NBER WORKING PAPER SERIES

IF MY BLOOD PRESSURE IS HIGH, DO I TAKE IT TO HEART? BEHAVIORAL IMPACTS OF BIOMARKER COLLECTION IN THE HEALTH AND RETIREMENT STUDY

Ryan D. Edwards

Working Paper 19311
http://www.nber.org/papers/w19311

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2013

I developed this research while I was a visitor on sabbatical in the Department of Demography at the University of California, Berkeley, to whose faculty and staff I am indebted for good hospitality and lifelong learning. I am grateful to David Weir, Heidi Guyer, David Card, Jennifer Beam Dowd, Luis Rosero-Bixby, Will Dow, Ron Lee, Jesse Cunha, and Elda Pema for helpful remarks and advice, and to seminar participants at the University of San Francisco and the Naval Postgraduate School. All opinions and errors are mine alone. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

If My Blood Pressure Is High, Do I Take It To Heart? Behavioral Impacts of Biomarker Collection in the Health and Retirement Study

Ryan D. Edwards

NBER Working Paper No. 19311
August 2013
JEL No. D84,I1,J14

ABSTRACT

Starting in 2006, respondents in the U.S. Health and Retirement Study were asked to submit biomarkers and were notified of certain results. Respondents with very high blood pressure were given a card during the interview; all respondents were notified by mail of their BP, hemoglobin A1c, and total and HDL cholesterol readings alongside recommended thresholds. About 5.8 percent received the high blood pressure card, and 5.4 percent had high A1c levels, an indicator of diabetes. Rates of undiagnosed high BP and diabetes according to these biomarkers were 1.5 and 0.7 percent. Average treatment effects of biomarker collection on the panel overall were effectively zero, but notification of rare and dangerous readings triggered new diagnoses, increased pharmaceutical usage, and altered health behaviors among small subsamples of respondents and their spouses. Very high BP or A1c readings raised new diagnosis and medication usage by 20 to 40 percentage points. Uncontrolled high BP triggered reductions in own smoking and own and spouse's drinking. High A1c was associated with a 2.2 percent drop in weight and an increase in exercise among respondents without a previous diagnosis of diabetes, but with no changes among those already diagnosed, whose self-reported health and disability worsened.

Ryan D. Edwards
Department of Economics
Queens College, CUNY
Powdermaker Hall 300-S
Flushing, NY 11367
and NBER
redwards@qc.cuny.edu

1 Introduction

There is great interest in the ability of biomarkers to enhance our knowledge about the determinants of healthy aging, longevity, and disparities (Weinstein, Vaupel and Wachter, 2007).[1] Biomarkers convey rich information to researchers and potentially also to survey respondents as well. Principles surrounding the protection of human subjects, and in particular informed consent, may require a data collection team to inform the research subject about observed biomarkers when the withholding of such information could be harmful.[2] The natural question is whether notifying individuals about their biomarker readings causes any changes in behavior or circumstances that are distinct from any effects of the underlying conditions indexed by the biomarkers.[3]

In 2006, the fourteen year-old Health and Retirement Study (HRS) of the University of Michigan's Institute for Social Research expanded to include biomarkers and several other new measures in an Enhanced Face-to-Face interview conducted on randomly selected rotating halves of the sample.[4] During the eighth and each subsequent biennial HRS wave

[1] Biomarkers vary widely in what they measure, running from relatively standard objective measurements of body characteristics, such as height, weight, and waist circumference, but also extending to genetic analysis of the DNA in blood samples. Between these two extremes are metrics that physicians commonly collect during routine physicals in addition to height and weight, such as blood pressure and blood characteristics like the levels of cholesterol, triglycerides, and blood sugar. Recent years have brought the addition of biomarkers to several preexisting longitudinal surveys, including the National Longitudinal Study of Adolescent Health (Add Health) in its third wave, and the Health and Retirement Study in its eighth wave.

[2] The National Health and Nutrition Examination Survey (NHANES) also notifies participants of a wide array of biomarker levels, including overweight, blood pressure, oral health, and so on. The Add Health collected biomarkers in its third wave and notified respondents of results of testing for HIV, chlamydia, and gonorrhea. By contrast, the Demographic and Health Surveys (DHS) measure HIV status in developing countries but explicitly do not inform respondents about results, ostensibly to preserve anonymity and respondents' safety. Instead, DHS participants are offered referrals for free counseling and testing.

[3] Researchers are increasingly aware that participation in a panel study can change some types of behavior under some conditions (Halpern-Manners and Warren, 2012), no matter how much the researcher would like to minimize his or her footprint. In the present context, the possibility that observation becomes a treatment seems considerably more likely than in standard cases of panel conditioning, because the motivating presumption of the IRB is that information about risky levels of biomarkers has measurable impacts on behavior and outcomes. If it did not, withholding it would cause no harm.

[4] HRS ostensibly chose to collect biomarkers and other measures on half the panel once every other wave, so that every respondent would submit biomarkers every four years, because the core HRS survey was already lengthy, and because of the increased costs involved with in-person collection of biomarkers. As discussed by Weir (2007), the original plan was for a third of the sample to be biomarked every wave, for a six-year gap between readings. A pilot consisting of a subset of the 2006 biomarkers was conducted in the 2004 wave, during which 3,734 HRS respondents were asked for consent. Of those, 3,447 complied and most completed breath and hand strength tests. About half completed walking tests, and 515 completed height and weight measurements. No measures of blood pressure or blood composition were attempted. As shown in Figure 5, the 2004 wave was abnormal in that most interviews were conducted in person in hopes of raising rates of consenting to the Social Security earnings match.

respondents in one rotating half or "biomarker group" were asked their consent to physical measures including three measures of blood pressure (BP), a saliva sample, and a blood sample, all collected by an interviewer roughly in the middle of the core interview. For respondents whose minimum blood pressure was greater than 160 systolic or 110 diastolic, interviewers left behind a "high blood pressure card," which recommended that the respondent immediately see a physician for a recheck of blood pressure. HRS later notified all participants of up to four results by mail, accompanied by suggestions to see a physician if their biomarkers were outside normal ranges: blood pressure, hemoglobin A1c, total cholesterol, and HDL or "good" cholesterol.

Whether information about biomarkers that is new to the researcher may trigger any behavioral change by the respondent is unclear. First, the information may not be new to the respondent. Individuals who already know their biomarkers, even if they are at risky levels, either may have already attempted to change their behavior without successfully altering the biomarker, may not be able to change their behavior, or may have chosen not to do so. Second, even new information may not trigger behavioral change if respondents do not understand or believe it, do not visit or believe their doctors, or if their doctors do not adequately help control the condition (Berlowitz et al., 1998; Hyman and Pavlik, 2001). Third, there are measurement issues. The "white coat effect" in a physician's office (Verdecchia et al., 1997) may translate to biomarker collection during a home visit. Readings could be influenced by events prior to the interview,[5] and lab irregularities could produce classical measurement error or missing data. Pre-existing knowledge about biomarker levels, or other characteristics that are correlated with biomarkers may determine a subject's willingness or ability to be biomarked, and it could alter observed relationships between biomarkers and outcomes.[6] Fourth, it is challenging to distinguish statistically between effects of the

[5] This is not true of hemoglobin A1c, which measures average blood sugar over several preceding months (Weir, 2007). Likewise, total and HDL cholesterol appear to be unaffected by fasting. But there are well-known daily patterns in blood pressure, which may also respond to the stress of an in-person interview.

[6] If respondents are aware of their health conditions and are taking medication to return biomarkers to normal ranges, like high blood pressure medication or insulin shots, then one would expect the biomarkers not to register outside of normal ranges and thus not prompt behavioral change. Thus in terms of the biomarker, an individual who is successfully managing the biomarker via medication should be indistinguishable from one whose biomarker was normal all along. But health outcomes may well differ between these two groups, even if behavioral responses seem likely not to differ.

notification versus effects of the abnormal biomarker level, because informed consent does not permit some individuals to submit biomarkers but not be informed in order to serve as a control group. Whether the revelation of an objectively risky biomarker reading in a household panel survey may change behavior is thus an open question, and it is the focus of this paper.

Two key aspects of the data are the prevalence of biomarker readings outside normal range and the extent of preexisting knowledge of the underlying condition. The HRS is nationally representative of Americans over age 50, a group that has relatively good access to health insurance and care and seems likely to be aware of their health. Of the roughly 7,000 individuals who submitted biomarkers in the eighth wave of the HRS in 2006 and were interviewed again in 2008, 5.8 percent received the high blood pressure card, and 5.4 percent had A1c levels of 7.0 percent or higher, an indicator of diabetes. A majority of individuals in these two interesting subgroups had also reported a preexisting diagnosis of the underlying disease, but 25 and 12 percent of them respectively had not, suggesting rates of undiagnosed high blood pressure and diabetes of 1.5 and 0.7 percent. Individuals with HDL cholesterol below the recommended threshold of 40 mg/dL were only 8.5 percent of the sample, about half of whom were already on cholesterol medication. As one might expect, the prevalence of high blood pressure more broadly defined, in this case anything over 120/80, high total cholesterol above 200 mg/dL, and previous diagnoses of those conditions and usage of associated medications were much higher, often approaching half the sample. Those most at risk for screening outside of normal range on these four biomarkers were disproportionately male, African-American, Hispanic, not homeowners, and did not report health insurance coverage.

Probably because dangerous levels of biomarkers were relatively rare in the sample, and rarer still among those previously undiagnosed, an intent-to-treat (ITT) analysis reveals that average treatment effects of collecting biomarkers and notifying participants of results tend to be indistinguishable from zero. The only statistically significant findings are that biomarker collection may have reduced rates of doctor visits and prescription medication usage by about 1.5 percent on average. This is a small amount relative to average rates of

around 80 to 90 percent reported in the sample, and the reduction does not extend to the average number of doctor visits, but it is an interesting result that could reflect a perception that biomarker collection substitutes for a physical examination. Thus a blanket policy of screening Americans over age 50 in households seems unlikely to produce tangible health benefits. Given the high rates of insurance coverage and care utilization reported by HRS respondents, this result is not surprising.

More interesting findings are that conditions and behavior among subgroups with biomarkers outside normal range appear to have responded to notification. This is especially true among the smaller subsets of those not reporting a previous diagnosis of the underlying condition, who presumably were surprised by the news. For those who did not know they had high blood pressure, receiving the high blood pressure card raised the prevalence of diagnosis and medication usage two years later by around 20 percentage points. Effects for respondents with A1c above 7.0 were larger, around 40 percentage points in increased diagnosis and medication usage for the previously undiagnosed, and self-reported weight appears to fall among these individuals by around 2 percent while frequency of physical activity increases. Perhaps most interesting is that spouses seem to react to the abnormal biomarker readings of their partners in addition to their own, revealing some nuanced household-based approaches to health maintenance and production. In particular, an undiagnosed spouse's high A1c reading also increases own exercise, while a diabetic spouse's reading appears to reduce own weight, suggesting responses of a shared input like diet, which unfortunately the HRS does not measure. Spouses of respondents with uncontrolled high blood pressure appear to reduce their drinking, and in particular their binge drinking, at the same time as the respondents themselves.

These results are interesting in view of their implications for behavior and policy, and for interpreting dynamics in panel studies with biomarker collection. Estimated impacts of abnormal biomarker notification can be statistically significant and meaningful for the individual but are often but not always relegated to the 1 or 2 percent of the eligible sample who screened positive without prior knowledge of the condition. Biomarker collection and notification, or by extension visits by trained medical professionals, thus has real but

extremely circumscribed average effects on outcomes among Americans over 50. Compared to the likely costs involved with a blanket extension of such services, the benefits seem minimal to nonexistent without additional targeting. A strategy of screening individuals whose characteristics place them more at risk of screening outside normal range might maximize the benefits. Analysis of panel datasets that include biomarker collection and notification is likely prone to small amounts of bias unless the panel information is utilized to control for the effects of biomarking, which appear real but certainly not as large as the effects of mode of interview on self-reports. The tight correlation between mode of interview and biomarker collection in the HRS implies that an analysis of either should account for both whenever that correlation is present.

In the sections that follow, I describe HRS biomarker collection and notification in greater detail and briefly discuss my statistical approaches. Then I present the basic characteristics of the panel interviewed in 2006 before revealing the main results and discussing robustness and persistence. The final section concludes.

2 Biomarkers in the Health and Retirement Study

The Health and Retirement Study (HRS) is a biennial panel survey of U.S. households sponsored by the National Institute on Aging and conducted by the University of Michigan's Institute for Social Research (Juster and Suzman, 1995). Originally begun in 1992 with a representative sample of Americans born between 1931 and 1941, the HRS was merged with a sister dataset and expanded in its fourth wave in 1998 to represent Americans aged 50 and over, and it has periodically added new birth cohorts in order to maintain representative coverage. In the eighth wave of data collection in 2006, when biomarker collection was added, there were a total of 18,469 respondents.

2.1 Biomarker collection and notification

Starting in 2006, the HRS expanded to include biomarkers and several other new measures in Enhanced Face-to-Face (EFTF) interviews conducted on randomly selected rotating halves

of the panel each wave. Consenting respondents would thus be biomarked once every four years.[7] During EFTF interviews, interviewers measured and collected biomarkers including blood pressure, pulse, saliva and blood samples; conducted tests of grip strength, breath, balance, and walking; and left behind a questionnaire on psychosocial topics. Weir (2007) describes biomarking as having occurred roughly in the middle of the EFTF interview, after the self-reports of health, height and weight, and disability.[8]

A key feature of biomarker collection in the HRS was the commitment to notifying respondents about four biomarker results: blood pressure, A1c, total cholesterol, and HDL cholesterol.[9] Notification of very high blood pressure occurred immediately, and notification by mail of all four results followed within about a month for all respondents who submitted biomarkers. When the minimum of three measures of blood pressure exceeded 160 mmHg systolic or 110 diastolic, HRS interviewers left behind a "high blood pressure card" similar to that depicted in Figure 1. The card, which here shows the average readings for the subsample that received it in 2006, recommends that the respondent see a physician immediately, with bold and underline emphasis on the word "immediately." Later, all recipients received a mailed notification letter reporting all four biomarker levels, the recommended normal ranges, and suggestions to see a physician if the biomarkers were outside normal range. A representative mockup of the notification letter is shown in Figure 2, which lists biomarker levels for the sample average and the thresholds as they were specified in 2006.[10]

[7] Both respondents in a couple household in HRS would be measured in the same wave; new spouses who enter the survey are asked to biomark with the other spouse. In addition to new spouses, new additions to the biomarked pool include non-respondents in previous waves who are asked to biomark when they reenter the panel. The method of assignment of previous non-responders to biomarking waves is unknown.

[8] Section C of the computerized HRS questionnaire asks about health, height, and weight, while disability questions appear in section G. The biomarkers were in section I, presumably followed by questions on employment (J), work history (L), work-related disabilities (M), and so on. An open question is whether respondents knew in advance of the self reports what biomarker collection was going to entail and changed their responses in a special way. I cannot rule this out, but I think it is more likely their self reports changed only because it was an in-person rather than telephone interview, an effect that I find in the data and that I control for as I discuss later.

[9] Consent forms for biomarker collection included language indicating that HRS would communicate blood pressure, cholesterol, and blood sugar results to the respondent, as well as an indication if they are outside normal range, and an instruction to share the information with a doctor.

[10] The recommended thresholds that appear here and in the instructions for the high blood pressure card roughly match official guidelines from the National Heart Lung and Blood Institute (NHLBI), Mayo Clinic, and other sources that can be found online. Thresholds for the high BP card correspond to stage-2 hypertension as defined by the NHLBI. The A1c threshold of 7.0 is higher than that of 6.5 recommended by International Expert Committee (2009), but the Mayo Clinic's target for diabetes control in diagnosed populations is 7.0 or lower. Respondents whose blood work was incomplete due to lab errors or other considerations received a letter including text to that effect on the first page. In the 2012 wave, the high blood

On average, mailed notification letters followed collection by about 10 days. More than half of the reports occurred before July 2006, and over ninety percent were reported before November 2006, but about 50 were delayed into the first two months of 2007. Most reports were marked either the same month as the collection or in the following month. For comparison, the earliest core interview in wave 9 was conducted in February 2008, or a full year after the latest blood spot report. The average gap between the report and the starting date of the core interview in the following wave was about 23 months. The timing of data collection and biomarker notification in the HRS panel is depicted in two timelines in Figure 3, one for each of the two rotating biomarker groups in the HRS. The black arrows show the timing of the five information flows: z_0, the high blood pressure card; and z_1 through z_4, the four biomarkers in the notification letter. The gray arrows depict collection of self-reported outcomes y and covariates x as well as the objective biomarker data b measured every other wave.

2.2 Treatment and control groups

The structure of the HRS allows me to test two hypotheses about the effects of biomarker collection and notification:

1. Average treatment effects (ATE) of submitting biomarkers and receiving notification among the entire 2006 biomarker group are nonzero.

2. Average treatment effects on the treated (ATET) of screening outside normal range in 2006 are nonzero.

I can assess the average treatment effects of biomarker collection and notification using the random assignment of respondents to biomarker collection in 2006 like an instrumental variable (Imbens and Angrist, 1994); here, it suffices to estimate the reduced form equation

pressure thresholds in the notification letters were changed to 140/90, and the total cholesterol threshold was changed to 240. Thresholds for A1c and HDL cholesterol have remained the same. Starting with the 2008 wave, HRS ceased mailing results to respondents in the states of New York and California based on a determination by legal advisors to HRS. Respondents in those states were sent a different notification letter listing a call-in number they could use to obtain results. I obtained information about these thresholds via email discussions with HRS investigators; aside from the high BP card instructions, they are otherwise not mentioned in the public data releases.

of the outcome on the instrument. This approach benefits from the rigorous identification provided by randomization, but if behavioral changes are concentrated among the small subgroups who screen outside normal range, the average treatment effect could be near zero even when there are interesting but rare reactions in the sample. I can test for these average treatment effects on the treated by exploiting the panel nature of the HRS. Analysis of pooled longitudinal data with individual fixed effects (FE) is a generalized difference-in-differences approach where past observations of treated cases serve as additional controls.

Figure 4 provides a visualization of treatment and control groups. Within the group of respondents C_{2006} shown at left who submitted biomarkers in 2006 are the partially overlapping subgroups A_{2006} and B_{2006}, those who screened outside normal range and those who had a preexisting diagnosis. It is natural to expect behavioral change to be concentrated here. Subgroups D_{2006}, E_{2006}, and F_{2006} were assigned to the 2006 biomarker group but did not or could not comply with the treatment. From an intent-to-treat perspective, the average treatment effects of biomarker collection in hypothesis 1 are revealed by comparing the entire 2006 biomarker group shown at left with the entire 2008 biomarker group shown at right, in which there are analogous subgroups that are not all observable in 2006. If randomization is strong, this can be done by simply comparing means in 2008; a generalized approach is a panel FE regression using pooled data up to 2008:

$$y_{it} = \alpha_i + \sum_t D_t + \beta^{ITT} b_{it-1}^{2006} + X_{it}B + \epsilon_{it}, \tag{1}$$

where β^{ITT} is the intent-to-treat estimate of the effect of being randomly assigned to the 2006 biomarker group in the prior wave, when the indicator $b_{it-1}^{2006} = 1$; the α_i are individual fixed effects; the D_t are wave or time dummies; the X_{it} are additional time-varying controls like age and marital status; and ϵ_{it} is a white-noise error.

For testing hypothesis 2, the ideal comparison would be between those who screened out of normal range in 2006, A_{2006}, and the unobservable group of respondents who would have screened out of normal range, A_{2008}. This comparison is infeasible because everyone who

submitted biomarkers in 2006 was also notified.[11] With a pooled dataset of all respondents, a panel FE estimator of the effects of screening outside normal range, when the indicator $z_k = 1$, draws identification from the behavior in the panel of all subgroups, including the unobservable control group A_{2008} and the treated group A_{2006}. In particular, past observations of A_{2006} serve as additional controls in a panel regression. This equation takes a similar form:

$$y_{it} = \alpha_i + \sum_t D_t + \sum_{k=0}^{4} \beta_k \, z_{kit-1} + X_{it}B + \epsilon_{it}, \tag{2}$$

where the z_{kit-1} are the biomarker notifications sent out after the previous wave.

2.3 Mode of interview

A complication is that biomarker collection and telephone interviewing are inversely correlated in the HRS, and both appear to affect self-reported outcomes. Biomarkers are collected in person via EFTF interviews, but the HRS has historically been a telephone-based survey, with in-person interviews limited to first-time respondents, nursing home residents, and others for whom a telephone interview was inadequate. As revealed by Figure 5, which plots the share of the groups assigned to biomarking in 2006 and 2008 interviewed by telephone in each wave starting with 1998, mode of interview has shifted dramatically first in 2004 and then again in 2006 with the initiation of regular biomarker collection.[12] Since then, the group not asked for biomarkers is almost exclusively telephoned each wave, while the other group requires EFTF interviews for biomarker collection.

Patterns in the data suggest that mode of interview affects several important self reports in the HRS. Figure 6 shows averages and confidence intervals in self-reported weight since 1998 separately for the two groups who submitted biomarkers in 2006 and in 2008. The two track each other closely prior to 2006, when they sharply diverge. Before they submitted

[11] Although respondents in group A_{2008} submitted biomarkers two years later in 2008, it is unclear whether their 2008 biomarkers would be good proxies for what their 2006 biomarkers would have been. The key unknown is the persistence of biomarkers over the life course, which likely depends both on biological and on socioeconomic conditions. If high blood pressure, A1c, and cholesterol are chronic and stable among those undiagnosed, and if those undiagnosed in the HRS tend to remain undiagnosed, then the 2008 biomarker levels would be good proxies for the 2006 levels among the key subgroup $A_{2008} - B_{2008}$. But it is easy to speculate otherwise.

[12] See footnote 4 for a discussion of why 2004 was different.

biomarkers during the 2006 wave, respondents told their EFTF interviewers that their weight had risen by about 0.25 kg, while the rest of the sample on telephone interview indicated they had not gained any weight on average. A similar pattern appears in 2008, when the group on telephone interview indicated they lost an average of 1 kg while those submitting biomarkers in person said they had gained about 0.5 kg. Similar fluctuations are apparent in self-reported current smoking behavior. The most plausible explanation for this is that face-to-face interviews elicit more accurate responses about characteristics with a visible or otherwise apparent component.

When mode of interview and biomarker collection both affect self-reported outcomes and are themselves tightly correlated, not only are simple mean comparisons between the rotating biomarker groups confounded, so too are simple differences in differences.[13] One could simply avoid this problem by comparing outcomes within rather than between rotating biomarker groups, but hypothesis 1 could not be directly tested in this way, and rigorously testing hypothesis 2 requires a better control group than the members of the 2006 biomarker group who did not screen outside normal range. A better alternative is to alter the panel FE regression equations to control for effects of interview mode, which will be separately identified by the exogenous variation not perfectly correlated with biomarker collection that is apparent in Figure 5. Inserting an indicator for interview mode m_{it} into my regression equations produces

$$y_{it} = \alpha_i + \sum_t D_t + \beta^{ITT} b_{it-1}^{2006} + \gamma\, m_{it} + X_{it}B + \epsilon_{it}, \tag{3}$$

and

$$y_{it} = \alpha_i + \sum_t D_t + \sum_{k=0}^{4} \beta_k\, z_{kit-1} + \gamma\, m_{it} + X_{it}B + \epsilon_{it}. \tag{4}$$

[13] Suppose in person interview mode changes outcome y contemporaneously by an amount m, while biomarking notification changes y by an amount b with a one-wave lag. Assume no other trending effects. Then in 2006, biomarked group A_{2006} will register $y + m$, while telephone-interviewed group A_{2008} will register y. In 2008, group A_{2006} is interviewed by telephone but feels the effect of notification and thus will register $y+b$, while group A_{2008} is biomarked and will register $y+m$. In 2008, the simple difference in means $DIM = b-m$, while the difference in differences $DID = [(y+b)-(y+m)] - [(y+m)-y] = b-m-m = b-2m$. Neither method separately recovers b nor m, although one could proceed by subtracting those measures: $m = DIM - DID$. This is sufficiently nonstandard that I elect to use panel FE with mode-of-interview effects instead.

A final helpful modification is to interact the biomarker notification indicators z_{ik} with indicators of preexisting diagnoses at $t-1$ of the health conditions c_{ik} that indexed by the biomarkers, such as high blood pressure or diabetes. With these interactions, equation (4) becomes

$$y_{it} = \alpha_i + \sum_t D_t + \sum_{k=0}^{4} \beta_k \, z_{kit-1} \, c_{kit-1} + \gamma \, m_{it} + X_{it}B + \epsilon_{it}, \tag{5}$$

where the treatments are now defined as screening outside normal range either with or without a previous diagnosis of the condition. As I show, this distinction is important.

3 Results

3.1 The sample submitting biomarkers in 2006

The rows in Table 1 list an array of salient characteristics of the 2006 biomarker group stratified across the columns according to their biomarker results.[14] Of the five subgroups shown, three small groups of around 400 respondents each screened positive for what I loosely term "rare and dangerous" conditions: blood pressure high enough to receive the high BP card (above 160/110 mmHg), A1c above 7.0 percent, or HDL cholesterol below 40 mg/dL. Between 2,400 and 3,800 screened positive for high BP (above 120/80 but not over 160/110) or high total cholesterol (above 200 mg/dL), shares that approach half the sample. In 2010, HRS revised upward their normal-range thresholds for these two biomarkers.

The broad messages in Table 1 are that screening out of normal range on these four biomarkers is associated with many preexisting characteristics; undiagnosed health conditions apparently exist even though insurance coverage rates and utilization tend to be high; and although health care utilization and behavior may not vary much in the sample, there is room for behavioral responses. The top of the table reveals familiar socioeconomic patterns in disease incidence, with African American and Hispanic males overrepresented

[14] I restrict attention to respondents who appeared in both the 2006 and 2008 waves, which effectively drops about 10 percent of the sample submitting biomarkers across the board, of whom typically 6 percent had died and the other 4 percent HRS apparently could not reach. As I discuss shortly, there did not appear to be substantial differentials in attrition or mortality across these groups defined by biomarker results by 2008, a finding that seems surprising.

among those with very high blood pressure or high A1c. Depending on the biomarker, screening out of normal range may also be associated with marital status, homeownership, and health insurance coverage. The middle of the table shows that self-reported rates of physicians' diagnoses of health conditions certainly vary across these groups but are not universally reported by those who screened out of normal range on the relevant biomarker, and that rates of new diagnosis for these conditions hover around 2–6 percent between biennial waves.[15] Health care utilization is high across the board, with 90 percent or more reporting at least one visit to a doctor, ER, or clinic in the past 2 years, and pharmaceutical usage rates above 75 percent. The bottom of the table shows reduced self-reported health and increased disability among respondents with high A1c, which appear to correlate with weight and less exercise, but also less drinking and somewhat less smoking. Aside from that and some limited evidence that respondents with high blood pressure are behaving less healthily, there is much similarity across these groups in these behavior metrics. Smoking is not very prevalent at around 14 percent, but exercise and drinking could probably be altered. Unfortunately the HRS does not measure diet.

With five distinct types of "morbidities" represented by screens outside normal range, the structure of comorbidity is challenging to summarize but potentially important for understanding behavioral responses. In addition, comorbidity across spouses is potentially relevant because some responses to biomarker notification may be made at the level of the household. Table 2 shows pairwise correlations between respondents' and spouses' screens outside of normal range.[16] Notable results here are the significant but small correlations between the high BP card and high A1c, and between high A1c and low HDL, which are around 0.05. Other significant correlations appear between high A1c and high total cholesterol, and between high total cholesterol and low HDL cholesterol, which range from -0.04 to -0.13. Correlations across spouses are also small and follow an interesting matching pattern: the correlations between spouses' high A1c screens or between spouses' high total

[15] For the current level of condition diagnosis here and in the panel FE analysis later, I use the raw responses, in which respondents can dispute records from past waves. In this table, the change in diagnosis is calculated using the current statements about present and past diagnosis. Differences between these data definitions are minimal and do not appreciably affect results.

[16] Pairwise correlations between levels of these biomarkers are qualitatively similar to the correlations between screens out of normal range shown in Table 2.

cholesterol screens are around 0.1 and significant, and there is some weak evidence that blood pressure may be jointly elevated among spouses, but both typically do not receive the high BP card.[17] Altogether, while correlations in screens among respondents and between respondents and their spouses are sometimes significant, they are also small, indicating that notification of an abnormal screen is often an independent and isolated event.[18]

3.2 Mortality and sample attrition

The HRS tracks mortality well (Weir, 2010), and attrition is low compared to other household surveys (Banks, Muriel and Smith, 2010). An important question is whether respondents who screen out of normal range on these biomarkers might have experienced differential mortality or panel attrition, which could inject bias into estimates of the effects of biomarker notification. I tested for differential mortality and attrition among the 2006 biomarker group by separately modeling death or nonresponse in 2008 as functions of indicator variables for screens out of normal range and an array of socioeconomic control variables using probit and logit specifications.[19] Once I controlled for self-reported health conditions, scores outside normal range were not significantly associated with either mortality or attrition after 2 years.[20]

[17] Similarity in spousal characteristics probably due to assortative mating suggest that if anything, we might expect to see more correlation in these notifications than we do. Correlations between spouses' weight or body mass indexes, smoking, and frequencies of physical activity are between about 0.12 and 0.4 in this sample.

[18] Another way to summarize the comorbidity structure is by cross tabulation. Within each group of roughly 400 respondents in Table 1 with one of the three rare screens, the high BP card, high A1c, or low HDL, only between 21 and 41 individuals received a second notification of a rare screens, such as a high BP card combined with high A1c or low HDL, or high A1c combined with low HDL. Only 5 respondents scored outside normal range on all three. Respondents with spouses who also screened outside normal range on the same one of these three rare measures similarly numbered between 16 and 33 in total depending on the biomarker.

[19] I used the same covariates that are shown in Table 3 plus 9 indicators of diagnosed health conditions: high blood pressure, diabetes, cancer, lung disease, heart problems, stroke, psychological problems, memory-related disease, and arthritis. Results are available from the author upon request.

[20] This is not to say that biomarkers do not affect mortality. Rosero-Bixby and Dow (2012) reveal significant effects of biomarkers on mortality in a Costa Rican panel. Rather, this is a statement about how scoring above particular thresholds of biomarkers does or does not affect mortality and other forms of attrition over two years in the HRS. In these HRS data, the levels of A1c and total cholesterol are significant predictors of death by 2008 even though screens out of normal range on those biomarkers are not. Presumably the information contained in the thresholds alone is picked up well by other covariates, likely the health conditions in particular.

3.3 Predictors of screens outside normal range

Table 3 displays selected marginal effects of respondents' characteristics on the probability of screening outside normal range in 2006 on the five biomarker categories either with or without a previous diagnosis of the disease. HRS directly asks about high blood pressure and diabetes but not about diagnoses of high or low cholesterol per se. To proxy the latter, I experimented with using diagnoses of heart problems, the closest match among the questions, and results were comparable to what I show here, which instead differentiates by whether the respondent was taking cholesterol medication.[21]

Differences according to preexisting knowledge of the disease are evident in Table 3, as are familiar correlations between socioeconomic status and disease incidence, especially when uncontrolled. There are large and significant effects of many covariates, including most prominently sex, race, Hispanic origin, and weight, on the probability of rare and dangerous screens resulting in the high BP card or high A1c among those with preexisting diagnoses. These respondents could be termed "noncompliers," because they know they have the underlying condition but are unable or unwilling to control the biomarker. In contrast, there are small and insignificant effects of most characteristics on the incidence of rare and dangerous screens among those whom we might call the "undiagnosed." Although the latter are not unpredictable, the patterns in Table 3 imply that such screens are probably more unexpected. Finally, the marginal effects of not having health insurance on undiagnosed moderately high blood pressure and on the cholesterol screens are interesting but apparently do not reflect a universal trend.

Results here also bolster the argument that there is room for behavioral change. Unfortunately they also reveal that unhealthy behaviors are more significantly linked to the noncompliers than to the undiagnosed.

[21] I can similarly define previous knowledge of high blood pressure and diabetes by whether or not the respondent is taking the associated medicine. Results are similar to those reported here.

3.4 Average treatment effects

Findings thus far suggest that any new information conveyed by biomarker notification is likely to be very limited in scope within the sample. Tables 1 and 3 report that only around 1.5 percent of participants in the biomarker wave received a high blood pressure card and did not know they had high blood pressure, and perhaps 0.7 percent screened positive for high A1c and did not already have a diabetes diagnosis. If these notifications are most salient, one would expect the average treatment effects of biomarker collection and notification on all panel respondents to be small.

Table 4 confirms this, reporting insignificant effects of assignment to the 2006 biomarker group on 18 of 20 outcomes in 2008 obtained by estimating equation (3) with a linear panel FE estimator on the pooled sample of all HRS respondents observed between 1998 and 2008.[22] Two outcomes appear to respond significantly to assignment: having seen a doctor and regularly taking prescription drugs since the previous wave, which fall by 1.5 and 1.2 percentage points respectively. The average number of doctor visits since the previous wave does not respond significantly (not shown), which could be consistent with small changes in the prevalence of a common event like this. Both of these results pass a falsification test of regressing the variable contemporaneously on the assignment rather than on the lag assignment. That does not reproduce these coefficients, and it should have if the association were preexisting or due to an third variable. Simple difference-in-differences estimation, valid here because mode of interview is insignificant in these two cases, roughly confirms these findings. Given later results that indicate elevated medication usage among some subgroups after biomarker notification and no effects on doctor visits, it is hard to know what to make of these two outliers. It is possible that respondents may have viewed biomarker collection, which could not have resulted in a new or refilled prescription, as a substitute for a doctor visit or routine physical examination, which could have.

[22] All regressions in Tables 4–7 are linear, including those modeling dichotomous outcomes. As shown in Table 1, these outcomes tend to be quite common in the sample, removing a typical concern about applying the linear probability model to limited dependent variable analysis. An additional motivation for using the linear model, aside from its more straightforward asymptotic characteristics in the presence of fixed effects, is that interaction terms are better defined. Results of nonlinear logit models tended to be consistent with those of the linear probability model.

In 13 of the 20 regressions, mode of interview is significantly associated with the self-reported outcome. On average, these effects are not particularly large, but due to the tight correlation between biomarker collection and mode of interview, they are important here. Self-reported weight is 0.3 percent lower for those on telephone interview, significant at the 1 percent level, but this is only about one sixth the size of the average percent understatement of self-reported weight relative to objective weight as measured in the EFTF, which is about 1.8 percent in the data and has been discussed elsewhere (Weir, 2007). Without controlling for mode of interview, however, estimating equation (1) with log weight as the endogenous variable attaches that highly significant estimate of -0.003 to lagged biomarker assignment, a clear case of omitted variable bias.

3.5 Average treatment effects on the treated

While the average treatment effects on the entire 2006 biomarker sample tend to be insignificant, a different picture emerges when I model effects on individuals specifically treated by notifications of screens outside normal range. In particular, I observe statistically significant and interesting responses among individuals who screened positive for either of two of the three rare and dangerous conditions, very high BP or high A1c, and especially among those without a preexisting diagnosis of the underlying condition.[23] These results pass a falsification test, and they are accompanied by some interesting effects of spouses' notifications.

3.5.1 Respondents' notifications

Table 5 reports estimates of the marginal effects of biomarker notification interacted with preexisting conditions on 20 outcomes using panel fixed effects applied to equation (5). To no great surprise, the largest effects here of biomarker screens out of normal range are on condition diagnoses and related pharmaceutical use by the 2008 wave, particularly among respondents who did not previously report having the condition. The second-largest coefficient in the table is the 42.8 percentage point increase in diabetes diagnosis among those

[23] For brevity, I skip over the results of estimating equation (4), without condition interactions. These results are close to the weighted averages of the effects shown in Table 5 and are thus often less clear and less interesting. They are available from the author upon request.

who screened positive for high A1c and who did not report having diabetes, followed by the 39.7 percent increase in the rate of diabetes medication usage among that group. Among those who received the high BP card without a previous diagnosis of high BP, diagnoses increase 17 percent and drug usage increase 16.1 percent. Respondents with high BP but below the card threshold were 5.9 percentage points more likely to have a high BP diagnosis and 3.9 percent more likely to be on BP medication by 2008, increases that roughly double the usual rates shown in Table 1. Patterns in other diagnoses and medication usage are less clear probably due in part to the lack of alignment between high cholesterol and a condition the HRS asks. Noteworthy is the elevated risk of diagnosis of heart problems or stroke, by 4.7 and 4.6 percent, among those who received the high BP card and who already knew they had high blood pressure.

Notifications are correlated with some measures of health and disability, but many of these patterns seem to reflect preexisting characteristics rather than plausible effects of the notification. This is revealed by the falsification tests in Table 6 where I have modeled outcomes in 2006 rather than 2008 as functions of biomarker notifications from the 2006 wave. The marginal effects on contemporaneous diagnoses and medication usage here are as they should be: positive for the group that had them and negative for the group that did not. But coefficients on the disability indexes for those with high BP but no card are similar in Tables 5 and 6, suggesting that those are not effects of notification. Positive and significant coefficients on 2008 disability measures among those with high A1c and a diabetes diagnosis in Table 5 are not present in 2006, however, nor is the protective effect against IADL disability for respondents with high A1c but without a diabetes diagnosis, implying those effects may be real.

There appear to be relatively few significant effects for the groups with high BP but no card or for those with cholesterol screens out of normal range that can pass the falsification tests shown in Table 6. Obvious effects are the 14–15 percent increases in cholesterol medication usage rates among those with such screens who had not already been taking it.[24] Another effect that passes the falsification test is the 4.9 percent additional increase in

[24] The large and significant negative coefficients here on abnormal cholesterol screens among those who were already taking those medications are real. These subgroups report less than 100 percent usage in 2008

diabetes diagnosis among those with low HDL cholesterol who were already taking cholesterol medication. This is roughly a doubling of the usual incidence reported by Table 1 and is striking in light of concerns that using statins to treat cholesterol may raise the risk of diabetes onset (Sattar et al., 2010). Statins are thought to be more effective at lowering high cholesterol rather than raising HDL cholesterol, which makes this result and the lack of an association with high screens on total cholesterol somewhat puzzling. But as shown by Table 1, the subsample in 2006 with low HDL cholesterol was more likely to be taking cholesterol medication than those with high total cholesterol.

Probably the most interesting results in Table 5 concern weight, drinking, and exercise among undiagnosed diabetics, those respondents with high A1c but without a diagnosis. The coefficient on log weight for this subgroup in 2008 is -0.022 and is significant at the 4.9 percent level (t-stat of -1.97). There is no sign of any contemporaneous association from the falsification test in Table 6. In addition to large increases in diabetes diagnosis and medication usage, this group also reports 0.306 fewer drinking days per week, the other insignificant coefficients on their drinking behavior are all negative, and while they may report less frequent vigorous physical activity, the frequencies of moderate and especially light activity appear to rise. These outcomes are fully consistent with practitioners' common recommendations for diabetes control and appear to be the clearest evidence of behavioral responses here.

The other behavioral changes of note that we see in Table 5 are concentrated among the "noncompliers" who received the high BP card and already knew they had high BP. Among this group, and also among those with high BP but no card who did not already have the diagnosis, the coefficient on smoking is negative and significant, here -0.023 (t-stat of -2.06). Coefficients on drinking are also negative but insignificant among the noncompliers with the high BP card, with the largest coefficient of any significance in the table, -0.523 (t-stat of -1.91), appearing here on the number of binge drinking days (4+ drinks on one occasion) in the past 90 days, fully one third of the average response. Frequency of light exercise also falls among this group, which may reflect the redistribution of household chores,

for unknown reasons. Given that the notification thresholds in question are not extreme, one interpretation is that the notification is not a good indicator of a persistent condition in these cases.

the primary examples of light exercise given in the question, away from the respondent with a history of high blood pressure following receipt of the high BP card.

3.5.2 Spouses' notifications

Although doing so will require dropping the third of the biomarker sample who are not coupled, I can also examine the effects on respondents' outcomes and behaviors of their spouses' receiving notification of screens outside normal ranges using a similar methodology. Adding spouses' notifications and their interactions with spouses' preexisting conditions to equation 5 and reestimating produces a set of coefficients on own-notifications that are similar to those in Table 5, because own and spousal notifications are largely independent per Table 2. It also produces a new set of coefficients on spousal notifications, which appear in Table 7.

As one might expect, there are far fewer significant results here, especially among diagnoses and other variables that measure own status. When effects on diagnoses and disability are significant here, they are often also similarly signed and significant in the falsification regression (not shown) and thus are probably not real effects.

One of the results that passes the falsification test is the marginal effect of a known diabetic spouse's high A1c screen on the respondent's own log weight of -0.015, significant at the 1.2 percent level (t-stat of -2.52). This result is not mirrored by any reaction in own physical activity, suggesting that this weight loss may have been obtained through shifts in diet, which the HRS does not measure. For respondents whose spouses screened high on A1c but did not have a preexisting diabetes diagnosis, there was no response of own weight, while the frequency of light exercise increased significantly, roughly as much as reported by the spouse in question, with a coefficient here of -0.425 (t-stat of -3.13) compared with -0.372 (t-stat of -2.01) shown in Table 5.

The other behavioral responses of note by spouses in Table 7 concern drinking behavior. Relatively strong evidence suggests a spouse's high blood pressure causes reductions in own binge drinking. The coefficient of -0.962 (t-stat of -2.71) in the first column, for spouses of respondents who received the high BP card and had already had a diagnosis of high

blood pressure, is greater in size and significance than the own-coefficient of −0.523 (t-stat of −1.91) in Table 5. If binge drinking is a shared activity between spouses, it stands to reason that behaviors might shift together in response to a notification.

3.5.3 Robustness and persistence

Several other estimation strategies produced results similar to those shown here.[25] Simple differences in differences (DID) are unbiased within the 2006 biomarker group because mode of interview is changing uniformly. Those estimates mirrored the results of linear panel FE regression, which is a generalization of DID. I also explored balancing the panel by selecting only treated respondents and choosing controls from individuals in the 2008 biomarker group using nearest neighbors matching from a propensity score algorithm. That technique, along with restricting the panel regression sample to just the 2006 biomarker group, also produced panel FE estimates that were very similar to those presented here. The broad conformity of results suggests that much of the identification stems from changes among the treated observations.

Testing for the persistence of results is hampered by the fact that while the other half of the sample submitted biomarkers in 2008, most of their notifications are currently omitted variables because the data have been unavailable.[26] This motivated me to simply drop the 2010 wave in my analysis thus far, although including it does not appreciably alter results. Testing for twice-lagged effects of biomarker notification requires including the 2010 wave, and results of that analysis suggest that the effects on diagnoses and medication usage appeared to persist and possibly strengthen. But those results did not support the hypothesis that effects on self-reported weight, drinking, and exercise persisted.

[25] An exception was an approach based on regression discontinuities, and the likely reason is that the HRS is too small for the approach to work well in this context. In the case of A1c readings and weight, for example, there was too much noise in the relationship between the two among the sample above the reading of 7.0 percent to generate a clear discontinuity in behavior as one would expect there to be. Graphs of diabetes diagnosis against A1c similarly did not depict sharp cutpoints around the threshold. Panel FE results suggest that conditioning on preexisting diagnosis appears to be important for revealing many effects, but doing so unfortunately reduces sample size even further and increases noise enough so that clear regression discontinuities remain elusive. Multivariate regression seems better suited to understanding behavior when sample size is limited and many characteristics affect outcomes.

[26] Email correspondence with the HRS team revealed that the 2008 biomarker group was also notified in a staggered fashion because of laboratory issues and legal constraints in New York State and California.

Another interesting and feasible test of the persistence of behavioral effects is whether objective weight had fallen between 2006 and 2010, the second round of biomarker collection for the group that went first, among the undiagnosed subgroup with high A1c. As shown in Table 8, objective weight fell rather uniformly among respondents who submitted biomarkers in both waves. Although weight fell between 1.6 and 1.7 percent faster among those with high A1c in 2006, the difference was not statistically significant mostly because of small subsample size and the strength of the background trend, the latter of which presumably reflected the aging of the panel.

4 Discussion

The evidence suggests that notifications of biomarker screens outside normal range in the 2006 wave of the Health and Retirement Study (HRS) triggered significant changes in outcomes and behavior two years later among select subgroups of respondents under certain circumstances. After biomarker collection, HRS interviewers immediately notified respondents of very high blood pressure by leaving behind a high BP card, and HRS later mailed out to respondents the full results of their blood pressure, hemoglobin A1c, total cholesterol, and HDL cholesterol readings, as well as indications of whether those readings were outside normal range. Three types of the five possible notifications were "rare and dangerous": the high BP card received by 5.8 percent of respondents, high A1c registered by 5.4 percent, and low HDL cholesterol registered by 5.4 percent. Two of these notifications, very high blood pressure above 160/110 and A1c above 7.0, each produced the most extensive and interesting effects on outcomes and behavior two years later in the panel.

By far the largest and most statistically significant impacts of these screens were on rates of new disease diagnosis, which by definition affected the previously undiagnosed. Respondents who already knew they had the disease and either were unable or unwilling to control the biomarker, who can be termed "noncompliers" for simplicity, also responded but in different ways. Rates of diagnosis among the previously undiagnosed jumped 17 percent for recipients of the high BP card and almost 40 percent for high A1c screens.

Connected with these were similarly large increases in usage rates of associated medications. The prevalence of such undiagnosed cases among the 2006 biomarker group was small, only about 1.5 percent in the case of the high BP card and 0.7 percent with high A1c. Subsequent patterns in self-reported doctor's diagnosis suggest either that significant shares were false positives or that respondents or their doctors did not find the information salient. Applying the observed increases in diagnosis rates by 2008 to the prevalence of abnormal screens among those without a diagnosis produces estimates of the lower bound of undiagnosed disease prevalence in the 2006 sample overall of only 0.3 percent for each of the two.

Behavioral adjustment was apparent and arguably strongest among undiagnosed diabetics with high A1c, and it also appeared among spouses of these and other respondents. The 0.7 percent of respondents with high A1c and without a diabetes diagnosis reported losing an average of 2.2 percent of their body weight by the following wave, and they also reported less drinking and more frequent exercise. Trends in objective measures of weight, collected in 2006 and again in 2010, lend support to this finding but are not definitive. Spouses of undiagnosed respondents with high A1c also reported increased frequency of light exercise, roughly in line with what their partners reported. These patterns were not found among the 4.7 percent with high A1c who were already known diabetics, who instead suffered deteriorations in self-reported disability rates and health status. Interestingly, spouses of these diabetics reported reductions in their own weight of 1.5 percent, even while spouses of the undiagnosed with high A1c reported effectively no weight loss. Exercise and drinking among these spouses of noncompliers was unaffected, which could indicate that their weight loss came through a change in diet, which the HRS does not measure. Were that the case, it would suggest households may take a multi-staged approach to diabetes management, focusing at first on exercise and then later on diet. Signs of rising disability among those with uncontrolled diabetes imply that such a strategy may become more feasible than exercise promotion as the disease progresses.

In contrast, changes in behavior were more visible among noncompliers who received the high BP card than among the undiagnosed who received it. This might be expected if high blood pressure is viewed as less serious at onset than is type 2 diabetes, which

seems plausible. Respondents already diagnosed with high blood pressure who received the high BP card appeared to significantly reduce their smoking, by 2.3 percentage points or about one fifth of the baseline prevalence, and also reduced their drinking intensity. Their frequency of light exercise actually fell, possibly reflecting a reallocation of household chores away from the respondent. Although they did not report picking up the slack, spouses of these respondents did report less binge drinking, mirroring the effects of the screen on own behavior and suggesting that binge drinking tends to be a joint activity.

Exactly why uncontrolled high blood pressure appears to trigger changes in own behavior while uncontrolled diabetes apparently does not remains somewhat of an open question. Only a very small share of respondents with high A1c are likely to be type 1 or lifelong diabetics, probably less than 1 percent. It is conceivable that diabetic HRS respondents and their doctors may have set A1c targets above 7.0. The reason for the asymmetry may well be rising disability among those with uncontrolled diabetes, which presumably impedes exercise. Smoking is addictive, which may explain why it does not appear to respond even though it probably should be a part of diabetes control (Gunton et al., 2002). Drinking is already reduced among this subsample, per Table 1, and may not be easy to lower any further.

As Tables 1 and 3 reveal, the characteristics of respondents with high A1c suggest that uncontrolled diabetes is a condition associated with minority identity and lower socioeconomic status. Whether these characteristics are important for behavior because they constrain choices or because they alter preferences is a perennial question in health economics, and the present study offers no new insights. With added detail from the restricted HRS geocodes file, perhaps elucidating neighborhood characteristics, some headway might be made in understanding the determinants of response and nonresponse to notifications. More work could also be done to examine psychological covariates of these behaviors, since the EFTF interviews during biomarker waves ended with the interviewers' leaving behind new "lifestyle" questionnaires on psychosocial topics.

Ultimately, these data may thus provide insights about preventable hospitalizations among older Americans, some of which are attributable to uncontrolled hypertension or

diabetes, especially among poor communities, and to access to care (Bindman et al., 1995; Jiang, Russo and Barrett, 2009). Although it does not ask separately about emergency room visits, the HRS does ask about the frequency of overnight hospitalization between waves, which was highest in 2006 among subsets of "noncompliers" with very high BP or high A1c and a preexisting diagnosis. Known diabetics with high A1c in 2006 were significantly more likely to report overnight hospitalization (panel FE coefficient of 0.096, t-stat of 3.58) and more stays (0.170, t-stat of 2.11) in 2008. But it is unclear what to draw from this, besides motivation for a fuller assessment of their characteristics and what drives them to hospitalization compared with other diabetics. I leave such efforts for future inquiry.

One coefficient in Table 5 suggests that low HDL cholesterol screens may trigger a doubling of the typical 2-year increase in new diabetes prevalence. It seems wise not to place much emphasis on a single result like this, especially when there appears to be no connection between high cholesterol screens and subsequent diabetes. But it is interesting in light of concerns about statins (Sattar et al., 2010) and is probably worth future study.

Stepping back, important questions are whether these behavioral responses to biomarker notification are empirically important on average and thus potentially policy relevant. The intent-to-treat estimates of the average impacts of biomarker collection in the sample are effectively zero, except insofar as they may reduce doctor visits and medication usage by 1 or 2 percent, and this must be because screening does not generate much salient information on average. Undiagnosed "rare and dangerous" conditions are uncommon in the sample, and while uncontrolled conditions are less rare, it appears that notification of their uncontrolled nature does not change behavior by a lot. Averaged across all respondents, a policy of collecting biomarkers and informing about results and normal ranges will have few tangible effects because those treated by salient information are a small subset, and while their conditions change, they do not change drastically. A more targeted intervention would have larger average effects. A question I leave for future study is whether a policy of screening based on observable characteristics like those in Table 3 could produce average treatment effects that would be worth the costs. Certainly these findings suggest there are unmet surveillance and care needs among Americans aged 50 and over, but they appear to be

concentrated among relatively small subgroups.

A key methodological question is whether these effects on those treated by the notifications are in fact attributable to notification and thus to the intervention as defined by collecting biomarkers and notifying abnormal screens, or whether they instead simply reflect the impacts of the biomarker levels themselves and thus would have happened regardless. Because all respondents who screen abnormally were also notified, there is no way to answer this question to the standards of a randomized controlled trial. But my reading of the evidence is that these effects are indeed associated with the arrival of information and not the biomarker level. Key results pass the falsification test of contemporaneous association, and the robustness of results across specifications and samples lends further credence to the view that the effects are causal. Further examination of this question with the 2008 results may shed more light on the topic.

The impacts of biomarker notification on economic outcomes presumably via their intermediate effects on health may be interesting to explore. Labor supply and spending might respond to screens out of normal range if they produce new health conditions, and the HRS panel includes respondents who are not yet retired. Labor force participation among HRS respondents with high A1c screens was about the sample average of 54 percent, while it was about 62 percent among those who received the high BP card. For those under age 65 in the U.S., of course, the onset of a new health condition might actually increase participation for insurance purposes, as long as disability permits working. New disease onset before retirement seems likely either to generate precautionary saving or financial stress.

Implications of these results range from practical insights to new understanding of behavior and unmet health services needs in the U.S. population around the age at retirement, and ultimately to informing policy. Methodologically, this research contributes the insight that the strong correlation between biomarker collection and mode of interview will probably affect self-reported measures. Collecting biomarkers in a cross-sectional survey like the core NHANES will affect self reports relative to a telephone survey but not relative to itself, if all respondents are interviewed in person. Panel studies like HRS, Add Health, and others that include biomarking in some but not all waves are more likely to be impacted.

References

Banks, James, Alastair Muriel and James P. Smith. 2010. "Attrition and Health in Ageing Studies: Evidence from ELSA and HRS." *Rand Labor and Population Working Paper Series* WR-784:1–40.

Berlowitz, Dan R., Arlene S. Ash, Elaine C. Hickey, Robert H. Friedman, Mark Glickman, Boris Kader and Mark A. Moskowitz. 1998. "Inadequate Management of Blood Pressure in a Hypertensive Population." *New England Journal of Medicine* 339(27):1957–1963.

Bindman, Andrew B., Kevin Grumbach, Dennis Osmond, Miriam Komaromy, Karen Vranizan, Nicole Lurie, John Billings and Anita Stewart. 1995. "Preventable Hospitalizations and Access to Health Care." *JAMA* 274(4):305–311.
URL: *+ http://dx.doi.org/10.1001/jama.1995.03530040033037*

Gunton, Jenny E., Linda Davies, Errol Wilmshurst, Greg Fulcher and Aidan McElduff. 2002. "Cigarette Smoking Affects Glycemic Control in Diabetes." *Diabetes Care* 25(4):796–797.
URL: *http://care.diabetesjournals.org/content/25/4/796.2.short*

Halpern-Manners, Andrew and John Robert Warren. 2012. "Panel Conditioning in Longitudinal Studies: Evidence From Labor Force Items in the Current Population Survey." *Demography* 49:1499–1519.

Hyman, David J. and Valory N. Pavlik. 2001. "Characteristics of Patients with Uncontrolled Hypertension in the United States." *New England Journal of Medicine* 345(7):479–486.

Imbens, Guido W. and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2):467–475.

International Expert Committee. 2009. "International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes." *Diabetes Care* 32(7):1327–1334.

Jiang, H. Joanna, C. Allison Russo and Marguerite L. Barrett. 2009. "Nationwide Frequency and Costs of Potentially Preventable Hospitalizations, 2006." *HCUP Statistical Brief* 72.

U.S. Agency for Healthcare Research and Quality, Rockville, MD.
URL: *http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.pdf*

Juster, F. Thomas and Richard Suzman. 1995. "An Overview of the Health and Retirement Study." *Journal of Human Resources* 30, supplement:S7–S56.

Rosero-Bixby, Luis and William H. Dow. 2012. "Predicting mortality with biomarkers: a population-based prospective cohort study for elderly Costa Ricans." *Population Health Metrics* 10(11):1–15.

Sattar, Naveed, David Preiss, Heather M. Murray and et al. 2010. "Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials." *Lancet* 375(9716):735–42.

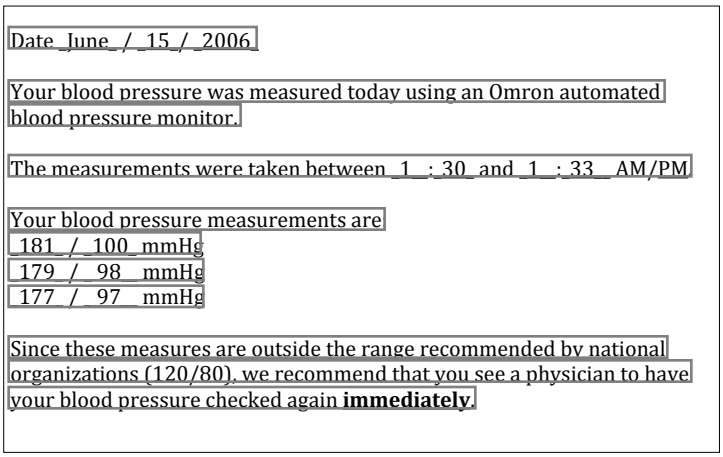
Verdecchia, Paolo, Giuseppe Schillaci, Claudia Borgioni, Antonella Ciucci and Carlo Porcellati. 1997. "Prognostic Significance of the White Coat Effect." *Hypertension* 29(6):1218–1224.

Weinstein, Maxine, James W. Vaupel and Kenneth W. Wachter, eds. 2007. *Biosocial Surveys*. Washington: National Academies Press.

Weir, David. 2007. Elastic Powers: The Integration of Biomarkers into the Health and Retirement Study. In *Biosocial Surveys*, ed. Maxine Weinstein, James W. Vaupel and Kenneth W. Wachter. Washington: National Academies Press chapter 4, pp. 78–95.

Weir, David R. 2010. "Socio-economic Status and Mortality: Perceptions and Outcomes." Unpublished manuscript presented at the 2010 annual meeting of the Population Association of America.

Figure 1: A high blood pressure card from HRS



Notes: The appearance of this card is derived from correspondence with Health and Retirement Study investigators. The listed levels are the average readings in the 2006 sample for respondents who received the high blood pressure card. Emphasis in the actual cards as shown. Interviewers were instructed to leave this card behind with respondents whose minimum systolic BP was greater than 160 or whose minimum diastolic BP was greater than 110.

HRS | HEALTH AND RETIREMENT STUDY

UNIVERSITY OF MICHIGAN
Institute for Social Research
426 Thompson St.
Ann Arbor, Michigan 48104
Ph. (866) 611-6476
http://hrsparticipants.isr.umich.edu

John Q. Respondent
1000 Main Street
Some Town, XY 12345

DATE OF TEST: June 15th, 2006

Dear Mr. John Q. Respondent,

Thank you for participating in the Health and Retirement Study. The test results of your cholesterol and hemoglobin A1c tests have been returned from the laboratory. On the enclosed page you will find your results for your blood pressure readings, cholesterol (both total and HDL cholesterol), and hemoglobin A1c tests. As you know we cannot provide any medical advice to you. If you have any questions about these results please contact your doctor.

We look forward to your continued interest and participation in all aspects of this nationally recognized and highly respected study. If you have questions about this study, please contact us at **1-800-123-4567**.

Sincerely,

Jane Z. Scientist, Ph.D.
Director – Health and Retirement Study

If you have questions about your rights as a research participant, or wish to obtain information, ask questions or discuss any concerns about this study with someone other than the researcher(s), please contact the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board, 540 E Liberty St., Ste 202, Ann Arbor, MI 48104-2210, (734) 936-0933 (or toll free, (866) 936-0933), irbhsbs@umich.edu. Please refer to protocol HUM00061128.

Blood Pressure Readings:

	Systolic		Diastolic	
Time 1:	131	/	80	mmHg
Time 2:	129	/	79	mmHg
Time 3:	128	/	78	mmHg

N/A – reading error or invalid result

According to the American Heart Association, systolic pressure (top number) of 120 mmHg or higher, and a diastolic pressure (bottom number) of 80 mmHg or higher may indicate hypertension (high blood pressure). The Health and Retirement Study may not measure blood pressure in the same way that your blood pressure is measured in your doctor's office. However, if your blood pressure is 120/80 mmHg or higher, you should see your physician or other health professional to recheck this result and discuss it with them.

Total Blood Cholesterol: 190 mg/dl

According to the American Heart Association, total cholesterol of 200 mg/dl or higher may indicate hypercholesterolemia (high cholesterol). If your total cholesterol is 200 mg/dl or higher you should see your physician or other health professional to recheck this result and discuss it with them.

HDL Blood Cholesterol: 55 mg/dl

According to the American Heart Association, high-density lipoprotein (HDL) cholesterol level of greater than 40 mg/dl helps to reduce the risk of heart disease. If your HDL cholesterol is lower than 40 mg/dl you should see your physician or other health professional to recheck this result and discuss it with them.

Hemoglobin A1C: 5.6%

The hemoglobin A1c test shows the average amount of sugar in your blood over the last three months. The American Diabetes Association recommends that the goal of this result should be less than 7%. If your result is 7% or higher you should see your physician or other health professional to recheck this result and discuss it with them.

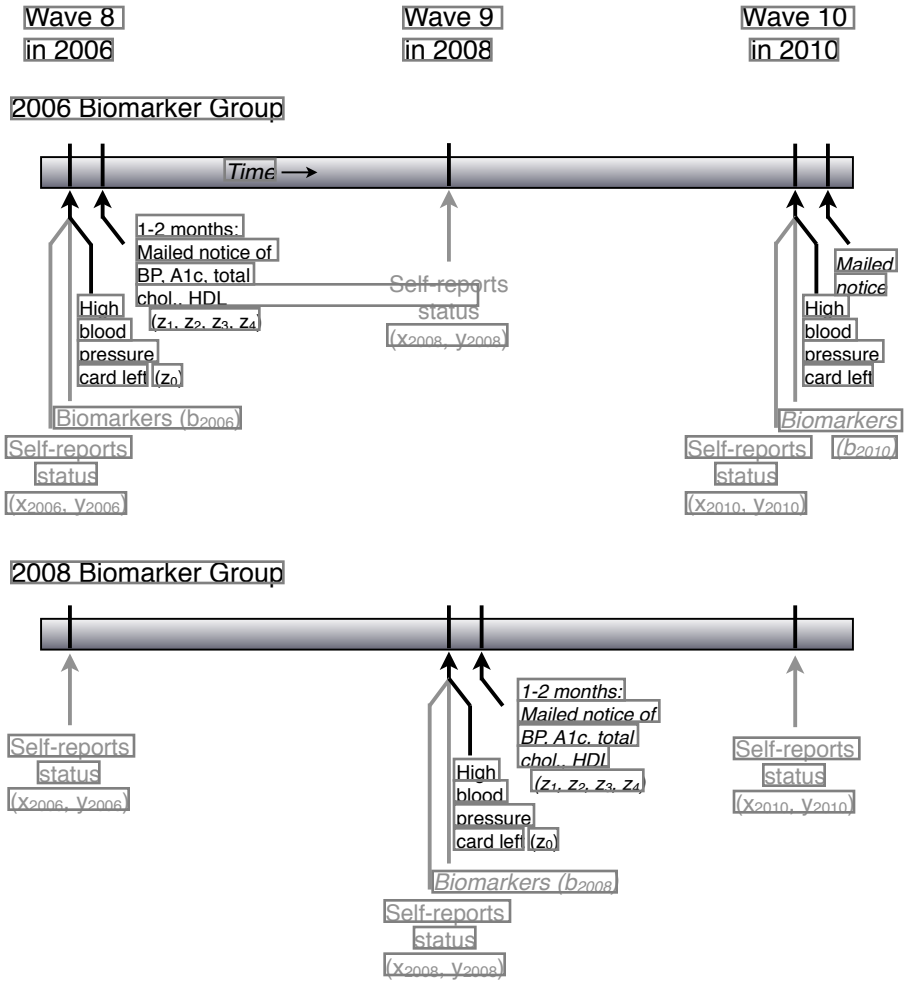
For more information, please refer to the following web sites:

American Heart Association	http://www.heart.org
American Diabetes Association	http://www.diabetes.org
National Institutes of Health	http://nih.gov
National Institute on Aging	http://www.nia.nih.gov

Figure 2: A biomarker notification letter from HRS

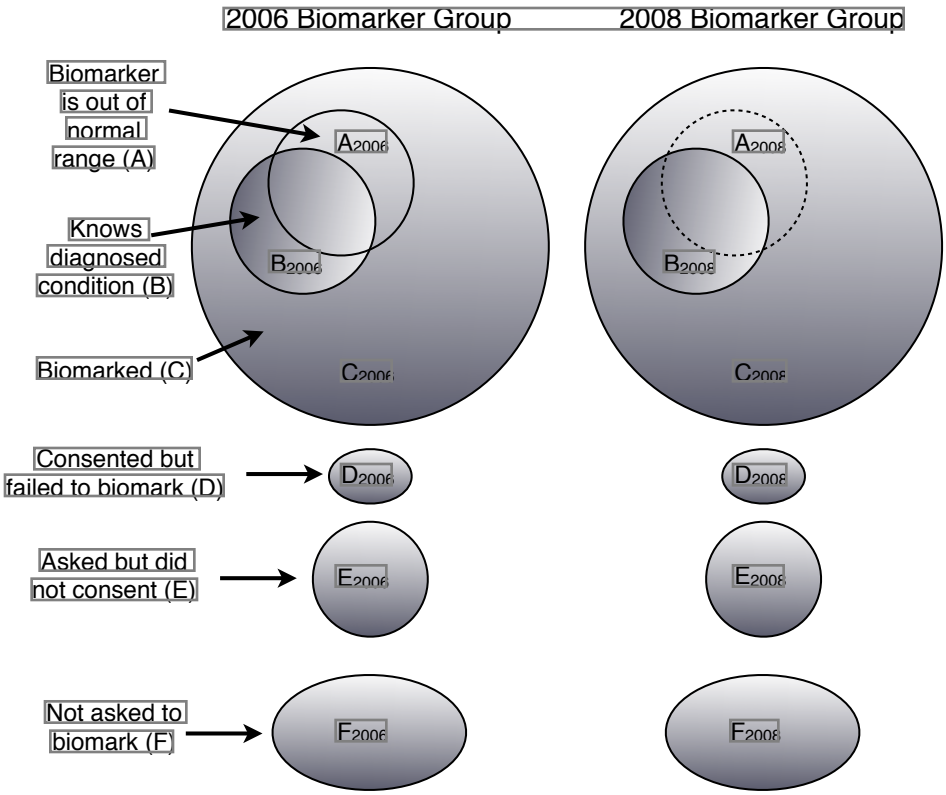
Notes: This letter is based on a Word template provided by Health and Retirement Study investigators. The listed levels are the average readings in the 2006 sample. All respondents who submitted biomarkers received a letter like this one. See the text for details.

Figure 3: Timing of measurements and treatments during biomarking in the HRS



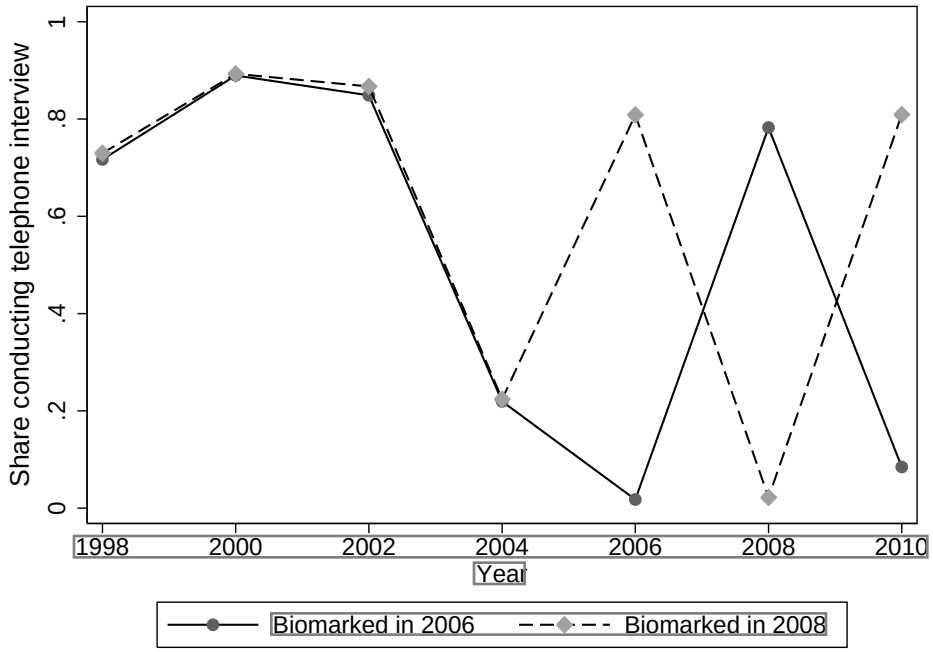
Source: Health and Retirement Study, various waves.

Figure 4: Disposition of the HRS panel during biomarking in 2006



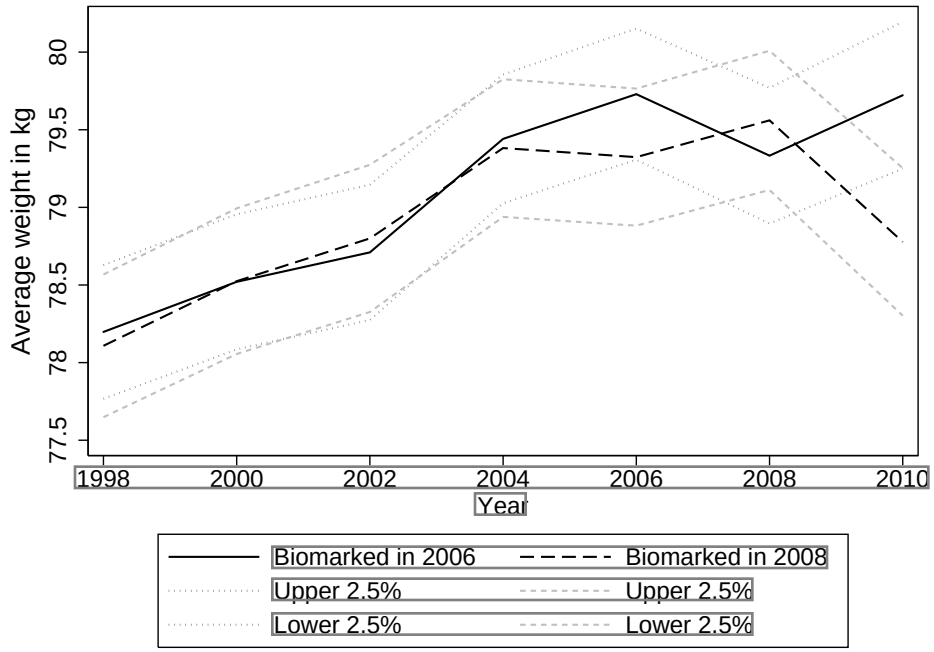
Source: Health and Retirement Study, 2006 various wave. All depicted subgroups are present in both the 2006 and 2008 waves. The collection of subgroups on the left are those randomly selected to submit biomarkers in 2006; those on the right are the remaining half selected to submit in 2008. The intent-to-treat (ITT) estimate of the effect of biomarker collection is the simple comparison of the two groups in 2008. An average treatment effect on the treated associated with screening outside normal range would be revealed by comparing group A_{2006} with the unobservable group A_{2008}.

Figure 5: Mode of interview in the Health and Retirement Study



Source: Health and Retirement Study, various waves, and the RAND HRS file version L

Figure 6: Self-reported weight across biomarker groups in the Health and Retirement Study



Source: Health and Retirement Study, various waves, and the RAND HRS file version L

Table 1: Characteristics of HRS respondents asked to submit biomarkers in 2006

	All assigned resp's	Biomarker submitters	High BP card	High BP, no card	High A1c	High total chol.	Low HDL chol.
Number	8,587	7,127	412	3,809	387	2,399	383
Average age	67.2	67.0	71.6	67.7	66.8	65.4	67.5
Percent							
Male	41.1	40.6	41.0	45.6	46.0	33.7	73.1
African American	14.1	13.2	23.1	13.9	22.5	10.9	11.2
Hispanic	9.1	8.0	9.2	7.8	16.0	8.6	8.6
Married or coupled	67.9	68.8	59.5	69.4	67.7	70.4	77.5
Does not own home	19.7	18.0	24.8	18.0	29.7	16.5	18.3
Does not report health insurance	5.7	5.3	5.8	5.2	7.8	7.3	7.6
Percent with condition this wave							
High blood pressure	56.2	55.8	74.5	61.7	68.7	49.7	61.9
Diabetes	19.6	19.3	26.9	20.0	87.9	13.0	28.2
Heart problems	23.5	23.6	26.7	22.5	28.9	16.1	32.9
Percent with new condition this wave							
High blood pressure	5.1	5.1	4.9	5.3	3.9	5.1	5.5
Diabetes	2.8	2.8	3.2	2.8	6.0	2.1	5.2
Heart problems	3.5	3.4	3.0	3.1	5.7	1.8	3.9
Percent							
Visited doctor past 2 years	94.8	95.0	90.3	94.7	94.3	93.4	93.2
Regularly taking Rx past 2 years	80.9	81.3	81.8	81.8	91.7	74.7	81.7
Taking BP medication	50.4	50.0	65.2	54.8	62.3	42.9	55.6
Taking diabetes medication	16.3	16.0	22.8	16.5	80.9	10.0	22.7
Taking cholesterol medication	38.9	39.3	38.7	40.8	55.4	25.7	45.5
Self-reported health status 1-5							
1 = excellent, 5 = poor	2.81	2.76	2.98	2.75	3.40	2.64	2.96
Sum of some difficulty on 0-5 ADLs	0.31	0.24	0.26	0.23	0.39	0.21	0.34
Self-reported weight in kg	79.8	79.9	80.6	81.4	88.8	78.6	91.3
Objective weight in kg	80.5	80.5	81.1	82.0	88.3	79.7	90.9
Percent smokes now	13.7	13.5	13.0	13.0	12.0	14.8	18.8
Cigarettes per day	2.16	2.12	1.97	2.03	1.79	2.44	3.48
Percent drinking alcohol now	69.4	70.3	72.4	75.2	47.7	75.0	49.0
Drinking days per week	1.15	1.19	1.16	1.25	0.41	1.29	0.74
Drinks per drinking day	0.69	0.70	0.72	0.75	0.48	0.75	0.49
Binge drinking days in last 90	1.28	1.28	1.49	1.53	0.77	1.33	0.97
Phys. activity freq. 1-5, 1 = most, 5 = least							
Vigorous	4.06	4.02	4.16	4.03	4.42	3.99	4.14
Moderate	2.85	2.78	2.92	2.77	3.16	2.73	2.98
Light	2.52	2.45	2.53	2.45	2.78	2.31	2.69

Notes: Data are from the 2006 wave of the HRS, distributed via the HRS website and from the RAND HRS file version L. The universe is all individuals randomly selected to submit biomarkers in 2006, including blood pressure (BP), blood sugar (A1c), total and HDL or "good" cholesterol, and other measures, who were also present in the 2008 wave. The second column shows respondents who submitted any biomarkers at all, including the physical measures such as grip, breath, and balance. Respondents in the third column received the High BP card if the minimum blood pressure out of three measures exceeded 160 mmHg systolic or 110 diastolic. Those in the fourth column had minimum blood pressure outside the recommended ranges of 120 systolic or 80 diastolic but did not receive the High BP card. Those in the fifth column had hemoglobin A1c readings of 7.0 percent or higher. Respondents in the sixth column had total cholesterol readings of 200 mg/dL or higher. Those in the seventh column had HDL cholesterol readings below 40 mg/dL. For the 2006 prevalence of high BP, diabetes, and heart problems, I report the raw responses, in which respondents can dispute reports from previous waves. New incidence is calculated using respondents' current statements about present and past conditions, which they may have updated. Average intensive measures of smoking and drinking are calculated by assigning zeros to abstentions.

Table 8: Trends in objective weight between 2006 and 2010 by A1c reading in 2006

	Had high A1c	Had normal A1c	Had high A1c and no diabetes condition	Had normal A1c and no diabetes condition	Diff in diffs	Diff in diffs
	(1)	(2)	(3)	(4)	(1)–(2)	(3)–(4)
Δ log weight	−0.028***	−0.012***	−0.029***	−0.011***	−0.016	−0.017
	(0.011)	(0.002)	(0.012)	(0.002)	(0.011)	(0.012)
N	233	4,234	33	3,623		

Notes: Data are from the 2006 and 2010 waves of the HRS. Each element in the first four columns is the change between 2006 and 2010 in the log objective weight for the group indicated. The last two columns show difference-in-differences estimates of the effect of a high A1c screen above 7.0 percent in 2006.