar to many social scientists, it is an area in which biomedical researchers have a great deal of expertise and experience. Many existing documents describe recommended procedures and laboratory practices for the handling of biospecimens. These documents provide an excellent starting point for any social scientist who is interested in adding biospecimens to survey research.

Recommendation 1: Social scientists who are planning to add biological specimens to their survey research should familiarize themselves with existing best practices for the collection, storage, use, and distribution of biospecimens. First and foremost, the design of the protocol for collection must ensure the safety of both participants and survey staff (data and specimen collectors and handlers).

Although existing best-practice documents were not developed with social science surveys in mind, their guidelines have been field-tested and approved by numerous IRBs and ethical oversight committees. The most useful best-practice documents are updated frequently to reflect growing knowledge and changing opinions about the best ways to collect, store, use, and distribute biological specimens. At the same time, however, many issues arising from the inclusion of biospecimens in social science surveys are not fully addressed in the best-practice documents intended for biomedical researchers. For guidance on these issues, it will be necessary to seek out information aimed more specifically at researchers at the intersection of social science and biomedicine.

COLLECTING, STORING, USING, AND DISTRIBUTING BIOSPECIMENS

As described in Chapter 2, the collection, storage, use, and distribution of biospecimens and biodata are tasks that are likely to be unfamiliar to many social scientists and that raise a number of issues with which even specialists are still grappling. For example, which biospecimens in a repository should be shared, given that in most cases the amount of each specimen is limited? And given that the available technology for cost-efficient analysis of biospecimens, particularly genetic analysis, is rapidly improving, how much of any specimen should be used for immediate research and analysis, and how much should be stored for analysis at a later date? Collecting, storing, using, and distributing biological specimens also present significant practical and financial challenges for social scientists. Many of the questions they must address, such as exactly what should be held, where it should be held, and what should be shared or distributed, have not yet been resolved.

Developing Data Sharing Plans

An important decision concerns who has access to any leftover biospecimens. This is a problem more for biospecimens than for biodata because in most cases, biospecimens can be exhausted. Should access be determined according to the principle of first funded, first served? Should there be a formal application process for reviewing the scientific merits of a particular investigation? For studies that involve international collaboration, should foreign investigators have access? And how exactly should these decisions be made? Recognizing that some proposed analyses may lie beyond the competence of the original investigators, as well as the possibility that principal investigators may have a conflict of interest in deciding how to use any remaining biospecimens, one option is for a principal investigator to assemble a small scientific committee to judge the merits of each application, including the relevance of the proposed study to the parent study and the capacities of the investigators. Such committees should publish their review criteria to help prospective applicants. A potential problem with such an approach, however, is that many projects may not have adequate funding to carry out such tasks.

Recommendation 2: Early in the planning process, principal investigators who will be collecting biospecimens as part of a social science survey should develop a complete data sharing plan.

This plan should spell out the criteria for allowing other researchers to use (and therefore deplete) the available stock of biospecimens, as well as to gain access to any data derived therefrom. To avoid any appearance of self-interest, a project might empower an external advisory board to make decisions about access to its data. The data sharing plan should also include provisions for the storage and retrieval of biospecimens and clarify how the succession of responsibility for and control of the biospecimens will be handled at the conclusion of the project.

Recommendation 3: NIA (or preferably the National Institutes of Health [NIH]) should publish guidelines for principal investigators containing a list of points that need to be considered for an acceptable data sharing plan. In addition to staff review, Scientific Review Panels should read and comment on all proposed data sharing plans. In much the same way as an unacceptable human subjects plan, an inadequate data sharing plan should hold up an otherwise acceptable proposal.

Supporting Social Scientists in the Storage of Biospecimens

The panel believes that many social scientists who decide to add the collection of biospecimens to their surveys may be ill equipped to provide for the storage and distribution of the specimens.

Conclusion: The issues related to the storage and distribution of biospecimens are too complex and involve too many hidden costs to assume that social scientists without suitable knowledge, experience, and resources can handle them without assistance.

Investigators should therefore have the option of delegating the storage and distribution of biospecimens collected as part of social science surveys to a centralized biorepository. Depending on the circumstances, a project might choose to utilize such a facility for immediate use, long-term or archival storage, or not at all.

Recommendation 4: NIA and other relevant funding agencies should support at least one central facility for the storage and distribution of biospecimens collected as part of the research they support.

PROTECTING PRIVACY AND CONFIDENTIALITY: SHARING DIGITAL REPRESENTATIONS OF BIOLOGICAL AND SOCIAL DATA

Several different types of data must be kept confidential: survey data, data derived from biospecimens, and all administrative and operational data. In the discussion of protecting confidentiality and privacy, this report has focused on biodata, but the panel believes it is important to protect all the data collected from survey participants. For many participants, for example, data on wealth, earnings, or sexual behavior can be as or more sensitive than genetic data.

Conclusion: Although biodata tend to receive more attention in discussions of privacy and confidentiality, social science and operational data can be sensitive in their own right and deserve similar attention in such discussions.

Protecting the participants in a social science survey that collects biospecimens requires securing the data, but data are most valuable when they are made available to researchers as widely as possible. Thus there is an inherent tension between the desire to protect the privacy of the participants and the desire to derive as much scientific value from the data as possible, particularly since the costs of data collection and analysis are so high. The following recommendations regarding confidentiality are made in the spirit of balancing these equally important needs.

Genomic data present a particular challenge. Several researchers have demonstrated that it is possible to identify individuals with even modest amounts of such data. When combined with social science data, genomic data may pose an even greater risk to confidentiality. It is difficult to know how much or which genomic data, when combined with social science data, could become critical identifiers in the future. Although the problem is most significant with genomic data, similar challenges can arise with other kinds of data derived from biospecimens.

Conclusion: Unrestricted distribution of genetic and other biodata risks violating promises of confidentiality made to research participants.

There are two basic approaches to protecting confidentiality: restricting data and restricting access. Restricting data—for example, by stripping individual and spatial identifiers and modifying the data to make it difficult or impossible to trace them back to their source—usually makes it possible to release social science data widely. In the case of biodata, however, there is no answer to how little data is required to make a participant uniquely identifiable. Consequently, any release of biodata must be carefully managed to protect confidentiality.

Recommendation 5: No individual-level data containing uniquely identifying variables, such as genomic data, should be publicly released without explicit informed consent.

Recommendation 6: Genomic data and other individual-level data containing uniquely identifying variables that are stored or in active use by investigators on their institutional or personal computers should be encrypted at all times.

Even if specific identifying variables, such as names and addresses, are stripped from data, it is still often possible to identify the individuals associated with the data by other means, such as using the variables that remain (age, sex, marital status, family income, etc.) to zero in on possible candidates. In the case of biodata that do not uniquely identify individuals and can change with time, such as blood pressure and physical measurements, it may be possible to share the data with no more protection than stripping identifying variables. Even these data, however, if known to intruders, can increase identification disclosure risk when combined with enough other data. With sufficient characteristics to match, intruders can uniquely identify individuals in shared data if given access to another data source that contains the same information plus identifiers.

Conclusion: Even nonunique biodata, if combined with social science data, may pose a serious risk of reidentification.

In the case of high-dimensional genomic data, standard disclosure limitation techniques, such as data perturbation, are not effective with respect to preserving the utility of the data because they involve such extreme alterations that they would severely distort analyses aimed at determining gene–gene and gene–environment interactions. Standard disclosure limitation methods could be used to generate public-use data sets that would enable low-dimensional analyses involving genes, for example, one gene at a time. However, with several such public releases, it may be possible for a key match to be used to construct a data set with higher-dimensional genomic data.

Conclusion: At present, no data restriction strategy has been demonstrated to protect confidentiality while preserving the usefulness of the data for drawing inferences involving high-dimensional interactions among genomic and social science variables, which are increasingly the target of research. Providing public-use genomic data requires such intense data masking to protect confidentiality that it would distort the high-dimensional analyses that could result in ground-breaking research progress.

Recommendation 7: Both rich genomic data acquired for research and sensitive and potentially identifiable social science data that do not change (or change very little) with time should be shared only under restricted circumstances, such as licensing and (actual or virtual) data enclaves.

As discussed in Chapter 3, the four basic ways to restrict access to data are licensing, remote execution centers, data enclaves, and virtual data enclaves. Each has its advantages and disadvantages.[1] Licensing, for example, is the least restrictive for a researcher in terms of access to the data, but the licensing process itself can be lengthy and burdensome. Thus it would be useful if the licensing process could be facilitated.

Recommendation 8: NIA (or preferably NIH) should develop new standards and procedures for licensing confidential data in ways that will maximize timely access while maintaining security and that can be used by data repositories and by projects that distribute data.

Ways to improve the other approaches to restricted access are needed as well. For example, improving the convenience and availability of virtual data enclaves could increase the use of combined social science and biodata without

[1]See the discussion on "Choosing a Data Sharing Strategy" in Chapter 3.

a significant increase in risk to confidentiality. The panel notes that much of the discussion of the confidentiality risk posed by the various approaches is theoretical; no one has a clear idea of just what disclosure risks are associated with the various ways of sharing data. It is important to learn more about these disclosure risks for a variety of reasons—determining how to minimize the risks, for instance, or knowing which approaches to sharing data pose the least risk. It would also be useful to be able to describe disclosure risks more accurately to survey participants.

Recommendation 9: NIA and other funding agencies should assess the strength of confidentiality protections through periodic expert audits of confidentiality and computer security. Willingness to participate in such audits should be a condition for receipt of NIA support. Beyond enforcement, the purpose of such audits would be to identify challenges and solutions.

Evaluating risks and applying protection methods, whether they involve restricted access or restricted data, is a complex process requiring expertise in disclosure protection methods that exceeds what individual principal investigators and their institutions usually possess. Currently, not enough is known to be able to represent these risks either fully or accurately. The NIH requirement for data sharing necessitates a large investment of resources to anticipate which variables are potentially available to intruders and to alter data in ways that reduce disclosure risks while maintaining the utility of the data. Such resources are better spent by principal investigators on collecting and analyzing the data.

Recommendation 10: NIH should consider funding Centers of Excellence to explore new ways of protecting digital representations of data and to assist principal investigators wishing to share data with others. NIH should also support research on disclosure risks and limitations.

Principal investigators could send digital data to these centers, which would organize and manage any restricted access or restricted data policies or provide advisory services to investigators. NIH would maintain the authority to penalize those who violated any confidentiality agreements, for example, by denying them or their home institution NIH funding. Models for these centers include the Inter-university Consortium for Political and Social Research (ICPSR) and its projects supported by NIH and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and the UK data sharing archive. The centers would alleviate the burden of data sharing as mandated of principal investigators by NIH and place it in expert hands. However, excellence in the design of data access and control systems

is likely to require intimate knowledge of each specific data resource, so data producers should be involved in the systems' development.

INFORMED CONSENT

As described in Chapter 4, informed consent is a complex subject involving many issues that are still being debated; the growing power of genetic analysis techniques and bioinformatics has only added to this complexity. Given the rapid pace of advances in scientific knowledge and in the technology used to analyze biological materials, it is impossible to predict what information might be gleaned from biological specimens just a few years hence; accordingly, it is impossible, even in theory, to talk about perfectly informed consent. The best one can hope for is relatively well-informed consent from a study's participants, but knowing precisely what that means is difficult. Determining the scope of informed consent adds another layer of complexity. Will new analyses be covered under the existing consent, for example? There are no clear guidelines on such questions, yet specific details on the scope of consent will likely affect an IRB's reaction to a study proposal.

What Individual Researchers Need to Know and Do Regarding Informed Consent

To be sure, there is a wide range of views about the practicality of providing adequate protection to participants while proceeding with the scientific enterprise, from assertions that it is simply not possible to provide adequate protection to offers of numerous procedural safeguards but no iron-clad guarantees. This report takes the latter position—that investigators should do their best to communicate adequately and accurately with participants, to provide procedural safeguards to the extent possible, and not to promise what is not possible.[2] Social science researchers need to know that adding the collection of biospecimens to social science surveys changes the nature of informed consent. Informed consent for a traditional social science survey may entail little more than reading a short script over the phone and asking whether the participant is willing to continue; obtaining informed consent for the collection and use of biospecimens and biodata is generally a much more involved process.

[2]In a few cases, it may be necessary to deceive participants about the purposes of a study—for example, in field tests of labor market discrimination—but these situations are unlikely to occur in biosocial studies. However, the Common Rule (45 CFR 46: 46.116.c.2, 46.116.d.3) explicitly permits such exceptions when they are scientifically necessary.

Conclusion: Social scientists should be made aware that the process of obtaining informed consent for the use of biospecimens and biodata typically differs from social science norms.

If participants are to provide truly informed consent to taking part in any study, they must be given a certain minimum amount of information. They should be told, for example, what the purpose of the study is, how it is to be carried out, and what participants' roles are. In addition, because of the unique risks associated with providing biospecimens, participants in a social science survey that involves the collection of such specimens should be provided with other types of information as well. In particular, they should be given detail on the storage and use of the specimens that relates to those risks and can assist them in determining whether to take part in the study.

Recommendation 11: In designing a consent form for the collection of biospecimens, in addition to those elements that are common to social science and biomedical research, investigators should ensure that certain other information is provided to participants:

- **how long researchers intend to retain their biospecimens and the genomic and other biodata that may be derived from them;**
- **both the risks associated with genomic data and the limits of what they can reveal;**
- **which other researchers will have access to their specimens, to the data derived therefrom, and to information collected in a survey questionnaire;**
- **the limits on researchers' ability to maintain confidentiality;**
- **any potential limits on participants' ability to withdraw their specimens or data from the research;**
- **the penalties[3] that may be imposed on researchers for various types of breaches of confidentiality; and**
- **what plans have been put in place to return to them any medically relevant findings.**

Researchers who fail to properly plan for and handle all of these issues before proceeding with a study are in essence compromising assurances under informed consent. The literature on informed consent emphasizes the importance of ensuring that participants understand reasonably well what they are consenting to. This understanding cannot be taken for granted, particularly as it pertains to the use of biological specimens and the data derived therefrom.

[3]Penalties might include NIH eliminating researchers' eligibility for funding and institutions eliminating research privileges of faculty.

While it is not possible to guarantee that participants have a complete understanding of the scientific uses of their specimens or all the possible risks of their participation, they should be able to make a relatively well-informed decision about whether to take part in the study. Thus the ability of various participants to understand the research and the informed consent process must be considered. Even impaired individuals may be able to participate in research if their interests are protected and they can do so only through proxy consent.[4]

Recommendation 12: NIA should locate and publicize positive examples of the documentation of consent processes for the collection of biospecimens. In particular, these examples should take into account the special needs of certain individuals, such as those with sensory problems and the cognitively impaired.

Participants in a biosocial survey are likely to have different levels of comfort concerning how their biospecimens and data will be used. Some may be willing to provide only answers to questions, for example, while others may both answer questions and provide specimens. Among those who provide specimens, some may be willing for the specimens to be used only for the current study, while others may consent to their use in future studies. One effective way to deal with these different comfort levels is to offer a tiered approach to consent that allows participants to determine just how their specimens and data will be used. Tiers might include participating in the survey, providing specimens for genetic and/or nongenetic analysis in a particular study, and allowing the specimens and data to be stored for future uses (genetic and/or nongenetic). For those participants who are willing to have their specimens and data used in future studies, researchers should tell them what sort of approval will be obtained for such use. For example, an IRB may demand reconsent, in which case participants may have to be contacted again before their specimens and data can be used. Ideally, researchers should design their consent forms to avoid the possibility that an IRB will demand a costly or infeasible reconsent process.

Recommendation 13: Researchers should consider adopting a tiered approach to obtaining consent. Participants who are willing to have their specimens and data used in future studies should be informed about the process that will be used to obtain approval for such uses.

[4]Note that this report does not address the issue of obtaining informed consent from children.

What Institutions Should Do Regarding Informed Consent

Because the details of informed consent vary from study to study, individual investigators must bear ultimate responsibility for determining the details of informed consent for any particular study. Thus researchers must understand the various issues and concerns surrounding informed consent and be prepared to make decisions about the appropriate approach for their research in consultation with staff of survey organizations. These decisions should be addressed in the training of survey interviewers. As noted above, however, the issues surrounding informed consent are complex and not completely resolved, and researchers have few options for learning about informed consent as it applies to social science studies that collect biospecimens. Thus it makes sense for agencies funding this research, the Office for Human Research Protection (OHRP), or other appropriate organizations (for example, Public Responsibility in Medicine and Research [PRIM&R]) to provide opportunities for such learning, taking into account the fact that the issues arising in biosocial research do not arise in the standard informed consent situations encountered in social science research. It should also be made clear that the researchers' institution is usually deemed (e.g., in the courts) to bear much of the responsibility for informed consent.

Recommendation 14: NIA, OHRP, and other appropriate organizations should sponsor training programs, create training modules, and hold informational workshops on informed consent for investigators, staff of survey organizations, including field staff, administrators, and members of IRBs who oversee surveys that collect social science data and biospecimens.

The Return of Medically Relevant Information

An issue related to informed consent is how much information to provide to survey participants once their biological specimens have been analyzed and in particular, how to deal with medically relevant information that may arise from the analysis. What, for example, should a researcher do if a survey participant is found to have a genetic disease that does not appear until later in life? Should the participant be notified? Should participants be asked as part of the initial interview whether they wish to be notified about such a discovery? At this time, there are no generally agreed-upon answers to such questions, but researchers should expect to have to deal with these issues as they analyze the data derived from biological specimens.

Recommendation 15: NIH should direct investigators to formulate a plan in advance concerning the return of any medically relevant findings to

survey participants and to implement that plan in the design and conduct of their informed consent procedures.

INSTITUTIONAL REVIEW BOARDS

Investigators seeking IRB approval for biosocial research face a number of challenges. Few IRBs are familiar with both social and biological science; thus, investigators may find themselves trying to justify standard social science protocols to a biologically oriented IRB or explaining standard biological protocols to an IRB that is used to dealing with social science—or sometimes both. Researchers can expect these obstacles, which arise from the interdisciplinary nature of their work, to be exacerbated by a number of other factors that are characteristic of IRBs in general (see Chapter 4).

Recommendation 16: In institutions that have separate biomedical and social science IRBs, mechanisms should be created for sharing expertise during the review of biosocial protocols.[5]

What Individual Researchers Need to Do Regarding IRBs

Because the collection of biospecimens as part of social science surveys is still relatively unfamiliar to many IRBs, researchers planning such a study can expect their interactions with the IRB overseeing the research to involve a certain learning curve. The IRB may need extra time to become familiar and comfortable with the proposed practices of the survey, and conversely, the researchers will need time to learn what the IRB will require. Thus it will be advantageous if researchers conducting such studies plan from the beginning to devote additional time to working with their IRBs.

Recommendation 17: Investigators considering collecting biospecimens as part of a social science survey should consult with their IRBs early and often.

What Research Agencies Should Do Regarding IRBs

One way to improve the IRB process would be to give members of IRBs an opportunity to learn more about biosocial research and the risks it entails.

[5]Sharing expertise between biomedical and social science IRBs does not require a return to the days when there was only one IRB at each institution, a situation that still exists at many small institutions. For example, the Social and Behavioral Science IRB at the University of Wisconsin, Madison, has asked a geneticist to serve as an ex officio member of the IRB when it considers protocols that use genetic data.

This could be done by individual institutions, but it would be more effective if a national funding agency took the lead (see Recommendation 14).

CONCLUSION

It is the panel's hope that its recommendations will support the incorporation of social science and biological data into empirical models, allowing researchers to better document the linkages among social, behavioral, and biological processes that affect health and other measures of well-being while avoiding or minimizing many of the challenges that may arise. Implementing these recommendations will require the combined efforts of both individual investigators and the agencies that support them.

References

Amber, D. (2000). Case at VCU brings ethics to forefront: Genetics researchers who elicit family medical histories cautioned about informed consent. *The Scientist, 14*(9), 1. Available: http://www.the-scientist.com/article/display/11831/ [accessed July 12, 2009].

Aubert, G., and Lansdorp, P.M. (2008). Telomeres and aging. *Physiological Reviews, 88*, 557–579.

Australasian Biospecimen Network. (2007). *Guidelines for Biorepository Protocols.* (Revision 4, March). Available: http://www.abrn.net/pdf/ABN_SOPs_Review_Mar07_final.pdf [accessed February 18, 2010].

Bankhead, C. (2004). Privacy regulations have mixed impact on cancer research community. *Journal of the National Cancer Institute, 96*(23), 1,738–1,740.

Burman, W.J., Reves, R.R., Cohn, D.L., and Schooley, R.T. (2001). Breaking the camel's back: Multicenter clinical trials and local institutional review boards. *Annals of Internal Medicine, 134*, 152–157.

Church, G.M. (2005). The Personal Genome Project. *Molecular Systems Biology*, 1:2005.0030 doi:10.1038/msb4100040. Available: http://thepersonalgenome.com/2005/12/church_on_the_p/ [accessed June 22, 2010].

Clayton, E.W. (2008). *Data Access vs. Confidentiality: Balancing Risks and Benefits*. Presentation at the Workshop on Collecting, Storing, Protecting, and Accessing Biological Data Collected in Social Surveys, November 17–18, National Research Council, Washington, DC.

Dalton, R. (2004). When two tribes go to war. *Nature, 430*(6999), 500–502.

DeRenzo, E.G., Biesecker, L.G., and Meltzer, N. (1997). Genetics and the dead: Implications for genetics research with samples from deceased persons. *American Journal of Medical Genetics, 69*(3), 332–334.

Deschênes, C., Cardinal, G., Knoppers, B.M., and Glass, K.C. (2001). Human genetic research, DNA banking and consent: A question of "form"? *Clinical Genetics, 59*(4), 221–239.

Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D.B. (2010). Rare variants create synthetic genome-wide associations. *PLoS Biology, 8*(1), e1000294.

Doyal, L., and Tobais, J. (Eds.). (2001). *Informed Consent in Medical Research*. London: BMJ Books.

Elger, B.S., and Caplan, A.L. (2006). Consent and anonymization in research involving biobanks: Differing terms and norms present serious barriers to an international framework. *European Molecular Biology Organization Reports, 7*(7), 661–666.

Epel, E.S., Blackburn, E.H., Lin, J., Dhabhar, F.S., Adler, N.E., Morrow, J.D., and Cawthon, R.M. (2004). Accelerated telomere shortening in response to life stress. *Proceedings of the National Academy of Sciences, 101*(49), 17,312–17,315.

Eriksson, S., and Helgesson, G. (2005). Potential harms, anonymization, and the right to withdraw consent to biobank research. *European Journal of Human Genetics, 13*(9), 1,071–1,076.

Faden, R.R., and Beauchamp, T.L. (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press.

Fost, N., and Levine, R.J. (2007). The dysregulation of human subjects research. *Journal of the American Medical Association, 298*(18), 2,196–2,198.

Greely, H.T. (1997). The control of genetic research: Involving the groups between. *Houston Law Review, 33*(5), 1,397–1,430.

Greely, H.T. (1999). Breaking the stalemate: A prospective regulatory framework for unforeseen research uses of human tissue samples and health information. *Wake Forest Law Review, 34*(3), 737–766.

Greely, H.T. (2007). The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annual Review of Genomics and Human Genetics, 8*, 343–364.

Greely, H.T. (2009). Collecting biomeasures in the PSID: Ethical and legal concerns. *Biodemography and Social Biology, 55*(2), 270–288.

Hao, K., Li, C., Rosenow, C., and Hung Wong, W. (2004). Estimation of genotype error rate using samples with pedigree information: An application on the GeneChip Mapping 10K array. *Genomics, 84*(4), 623–630.

Hauser, R. M., and Weir, D. (in press). Recent developments in longitudinal studies of aging in the United States. Submitted to *Demography*.

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., and Craig, D.W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics, 4*(8), e1000167.

Huizinga, D., Haberstick, B.C., Smolen, A., Menard, S., Young, S.E., Corley, R.P., Stallings, M.C., Grotpeter, J., and Hewitt, J.K. (2006). Childhood maltreatment, subsequent antisocial behavior, and the role of monoamine oxidase A genotype. *Biological Psychiatry, 60*(7), 677–683.

Institute of Medicine. (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. S.J. Nass, L.A. Levit, and L.O. Gostin, Eds, Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule. Board on Health Sciences Policy and Board on Health Care Services. Washington, DC: The National Academies Press.

International Society for Biological and Environmental Repositories. (2005). Best practices for repositories I: Collection, storage, and retrieval of human biological materials for research. *Cell Preservation Technology, 3*(1), 5–48.

International Society for Biological and Environmental Repositories. (2008). 2008 best practices for repositories: Collection, storage, retrieval, and distribution of biological materials for research. *Cell Preservation Technology, 6*(1), 3–58.

Jones, N.L., and Sarata, A.K. (2008). *Genetic Information: Legal Issues Relating to Discrimination and Privacy*. CRS Report for Congress. Washington, DC: Congressional Research Service.

Kapp, M.B. (2008). Biobanking human biological materials: Issues surrounding the collection of samples for use in future research. *Pharmaceutical Medicine, 22*(2), 75–84.

Karr, A.F., Kohnen, C.N., Oganian, A., Reiter, J.P., and Sanil, A.P. (2006). A framework for evaluating the utility of data altered to protect confidentiality. *American Statistician, 60*(3), 224–232.

Levine, R.J. (1982). Commentary on the readability of consent forms. *IRB: A Review of Human Subjects Research, 4*(1), 8–9.

Levine, R.J. (1988). *Ethics and Regulations of Clinical Research* (2nd ed.). New Haven, CT: Yale University Press.

Levine, R.J. (2001). Institutional review boards: A crisis in confidence. *Annals of Internal Medicine, 134*(2), 161–163.

Levine, R.J. (2006). Empirical research to evaluate ethics committees' burdensome and perhaps unproductive policies and practices: A proposal. *Journal of Empirical Research on Human Research Ethics, 1*(3), 1–4.

Lindau, S.T., Hoffmann, J.N., Lundeen, K., Jaszczak, A., McClintock, M.K., and Jordan, J.A. (2009). Vaginal self-swab specimen collection in a home-based survey of older women: Methods and applications. *Journal of Gerontology: Social Sciences, 64B*(S1), i106–i118.

Littenberg, B., and MacLean, C.D. (2006). Passive consent for clinical research in the age of HIPAA. *Journal of General Internal Medicine, 21*(3), 207–211.

Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Malden, MA: Wiley-Interscience.

Lowrance, W.W. (2006a). *Privacy, Confidentiality, and Identifiability in Genomic Research*. Discussion document for workshop convened by the National Human Genome Research Institute, October, Bethesda, MD. Available: http://www.genome.gov/Pages/About/OD/ReportsPublications/IdentifiabilityWorkshopWhitePaper.pdf [accessed June 15, 2008].

Lowrance, W.W. (2006b). *Summary of the NHGRI Workshop on Privacy, Identifiability, and Confidentiality in Genomic Research*. Workshop convened by the National Human Genome Research Institute, October, Bethesda, MD. Available: http://www.genome.gov/19519198 [accessed June 15, 2008].

Lunshof, J.E., Chadwick, R., Vorhaus, D.B., and Church, G.M. (2008). From genetic privacy to open consent. *Nature Reviews Genetics, 9*(5), 406–411.

Malin, B., and Sweeney, L. (2004). How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems. *Journal of Biomedical Informatics, 37*(3), 179–192.

Malinowski, M.J. (2005). Taking genomics to the biobank: Access to human biological samples and medical information. *Louisiana Law Review, 66*(43), 43–67.

Marmot, M., and Steptoe, A. (2008). Whitehall II and ELSA: Integrating epidemiological and psychobiological approaches to the assessment of biological indicators. In National Research Council, *Biosocial Surveys* (pp. 42–59). Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys, M. Weinstein, J.W. Vaupel, and K.W. Wachter, Eds. Committee on Population, Division of Behavioral and Social Sciences and Education. Washington DC: The National Academies Press.

Martin, J., and Marker, D.A. (2007). Informed consent: Interpretations and practice on social surveys. *Social Science and Medicine, 65*(11), 2,260–2,271.

McGuire, A.L., Fisher, R., Cusenza, P., Hudson, K., Rothstein, M.A., McGraw, D., Matteson, S., Glaser, J., and Henley, D.E. (2008). Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: Points to consider. *Genetics in Medicine, 10*(7), 495–499.

Meslin, E.M. (2006). Shifting paradigms in health services research ethics: Consent, privacy, and the challenges for IRBs. *Journal of General Internal Medicine, 21*(3), 279–280.

Meslin, E.M., and Quaid, K.A. (2004). Ethical issues in the collection, storage, and research use of human biological materials. *Journal of Laboratory and Clinical Medicine, 144*(5), 229–234.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. Committee on Science, Engineering, and Public Policy, Policy and Global Affairs Division. Washington, DC: The National Academies Press.

National Cancer Institute. (2007). *National Cancer Institute Best Practices for Biospecimen Resources*. Washington, DC: Author.

National Human Genome Research Institute. (2008). *Cases of Genetic Discrimination*. Available: http://www.genome.gov/12513976 [accessed June 26, 2009].

National Institutes of Health. (2006). *Data Sharing Workshop for Behavioral and Social Studies that Collect Genetic Data: Workshop Highlights*. Workshop of the Behavioral and Social Research Program of the National Institute on Aging, Bethesda, MD. Available: http://www.nia.nih.gov/NR/rdonlyres/AF0997F6-0C16-4A76-96C0-D3780F00E6D4/7292/2006Aug_Genetics_Data_Sharing_Highlights_09_01_06_.PDF [accessed June 23, 2009].

National Institutes of Health. (2007). *NIH Data Sharing Policy*. Available: http://grants.nih.gov/grants/policy/data_sharing/ [accessed June 20, 2008].

National Institutes of Health. (2009). *Certificates of Confidentiality Kiosk*. Available: http://grants.nih.gov/grants/policy/coc/ [accessed June 16, 2009].

National Institutes of Health, Office for Human Research Protections, Association of American Medical Colleges, and American Society of Clinical Oncology. (2005). *Alternative Models of IRB Review: Workshop Summary Report*. Available: http://www.hhs.gov/ohrp/sachrp/documents/AltModIRB.pdf [accessed July 12, 2009].

National Research Council. (1993). *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics.* Panel on Confidentiality and Data Access, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2000). *Improving Access to and Confidentiality of Research Data: Report of a Workshop.* Committee on National Statistics, C. Mackie and N.Bradburn, Eds. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2003). *Protecting Participants and Facilitating Social and Behavioral Science Research.* Panel on Institutional Review Boards, Surveys, and Social Science Research. C.F. Citro, D.R. Iigen, and C.B. Marrett, Eds. Committee on National Statistics and Board on Behavioral, Cognitive, and Sensory Sciences. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2005). *Expanding Access to Research Data: Reconciling Risks and Opportunities*. Panel on Data Access for Research Purposes, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2007). *Putting People on the Map: Protecting Confidentiality with Linked Social–Spatial Data*. Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self-Identifying Data. M.P. Gutmann and P.C. Stern, Eds. Committee on the Human Dimensions of Global Change. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2008). *Biosocial Surveys.* Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys. M. Weinstein, J.W. Vaupel, and K.W. Wachter, Eds. Committee on Population, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nolte, M.A., and Keller, J.J. (2004). *Research Use of Restricted Data: The HRS Experience*. Paper presented at Joint Statistical Meetings 2004, August, Toronto, Canada. Available: http://www.amstat.org/Sections/Srms/Proceedings/y2004/Files/Jsm2004-000274.pdf [accessed June 24, 2009].

Nosowsky, R., and Giordano, T.J. (2006). The Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule: Implications for clinical research. *Annual Review of Medicine, 57*, 575–590.

Office for Human Research Protections. (2008). *Guidance on Research Involving Coded Private Information or Biological Specimens*. Available: http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm [accessed July 14, 2009].

Office of Biorepositories and Biospecimen Research. (2008). *Custodianship and Ownership Issues in Biospecimen Research: Symposium–Workshop, Summary*. Paper prepared by Rose Li and Associates for the National Cancer Institute.

Offit, K. (2008). Genomic profiles for disease risk: Predictive or premature? *Journal of the American Medical Association, 299*(11), 1,353–1,355.

Organisation for Economic Co-operation and Development. (2007). *OECD Best Practice Guidelines for Biological Resource Centres*. Paris: Author.

Rothstein, M.A. (2005). Expanding the ethical analysis of biobanks. *Journal of Law, Medicine, & Ethics, 33*(1), 89–101.

Rothstein, M.A. (2008). Is GINA worth the wait? *Journal of Law, Medicine & Ethics, 36*(1), 174–178.

Saunders, I.W., Brohede, J., and Hannan, G.N. (2007). Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference. *Genomics, 90*(3), 291–296.

Shickle, D. (2006). The consent problem within DNA biobanks. *Studies in History and Philosophy of Science Part C—Studies in History and Philosophy of Biological and Biomedical Sciences, 37*(3), 503–519.

Shostak, J. (2006). *De-identification of Clinical Trials Data Demystified*. Paper presented at PharmaSUG SAS conference, May, Bonita Springs, FL.

Strouse, D.S. (2005). Informed consent to genetic research on banked human tissue. *Jurimetrics Journal, 45*, 135–152.

Twitchell, K.T. (2003). Bloodborne pathogens: What you need to know—Part I. *American Association of Occupational Health Nurses Journal, 51*(1), 38–45.

UK Biobank. (2007). *UK Biobank Ethics and Governance Framework*. Available: http://www.ukbiobank.ac.uk/docs/EGFlatestJan20082.pdf [accessed June 20, 2008].

United Nations Educational, Scientific and Cultural Organization. (2002). *Human Genetic Data: Preliminary Study by the IBC on its Collection, Processing, Storage and Use*. Paris: Author. Available: http://portal.unesco.org/shs/en/files/2138/10563744931Rapfinal_gendata_en.pdf/Rapfinal_gendata_en.pdf [accessed June 16, 2009].

U.S. Department of Health and Human Services. (2000). *Standards for Privacy of Individually Identifiable Health Information*. Available: http://aspe.hhs.gov/admnsimp/final/pvcguide1.htm [accessed June 26, 2009].

U.S. Department of Health and Human Services. (2003). *Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule*. Washington, DC: Author. Available: http://privacyruleandresearch.nih.gov/pdf/HIPAA_Booklet_4-14-2003.pdf [accessed June 26, 2009].

U.S. Department of Health and Human Services. (2005a). *Are You a Covered Entity?* Available: http://www.cms.hhs.gov/HIPAAGenInfo/06_AreYouaCoveredEntity.asp [accessed July 10, 2009].

U.S. Department of Health and Human Services. (2005b). *Protection of Human Subjects*. Available: http://www.hhs.gov/ohrp/documents/OHRPRegulations.pdf [accessed June 26, 2009].

U.S. House of Representatives. (2008). *Genetic Information Nondiscrimination Act of 2008*. H.R. 493, 110th Congress. Washington, DC: U.S. Government Printing Office.

Vaught, J.B., Lockhart, N., Thiel, K.S., and Schneider, J.A. (2007). Ethical, legal, and policy issues: Dominating the biospecimen discussion. *Cancer Epidemiology, Biomarkers, & Prevention, 16*(12), 2,521–2,523.

Vollman, J., and Winau, R. (1996). Informed consent in human experimentation before the Nuremberg code. *British Medical Journal, 313*(7070), 1,445–1,449.

Wade, N. (2010). A new way to look for diseases' genetic roots. *The New York Times*. Available: http://www.nytimes.com/2010/01/26/science/26gene.html [accessed March 29, 2010].

Weir, R.F., and Olick, R.S. (2004). *The Stored Tissue Issue: Biomedical Research, Ethics, and Law in the Era of Genomic Medicine*. New York: Oxford University Press.

Wellcome Trust. (2003). *Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility*. Report of a meeting organized by the Wellcome Trust and held on January 14–15, Fort Lauderdale, FL. Available: http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf [accessed June 22, 2010].

Westin, A.F. (2007). *How the Public Views Privacy and Health Research.* Available: http://www.iom.edu/Object.File/Master/48/528/%20Westin%20IOM%20Srvy%20Rept%2011-1107.pdf [accessed November 11, 2008].

Westin, A.F. (2008, November). *Americans' Changing Concerns About Health Privacy*. Presentation at the Workshop on Collecting, Storing, Protecting, and Accessing Biological Data Collected in Social Surveys, November 17–18, National Research Council, Washington, DC.

Widom, C.S., and Brzustowicz, L.M. (2006). MAOA and the "cycle of violence": Childhood abuse and neglect, MAOA genotype, and risk for violent and antisocial behavior. *Biological Psychiatry, 60*(7), 684–689.

Wolf, L.E., Croughan, M., and Lo, B. (2002). The challenges of IRB review and human subjects protections in practice-based research. *Medical Care, 40*(6), 521–529.

World Health Organization. (1998). *Proposed International Guidelines on Ethical Issues in Medical Genetics and Genetic Services*. Report of a WHO meeting on ethical issues in medical genetics. Geneva: Author. Available: http://www.who.int/genomics/publications/en/ethicalguidelines1998.pdf [accessed June 16, 2009].

World Medical Association. (1964). *Declaration of Helsinki: Recommendations Guiding Physicians in Biomedical Research Involving Human Subjects*. Available: http://www.cirp.org/library/ethics/helsinki/ [accessed June 17, 2009].

Zimmern, R.L., and Kroese, M. (2007). The evaluation of genetic tests. *Journal of Public Health, 29*(3), 246–250.

Appendix A

Agenda for the Workshop on Collecting, Storing, Protecting, and Accessing Biological Data Collected in Social Surveys

November 17–19, 2008

The Keck Center of the National Academies
500 Fifth Street NW
Washington, DC

November 17, 2008—Room 100

8:30–8:45 am — Welcome and Opening Remarks
NRC representative
Bob Hauser, Committee chair
Richard Suzman, NIA
- Background and purpose
- Goals of the workshop
- Sponsor perspective

SESSION 1: Overview
Session Chair: Bob Hauser, *University of Wisconsin*

8:45–9:05 — Data Access Versus Confidentiality: Balancing Risks and Benefits
Ellen Wright Clayton, *Vanderbilt University*

9:05–9:25 — Where Are We Now? Where Do We Want to Go?
George Church, *Harvard University*

9:25–9:45 — Overview of Biobanks and Data Harmonization
Jennifer Harris, *The Norwegian Institute of Public Health, Oslo, Consultant to NIA, NIH*

9:45–10:45	Discussion
10:45–11:00	**Break**
	SESSION 2: Risks and Rewards of Data Linking Session Chair: Hank Greely, *Stanford University*
11:00–11:20	Risks and Rewards Katherine Mullan Harris, *University of North Carolina*
11:20–11:40	Legal and Ethical Issues in Using Biological Data and How They Differ from Social Science Data Barbara Koenig, *Mayo Clinic, University of Minnesota*
11:40–12:30 pm	Discussion
12:30–1:30	**Lunch**
	SESSION 3: Storing and Maintaining Biological Data Session Chair: Maxine Weinstein, *Georgetown University*
1:30–1:45	Case Study 1—PENN Biomarker Core of the Alzheimer's Disease Neuroimaging Initiative Leslie M. Shaw, *University of Pennsylvania*
1:45–2:00	Case Study 2—SWAN: Current Practices of Receipt and Storage of Biospecimens Kathi Shea, *SeraCare Life Sciences, Inc.*
2:00–2:15	Case Study 3—Laws and Regulations on Biobanks: Present Status and Future Directions Bartha Knoppers, *University of Montreal*
2:15–3:15	Discussion
3:15–3:30	**Break**
	SESSION 4: Informed Consent Session Chair: Barbara Stanley, *Columbia University*
3:30–4:00	Informed Consent: Best Practices Holly Taylor, *Johns Hopkins University*

4:00–4:15	Informed Consent: Recent Developments in Legal and Ethical Requirements for Data Collection and Use Paul Appelbaum, *Columbia University*
4:15–4:30	Research with Biological Data Collected in Social Surveys: The Role of IRBs in Informed Consent Karen Maschke, *The Hastings Center*
4:30–5:30	Discussion
6:30–8:30	**Working Dinner**

November 18, 2008—Room 100

	SESSION 5: Protecting Data and Confidentiality Session Chair: Bob Wallace, *University of Iowa*
8:30–8:50 am	Americans' Changing Concerns About Health Privacy Alan Westin, *Columbia University (Emeritus)*
8:50–9:10	Quantifying Disclosure Risks Jerry Reiter, *Duke University*
9:10–9:30	A Privacy Preserving Framework for Integrating, Storing, and Querying Biological Data Murat Kantarcioglu, *University of Texas*
9:30–9:50	Assessing the Utility of Statistical Methods for Limiting Disclosure Risk: Value of Synthetic Data Sets John Abowd, *Cornell University*
9:50–11:00	Discussion
11:00–11:15	**Break**
	SESSION 6: Accessing and Sharing Data Session Chair: Myron Gutmann, *University of Michigan*
11:15–11:35	Data Licensing Agreements/Restricted Access Alan Karr, *National Institute of Statistical Sciences*

11:35–11:55	Repository Issues and Practices MaryFran Sowers, *University of Michigan*
11:55–12:45 pm	Discussion
12:45–1:45	**Lunch**
1:45–3:00	SESSION 7: Collecting Our Thoughts Session Chair: Bob Hauser, *University of Wisconsin* Format: Brief comments from the chair followed by a period of general discussion. • What have we learned? • Remaining questions • Conclusions • Future directions
3:00 pm	**Public Workshop Adjourn**
3:15–5:30	CLOSED SESSION (COMMITTEE MEMBERS ONLY)—Room 213 • Report preparation
6:30–8:30	**Working Dinner**

November 19, 2008

9:00–5:30 pm	CLOSED SESSION (COMMITTEE MEMBERS ONLY)—Room 208 • Report preparation (cont.)

Appendix B

Participants in the Workshop on Collecting, Storing, Protecting, and Accessing Biological Data Collected in Social Surveys

NOVEMBER 17–18, 2008

John Abowd, *Cornell University*
Paul S. Appelbaum, *Columbia University*
Timothy J. Beebe, *Mayo College of Medicine*
Partha Bhattacharyya, *National Institute on Aging*
Stephen J. Blumberg, *Centers for Disease Control and Prevention*
Laura Branden, *Westat*
Michelle Brotzman, *Westat*
Somnath Chatterji, *World Health Organization*
George M. Church, *Harvard Medical School*
Ellen Wright Clayton, *Vanderbilt University*
George T. Duncan, *Carnegie Mellon University*
Henry T. Greely, *Stanford University*
Myron P. Gutmann, *University of Michigan*
John Haaga, *National Institute on Aging*
Elizabeth Hamilton, *National Institute on Aging*
Jennifer Harris, *Norwegian Institute of Public Health, Oslo, and Consultant to the National Institute on Aging*
Kathie Mullan Harris, *University of North Carolina*
Robert M. Hauser, *University of Wisconsin*
John Milburn Jessup, *National Cancer Institute*
Clifford Johnson, *Centers for Disease Control and Prevention*
Murat Kantarcioglu, *University of Texas*
Alan F. Karr, *National Institute of Statistical Sciences*

Jonathan King, *National Institute on Aging*
Bartha M. Knoppers, *University of Montreal*
Barbara A. Koenig, *Mayo College of Medicine*
Laura Levit, *Institute of Medicine*
Stacy Tessler Lindau, *University of Chicago*
Jennifer H. Madans, *Centers for Disease Control and Prevention*
Karen J. Maschke, *Hastings Center*
James W. McNally, *University of Michigan*
Geraldine M. McQuillan, *Centers for Disease Control and Prevention*
Sharyl Nass, *Institute of Medicine*
Lis Nielsen, *National Institute on Aging*
Randall J. Olsen, *Ohio State University*
Georgeanne Patmios, *National Institute on Aging*
John Phillips, *National Institute on Aging*
Jerome P. Reiter, *Duke University*
Jeanne Rosenthal, *Westat*
Jane Schulman, *Westat*
Leslie Shaw, *University of Pennsylvania*
Kathi Shea, *SeraCare, Inc.*
Sherry Sherman, *National Institute on Aging*
Janet M. Eisenhauer Smith, *University of Wisconsin, Madison*
Mary Fran Sowers, *University of Michigan*
Erica Spotts, *National Institute on Aging*
Barbara Stanley, *Columbia University*
Richard Suzman, *National Institute on Aging*
Holly Taylor, *Johns Hopkins University*
Melissa Thomas, *Mathematica Policy Research, Inc.*
Arti Varanasi, *Westat*
Ulyana Vjugina, *American Society of Hematology*
Robert B. Wallace, *University of Iowa*
Maxine Weinstein, *Georgetown University*
Alan Westin, *Columbia University (emeritus)*
Louise Wideroff, *National Cancer Institute*
Gooloo Wunderlich, *Division of Behavioral and Social Sciences and Education*

Appendix C

Biographical Sketches of Panel Members and Staff

Robert M. Hauser (*Chair*) is Vilas Research professor of sociology and director, Center for Demography of Health and Aging at the University of Wisconsin, Madison. His research in sociology, statistics, and demography has addressed aging and the life course, social and economic inequality, educational attainment, and cross-national and cross-temporal comparisons of intergenerational mobility. Since 1980, he has led the Wisconsin Longitudinal Study (WLS), which has followed the life course of more than 10,000 Wisconsin high school graduates. Dr. Hauser is a fellow of the American Statistical Association and a member of the National Academy of Sciences. He has served on numerous National Research Council (NRC) committees, including the Panel on Institutional Review Boards, Surveys, and Social Science Research; the Panel to Review the 2000 Census; and the Committee on Performance Levels for Adult Literacy. He holds a B.A. in economics from the University of Chicago and an M.A. and Ph.D., both in sociology, from the University of Michigan.

George M. Church is professor of genetics at Harvard Medical School and director of the Center for Computational Genetics. With degrees from Duke University in chemistry and zoology, he coauthored research on 3-D software and RNA structure with Sung-Hou Kim. His work on his Ph.D. from Harvard in biochemistry and molecular biology with Wally Gilbert included the first direct genomic sequencing method in 1984. Dr. Church initiated the Human Genome Project as a research scientist at newly formed Biogen Inc. and a Monsanto life sciences research fellow at the University of California, San Francisco, with Gail Martin. He invented the broadly applied concepts

of molecular multiplexing and tags, homologous recombination methods, and array DNA synthesizers. Technology transfer of automated sequencing and software to Genome Therapeutics Corp. resulted in the first commercial genome sequence (the human pathogen *H. pylori*, 1994). Dr. Church has served in advisory roles for 12 journals (including *Nature Molecular Systems Biology*), five granting agencies, and 24 biotech companies (founding Knome, Joule, and LS9). His current research focuses on integrating biosystems modeling with the Personal Genome Project and synthetic biology.

Barney Cohen (*Study Director*) is director of the Committee on Population of the National Academies/NRC. His work at the NRC has encompassed a wide variety of domestic and international projects, including studies on fertility, morbidity, mortality, housing, urbanization, migration, aging, and HIV/AIDS. Currently, he is also serving as the liaison of the National Academies to the Academy of Science of South Africa and the Ghanaian Academy of Arts and Sciences as part of a larger project aimed at supporting the development of academies of science in Africa. Dr. Cohen holds an M.A. in economics from the University of Delaware and a Ph.D. in demography from the University of California, Berkeley.

George T. Duncan joined the Carnegie Mellon University faculty in the Department of Statistics in 1974 and the Heinz College faculty in 1978. He became professor emeritus in 2008. He has served as director of the Heinz College's M.S., M.P.M., and Ph.D. programs. He served as associate dean for faculty from 2001 to 2002. Prior to coming to Carnegie Mellon, Dr. Duncan taught in the mathematics department at the University of California, Davis. He is a visiting faculty member at Los Alamos National Laboratory, has been a visitor at Cambridge University, and was Lord Simon visiting professor at the University of Manchester in 2005. He is a fellow of the American Statistical Association, a fellow of the American Association for the Advancement of Science, a fellow of the Royal Statistical Society, and an elected member of the International Statistical Institute. Dr. Duncan's general research interests are in Bayesian decision making and information technology and social accountability. His primary focus is on confidentiality of statistical databases. His work has appeared in leading journals, including the *Journal of the American Statistical Association*, *Management Science*, *Econometrica*, *Operations Research*, *Psychometrika*, and *Biometrika*. He holds a B.S. and an M.S. in statistics from the University of Chicago and a Ph.D. in statistics from the University of Minnesota.

Henry T. Greely is Deane F. and Kate Edelman Johnson professor of law in Stanford University's Law School. He also holds an appointment (by courtesy) with the Stanford University Department of Genetics. Professor Greely specializes in the legal implications of new biomedical technologies, especially those

related to genetics, neurosciences, and stem cells. He frequently serves as an advisor on California, national, and international policy issues. He chairs the California Advisory Committee on Human Embryonic Stem Cell Research, is a codirector of the Law and Neuroscience Project, and is a founder and executive committee member of the Neuroethics Society. Active in university leadership, Professor Greely chairs the steering committee for the Stanford Center for Biomedical Ethics, directs the law school's Center for Law and the Biosciences and the Stanford Center for Biomedical Ethics' Program on Stem Cells in Society, and serves on the leadership council for the university's interdisciplinary Bio-X Program. He is a fellow of the American Association for the Advancement of Science. He holds a J.D. from Yale Law School.

Myron P. Gutmann is assistant director of the National Science Foundation, with responsibility for the Directorate for the Social, Behavioral, and Economic Sciences. He is also research professor at the Inter-university Consortium for Political and Social Research (ICPSR) and professor of history and information, University of Michigan. From 2001 to 2009 he served as director of ICPSR. He has a broad range of interests in interdisciplinary historical population studies, especially relating population to agriculture, the environment, and health. He also studies ways in which digital materials can be properly preserved and shared and how the confidentiality of research subjects can be protected when data about them is made available for secondary use. Dr. Gutmann teaches about historical demography and about the social, demographic, and economic history of Europe and the Americas. He holds an M.A. and a Ph.D. from Princeton University. He has served on a number of national and international advisory committees, including the U.S. Committee for CODATA and the National Academies' Board on Research Data and Information. Dr. Gutmann's work on this panel was completed before he assumed his position at the National Science Foundation.

Robert J. Levine is professor of medicine and lecturer in pharmacology; director of the Law, Policy and Ethics Core of the Center for Interdisciplinary Research on AIDS; and senior fellow in bioethics at Yale University. He is a fellow of the Hastings Center and the American College of Physicians; a member of the American Society for Clinical Investigation and American Society for Pharmacology and Experimental Therapeutics; a director and former vice chair of PRIM&R (Public Responsibility in Medicine and Research); past president of the American Society of Law, Medicine and Ethics (two terms); and past chairman of the Connecticut Humanities Council. In the past he was also chair of the Institutional Review Board at Yale-New Haven Medical Center (1969–2000), founding codirector of Yale University's Interdisciplinary Bioethics Center, chief of the Section of Clinical Pharmacology at Yale, chairman of the Section on Medico-Legal Matters and R&D Administration of the American Society

for Clinical Pharmacology and Therapeutics, Associate Editor *of Biochemical Pharmacology*, and editor of *Clinical Research*. Dr. Levine is the founding editor of *IRB: A Review of Human Subjects Research* (editor 1979–2000 and currently chair of the Editorial Board) and has served as consultant to several federal and international agencies involved in the development of policy for the protection of human subjects (including twice serving as chair of the Council for International Organizations of Medical Sciences Steering Committee to revise its *International Ethical Guidelines for Biomedical Research Involving Human Subjects*). He is the author of numerous publications, including the book *Ethics and Regulation of Clinical Research* (two editions). In the last 35 years, most of Dr. Levine's research, teaching, and publications have been in the field of medical ethics, with particular focus on the ethics of research involving human subjects.

John Quackenbush is professor of computational biology and bioinformatics with appointments in the Department of Biostatistics, Harvard School of Public Health, and the Dana-Farber Cancer Institute. His work focuses on the challenges of how best to collect, manage, and analyze genomics data, with an emphasis on methods, spanning the laboratory to the laptop, for using genomic and computational approaches to reveal the underlying biology. Recently he has been looking at patterns of gene expression in cancer with the goal of elucidating the networks and pathways that are fundamental in the development and progression of the disease. He holds a Ph.D. in theoretical particle physics from the University of California, Los Angeles.

Jerome P. Reiter is associate professor of statistical science, Department of Statistical Science, Duke University. His methodological research focuses mainly on ways of protecting confidentiality in public-use data, handling missing data in large surveys, and drawing causal inferences in observational data. Dr. Reiter is chair of the Privacy and Confidentiality Committee of the American Statistical Association. He is associate editor of several publications, including the *Journal of Privacy and Confidentiality*, the *Journal of the American Statistical Association*, and *Survey Methodology*. He holds a B.S. from Duke University and a Ph.D. in statistics from Harvard University.

Robert B. Wallace is professor of epidemiology, Department of Epidemiology, and director of the Center on Aging at the University of Iowa. His current research interests include the epidemiology of aging, cancer epidemiology and control, and survey research. He presently is leading the Women's Health Initiative in Iowa, and he was recently named chair of the Board on Select Populations of the Institute of Medicine. He is a member of the Institute of Medicine and has served on several NRC committees. Dr. Wallace holds an M.D. from

Northwestern University and an M.Sc. in epidemiology from the State University of New York, Buffalo.

Maxine Weinstein joined the faculty of Georgetown University in 1987; she holds an appointment in the Graduate School of Arts and Sciences as distinguished professor of population and health. Her work explores the behavioral and biological dimensions of reproduction and aging. She is an investigator on the Taiwan project, a study that explores the reciprocal relations among stress, health, and the social environment among the elderly. She is also an investigator on the MIDUS II study. Dr. Weinstein holds a B.S. from Antioch College and a Ph.D. from Princeton University.

Appendix D

Acronyms

AAMC	Association of American Medical Colleges
ASCO	American Society of Clinical Oncology
BNSF	Burlington Northern Santa Fe Railroad
BSR	Behavioral and Social Research Program
CFR	Code of Federal Regulations
CLIA	Clinical Laboratory Improvement Amendments
COI	conflict of interest
DASL	Data and Story Library
EEOC	Equal Employment Opportunity Commission
ELSA	English Longitudinal Study of Aging
FFPE	formalin-fixed, paraffin-embedded tissue
FOIA	Freedom of Information Act
GINA	Genetic Information Nondiscrimination Act
GWASs	genome-wide association studies
HDL	high-density lipoprotein
HHS	U.S. Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act

HRS	Health and Retirement Study
ICPSR	Inter-university Consortium for Political and Social Research
IOM	Institute of Medicine
IP	intellectual property
IRB	Institutional Review Board
ISBER	International Society for Biological and Environmental Repositories
ISO	International Organization for Standardization
LBL	Lawrence Berkeley Laboratory
MiCDA	Michigan Center on the Demography of Aging Data
MTA	material transfer agreement
NCES	National Center for Education Statistics
NCHS	National Center for Health Statistics
NCI	National Cancer Institute
NHANES	National Health and Nutrition Examination Survey
NHGRI	National Human Genome Research Institute
NIA	National Institute on Aging
NICHD	Eunice Kennedy Shriver National Institute of Child Health and Human Development
NIH	National Institutes of Health
NORC	National Opinion Research Center
NRC	National Research Council
OBBR	Office of Biorepositories and Biospecimen Research
OECD	Organisation for Economic Co-operation and Development
OHRP	Office of Human Research Protections (formerly OPRR)
OPRR	Office of Protection from Research Risks
QA	quality assurance
QC	quality control
SDL	statistical disclosure limitation
SNP	single nucleotide polymorphism
SOP	standard operating procedure
SWAN	Study of Women's Health Across the Nation
UNESCO	United Nations Education, Scientific and Cultural Organization

UPS	uninterruptible power supply
VCU	Virginia Commonwealth University
WHO	World Health Organization

COMMITTEE ON NATIONAL STATISTICS

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

COMMITTEE ON POPULATION

The Committee on Population was established by the National Research Council in 1983 to bring the knowledge and methods of the population sciences to bear on major issues of science and public policy. The committee's work includes basic studies of fertility, health and mortality, and migration aimed at improving programs for the public health and welfare in the United States and developing countries. The committee also fosters communication among researchers in different disciplines and countries and policy makers in government, international agencies, and private organizations. The work of the committee is made possible by funding from several government agencies and private foundations.