Is it a Biomarker, a Biomeasure, or a Bioindicator? – Research Models to Organize Biomedical Data for Demographic Analysis

Submitted for consideration as part of the Meetings of the 2008 European Population Conference.

James W. McNally, PhD
NACDA, Program on Aging
National Archive of Computerized Data in Aging
Institute for Social Research
University of Michigan
330 Perry Building
734-615-9520
jmcnally@umich.edu

Martha Sayre, BS
NACDA, Program on Aging
National Archive of Computerized Data in Aging
Institute for Social Research
University of Michigan
330 Perry Building
734-615-9516
msayre@umich.edu

Introduction

The growing recognition of the importance of collecting biological and medical information along with traditional demographic and socioeconomic indicators has broadened our understanding of the relationships of biological processes on social outcomes. Many surveys have routinely collected basic variables on health such as Body Mass Index (BMI) but now we are seeing an increase in the collection of more sophisticated measures such as those derived from blood spots, plasma draws and vaginal swabs. Established projects with a biomedical focus such as the Framingham Study, the Woman's Health Initiative (WHI), Social Environment and Biomarkers of Aging Study (SEBAS), National Health and Nutrition Examination (NHANES) and Study of Women's Health Across the Nation (SWAN) routinely collect biomarker data because of their specific research focus. More recently the collection of biomarker information as part of socio-economic studies such as the Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA) reflects the acceptance of this approach within mainstream social survey research. The growth in both the collection of biomarker information and the development of analytic models that incorporate biological and social information argues that the science will benefit from the systematic organization of information on biomarker data into a searchable registry. By identifying datasets and

specific variables that contains biomarker information and making this information available to the community we can encourage the use of this data for research.

There has been an explosion in both the collection and use of biomarker information in social research during the past decade (Seplaki et al., 2004 and 2006; Goldman et al, 2003; Ottenbacher et al., 2002, Newell et al., 2001; Brown et al., 2001; Sempos et al., 2000). This increase reflects both the growing recognition of the important of understanding the relationships between biomedical measures of health and individual outcomes and the growth of multidisciplinary research frameworks that blend social and medical paradigms of knowledge. Unfortunately, the systematic organization of biomarker information into an effective knowledge base has lagged far behind its collection. NACDA has identified publicly available data with biomedical information data going back as far as the 1700's and earlier including studies such as the such as Fogel's collection of Surgeons' Certificates for Aging Veterans of the Union Army (Fogel, 2002) and the Descriptors and Measurements of the Height of Runaway Slaves and Indentured Servants in the United States for 1700 to 1850 by Komlos (2001). The work bridging biomedical and social research has expanded our understanding of issues of health, longevity and quality of life, but organizing the universe of data underlying this growing body of literature in a manner that facilitates the use of biomarker data as a resource for secondary analysis is an essential step still unaddressed. In response to the growing need for guidance, a number of recent initiatives have begun the process of establishing centralized resources for researchers and analysts interested in biomedical and biosocial research.

More and more research organizations that focus on the intersections of biomedical research are being funded by federal agencies such as the National Institute on Aging, and then now play a prominent role in the development of research agendas. Groups based in the United States such as the Center on Biodemography and Population Health (CBPH) at USC/UCLA and the Chicago Core for Biomarkers in Population-Based Aging Research (CCBAR) at the University of Chicago, and associations such as the "Network on Measurement of Biological Risk" all play a role in generating interest and education among demographers, economists and other social scientists interested in performing the best science when using biomedical indicators in research.

This paper reviews this research advances and attempts to enhance these efforts by introducing the unique focus on data management offered by the NACDA Program on Aging based at the University of Michigan. Due to our research mission that emphasizes a better understanding of how research data is accessed and how the use of these data impact the history of science, this paper offers an overview of the emerging integration of biomedical data with traditional demographic and population analysis approaches. The present analysis reviews what we know about existing biomarker data in the public domain and how these data have used the research process or overlooked as potential resources for a better understanding of how health impacts social outcomes.

Research Issues
This research was initiated because, despite the growing interest and funding support for the collection of biomedical information, no single organization has yet created a comprehensive listing of datasets that collect biomarker and medically based indicators no less the specific variables collected. Because of our unique mission, NACDA is well

suited to addressing the tasks associated with identifying the body of datasets that collect biomarker and medically based indicators. Even more important than the simple cataloguing of potentially useful data collections is the need for a system or registry that systematically organizes the content of datasets with a medical or health focus and identifies the specific type of variables they collect. This takes on greater importance as the debate on research biosocial frameworks has evolved dramatically in recent years as researchers use emerging data to explore new approaches measuring the interrelationships of physical and social measures (Seplaki et al., 2004).

While some researchers have suggested, that the wealth of research emerging from collaborative bio-social research will continue without a formal review of the existing universe of biomedical data collections because a limited number of well established studies are intensely used and suffice for more research designs. While we fully agree with the assessment regarding the productivity use of studies such as Framingham Study or the Baltimore Longitudinal Study, we would also suggest that the lack of organized alternative sources of data that allow for new research and the validation of existing research might be a factor in the intensive use of a few core datasets. Researchers cannot fully access all potential applications of a research design when they lack a thorough accounting of the array of biomarker variables available for analysis. Lacking a working registry of biomarkers both within specific data and across multiple data collections, many if not most datasets containing variables of interest to the research community will remain underutilized. More importantly, biosocial research designs would clearly benefit if existing studies based upon core datasets could be replicated using data that introduces different populations, time periods and geographies.

Analysis Model

This paper argues that the existing body of biomarker research contained within secondary data collections represents a rich treasure trove for biosocial research that needs to be productively explored. Unfortunately, the presentation format and structure of such information can vary dramatically from study to study creating complexities in the creation of comparative measures across studies and across time. This paper reviews these issues by developing four broad classification categories that capture most biomarker and medical variables.

Category 1 represents direct measures of the specific biomarker or specimen information. This category would include measures drawn directly from specimen data such as blood and swabs found in both historical health surveys and the more recent studies that are better known in the research literature. Biomarker indicators would include measures of serum protein (gm/100ml) or urinary iodine/creatinine (micrograms/gm) such as found in the NHANES studies or epinephrine-induced platelet aggregation blood tests such as performed in the 1975 to 1979 Aspirin Myocardial Infarction Study. It would also include physical indicators or "biomeasures" such as height, weight, body mass or skin-fold commonly found in many studies such as the Health Interview Study, Hispanic EPESE and HRS among others. NACDA, outside of NHBLI, represents the single largest source of studies with direct biomarker and biomeasure indicators and the only public source of historic studies containing such information allowing for the analysis of biosocial outcomes across time.

Categories 2 through 4 represent measures that are more indirect and that present different challenges to both the modeling and the interpretation process of biosocial research frameworks.

Category 2 biomarker measures would include variables that reference specific biomedical tests but only indicate the taking of the test as opposed to the recorded outcome of the test itself. These kinds of variables are common in data that reflect medical examination information such as found in the National Hospital Discharge Surveys (NHDS) or the National Ambulatory Medical Care Surveys (NAMCS). These indirect measures would include series questions that report as to the administration of specific tests including cholesterol tests, CBC, and levels of triglycerides, insulin, glucose or other diagnostic tests. The data does not report the test results in detail, only indicating the administration of the specific tests. While limiting, these all represent informative biomarker measures analysis of this information appears routinely in research examining disease processes and health outcomes such as smoking, diabetes or obesity.

Category 3 and 4 variables represent simpler and less informative measures based upon the reported presence of a disease process without any evidence of a testing either direct or indirect.

Category 3 variables are those where a respondent reports the presence of disease process based upon a reported diagnosis arrived at by a competent health professional. These are questions obtained in a wide array of surveys that ask health questions where the respondent is queried regarding their having been specifically informed of having a health issue. "Have you ever been by physician told that you have…(specific disease)?".

Category 4 variables are the same but lack the explicit requirement for a physician diagnosis. These are questions normally collect retrospective information on a disease process using questions like: "Have you ever had…(specific disease)?" These kinds of questions, while far from ideal, remain informative when the research infers biomarkers results or medical outcome ranges from the responses. Thus while recognizing the potential value of these two forms of health measures the current application will focus on Categories 1 and 2 which are stronger measures for biosocial research and are present in the majority of secondary data collections reviewed to date.

Each of these categories requires different methodological approaches when incorporating into demographic and population based models from direct estimates of prevalence and incidence to more exploratory methods that are traditionally applied to deficient or incomplete information sources. The demographic toolkit has ways of approaching these kinds of challenges, but they need to be refined for these new applications. This paper broadly addresses these issues and offers examples of how they have been used in the future.

Language and definitional issues

Another area of incorporating biomedical research within a population based social research framework is the issue of language. While the term "biomarker" has become commonly used to generically refer to all biomedical indicators, the term increasingly is being used to refer to a specific measures of health and functioning at the molecular level while the term biomeasure refers to the measurement of functioning such as blood

pressure. Bioindicators is a relatively new refinement that can be used to define gross measures such as BMI, waist hip ratio and other measurements that are suggestive of health functioning. Increasingly to work effectively with biomedical data population researchers need to familiarize themselves with the terminology used by potential collaborators from the physical and clinical health disciplines. As the field is evolving rapidly a common language is a necessity for effective and valid research.

Outcomes

This paper surveys the emerging interest in biomedical indicators and their incorporation with existing population based models of health and social behaviors. We review the development and growth of biomedical data collections as a part of socioeconomic surveys and we provide an overview of existing public use data collections that contain biomedical indicators. The paper discusses the scientific challenges of incorporating biomedical and population based research models and some of the issues the impeded research development including issues of focus, theoretical direction and differences in the scientific method across these diverse disciplines. The paper suggests future directions for research and emerging models, approaches and analysis tools that will facilitate this exciting new research strategy in the years to come.