1

Collecting Biological Indicators in Household Surveys

Caleb E. Finch and James W. Vaupel

In 1996, the Committee on Population of the National Research Council sponsored a novel workshop on the biodemography of aging, which resulted in the volume *Between Zeus and the Salmon*: *The Biodemography of Longevity* (National Research Council, 1997). The workshop and its report, which considered the continuing increases in human life span in a broad biological context, launched a new phase in studies of human aging at the population level. To no one's surprise, a mere four years later, we still lack basic understanding of why human life expectancies continue to increase at roughly similar rates around the world, despite huge differences in the patterns of disease and lifestyle.

The present volume, and its antecedent workshop in February 2000, was organized to pursue certain questions raised by *Zeus and the Salmon*. In particular we wanted to examine the issues and prospects for collecting biological data from individuals in household surveys. This volume, with its rich collection of essays, is a guide or a handbook to many emerging and portentous questions. The basic question remains open: Should large population-based household surveys consider instituting the collection of biological material (e.g., blood or urine), physiological measurements (e.g., blood pressure or handgrip strength), and environmental measurements (e.g., cadmium exposure or radon levels) in addition to the usual demographic, socioeconomic, and/or health data? This information would soon be integrated with the huge database from household surveys that describes many social, economic, educational, and even behavioral data on millions of individuals in the United States.

In the ensemble, the essays cover the questions of what, why, whether, who, and how. What kinds of biological materials, physiological measurements, and environmental measurements could and should be included in household surveys? Which researchers would be interested in this kind of information? Why should we or shouldn't we collect these data—what are the costs and disadvantages? What kind of personnel is needed for these new goals, with what kind of background and training? It is far from clear how to successfully include biological specimens and physiological and environmental measurements in household surveys. Collecting this information poses ethical concerns about privacy, confidentiality, and potential consequences for individuals and social groups.

The chapters are written by researchers and scholars with very differing perspectives. The topics addressed are diverse, as are the styles and structures of the essays. In exploring these new themes, each essay tackles multiple topics and sheds light on myriad questions. Several key issues are addressed in more than one essay, offering insights from different vantage points. The topics addressed in this volume are so new, and knowledge is evolving so rapidly, that it would be neither possible nor appropriate to expect the workshop to have produced a well-ordered set of concrete directions. Instead, readers will find much resonance across the diverse experience, values, and projections of the authors.

Household surveys have become a major research industry for demographers, economists, and other social and behavioral scientists. A prime example is the Health and Retirement Study (HRS), which is described and discussed in the chapter by Weinstein and Willis (see also Burkhauser and Gertler, 1995). This longitudinal study began in 1992 with a survey of 12,600 persons aged 51 to 61. In 1998 more than 22,000 persons were interviewed. Three further waves of the HRS are planned for 2000, 2002, and 2004. Besides this large, important study are scores of other studies of considerable size and significance. In the United States alone, there are at least ten other major population-based surveys with a focus on aging funded in full or in part by a single agency, the Behavioral and Social Science program of the National Institute on Aging (Wallace, 1997). Among other major surveys with an emphasis on specific age groups or the entire age range are the U.S. Panel Study of Income Dynamics and the German Socio-Economic Panel.

WHICH INDICATORS?

What kind of bioindicators can be gathered in household surveys? The chapter by Wallace is valuable in providing experienced, judicious guidance about practical issues. Weinstein and Willis, Christensen, and Crimmins and Seeman give helpful practical information in particular

contexts for incorporating bioindicators into surveys. The chapters by Christensen and Ewbank focus on DNA, which can be sampled in many ways, including from blood samples, blood dry spots obtained by the prick of a finger, cheek swabs, hair follicles, and urine. It is portentous that usable amounts of DNA might be gathered from the minute numbers of cells left unwittingly when a stamp is licked, or even when an envelope is handled during its return to an agency. Even if DNA is not available, studies of twins, siblings, and other relatives and studies of adopted children can be used to shed some light on genetic factors and their interaction with specific environmental factors. Several chapters, especially the one by Vogler, consider gene-environment issues.

Other authors focus on various physiological measurements. Wachter considers the evocative example of height as a very simple physiological measure. Halter and Reuben survey a wide range of function indicators. Indeed, they consider measures of functioning from the levels of molecules, cells, and organs up to the level of the whole organism. Crimmins and Seeman also consider an array of measurements about function, with emphasis on measures that capture aspects of the cumulative stress—the allostatic load—an organism has suffered. Several chapters mention performance tests. Lung capacity, for instance, can be simply measured by asking respondents to blow into a spirometer. Handgrip strength can be measured. Respondents can be asked to get up from a chair without using their arms. Performance testing is common in household surveys when information is needed on cognitive functioning: respondents are quizzed and their mental status assessed on the basis of how well they can remember and respond.

In some household surveys, administrative data from official records can be linked to respondents. The Health and Retirement Study obtains Social Security earnings and benefit histories from the Social Security Administration. Medical costs and diagnoses are obtained from Medicare records. In Denmark, a broad array of information can be garnered from administrative records. This means that considerable vital and personal information is on reserve for 100 percent of the population surveyed, even those who refuse to be interviewed.

Environmental measures that might be obtained in household surveys are considered by Wallace in his thought-provoking discussion. Here we briefly note some remarkable technical developments in miniaturized sensors that will soon enable characterization of local environments in unprecedented detail. In general, these devices are referred to as microelectromechanical systems, or MEMS. Miniaturized sensors on the scale of microns or even less (nanoscale) that can reside unobtrusively in a household are being developed for specific environmental parameters. Arrays of sensors (“electronic noses”), for example, can sample the local

atmosphere continuously for trace metals or specific organic toxins (Kovacs, 1998; Gardner and Bartlett, 1999). Even airborne microorganisms can be sampled and their DNA characterized through enzymatic amplification of DNA by polymerase chain reactions. Forthcoming generations of gene machines will be miniaturized, with mixing chambers, valves, and pumps as small as several microns. Nanoscale MEMS could be implanted in human volunteers, giving real-time correlations between environmental factors and their long-term health effects. Wireless technology will soon allow the possibility of collecting the most intimate information on body functions and activities, transmitted from within each household-based MEMS on a community basis. This technology could reveal the widely sought basis for the wide local variations in the incidence of cancers and vascular disease.

WHAT BENEFITS?

Genes, siblings, height, handgrip, air pollution, autopsies: this summary list conveys the broad range of bioindicators that can be added to household surveys. But why would a researcher want this kind of information? The chapters that describe the information that can be gathered also consider the reasons for gathering this information. For instance, Ewbank gives a penetrating discussion of the kind of genetic information that might be of use to a demographer. Vogler explains why data on siblings can shed valuable light on both genetic and environmental factors that determine health and behavior. Chapters by Christensen, Crimmins and Seeman, and Weinstein and Willis illustrate what can be learned from surveys that include genetic, physiological, and environmental information. In particular, Weinstein and Willis differentiate several themes: (1) obtaining population-representative data from nonclinical samples; (2) calibrating self-reports with other measures of health and disease; (3) explicating pathways and elaborating causal linkages between social environment and health; and (4) linking genetic markers with survey materials. The three chapters by Martin and Hu, McClearn, and Miller provide an accessible introduction to the needed conceptual framework. Wachter argues persuasively that social scientists should more actively consider bioindicators in tempering overly enthusiastic interpretations of genetic information that sometimes lead to crude biological determinism.

Autopsies are increasingly rare, and Martin and Hu make a cogent plea for their great value. An important precedent is the Nun Study, directed by David Snowdon at the University of Kentucky, which is yielding major insights into early indicators of Alzheimer disease (see, for example, Riley et al., 2000). Most of the Catholic sisters in this study have agreed to allow their brains to be autopsied. In general, funds for post-

mortem studies are difficult to obtain through the peer review process. It might be possible to ask survey respondents to allow autopsies at their demise as well as preservation of certain organs.

As the genome juggernaut grinds on, data sets with information on both an individual's genes and his or her physiology and environment will become more and more valuable. Soon we will know the location of all the human genes, which may tally 100,000, from which are transcribed even more types of messenger RNA. Rapidly advancing technology will then allow comprehensive analysis of genetic variations. We expect a huge number of individual gene differences, because each person's DNA code differs at intervals of about 100 to 1000 bases.

But getting to know the DNA variations, gene by gene, in different human populations will confront us with the next huge step in human genetics: the need to identify the *functional* significance of gene variations to the individual gene carrier. It is likely that most gene differences are neutral, with little-to-no tangible impact during development or aging. However, as we understand more about gene architecture, certain DNA variations may be predicted to be sensitive to the external environment. Present discussions of gene-environment interactions that alter aging are fundamental to the central problem in human biology during the 21st century: *to identify environmental factors that evoke harmful traits from particular sets of genes*.

We anticipate that ascertaining the adverse gene-environment interactions during development may be much easier than determining which have adverse impacts on health at later ages. Progress in these hugely complex problems of gene-environment interactions will synergize with the new field of functional genomics, which is addressing the functions of the huge number of new genes being discovered. With the fully detailed human gene map soon to be at hand, we may now consider the far more complex problem of *environmental maps* that will be needed to optimize individual health throughout the life course.

In the most general terms, three types of environmental factors can influence human health during aging: physical, chemical, and biological. Physical factors include temperature and solar radiation. Chemical factors from natural and biological sources include trace toxins (asbestos, lead, tobacco smoke), but also trace morphogens that can cause subtle abnormalities in development. Biological factors include diet and infectious organisms, but also stress from social interactions. We know little about the concentrations of a vast number of bioactive substances that may be present sporadically in the environment. It seems fair to say that our concept of the environment will evolve rapidly with new technical developments and may come to include multigenerational effects. For example, in the case of diabetes, the maternal physiological state existing before

pregnancy can influence fetal growth. Moreover, the ovary acquires its full stock of eggs in the fetus: thus, the egg cell from which all of our cells stem was exposed to the environment of our maternal grandmother (Finch and Loehlin, 1998). The depth of the transgenerational environment is a completely obscure aspect of human experience.

The huge number of variables being considered in this discussion calls for new statistical approaches to integrate all of these different kinds of parameters. Wachter discusses strategies of dimensionality reduction and new statistical models that consider whole sequences of life-course events or experiences as predictor variables, in place of the one-by-one predictor variables familiar in linear regression.

WHAT COSTS?

Given all the cogent reasons to add bioindicators to household surveys, what are the counter-arguments, what are the costs, the drawbacks, the disadvantages? In their insightful discussion, Weinstein and Willis consider three major categories of cost: (1) respondent burden, (2) financial and logistic constraints, and (3) the potential to compromise research objectives. The burden on respondents can be heavy, involving many hours of physical testing or uncomfortable invasions such as those required to draw blood. As a result, some people may refuse to participate in part or all of a survey. Additionally, some respondents who endured one survey round may decide not to participate in the next, which can be very disruptive to longitudinal analysis. Surveys in Denmark and in Taiwan suggest that respondent dropout may be modest. This concern, however, is so large that survey researchers may be wise to pilot test the collection of bioindicators on a subset of subjects before they risk challenging their entire population with the procedures. Moreover, bioindicator sampling can be expensive and logistically difficult, as conveyed by Wallace. Finally, research objectives may be compromised by some kinds of bioindicators if, as Weinstein and Willis put it, "the research process itself affects behaviors that we wish to study." In particular, providing respondents with information about their health and about their genetic risks may cause them to alter their behavior.

Throughout this volume lurk a number of ominous questions regarding ethical and legal issues that confront biological data collection, and data collection and use more generally. As one obvious example, the outfitting of households or people with arrays of wireless sensors raises enormous concerns about loss of privacy and the dissemination of lifestyle information. In addition to understanding how conceptually to add bioindicators to household surveys, researchers need to understand how ethically to do so.

Theoretical and practical questions about the ethics of biodemography and genetic research are changing as rapidly as the research methodologies themselves, if not more so. Seminars, workshops, and entire conferences are devoted to discussions of these multifaceted concerns, and the latter could easily constitute the focus of an entire volume. The chapters by Botkin and by Durfy provide practical, judicious guidance to the complicated ethical issues that face survey researchers who attempt to incorporate bioindicators. Botkin focuses on key factors in the area of informed consent. Durfy probes issues associated with the nature of, access to, and ownership of genetic information, and considers potential psychological and group-harm risks that may accrue to participation in research studies.

There are deep connections between demography as a discipline and both the broader social sciences and the biological sciences. The branch of demography known as biodemography continues to grow rapidly. Demographers are involved in many of the major household surveys, in part because of their training in statistics and their knowledge of concepts useful in studying large populations. It seems natural that demographers will be among the first who are able to design, run, and analyze surveys with bioindicators. Many other fields are likely to join these efforts. Although not strongly represented here, economists have begun to show an interest in biology and may be receptive to broadening their knowledge in ways that would help incorporate bioindicators into household surveys. Economists assume that people make decisions based on their preferences. Economists are now giving some thought to where preferences come from, including the childhood environment.

Lastly, we point out the need to consider training. If bioindicators are to be included in household surveys, then new kinds of personnel will be needed to conduct such surveys and to analyze the data collected. Training of field workers will be important in the success of adding bioindicators to field studies. Special skills are needed to explain the significance of complex tests for environmental factors or for particular gene variations. The shortage of a new generation of well-trained autopsy pathologists is also of great concern. Detailed postmortem histopathology is needed to adequately characterize morbid conditions, which are *always at some level the outcome of gene-environment interactions.* An expanded mindset, a broader vision, and enhanced biological thinking are needed for successful incorporation of bioindicators into household surveys. On a larger front, we state the obvious: A new approach to transdisiplinary training programs is needed to prepare future generations of scientists for charting the human life course on the emerging gene-environment maps.

REFERENCES

Burkhauser, R.V., and P.J. Gertler
1995 The health and retirement study. Data quality and early results. *The Journal of Human Resources*. 30, Supplement 1995.

Finch, C.E., and J. Loehlin
1998 Environmental influences that may precede fertilization: A first examination of the prezygotic hypothesis from maternal age influences on twins. *Behavior Genetics* 28:101-106.

Gardner, J.W., and P.N. Bartlett
1999 *Electronic Noses*. New York: Oxford University Press.

Kovacs, G.T.A.
1998 *Micromachined Transducers Sourcebook*. Boston: McGraw-Hill.

National Research Council
1997 *Between Zeus and the Salmon: The Biodemography of Longevity*. Committee on Population. K.W. Wachter and C.E. Finch, eds. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Riley, K.P., D.A. Snowdon, A.M. Saunders, A.D. Roses, J.A. Mortimer, and N. Nanayakkara
2000 Cognitive function and Apolipoprotein-E in the very old: Findings from the Nun Study. *Journal of Gerontology: Social Sciences* 55B(2):S69-S75.

Wallace, R.B.
1997 The potential of population surveys for genetic studies. Pp. 234-244 in *Between Zeus and the Salmon: The Biodemography of Longevity*. Committee on Population. K.W. Wachter and C.E. Finch, eds. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.