DIW Berlin

Deutsches Institut
für Wirtschaftsforschung

Data Documentation 14

Dean R. Lillard
Gert G. Wagner

The Value Added of Biomarkers in Household Panel Studies

IMPRESSUM

DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Königin-Luise-Str. 5
14195 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1861-1532

DIW Berlin

Data Documentation 14

Dean R. Lillard*

Gert G. Wagner**

The Value Added of Biomarkers in Household Panel Studies

Berlin, August 2006

* Cornell University and DIW Berlin; DRL3@cornell.edu

** SOEP at DIW Berlin, Berlin University of Technology (TUB), and Cornell University, gwagner@diw.de (Corresponding Author)

Contents

1 Introduction

This note[1] is written from the perspective of economists who bring a collective experience gleaned from running a longitudinal multi-disciplinary survey (German Socio-Economic Panel Study – SOEP), and constructing and distributing a cross-national database that is extracted from ongoing longitudinal surveys (Cross-National Equivalent File – CNEF).[2] Although we speak as economists, we acknowledge and appreciate the value of the perspectives that other disciplines of social sciences offer on the topics we discuss below. And, of course, we do not claim to be life scientists. What we want to do is to highlight that life sciences can help social scientists better understand human behavior and how life scientists are already doing so.

Throughout the rest of this note we interchangeably use the terms "biomarkers" and "biometric data" to refer to data that measure physical characteristics of an *anonymous* respondent who gave an *informed consent*.

While there is a very broad debate about the role of biometric data in social science research (cf. Finch et al. 2001, Burkhauser and Lillard 2006), our purpose here is to argue a more narrow point – that social science research will be enhanced if ongoing household panel surveys can develop methods and get the resources to collect biometric data from their respondents. Panel surveys in general and household panel surveys in particular are ideal vehicles to use to maximize the potential insights into human behavior that biometric data can help reveal. Panel studies are ideal because they have accumulated a rich history of the social and demographic life of each respondent and his relatives across multiple generations. We argue that, by collecting biometric data of survey respondents, the value of existing data in household panel studies increases by several orders of magnitude.

The marriage of biometric and socio-economic data in panel studies is already a reality. Biometric data of various types are already routinely collected by, among other ongoing social science based data sets, the US Panel Study of Income Dynamics, the US Health and Retirement Study, the English Longitudinal Study of Ageing (ELSA), the Survey on Health,

1 We thank Richard Burkhauser for a careful reading of this document.

2 See for SOEP and a very brief description of CNEF Wagner et al. (2006).

Ageing, and Retirement in Europe, and in a more limited way (grip strength, cognitive ability) in the German Socio-Economic Panel. In addition, proposals are being developed to extend the set of biometric data that will be collected by them in the future. The set of data that will be collected is likely to expand over time both because research interest in these data are growing and because it is becoming easier and cheaper to extract different types of biometric data.

We argue that it is important for panel surveys to recognize and adapt to the promising insights and better understanding of human behavior that become possible when biometric data are collected from respondents to long running panel surveys.

2 Basic Rationale

The current push to combine biometric data with socio-demographic data reflects a growing recognition that biological factors may condition or even change how human beings respond to social and economic factors. If biological factors are important conditioning factors, then one needs biometric data to fully understand how human beings respond to social and economic factors over their life course.[3] For example, Udry (2000) analyzes the role of biological versus social determinants of “gender construction.” He argues that if social scientists do not recognize that biological factors shape behavior, then they risk becoming trapped in tautologies when they argue that notions of gender are social constructs. In addition, when biological factors condition how people react to a given social stimulus, all estimates of the effect of the social stimulus will be contaminated by classical omitted variable bias. In Caspi et al. (2002) for example, the authors examine not only the independent effects of parental education and certain genetic markers on the probability that parents are violent with their children, they also include an interaction term between parental education and the genetic markers. The finding of a statistically significant coefficient on the interaction term suggests that it may be important to control for the influence of “nature” in estimating the independent impact of a traditional social science variable.

In considering the types of biometric data to collect, it is useful to consider two (potentially) overlapping biometric data sets. Biometric data that:

- measure a person’s external physical characteristics (e.g. total body fat, height, waist circumference, grip strength, breathing, pulse, blood pressure, and perhaps more subjective characteristics such as quality of sleep and cognitive abilities)
- measure a person’s internal characteristics (saliva, blood, other biological specimens and all information that can derived from them)

Although one cannot draw sharp lines in how these two types of data might be used, one can generally say that measures of physical characteristics likely inform estimates of a person’s current or future health status. For example, there is well-documented evidence that people with higher total body fat are more likely to suffer a range of medical conditions such as

[3] On the other hand life sciences need socio-economic theory and empirical data (cf. Moffitt at al. 2006, p. 21).

diabetes now and in the future. The second type of data has more innovative and futuristic uses because biological scientists are able to extract more information from the samples and there is growing evidence of links between genetic markers and behavior.

3 Some Detailed Examples

There are well-known biomarkers one can collect to better document a person's current health status. This set includes things like fat-free body mass (total body fat), hip-waist-ratio, grip strength, height, and weight, etc. The decisions about which of these measures one should include will depend in part on the number of connections that have been identified by researchers between the biomarker of interest and health. For example, the medical community has reached broad consensus that many health conditions linked to obesity depend not on body mass but on the person's proportion of fat free mass. (cf. Burkhauser and Lillard 2006; Cawley and Burkhauser 2006).

This set of biomarkers allows for better characterizations of current health status that can then be used to understand how social determinants affect behavior of people with more precisely measured health conditions.

The second set of biometric data includes measures that likely predict future health status and that may directly determine (or at least condition) human behavior. Some of the biomarkers we mention above, such as fat free body mass, might be included here because future morbidity is known to be related to current obesity. However, other biomarkers, such as the presence or absence of genes known to be associated with (future) breast cancer risk, will affect future but not current health.

It is less obvious that people with these biomarkers act differently today in their social and economic behavior than do people without the biomarker. It is also the case that a person may alter his or her behavior today if he *learns* that he has the biomarker. Such was dramatically true before retroviral treatments were developed for HIV. Until that time, people who were diagnosed as HIV-positive dramatically altered their behavior, even when their current health status was not noticeably diminished by the disease.

It is difficult to anticipate all the ways in which social scientists might use the second type of biometric data to inform our understanding of human behavior. We gave a first example above (Caspi et al. 2002). More recently social scientists have shown that the presence of certain genetic markers can be used to predict health conditions that affect decisions of individuals to invest in education (Ding et al. 2006). Schnell (2006) provides another example in his review of the literature on fertility. He reports that the mother's testosterone level during pregnancy can predict a male child’s masculinity (attractiveness) and he predicts that a male‘s masculinity influences rates of fertility.

One can anticipate that the second type of biometric data would inform analyses of intergenerational mobility. Psychologists examine, for example, how genetic traits shared across generations are correlated with behavior that is also shared across generations. Researchers working in this strain of literature, known as “behavioral genetics,” do not necessarily require biometric samples because they rely on known relationships of gene transmission to identify groupings of respondents who share similar genetic makeup (for example they study siblings and identical twins). However, the precision of the findings in this type of research would improve with better information on the biological relations within family networks, especially the exact biological relationship between related persons.

The growing evidence that links human behavior to biometric data of the second type suggests that the set will continue to expand over time. Evidence now points to a relationship between self-reported well being and images of the brain (cf. Urry et al, 2004), and between risk aversion, personality traits, and non-cognitive ability (cf Dohmen et al. 2005).

There is also growing evidence that suggests that one can expand the biometric data included in the first set. For example, it has long been established that there is a relationship between labor market outcomes and the physical symmetry of one’s face (Biddle and Hamermesh 1994). One might conjecture that if the labor market rewards beauty, it might also reward a physical characteristic like the tenor of one’s voice.[4] Thus, one might also include digital copies of a person’s normal speaking voice when measuring a person’s physical characteristics.

As an aside we note that there exists a third type of data that is not clearly “biometric” but that may also be important for panel surveys. For example, evidence suggests that social scientists underestimated the role of good sleep on health and general well-being (see Siegrist et al 2006). Although most time use surveys measure treat time spent sleeping as a residual category, there are well established survey batteries on sleep and quality of sleep. One might consider using those to collect information on the quality of sleep. Data on cognitive ability is another example of data that have long been used in labor economics (cf. Heineck and Anger 2006 based on SOEP) and child development (cf. Knudsen et al. 2006,). The widespread use

[4] We thank Juergen Schupp for raising this question.

of measures of cognitive ability recommend them for inclusion as part of data collected in any modern household panel study. SOEP demonstrates that this is possible (Lang 2006, Solga 2005).

The above discussion ignores obvious and difficult questions about feasibility, costs, ethical issues, confidentiality, and whether data will be shared with respondents. These questions must be addressed in terms of potential payoffs and the costs associated with collection of the biometric data.

A strategy used by other panel studies in this regard is to ask experts to rank order those biometric data in terms of highest potential payoff and then gauge the costs associated with each. Such a process sometimes results in unanticipated orderings. For example, there is evidence of a high correlation between a mother's testosterone level during pregnancy and the ratio of the second and fourth finger of her male children (2D : 4D ratio; cf. Neave et al. 2003). Thus the image of the hand of male respondents may proxy for the mother's testosterone during pregnancy.

4 The Role of Household Panel Studies

Household panel studies, by design, generate a great deal of data on relationships between respondents, family members and other unrelated persons. These data are underutilized for those interested in the relative importance of nature and nurture. Properly structured, such data are a valuable resource to social scientists who want to appeal to shared genetic makeup and behavioral genetics. While data on intergenerational relationships are already being used to analyze intergenerational mobility in poverty, income or wage earnings mobility (Corak 2004; Ermisch et al. 2006; Francesconi et al. 2006), they also offer a solid platform for investigating a much wider array of intergenerational questions.

In the case of data needed for behavioral genetic analysis (cf. Schimmack and Lucas 2006), one might creatively design how samples are selected as a way to generate data without intrusive mechanisms. For example, if new samples are being drawn, one might use twin registries to over-sample sets of identical twins. Of equal interest can be an oversampling of dyads who share the same environment, but do not share genes. For example researchers have not fully exploited the fact that couples share an enviornment but are unrelated biologically. Therefore, panels studies may want to devote more resources to track both partners of a “couple” if they separate or divorce.

As far as we know, at present, ELSA is the only longitudinal panel study that collects biometric data through invasive procedures. When respondents agree to a nurse visit, ELSA collects blood pressure, lung function, anthropometric measures (height, weight, waist, hip), haemoglobin and ferritin, inflammatory markers of C-reactive protein and fibrinogen, lipids, fasting lipids, fasting glucose, and glycated haemoglobin. They also collect measures of physical functioning that includes balance tests, timed chair stands, and grip strength. ELSA also collects saliva samples over the course of one day, accompanied by a diary that they then process to measure levels of cortisol over the sample period. They also extract DNA for a genetic repository.

Aside from ELSA, there is no example for an invasive biomarker which is used in a fully fledged household panel study. However as noted before, many studies are collecting biometric data of various types. The SOEP and SHARE have, for instance, successfully collected the physical measurement of grip-strength (Hank et al. 2006).

Of course a proposal to establish a "bio-bank" will be complicated by additional ethical and data protection problems. The experience of ELSA is likely to be a valuable resource in this regard (e. g. Farah 2002).

5 Recommendations

Based on our reading of the literature which we summarized in this paper we strongly believe it is advisable to seriously consider the feasibility of collecting the following biomarkers in household surveys. Ethical considerations, data protection and costs will dictate which measures can feasibly be collected. As noted before, an expert panel should evaluate which measures promise the greatest scientific payoff.

The proposal to store biometric data in a strictly anonymous manner anticipates the development of new technology that will allow more information to be drawn from those samples at some future date.

- measured height and weight (starting with newborns and with some regular frequency)
- self-reported height and weight
- specific indicators for newborns (which are known by parents or may be included on birth records such as APGAR scores)
- measured body composition (fat free mass or total body fat)
- grip strength
- other body measurements related to obesity (e.g. hip-waist-ratio)
- blood samples for (1) instant analysis and (2) for freezing for future research
- saliva samples for (1) instant analysis and (2) for freezing for future research
- measure of sleep quality
- body images (face, hands, total body, possibly brain scans)
- speaking voice (digital copies)
- personality traits (measured by experiments)
- cognitive ability (tests)

For behavioral genetic analysis, one might creatively design how samples are selected as a way to generate data without intrusive mechanisms.

- for example, if new samples are being drawn, one might use twin registries to oversample sets of identical twins.

An incremental strategy may be recommended to collect some but not all of the above data. Certainly it will be possible to collect some of the data discussed above at relatively low cost and without as many of the ethical issues and consent problems involved with the rest.

Ultimately the decision about whether and which data to collect must be judged in terms of the relative costs and benefits to the research and policy communities and ultimately to the public at large.

References

Baker, Catherine. 2004. *Behavioral Genetics*, Washington, D.C.: http://www.aaas.org/spp/bgenes/publications.shtml.

Biddle, J. and D. Hammermesh. 1994. Beauty and the Labor Market. *American Economic Review*, December 1994

Burkhauser, Richard V. and Dean R. Lillard. 2006. The Case for NIA Leadership in Integrating State-of-the-Art Biomarkers into Next Generation Social-Science-Based Data, *Cornell University Working Paper*, Ithaca/NY.

Caspi, Avshalom et al. 2002. Role of Genotype in the Cycle of Violence in Maltreated Children. *Science* 297: 851-854.

Cawley, John and Richard V. Burkhauser. 2006. Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research, *Cornell Universit Working Paper*, Ithaca, NY (also NBER Working Paper No. W12291).

Corak, Miles (ed.). 2004. *Generational Income Mobility in North America and Europe*. Cambridge University Press, Cambridge,UK.

Ding, Weili, Steven R. Lehrer, J. Niels Rosequist, and Janet Audrain-McGovern. 2006. The Impact of Poor Health on Education: New Evidence Using Genetic Marklers, manuscript. Queens University.

Dohmen, T. et al. 2005. Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey, *DIW Discussion Paper No. 511*, Berlin.

Ermisch, John., Marco Francesconi, and Tom Siedler. 2006. Intergenerational Economic Mobility and Assortative Mating, *Economic Journal*, forthcoming (July 2006).

Farah, Matha J. 2002. Emerging ethical issues in neuroscience, *Nature Neuroscience* 5(11): 1123-1129.

Finch, C. E., J. W. Vaupel, and K. Kinsella (eds.). 2001. *Cells and Surveys – Should Biological Measures Be Included in Social Research?* Washington, D.C.: National Academic Press.

Francesconi, Marco, Stephen P. Jenkins, and Thomas Siedler. .006. Childhood Family Structure and Schooling Outcomes: Evidence for Germany. *DIW Discussion Paper 610*. Berlin.

Hank, Karsten, Hendrik Juerges, Jürgen Schupp, and Gert G. Wagner. 2006. Die Messung der Greifkraft als objektives Gesundheitsmaß in sozialwissenschaftlichen Bevölkerungsumfragen: erhebungsmethodische und inhaltliche Befunde auf der Basis von SHARE und SOEP. *DIW Discussion Paper* No. 577. Berlin.

Heineck, Guido, and Silke Anger. 2006. Cognitive Ability and Labor Market Outcomes in Germany – An Analysis Based on Pretest Data f SOEP. *DIW Discussion Paper* (forthcoming). Berlin.

Knudsen, Eric. I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff. 2006. Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce. *IZA Discussion Paper No. 2190*. Bonn.

Lang, Frieder R. 2006. Erfassung des kognitiven Leistungspotenzials und der „Big Five" mit Computer-Assisted-Personal-Interviewing (CAPI): Zur Reliabilität und Validität zweier ultrakurzer Test und des BFI-S, *DIW Research Note 9/2005*, Berlin

Moffitt, Terrie E., Avshalom Caspi, and Michael Rutter .2006. Measured Gene-Environment Interactions in Psychopathology. *Persepectives on Psychological Science* 181), 5-27.

Neave, Nick et al. .2003. Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society London B* 270: 2157-2172.

Schnell, Rainer. 2006. Some Examples for the Role of Biomarkers in Surveys. *DIW Research Note* (forthcoming). Berlin.

Schimmack, Ulrich and Richard Lucas. 2006. Marriage Matters: Spouses Similarity in Life Satisfaction. Paper Prepared for the 7th Conference of SOEP Users in Berlin, Mississauga and East Lansing.

Siegrist, Johannes et al. 2006. Work Stress, Quality of Sleep, and Self-rated Health. *DIW Discussion Paper* (forthcoming). Berlin.

Solga, Helga, Elsbeth Stern, Bernhard von Rosenbladt, Juergen Schupp and Gert G. Wagner. 2005. The measurement and importance of general reasoning potentials in schools and labor markets, *DIW Research Note 10/2005*, Berlin.

Udry, J. Richard. 2000. Biological Limits of Gender Construction. *American Sociological Review* 65: 443-457.

Urry, Heather L., Jack B. Nitschke, Isa Dolski, Daren C. Jackson, Kim M. Dalton, Corrina J. Mueller, Melissa A. Rosenkranz, Carol D. Ryff, Burton H. Singer, und Richard J. Davidson. 2004. Making a Life Worth Living, *Psychological Science* 15(6): 367-372.

Wagner, Gert G., Joachim R. Frick, and Juergen Schupp. 2006. Enhancing the Power of Household Panel Studies - The Case of the German Socio-Economic Panel Study (SOEP). *DIW Berlin Data Documentation No. 13*, Berlin.