

**Elastic Powers:  
The Integration of Biomarkers into the Health  
and Retirement Study**

*David Weir*

**Paper prepared for the National Academy Workshop on Collecting and Utilizing  
Biological Indicators and Genetic Information in Social Science Surveys**

**Draft: June 13, 2006**

## INTRODUCTION

The title of this paper, “Elastic Powers,” appeared as the last words of *Cells and Surveys*, the enormously influential 2001 volume from the National Academy of Science’s Committee on Population. In his closing reflections on the biosocial opportunities for surveys, Kenneth Wachter quoted from a Matthew Arnold poem that described the aging process as exhaustion from the cumulation of shocks and change—a literary anticipation of the concept of allostatic load that Wachter in turn stretched from a model of individual physiology in a social context into a model of paradigm shifts in the history of science (Wachter, 2001). He warned of the challenges to established modes of thinking that the integration of biology and social surveys would pose, and called for a renewal of what Arnold had termed “the elastic powers.”

Beginning the story of the integration of biomarkers into the Health and Retirement Study (HRS) with the closing words of *Cells and Surveys* is more than mere symbolism. Several contributors to that volume went on to be contributors to the HRS effort: Robert Wallace and Robert Willis as investigators, Eileen Crimmins and Douglas Ewbank as members of advisory groups, Teresa Seeman as an advisor to the study’s sponsor, the National Institute on Aging (NIA), and Jeffrey Halter as a consultant. While not acknowledged as a contributor in the list on p. vi, Richard Suzman’s considerable role in organizing the first volume was acknowledged by Jane Menken in her preface, and his leadership on the NIA side of the HRS cooperative agreement is unrelenting. Together, their own elastic powers in picking up where *Cells and Surveys* left off have been

considerable, and through their influence the elasticity of the entire field of population research has been renewed.

The scientific rationale for including biomarkers in HRS is not fundamentally different from the rationale for including them in any population survey concerned with health. They validate and add nuance to self-reports of health, they allow richer modeling of pathways of influence between the socio-economic and the physical, and they may capture aspects of health unknown to survey participants. This paper will give examples of how each of these are realized in the HRS. The development of biomarker data in other studies of older populations in the US such as the National Study of Midlife in the US (MIDUS), and the National Social Life, Health, and Aging Project (NSHAP) and outside the US in the English Longitudinal Study of Ageing (ELSA) and the Mexican Health and Aging Study (MHAS) has both provided models of what can be done and created great potential for comparative work with the addition of such data to the HRS.

Because of the unique place of the HRS in population surveys of aging, however, the decision to add biology to the HRS involved a number of other considerations, several of which were clearly anticipated by Weinstein and Willis in their chapter of *Cells and Surveys* (Weinstein and Willis, 2001). The HRS is a large longitudinal study that serves a large constituency of researchers from many different disciplines. At last count, there were over 6,000 registered users of the data, and over 1,000 unique authors of written research using the data. Putting its traditional aims at risk through attrition of respondents or elimination of critical established content would have been unacceptable.

Similarly, the confidentiality of respondents had to be protected, and the integrity of a longitudinal observation study not be transformed into an intervention study.

The ethical issues were considered carefully by the HRS investigators as well as the Institutional Review Board (IRB) governing the study. Notifying respondents of the results of well-established and commonly available diagnostic tests was deemed an ethical responsibility that overrides any concern that the information might alter future behavior. Because the tests contemplated by HRS assess familiar risk factors, and do not identify, for example, life-threatening cancers, the ethical conflict is not particularly difficult at this time. Biological material stored in repository for future use is governed by a separate IRB review. Respondents were asked to consent to having this material stored anonymously for future research. Ethical issues arising from any particular future test will need to be addressed at that time. For example, it is conceivable that some test of scientific value might not be permitted if it carried with it the ethical obligation to notify children or other non-participants of the possibility of an inherited disease, and if that notification were considered detrimental to the confidentiality of respondents.

Through the use of supplemental studies, some funded through peer review as competing supplements to the HRS, most of the elements of the HRS biomarker expansion were piloted on subsamples well before their introduction to the main survey. Those pilot efforts from 2001 through 2004 are thus a crucial part of the story.

Adding biomarkers without subtracting other things from an ongoing panel study leads to the question of cost. After the baseline interviews, the primary mode of interview in the HRS has been telephone. Although some biological material can be collected by mail, and the HRS had some success with this in a study of diabetes

described at some length below, it was clear that for a thorough integration of biology some form of personal contact would be needed: clinic visits, nurse visits to the home, or in-person interviewing. Because of the high cost of clinics or nurses, and because of technical innovations that have expanded what can be done by interviewers in the home, the HRS has designed its biomarker effort around conventional in-person interviewing.

First described in its 2005 renewal proposal, and now fully implemented in the ongoing 2006 data collection, the HRS has developed an integrated package of new content for a new model of in-home interview we describe as “enhanced” face-to-face. It includes anthropometrics, physical measures, blood spots, salivary DNA, and a self-administered psychosocial questionnaire. Most of the elements of this dramatically new development for the HRS were piloted in one way or another in smaller supplemental studies.

### **ADAMS**

The first effort at collecting biological specimens from HRS respondents came in a supplemental study of dementia, known as the Aging, Demographics, and Memory Study (ADAMS). The primary aim of the ADAMS study was to establish the prevalence of dementia in the population over 70 years of age from a nationally representative sample (Langa, et al, 2005). Because dementia is not a common condition even at that age, a simple random sample of the population would need to be fairly large to derive reliable estimates. The great virtue of using the HRS as a sampling frame for the

ADAMS was the ability to sample at higher rates from persons with higher likelihoods of being demented based on the cognitive assessments conducted by the HRS.

The stratification of sample selection for ADAMS was based on five cognitive categories from the HRS interview. These had to be established separately for persons who did their own cognitive assessments and those respondents for whom interviews were taken by proxy. In the case of proxies, the proxy reporter provides to the HRS interviewer an assessment of the cognitive function of the proxied respondent from the Jorm Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE). The cognitively normal group was further stratified by age (age 70–79 versus 80 or older) and sex in order to ensure adequate numbers in each of these subgroups. Finally, because of the long anticipated field period for the in-home ADAMS assessments, the recruitment was split between the 2000 wave of HRS and the 2002 wave, with some geographic areas drawn in one and some in the other, on a randomly sampled basis.

Because the rigorous in-home dementia assessments were conducted by collaborators at Duke University Medical School, the consent process for ADAMS was in two parts. First, an interviewer from the Institute for Social Research (ISR) at the University of Michigan contacted the respondent and an informed caregiver to obtain consent for Duke to contact them about the study. Duke then established contact and obtained consent for the home visit. This two-stage process worked against response rates in two ways. It extended the time interval between the HRS interview and the ADAMS visit, which resulted in loss due to mortality, and it provided respondents two opportunities to say no.

Overall, the response rate among survivors was 56%, yielding a final sample size of 856. Mortality rates (14% overall) were higher among the more cognitively impaired, but response rates conditional on survival were slightly higher for those groups. There were few significant predictors of non-response to the ADAMS study. Racial minorities participated at slightly higher rates than whites.

There are no established blood tests for dementia or cognitive impairment. Consequently, the ADAMS protocol relied primarily on extensive neuro-psychological testing and did not include any blood sampling. There is one well-established genetic risk factor for dementia, and that is the e4 allele of the ApoE gene. The ADAMS protocol did include collecting a buccal swab. Nearly everyone who participated in ADAMS agreed to provide this sample. Only 11 of the 856 refused (1.3%). The samples were sent to the pathology lab of the University of Michigan for extracting and typing of ApoE. Only 3 of the samples could not be genotyped.

The (weighted) distribution of genotypes found in the ADAMS respondents replicates fairly closely the expected population distributions (Table 1). As has been found elsewhere, the presence of any e4 allele is a risk factor for dementia, but not a particularly powerful one (Hyman et al., 1996). Preliminary analyses show that presence of the e4 allele (either homozygous 4/4 or heterozygous 3/4) was associated with approximately twice the odds of dementia compared to those with the e3/3 genotype, which is in the range of previously reported values (Breitner et al, 1998; Skoog et al., 1998). In the ADAMS data, the odds ratio for dementia associated with residing in a rural area compared with urban or suburban areas is nearly as high.

The ADAMS study demonstrated that HRS respondents would be willing to provide samples of DNA for research purposes. Within the group of respondents willing to participate in a three-hour home interview to assess dementia, cooperation with the DNA request was nearly universal. Combined with the 56% response rate to the ADAMS study overall, however, only 55% of the HRS respondents approached for ADAMS ultimately provided a DNA sample.

### **DIABETES STUDY**

In contrast to the ADAMS study, for which biomarker collection was a relatively small part of the overall assessment, the diabetes study was motivated in large part by the idea of collecting a clinically meaningful biomarker of the disease. Diabetes can result from a variety of underlying conditions (pancreatic failure to produce insulin, cellular resistance to absorbing insulin), but is always characterized by excessive levels of glucose in the blood. High levels of blood glucose cause damage to both large and small blood vessels and to nerves, potentially leading to many severe consequences (including cardiovascular disease). Among persons with diagnosed diabetes, the management of the disease targets the maintenance of lowered glucose levels (and, increasingly, the management of other CVD risk factors, especially hypertension). Thus, while there is considerable interest in understanding how people manage and cope with the disease, further study without a clinical marker for glucose levels seemed of relatively low priority.

A study on the scale of ADAMS, with an in-home assessment and blood draw, would have been quite expensive. Much of the non-biological information about diabetes could easily be collected by self-administered mail survey, which is far less expensive than even telephone interviews. The innovative aspect of the diabetes study was the attempt to gather dried blood spots through the mail for the analysis of glycosylated hemoglobin (HbA1c). A1c is an ideal measure for this study, as it is for the medical management of diabetes, because it summarizes the average levels of blood glucose over a two or three-month period. It also does not require fasting and can be done from blood collected at any time of the day. Glucose levels vary widely over the course of a day and in response to intake of food, making standard point-in-time readings very difficult to interpret in isolation.

Because of the reliance on A1c measures in the medical management of diabetes, commercial laboratories have developed assays for A1c that can be done in dried blood spots (DBS). This allows patients to take their own samples and mail them to a lab from which the results can be reported to their doctors, saving time and money. DBS assays for A1c require special proprietary pre-treatment of filter paper and utilize proprietary laboratory methods for analysis. Working with a commercial partner that has developed a DBS assay is therefore essential. Flexsite Diagnostics was the laboratory that did the HRS diabetes study and their support and cooperation were outstanding. They designed a specimen collection card specifically for the study. This allowed the respondents to mail their specimens directly to the lab with only an arbitrary numeric identifier so that the laboratory would not know the name or address of the respondent. Results were reported

to HRS by numeric identifiers and then merged with the questionnaire data and the usual HRS IDs.

The diabetes sample was selected from respondents to the 2002 wave of HRS. Only persons reporting a doctor diagnosis of diabetes were eligible. About 20% of the eligible sample was excluded because of their participation in another HRS mail survey (the Consumption and Activities Mail Survey). The eligible sample numbered 2,518. Of that group, 133 (5.3%) died prior to the beginning of the diabetes study in late 2003.

The diabetes study proceeded in two stages. First, respondents were sent a self-administered questionnaire, along with a check for forty dollars and an explanation that they would be receiving a second request to send a blood sample later (the usual HRS incentive for a mail survey is twenty dollars). Blood test kits were sent out to respondents when questionnaires were returned. A standard protocol of reminders was followed. After about six weeks, duplicate questionnaires and blood kits were mailed to persons who had not responded. After about eight weeks, follow-up telephone calls were placed to some respondents.

Questionnaires were returned by 1,897 sample members (79.7%). In contrast to both the core HRS interviews, and the ADAMS study, but quite consistent with other HRS mail surveys, there were substantial racial and ethnic differences in participation. Blacks and Hispanics had response rates about 10% lower than those of whites. Blood kits were returned by 1,233 respondents, which is 65% of those who returned the questionnaire. There was not much difference between Hispanic and other respondents on the blood test response rate conditional on participation in the questionnaire, but there was again a lower response rate among African-Americans. Combined with the

questionnaire response rate, the net biomarker rate was 52% of the eligible surviving sample.

The quality of the A1c data collected seems to be quite satisfactory. Figure 1 shows the level of A1c according to the type of treatment regime: 7.9 for those on insulin, 7.2 for those taking oral medication only, and 6.5 for those not taking medication (F-statistic=53.8,  $p<.0001$ ). The corresponding numbers for the 50+ population from the NHANES study for 2003/04 are 7.8, 7.0, and 6.2. Figure 2 shows the differentials by race and ethnicity: 7.1 for white non-Hispanics, 7.9 for African-Americans, and 7.9 for Hispanics (F-statistic=35.1,  $p<.0001$ ). That again differs only slightly from the comparable NHANES figures of 7.1, 7.9, and 7.8. Finally, Figure 3 shows that A1c also varies according to the respondent's self-assessed performance at managing the disease. Respondents giving themselves an "A" had A1c scores of 7.0, compared with 7.3 for "B" grades, and 7.8 for "C" or lower (F-statistic=15.1,  $p<.0001$ ).

Thus, in comparison with ADAMS, the diabetes study had a much higher overall participation rate but a fairly comparable net completion rate on the biomarker. Taken together, these two experiments suggested several important guidelines for future work on biomarkers in the HRS. First, multi-stage requests, in which the biomarker request is conditional on agreeing to one or more prior request, are bad for response rates. Second, self-administration and mailback of blood spots, while inexpensive, is unlikely to yield high response rates, and seems particularly ill-suited to maintaining high response rates of minorities. In-home requests, with a trained person present to take the sample, seemed to provide the best basis for administering biomarkers.

## 2004 FACE-TO-FACE INTERVIEWS

In 2004, the HRS was given additional funding from the Social Security Administration to use in-person interviewing to improve consent rates for linkage to Social Security records for two groups: all of the original 1992 HRS cohort (born 1931-41, plus spouses), and members of the 1998 War Baby cohort (born 1942-47, plus younger spouses) who had not yet given consent. This effort was successful. Seizing the opportunity created by in-person interviewing to pilot some other measures, the HRS obtained administrative supplements from the NIA to conduct in-person interviews with samples of the other cohorts to create a representative sample of the whole. From the combined set of in-person interviews, samples of about 100 persons from each single year of birth were assigned to do physical performance measures, with a subset getting height and weight measures.

Although the 2004 interviews did not include any blood or DNA work, or blood pressure testing, they were an important step in developing the 2006 strategy for biomarkers. We observed that there was a fairly high loss of sample due to respondents declining the in-person interview in favor of telephone—about 10% of those assigned. Thus, addressing “mode switches” is important for the HRS, given its history as a telephone survey. We also observed that failure to complete the physical performance measures (timed walk, puff test, grip strength) was related to self-reported physical limitations. Having good self-report indicators of those abilities would aid in understanding that censoring.

## **The 2006 Enhanced Face-to-Face Interview**

All the work of the various supplemental studies and pilot projects were brought together in the design of the enhanced face-to-face interview for 2006. The key elements of the 2006 HRS enhanced face-to-face interview are:

- Measured height and weight and waist circumference
- Blood pressure
- Timed walk, grip strength, puff test, balance test
- Dried blood spots for HbA1c, total cholesterol, HDL cholesterol, CRP and repository
- Salivary DNA for repository
- Self-administered mailback psychosocial questionnaire

### **Selection of Measures**

As a multi-disciplinary population survey serving a wide community of researchers, the decision process within HRS about any survey content, including biomarkers, must consider a wide range of potential uses and not focus narrowly on specific hypotheses or interests. Input was sought from a large number of experts. The choice of measures attempted to balance scientific value against cost and respondent and interviewer burden. There are two primary foci of the measures: the first being obesity and metabolic syndrome, for which the main goal is obtaining assessments now to model

risks of future events, and the second being frailty, for which the main goal is improving our characterization of the dynamics of disability and care needs of the elderly.

As an example of this selection process, the Quetelet body mass index (BMI) is obviously a critical measure for understanding obesity, and direct measures of height and weight should help to resolve any doubts about the accuracy of self-reports. But BMI is far from a perfect measure because of variability in muscle mass and other non-fat components of body weight. Waist circumference adds valuable complementary information about fatness and in particular about central adiposity. On the other hand, waist-hip ratio was considered to add relatively little to waist alone, and hip measurement is more intrusive for respondents and difficult for interviewers. Grip strength is a somewhat expensive measure because of the high price of the dynamometer devices, but when this is factored over an average of 50-60 EFTF interviews per interviewer this wave, and the potential for re-use in other waves, its value in assessing loss of muscle strength clearly outweighs its cost. A more difficult set of choices had to be made regarding the physical performance measures. There are a number of well-known assessments of lower body mobility, such as chair stands and “get-up-and-go”. We determined in 2004 that timed walk was the best single measure for our needs. For 2006, after consultation with other experts, we decided that what was needed to complement timed walk was not another measure focused on lower body strength, but rather measures that directly assessed balance because that can also be useful not only for modeling falls, but also in understanding cognition. We therefore added a ten-second semi-tandem stand, followed by a side-by-side stand, or a full tandem stand, depending on performance.

The most significant restriction imposed by cost constraints was on blood testing. Drawing of whole blood in the home or in a clinic would be extremely expensive in a dispersed national sample like the HRS. At the same time, the laboratory technologies for using dried blood spots (DBS) are advancing rapidly, making the scientific potential of this relatively inexpensive field collection protocol extremely attractive. At present, the HRS blood spots will be used to assay for glycosylated hemoglobin (A1c), total cholesterol, high-density lipoprotein (HDL) cholesterol, and C-reactive protein. All these measures are of course important in metabolic syndrome and cardiovascular risk. Having established the protocol for DBS collection, other assays can be added on as the technology improves and as scientific interest and funding develop.

Despite the restrictions imposed by both cost concerns and scientific focus, the new measures added to HRS cover a lot of ground. In their paper in Cells and Surveys, Eileen Crimmins and Teresa Seeman outlined a table of 17 measures on eight different physiological systems that were related to social and behavioral influences and health outcomes (p. 20). Of these, the HRS covers 8 measures in four of the systems. The two most significant systems not covered are the sympathetic nervous system and the hypothalamic pituitary adrenal axis. Good measures of functioning in these systems require either whole blood, urine, or multiple measures during the course of a day or over several days (e.g., cortisol). The HRS will continue to follow research using these measures and technologies for assessing them.

## **Sample Design**

The 2005 renewal proposal called for spreading the EFTF interviews over the next three waves of the HRS—randomly assigning one-third of the sample to each. Following the successful review of the proposal, the NIA recommended that this be accelerated to assign one-half of the sample in 2006 and the other half in 2008, creating the possibility of a four-year interval between biomarker collections rather than six. That recommendation was adopted. The assignments were made randomly at the household level. That means that both persons in a two-person household will get the biomarker interview in the same year. It also means that all sample clusters will get a mix of conventional telephone follow-ups and enhanced face-to-face interviews, and therefore that all interviewers must be trained for both types of interviews. While this increases training costs slightly, it allows for operational efficiencies when interviewers can mix the two types of activities and it allows for complete geographical representation in each wave.

### **Interviewer Training**

The HRS was fortunate to follow with about a year's lag the development of in-home biomarker interviews in the NSHAP study at NORC and the University of Chicago. Their success set a high standard. Critical to that success is the successful training of interviewers in both persuading respondents to participate, and in conducting the various measures successfully. The HRS had already developed protocols for height and weight measurement and physical performance measures. The ADAMS study had used buccal swabs to collect DNA. The switch to mouthwash samples offered both better quantities

of genetic material and an easier mode of administration. The two main areas for which new training materials had to be developed were blood pressure and blood spots. For the latter we were advised by Thomas McDade, as well as the commercial laboratory conducting the assays. For both we had input from Robert Wallace and Kenneth Langa, the two MDs on the HRS investigative team.

The HRS survey operations group developed a video on DVD that demonstrated the protocols for all aspects of the new content. This video was sent in advance to prospective interviewers interested in working on HRS (most of whom had worked for the study in previous years). It helped to screen out those interviewers who were too uncomfortable with the methods to do the work. It then also served as a training vehicle, and continues to serve as a refresher for interviewers in the field.

### **Consent and Reporting**

The HRS developed a booklet for the administration of physical measures and biomarkers. Respondents of course consent to participate in the HRS itself before interviewing begins. The biomarker assessments occur around the middle of the interview. For each of three sections, physical measures, blood spots, and DNA sample, the respondents are shown a printed information form, asked to read it and sign a consent before proceeding. Additionally, respondents are asked after signing the consent whether they feel it is safe for them to perform each measure immediately before doing it. Blood pressure results can be reported during the interview, and any respondent exceeding a specified threshold is given a card recommending that they see a doctor about their blood

pressure. Respondents are also told that their blood test results for HbA1c and cholesterol will be reported to them by mail. Both the blood test and the DNA consents permit future analysis to be done for HRS-related research purposes without reporting back to the respondent.

### **Early Results**

At this writing, the HRS is about ten weeks from the end of its 2006 field period, and has completed about ninety percent of its expected interviews. Early indications are that cooperation with the new EFTF interview is going well. Relatively few respondents have refused the face-to-face mode. Of those who have been interviewed, consent rates are over 94% for the physical measures, 82% for the DNA sample, and 81% for the blood spots. At present, older respondents are somewhat more likely to give DNA and less likely to give blood spots than younger respondents.

In the 2004 pilot work with physical performance measures, we noted a significant correlation between non-completion of the measures and self-reported physical limitations. The distribution of measured scores thus does not represent the true population distribution of abilities. In the case of timed walk, we have several good self-reports of lower body function that allow one to assess the function level of persons who decline to do the task, and potentially to impute a physical performance score. For the other measures, we do not. In 2006, therefore, we added self-rating questions on hand strength and on lung function to aid in understanding the functional abilities of those who do not do the measures.

Table 2 shows that self-rated hand strength correlates very strongly with measured grip strength, and that persons reporting weakness in the hand had substantially lower completion rates on the grip strength test. For lung function, shown in Table 3, the question on frequency of breathlessness is not quite as good at predicting response rates to the puff test. This may be due to the fact that (unwarranted) fear of infection from the device leads some well-functioning respondents to decline this test. Self-rating does correlate well with performance on the test.

There is some controversy about the quality of self-reports of height and weight, though the general finding seems to be that there is a general tendency to overstate height among the elderly (Ezzati et al., 2006; Gunnell et al., 2000). The most plausible explanation for this is that older people report their maximum adult height, not their current height after shrinkage due to age-related compression. Weight tends to be under-reported by the overweight, and over-reported by the underweight, leaving a relatively small bias on average. In preliminary results from 2006, as well as in a very small sample from 2004, the HRS tends also to find rather small errors in reported weight, and systematic over-reporting of height. The self-reports of height and weight are obtained before respondents are told they will be measured.

Figure 4 shows the pattern of heights found in 2006, graphing the mean measured height and mean self-reported height against the self-reported height. If self-reports were unbiased (equal to measured) on average, the graph should show a perfect 45-degree line. Instead, measured heights are lower than self-reports at every level of self-report, and more so at taller self-reported heights. The average differential is just under one inch. Measured heights are recorded to the nearest quarter-inch, and self-reports are in

round inches. In addition to the bias, there is some random error, as shown by the correlation coefficient of .89 between measured and self-reported height.

Figure 5 shows a similar graph for weights, grouping self-reports in ten-pound ranges on the horizontal axis, and graphing on the vertical axis the mean of measured and self-reported weight for each of those groups. The average error is about three pounds (self-reports below measured weight). The correlation is also impressively high at .97.

To put this in perspective, the average HRS respondent has a BMI of 29.1 using the measured data. Using instead the self-reported height lowers this to 28.2, while using instead the self-reported weight lowers it much less, to 28.6. Both together lower it to 27.8. Based on this evidence, the real scientific value from measuring weight in an older population, as opposed to relying on self-report, does not appear to be as great as the gain from measuring height. Height measurements are also less costly and less burdensome than using scales to measure weight.

Blood pressure data appear to be of good quality. The HRS protocol calls for three repeated measures, using an upper arm cuff with an automatic inflation device. These three measures are correlated at about .95, indicating good reliability. Table 4 shows that the mean of the three measures is reasonably well correlated with self-reported status. Interestingly, the mean blood pressure of persons who report a diagnosis and say it's under control is not much higher than those who say they do not have hypertension. Those who report their blood pressure is not under control do indeed have substantially higher measured levels.

## **CONCLUSION**

The integration of biomarkers into the HRS is very much a work in progress. The first big steps have been taken to transform a primarily telephone study into one using in-person interviewing to obtain direct physical measurements and collect biological samples in the home, challenging the “elastic powers” of the survey’s designers and its funders. HRS respondents have shown themselves willing to participate in this new survey experience, and the data they have provided appears to be of high quality. The HRS investigators hope to continue to expand and innovate in the inclusion of biomarkers as appropriate to the overall aims of the HRS. While this provides valuable new content to the HRS, and new points of contact with clinical and lab-based studies, a large population survey like HRS cannot replace the vastly greater biologic detail attainable in small clinical studies.

Soon the challenge to the elastic powers will shift from the design and implementation of the measures to the integration of them into longitudinal analyses using the data. It was this crucial intellectual transformation of how researchers conceive of problems that Wachter saw as the real challenge. The effort to collect such measures in population surveys will only be warranted by the new research insights they support. To support that challenge we will need to seek ways to encourage researchers to develop models to make use of them. And that in turn will stimulate new ideas and new measures for future waves of data collection.

TABLE 1 Distribution of ApoE Genotypes in ADAMS and Other Studies

Genotype	ADAMS* %	IOWA 65+ Population**	Framingham Population**
APOE e2/2	1	1	1
APOE e2/3	12	15	12
APOE e2/4	2	2	2
APOE e3/3	60	58	63
APOE e3/4	22	22	19
APOE e4/4	2	2	3

\*ADAMS percentages are weighted.

\*\*Data from Hyman et al, 1996.

TABLE 2 Self-Rating of Grip Strength, Response Rate to Grip Test, and Measured Grip Strength

Self-rated Hand Strength	Measured Grip Strength	Response Rate (%)	N Measured
Very Strong	36.7	86.5	1440
Somewhat Strong	31.9	88.1	4149
Somewhat Weak	24.7	78.8	1213
Very Weak	20.6	48.9	174
F-statistic	351.7	151.3	
p-value	<.0001	<.0001	

SOURCE: Preliminary HRS 2006 production data, unweighted.

TABLE 3 Self-Rated Shortness of Breath, Response Rate to Lung Test, and Measured Lung Function

Self-Rated Shortness of Breath	Expiratory Force	Response Rate (%)	N Measured
Often	272.2	72.3	391
Sometimes	313.2	85.3	1221
Rarely	367.2	87.7	2134
Never	378.0	87.6	3411
F-statistic	135.6	33.7	
p-value	<.0001	<.0001	

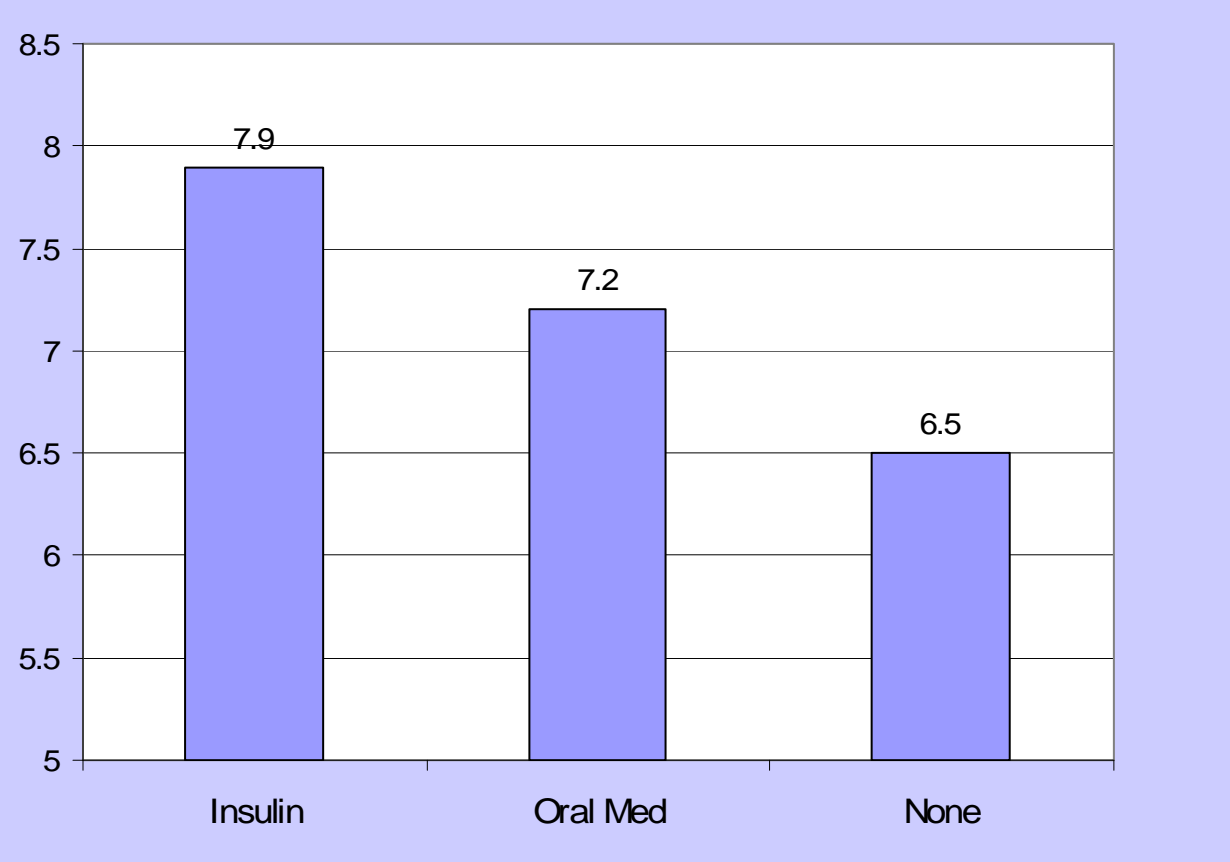
SOURCE: Preliminary HRS 2006 production data, unweighted.

TABLE 4 Biomarkers and Self-Reports: Measured Blood Pressure By Self-Report of HBP Dx and Control (mean of 3 measurements)

	Systolic	Diastolic
No HBP	126.4	78.1
Under Control No Meds	137.4	84.5
Under Control Using Meds	134.2	79.9
Not Under Control	148.4	87.6
F-statistic	138.7	71.2
p-value	<.0001	<.0001

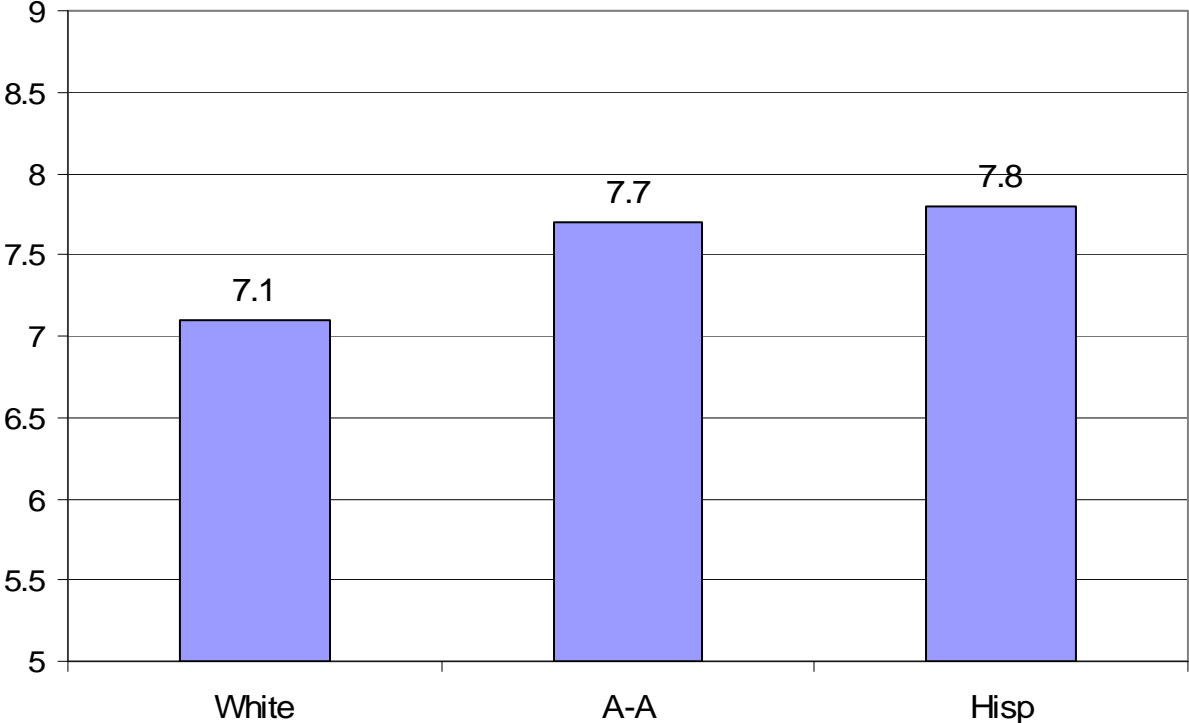
SOURCE: Preliminary HRS 2006 production data, unweighted.

**FIGURE 1: Mean HbA1c Score by Type of Medication Regime**



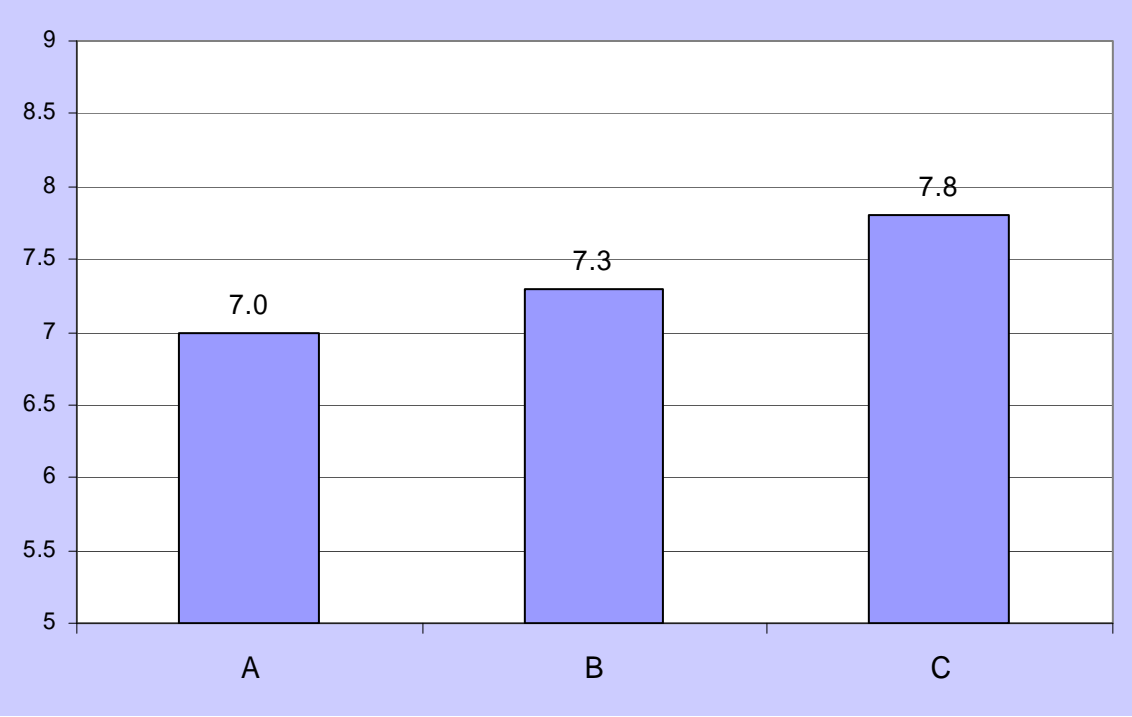
SOURCE: HRS Diabetes Study.

**FIGURE 2: Mean HbA1c Score by Race and Ethnicity**



SOURCE: HRS Diabetes Study.

**FIGURE 3: Mean HbA1c Score by Self-Rated Assessment (Letter Grade) of Self-Management of Diabetes**



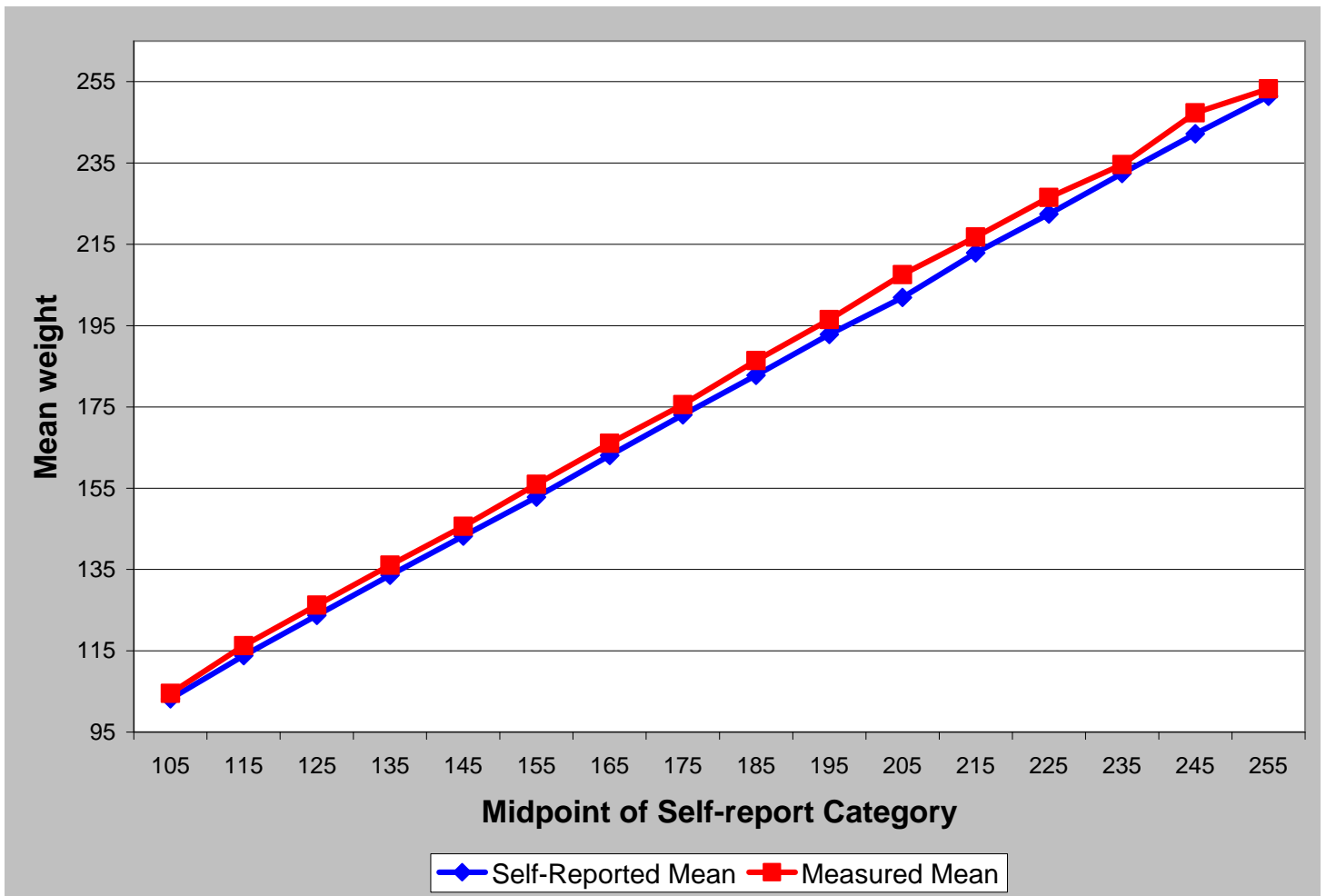
SOURCE: HRS Diabetes Study.

**FIGURE 4: Measured vs Self-Reported Height**



SOURCE: Preliminary HRS 2006 Production Data.

**FIGURE 5: Measured vs Self-Reported Weight**



SOURCE: Preliminary HRS 2006 Production Data.

## REFERENCES

- Breitner, John C. S., G. P. Jarvik, B. L. Plassman, A. M. Saunders, and K. A. Welsh. 1998. "Risk of Alzheimer Disease With the Epsilon-4 Allele for Apolipoprotein E in a Population-Based Study of Men Aged 62-73 Years." Alzheimer Disease and Associated Disorders 12(1):40-44
- Crimmins, Eileen, and Teresa Seeman, "Integrating Biology into Demographic Research on Health and Aging (With a Focus on the MacArthur Study of Successful Aging)", in Cells and Surveys Washington, D.C.: National Academy Press, 2001, pp. 9-41
- Ezzati, M., H Martin, S Skjold, S VanderHoorn, C Murray. 2006. , "Trends in national and state-level obesity in the USA after correction for self-report bias: evidence from health surveys," Journal of the Royal Society of Medicine 99: 250-257
- D Gunnell, L Berney, P Holland, M Maynard, D Blane, S Frankel and G Davey Smith. 2000. "How accurately are height, weight and leg length reported by the elderly, and how closely are they related to measurements recorded in childhood?" International Journal of Epidemiology 29(3):456-464
- Hyman, B. T., T. Gomez-Isla, M. Briggs, H. Chung, S. Nichols, F. Kohout, and R. Wallace. 1996. "Apolipoprotein E and Cognitive Change in and Elderly Population." Annals of Neurology 40(1):55-66.
- Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Ofstedal MB, Burke JR, Fisher GG, Fultz NH, Hurd MD, Potter GG, Rodgers WR, Steffens DC, Weir DR, Willis, RJ. "The Aging, Demographics and Memory Study: Design and Methods." Neuroepidemiology 2005;25:181-191

Skoog, I., C. Hesse, O. Aevansson, S. Landahl, J. Wahlstrom, P. Fredman, and K. Blennow. 1998.

"A Population Study of ApoE Genotype at the Age of 85: Relation to Dementia, Cerebrovascular Disease, and Mortality." Journal of Neurology, Neurosurgery, and Psychiatry 64(1):37-43.

Wachter, Kenneth "Biosocial Opportunities for Surveys," in Cells and Surveys: Should Biological Measures be Included in Social Science Research? Washington, D.C.: National Academy Press, 2001, pp. 329-336

Weinstein, M, and R. Willis, "Stretching Social Surveys to Include Bioindicators: Possibilities for the Health and Retirement Study, Experience from the Taiwan Study of the Elderly," in Cells and Surveys: Should Biological Measures be Included in Social Science Research? Washington, D.C.: National Academy Press, 2001, pp. 250-275.