

# Genetic Adverse Selection: Evidence from Long-Term Care Insurance and Huntington Disease

Emily Oster\*  
University of Chicago and NBER

Ira Shoulson  
University of Rochester

Kimberly Quaid  
Indiana University School of Medicine

E. Ray Dorsey  
University of Rochester

June 8, 2010

## Abstract

Individual, personalized genetic information is increasingly available, leading to the possibility of greater adverse selection over time, particularly in individual-payer insurance markets. We use data on individuals at risk for Huntington disease (HD), a degenerative neurological disorder with significant effects on morbidity, to estimate adverse selection in long-term care insurance. We find strong evidence of adverse selection: individuals who carry the HD genetic mutation are up to 5 times as likely as the general population to own long-term care insurance. This finding is supported both by comparing individuals at risk for HD to those in the general population and by comparing across tested individuals in the HD-risk population with and without the HD mutation.

## 1 Introduction

Personalized genetic information is increasingly available. Genes associated with increased risk of breast cancer, colon cancer, Parkinson disease and Alzheimer disease, among others, have been identified and testing for these genes is becoming much more common. Continued advances in technology and knowledge of the human genome are likely to bring even more sophisticated and precise testing, for these and other conditions.<sup>1</sup> This testing, in turn, is likely to increase the degree of private information that individuals have about their mortality and morbidity risks.<sup>2</sup> In this paper we explore the possible impact of this increased information on the markets for long-term care insurance.

---

\* Authors are writing on behalf of the Huntington Study Group PHAROS Investigators. We are grateful to Amy Finkelstein, Larry Katz, Lee Lockwood, Kathleen McGarry, Jesse Shapiro, Heidi Williams and participants in a seminar at the University of Chicago for helpful comments.

<sup>1</sup>For example, there has been significant recent work on genome wide association studies, which look at entire individual genomes to identify markers that are common in individuals with a given condition (e.g. Burton et al, 2007).

<sup>2</sup>Interestingly, it seems possible that much of this testing could take place outside of traditional clinical environments – for example, through companies like “23 and Me”.

There is a large literature in economics which estimates adverse selection in insurance markets.<sup>3</sup> Increasingly, the findings in this literature point to the importance of multiple dimensions of preference heterogeneity (Finkelstein and McGarry, 2006; Cohen and Einav, 2007; Cutler et al, 2008; Fang, Keene and Silverman, 2008; Einav et al, 2009). Although private information about risk type can and does drive insurance purchases, this may be outweighed (or reinforced) by other dimensions of heterogeneity. For example, private risk information about mortality may be counteracted by the fact that people with a lower preference for risk also own more life insurance, and these people are, on average, *less* likely to die early (McCarthy and Mitchell, 2003). However, this heterogeneity may also work in the opposite direction – to reinforce adverse selection – for a product like annuities (Einav et al, 2009).

For long-term care insurance, which we consider in this paper, it appears that preference heterogeneity serves to offset adverse selection due to private information; studies typically find little or no correlation between ownership and risk realization (Finkelstein and McGarry, 2006; Cutler et al, 2008). However, while there is limited reason to expect changes in preference heterogeneity over time, increased availability of genetic testing has the potential to dramatically increase private information in this market. In principle, this information could change the overall correlation between ownership and payouts and have large effects on the viability of this market in the long-term, particularly as in light of legislation which limits insurer ability to observe individual genetic information (e.g. United States House of Representatives, 2007).

Evaluating whether we will see increased adverse selection as genetic information increases requires a setting in which (a) individuals have a large amount of private information, which the researcher can observe and (b) that this is information individuals would like to act on. This paper takes advantage of a setting in which both requirements are satisfied, using a dataset on individuals at risk for Huntington disease (HD). We combine this with data on individuals without HD risk to estimate adverse selection in long-term care insurance.

HD is a degenerative neurological disorder caused by an inherited genetic mutation on chromosome 4 that affects roughly 1 in 10,000 individuals in Caucasian populations. Because of the inherited nature of the disease individuals have significant private information about their disease risk. Those with one parent with the disease know they have a 50% chance of developing it, and those who have taken a genetic test and carry the affected gene know they will develop the disease, assuming they do not die earlier from something else. A perfectly predictive genetic test for HD has been available since 1993.

---

<sup>3</sup>See, for example, Finkelstein and McGarry, 2006; Cawley and Philipson, 1999; Finkelstein and Poterba, 2004; McCarthy and Mitchell, 2003; Chiappori and Salanie, 2000; Cutler et al, 2008; Smart, 2000; De Meza and Webb, 2001.

Further, the information is extremely relevant. Individuals who carry the HD genetic mutation begin to deteriorate neurologically (mentally and physically), typically between the ages of 30 and 50, and death follows an average of 20 years after onset. Individuals become increasingly disabled over this period (Walker, 2007). There is no cure for HD, and only limited treatment options. The long, likely expensive, disability period suggests that long-term care insurance would have significant value.

We focus on long-term care insurance in this paper for a number of reasons. First, we argue that it is a particularly clear case for exploring adverse selection, since the primary difference between individuals with and without the genetic mutation is the probability of needing long-term care. This is in contrast to life insurance, where individuals with the genetic mutation are more likely to die early but also experience very different income streams during their lifetime *as a result of the disease*. Second, most long-term care insurance is purchased through individual policies, enhancing the possibility for adverse selection heavily affecting the aggregate functioning of the market. Finally, although the long-term care insurance market is small relative to, for example, health insurance, it is still large in absolute terms – 1.2% of GDP in 2004 – and in the absence of private insurance coverage most expenditures are covered by the government, through Medicaid (Brown and Finkelstein, 2009).

We test for the presence of adverse selection in long-term care insurance using data from a prospective cohort study (PHAROS) of approximately 1000 individuals at risk for HD. At enrollment into this sample, individuals had one parent who had HD, had not undergone genetic testing and were asymptomatic, so their chance of carrying the genetic mutation is approximately 50%.<sup>4</sup> Participants have been re-surveyed approximately every 9 months from the time of enrollment (1999 or 2000) to the present, and over this period approximately 10% have pursued testing for the genetic mutation.

We perform several tests using these data. First, we compare insurance ownership among the population at risk for HD to individuals in the general population, drawn from the Health and Retirement Survey (HRS). Second, for a subset of the tested individuals we are able to observe their genetic status<sup>5</sup> and identify individuals who know they have the genetic mutation (100% risk) and those that know they do not have the mutation (0% risk). We compare each of these two groups to the untested individuals with 50% risk. Finally, we can compare individuals who have been tested and know they carry the genetic mutation to those who have been tested and know they do not. This last analysis is our cleanest test. However, the sample size in this final test is small: we have only 71 tested individuals

---

<sup>4</sup>The fact that we observe only individuals who have not been tested at enrollment introduces possible selection issues, which are discussed in more detail in the context of the results.

<sup>5</sup>We infer individual test results based on subsequent disease development, or subsequent responses to questions about their self-perception of carrying the genetic mutation; this is described in more detail in the data section.

for whom we observe genetic status versus roughly 900 at-risk individuals who choose not to be tested. For this reason, the comparison of untested individuals with the general population in the HRS provides an important part of the analysis, as it reinforces the findings with a larger sample size and more statistically robust claim.

Using these data, we find evidence of significant adverse selection. In the general population (the HRS), 10% of individuals own long-term care insurance. In the PHAROS population, 27% of untested individuals own long-term care insurance (this proportion is extremely similar to the overall proportion in the PHAROS population, since most individuals are untested). Among those who pursued genetic testing and know they carry the mutation responsible for HD, ownership is close to 50%.<sup>6</sup> The rate of long-term care insurance ownership among those individuals who know they are at 100% risk for developing HD is 20 to 30 percentage points higher than those who also pursued genetic test and know they do not carry the genetic mutation for HD. This primary result of the paper can be seen simply in Figure 1.

An important issue in interpreting these results is whether individuals in the HD population are comparable to those in the HRS, other than their HD risk. The primary concern is that individuals in the HRS are, on average, older than individuals in the HD population. We attempt to address this by including controls for demographics and find that the effects of HD are not driven by demographic differences. In addition, comparing individuals within the HD at-risk population provides an extremely well identified test: these individuals are, in expectation, identical *ex ante* – nature “randomizes” which of them receive the affected copy of the gene. Finally, because long-term care insurance ownership tends to be higher among older individuals, the older population in the HRS suggest that, if anything, our results are an underestimate of adverse selection in long-term care.<sup>7</sup>

We add to a small existing literature on insurance purchases and genetic risk, which so far has had somewhat mixed results with small sample sizes (Armstrong et al, 2003; Aktan-Collan et al, 2001; Zick et al, 2000; Zick et al, 2005; Taylor et al, 2009). In the most closely related of these papers, Zick et al (2005) show increases in long-term care purchases as a result of an intervention informing individuals about an increased risk for Alzheimer disease. The analysis here provides sharper evidence on this question, because HD has a clearer genetic risk, our sample is much larger and the population contains individuals at widely different levels of risk.

Our results may be particularly policy relevant in light of the recent health care legislation (United

---

<sup>6</sup>Throughout the paper we will sometimes refer to individuals who are tested and know they carry the genetic mutation as having a “positive” test result.

<sup>7</sup>In an appendix we show estimates based on an alternative dataset (the Consumer Expenditure Survey), which is less well suited for many reasons, but does have a population with a closer age match. If anything, the results are stronger with this comparison group.

States Senate, 2010). The health care bill contains the “Community Living Assistance Services and Support Act” (CLASS Act), a provision which requires the government to provide a long term care insurance product. The premiums for this product can be based *only* on age, opening the door for significant insurance unraveling if there is adverse selection in purchases. Our results suggest this is a real concern; we discuss this briefly in the conclusion.

The rest of the paper is organized as follows. Section 2 gives background on Huntington disease and on the insurance markets considered here, as well as describing the data. Section 3 presents our empirical strategy. Section 4 shows our results and Section 5 concludes.

## 2 Background and Data

### 2.1 Background on Huntington Disease

In this section we provide only a brief overview of Huntington disease; for a fuller clinical discussion, please see Walker (2007).

Huntington disease (HD) is a degenerative neurological disorder affecting an estimated 30,000 individuals in the United States. Individuals with the disease typically begin to manifest symptoms in middle age (30-50) although age of onset varies from early childhood to as late as 80. The symptoms of HD include a movement disorder, impaired cognition, and psychiatric disturbances. The movement disorder consists of random, uncontrollable dance-like movements of the face, trunk, and extremities. The cognitive dysfunction includes memory loss and impaired higher order thinking, which affects an individual’s ability to work. The psychiatric disturbances are wide ranging and include depression, changes in personality, anxiety, and psychosis. The disease is progressive. Individuals will need increasing levels of supportive and often institutional care for many years. Death follows approximately 20 years after onset. Although it is difficult to be precise about the length of time individuals need long-term care, estimates suggest that on average they will need some type of care for at least ten years, and 70-80% end up in a nursing home at some point, for an average of around 6 years (Bolt, 1970; Nance and Saunders, 1996; Walker et al, 1981; Harper, 1996).

HD is a genetic disorder due to an inappropriate expansion in the huntingtin gene on chromosome 4. The disease is inherited in an autosomal dominant manner, such that individuals who have a parent with HD will have a 50% chance of inheriting the genetic mutation and subsequently developing the disease. Development of the disease without an affected parent is extremely rare. There is no cure for HD or treatment that slows the progression. In 2008, tetrabenazine was approved to treat the HD movement

symptoms, but it does not treat the cognitive or psychiatric symptoms, delay disease onset, or slow disease progression.

Since 1993, a test for the HD genetic mutation has been available. However, testing rates are fairly low: 5-10% of the at-risk population reports predictive testing (Meyers, 2004). Testing for HD is a significant decision for at-risk individuals and typically involves a period of pre-testing counseling. The lack of any treatment or cure, and the fear of being unable to live with knowing that one carries the mutation are significant barriers (Oster et al, 2008).

## 2.2 Background on long-term Care Insurance

Long-term care insurance is designed to cover expenditures for either home care or nursing home care for the elderly. This insurance is typically purchased through individual or small-group policies. Brown and Finkelstein (2009) provide an excellent summary of the current literature within economics on long-term care insurance markets. Long-term care insurance ownership is fairly limited in the general population, with ownership rates around 10% for individuals in the 60-85 age range.

Brown and Finkelstein (2007) describe characteristics of a typical long-term care insurance contract: an “elimination” period (analogous to a deductible), a maximum benefit period of 1-5 years, and a maximum daily benefit (which is typically below what a day in a nursing home costs, although probably above the cost of a day of home care). There is significant overlap between services offered by Medicaid and long-term care insurance (Brown and Finkelstein, 2007). However, in the overall population, one third of long-term care expenses are paid out of pocket, suggesting that neither Medicaid nor insurance is providing comprehensive coverage (Brown and Finkelstein, 2009).

From the perspective of our analysis, there are at least two features of this market which are important to understand: how attractive long-term care insurance is to someone with HD risk, and how we expect insurance pricing and availability to vary with HD risk. To address this first question, we can compare insurance usage for someone with and without a risk for HD. Brown and Finkelstein (2007) estimate that a healthy 65-year-old has about a 35% chance of ever living in a nursing home in their lifetime and a 47% chance of needing any nursing care (at home or in a facility). Conditional on entrance, individuals expect to spend about 1.7 years in a nursing home, or 3.5 years in any care. In contrast, the data suggests that someone who carries the HD mutation has about a 70% chance of living in a nursing home at some point, and a roughly 100% chance of needing any nursing care. Conditional on being in a nursing home, the average length of stay is around 6 years; overall time in any nursing care is closer to 10 years or more (Bolt, 1970; Nance and Saunders, 1996; Walker et al, 1981; Harper, 1996). Further, given

the young age of onset, individuals in this population should expect to pay premiums for a shorter period of time, further increasing the attractiveness of this insurance product. In general, we conclude that long term care insurance is significantly more valuable for someone with the HD mutation than someone in the general population.<sup>8</sup>

To address the second question, we have accessed and reviewed a number of long-term care insurance applications. The first thing to note is that, prior to asking anything detailed about medical history, these applications typically ask whether an individual “Currently has or has ever been diagnosed with” any of a list of conditions, which includes HD. Individuals who answer “yes” to anything on this list are advised that it is unlikely they will be insurable. Other than this initial screening, however, individuals are never asked specifically about their HD *risk* (or any genetic risk), nor are they asked about a family history or parental cause of death. The applications typically ask for medical records from any primary care physician that the individual has seen in the past eighteen months. This could, in principle, reveal HD risk, but even this is avoidable, by either never discussing HD risk with one’s primary care doctor or by getting a new doctor more than eighteen months before applying.

Subsequent to this written application, individuals typically meet with an insurance broker, and may be asked to undergo a physical screening by a doctor. Insurers we talked to suggested this would rarely be required for someone under 65, although a phone call to the client would be typical, as would a drug screen. Doctor’s records are sometimes ordered although, again, frequently not for younger individuals. Overall, this process suggests that currently healthy individuals – regardless of their HD risk – are likely to face the same long-term care insurance pricing. However, there may be a different story for someone experiencing symptoms – the symptoms of HD are quite noticeable, and it seems likely that a broker could pick up on this even in a casual screening or a phone call, at which point coverage would likely be denied. This suggests that delaying purchase until after onset of symptoms may not be optimal, or even possible. It is important to note that, given the structure of insurance applications, purchases by at risk individuals should not be viewed as “fraud”, since individuals are not asked about genetic risks.

---

<sup>8</sup>We can also use the methodology in Brown and Finkelstein (2007), along with information about HD disease progression, to calculate insurance loads. To do these calculations we use the basic framework provided by Brown and Finkelstein (2007) and used in their paper. We use the same assumptions that they use on interest rate progression, cost of nursing and payouts by insurers. The only change we make is in the transition matrices. For individuals at risk for HD we assume (a) more likely transition into needing home care and nursing care, (b) faster death once in those states, (c) no possibility of assisted living (this is quite unusual for individuals with HD, who are typically either at home or in a nursing home) and (d) no transition out of those states. We use information from a number of sources to calculate the HD transition matrices (Bolt, 1970; Nance and Saunders, 1996; Walker et al, 1981; Harper, 1996); the matrices we use are available from the authors. Brown and Finkelstein (2007) argue that the load is around \$0.18 for an individual in the general population, meaning the policy returns an average of 82 cents for each dollar paid in. In contrast, for someone at age 40 with a 50% HD risk we calculate a load of about -\$1.40: they expect to get \$2.40 back for each dollar paid in premiums. For someone who knows they have the genetic mutation, the load is -\$2.96, implying a payout of almost \$4 for each \$1 paid in.

## 2.3 Data

This paper makes use of two datasets. The first, data on individuals at risk for HD, comes from the PHAROS study; data on individuals in the general population is drawn from the Health and Retirement Survey.

### PHAROS Data

The PHAROS (Prospective Huntington At Risk Observational Study) study is a prospective, observational study of individuals at risk for HD conducted by the Huntington Study Group (Huntington Study Group PHAROS Investigators, 2006). The study began in 1999 and includes 1001 individuals at roughly 40 study sites in the United States and Canada. Individuals in the PHAROS study were interviewed at recruitment, and then approximately every nine months afterwards. The PHAROS study concludes in 2010. To be enrolled in PHAROS, individuals had to be at risk for HD: that is, they had one parent (or first-degree relative) with HD, but were not tested prior to enrollment. Participants in PHAROS are not a random sample of individuals at risk of HD. First, they needed to be willing to participate in the study, which may imply other differences. There is little we can do to address this. Second, enrolled individuals must not have been tested at the time of enrollment. We will discuss this second type of selection in more detail with the results. Third, participants had to be asymptomatic (not show symptoms of HD) at the time of study enrollment. The combination of the latter two points means that the chance of carrying the genetic mutation, among PHAROS participants overall, is slightly less than 50% (it is more like 40%).

Participant visits during PHAROS contained two primary sections. First, individuals responded to a set of questionnaires, some of which were given only once during the study period, and some of which were given at more than one visit, or at every visit. These questionnaires included psychological tests, questions about changes in life circumstances (marriage, children, etc) and basic demographics. Individuals were also asked about their disease experience – whether they had undergone genetic testing for HD, when they were tested, whether they had noticed any disease symptoms, what they thought their probability of having the HD genetic mutation was, etc. Second, visits included a doctor exam, at which doctors completed a series of motor tests with the individual to screen for signs of HD. At the end of this exam the doctors scored individuals on a scale from 0 to 4. A zero indicates “normal (no abnormalities)”, a 1 indicates “non-specific motor abnormalities (less than 50% confidence)”, a 2 indicates “motor abnormalities which may be a symptom of HD (50-89% confidence)”, a 3 indicates “motor abnormalities

that are likely signs of HD (90-98% confidence)” and a 4 indicates “motor abnormalities that are unequivocal signs of HD ( $\geq 99\%$  confidence)”.

Based on the discussion of long term care insurance above, it is important to distinguish between individuals who appear sick and those who do not. It seems likely that individuals who appear sick will not have the same access to insurance and, therefore, they should be excluded from the sample. We classify individuals with a 3 or 4 on the doctor scale as “symptomatic”, as they are already showing significant signs of HD. Because our unit of observation is an individual-year, some individuals are included in the sample for years in which they are not symptomatic, and then excluded from the sample later when they become symptomatic. We choose to include individuals with a score of 2 in our analysis, even though this score indicates some motor abnormalities. Our goal is to exclude people whose illness would be obvious to insurers. We argue this group with limited abnormalities would likely look similar to insurers, since their illness is not obvious even a doctor trained in evaluating this disease. However, we also show our results excluding these individuals.

Two questionnaires were administered which covered insurance ownership. The first, which was administered one or two times during the course of the study asked (a) whether the individual had any long-term care insurance and (b) if yes, how long they have owned it. The second questionnaire was intended to be administered at all visits and simply asked whether the individual had made any changes in their long-term care insurance since their last visit. Using this information, we coded insurance ownership for as many periods as possible for each individual. This involves some inference (i.e. if an individual reports having insurance and then in the next visit indicates they have not changed their ownership, we coded them as continuing to have insurance). Details are in Appendix A. We observe only the extensive margin of long-term care insurance ownership, although this may not result in much loss of data, since the range of long-term care insurance contracts is fairly limited (Brown and Finkelstein, 2007).

Our primary analysis involves comparing individuals in the PHAROS sample, who anticipated about a 50% chance of developing HD at enrollment, with individuals in a random sample of the population, who have approximately a 0% chance (the true chance, based on population prevalence, is 0.01%). In addition, we make use of variation within the HD population, using the sub-sample of individuals who pursued genetic testing outside of the study. Roughly 10% of participants did so and therefore must know their genetic status. All individuals in the study underwent genetic testing as part of the PHAROS study, but these results were not communicated to the research participant, and are not available to the investigators. However, for a large share of individuals who have been tested outside the study we can infer their test result either by using information from the doctor visit, or by the information they

provide on their self-reported probability of carrying the genetic mutation.

We do this inference as follows. We begin with using doctor evaluations to identify tested individuals who carry the mutation. As noted above, at each visit the individual is evaluated by a doctor, who indicates a level of confidence about their HD status. In addition, individuals are specifically evaluated on motor abnormalities. Based on these data, we code an individual as having carrying the mutation if either (a) doctor reports they have a 3 or 4 on the rating scale described above, or (b) the doctor reports significant motor abnormalities (motor abnormality score greater than 10). It is important to be explicit about the timing here. Individual-years in which individuals are already symptomatic are excluded from the analysis (as detailed above). However, in many cases we observe individuals both while they are *not* symptomatic, and then they later become symptomatic. We use the later data to indicate that the person likely received a positive test result, while using the earlier data for the analysis.

The procedure described above captures many individuals who carry the mutation. However, it will not capture individuals who are tested and learn they *do not* carry the mutation; it also will not capture individuals who learn they do carry the mutation but are never symptomatic over the period we observe. We therefore supplement this analysis with self-reported probability of having HD, which is asked at a subset of visits. In a number of cases, we observe these data after testing takes place. Seventy-five percent of time, individuals who are asked their HD status after testing report either 0% (presumably indicating they do not carry the HD mutation) or 100% (presumably indicating they carry the mutation). We code individuals as negative if they report a 0% chance of HD after testing, and positive if they report 100% chance. Using these combined procedures, we are able to infer testing status for 71 of the 91 individuals who are tested during the sample period. The individuals who cannot be coded are typically those who are not asked about their probability of carrying the genetic mutation after testing and do not appear sick during the period we observe them. Our data on positive test results is likely to be more comprehensive than the data on negative test results.<sup>9</sup>

Summary statistics on basic demographics appear in Panel A of Table 1. The PHAROS sample is predominantly female, and the majority are married with children. They are fairly highly educated, typically with some college, and mostly employed. The average age in this population is 40.

---

<sup>9</sup>In a very small number of cases (3) where individuals with HD symptoms who have been tested do not report that they carry the gene we rely on the doctor evaluations and code them as positive.

## Health and Retirement Survey Data

We compare individuals in the PHAROS data to individuals in the general population, drawn from the 2000 wave of the Health and Retirement Survey (HRS). The HRS surveys roughly 20,000 individuals over 50 every two years. In recent years the data has included some individuals under 50 as part of specific sub-studies or because they are spouses of sampled individuals; these younger people will not necessarily be representative of the general population, since they are more likely to be married to older individuals. The primary advantage of the HRS is that it asks about long-term care insurance ownership.

The primary disadvantage is that the individuals in the HRS are, on average, much older than the individuals in PHAROS. The average age in the PHAROS sample is around 45; in the HRS overall, it is 67. This is clearly problematic, particularly given the product we are considering. Evidence from the HRS data itself indicates that ownership of long-term care insurance increases as people age. It would be preferable to have a more closely age-matched dataset; unfortunately, there exists only very limited data on long-term care insurance ownership.

We address this mismatch in several ways. First, we limit the HRS data to individuals 65 and under. This represents the upper end of the age range in the PHAROS data. We also limit the HRS to individuals with at least a high school education, which increases the match with PHAROS, where education rates are higher due to the younger demographic. With this limitation there is complete overlap in the age distribution in the two datasets. However, the HRS is still much more heavily weighted to older people. This can be seen in Panel B of Table 1, which displays summary statistics for this younger subset of the HRS. Relative to the PHAROS individuals, those in the HRS are slightly less likely to be female, have more children and slightly less education; they are also older, with an average age of 58.

As a second step, all regressions include demographic controls – most importantly, fixed effects for age. The inclusion of age fixed effects means that any age cohorts which are represented only in the HRS and not in the PHAROS data will not drive the results. We will show our results excluding and including these age fixed effects. Since ownership of long-term care insurance increases over time in the general population, the results without age fixed effects should be a lower bound; the change in coefficients with inclusion of the fixed effects give us a sense of how important these differences across age groups are. We also include controls for employment.

In addition, in Appendix B we run our primary analyses using the Consumer Expenditure Survey which has information on expenditures on long-term care insurance. Although we argue in that Appendix that this dataset is, overall, a less ideal comparison than the HRS (due primarily to the fact that data is

collected at the household level only), it nevertheless provides a useful check. The results look very similar. If anything, the use of the Consumer Expenditure Survey suggests the HRS comparison understates the degree of adverse selection.

There are two important final notes. First, in contrast to the PHAROS data, there is only one observation per individual in the HRS. Second, we may wonder whether any of the individuals in the HRS are at risk for HD, which would mute the comparison. As noted, the rate of HD in the US is 1 in 10,000; in the sample of 7,000 HRS individuals we would therefore predict fewer than one individual with HD.

### 3 Empirical Strategy

Our primary empirical strategy amounts to a comparison of insurance ownership among individuals in the PHAROS population to those in the HRS. Define the variable  $HD_{risk}$  equal to 1 if the individual is in the PHAROS population and equal to 0 if they are in the HRS. Define  $Own_i$  as equal to one if the individual owns insurance, and zero otherwise. Our primary regression equation is below

$$Own_i = \alpha + \beta(HD_{risk})_i + \mathbf{A}\mathbf{X}_i + \epsilon_i \quad (1)$$

where  $\mathbf{X}_i$  is a vector of controls. We interpret the coefficient  $\beta$  as the extent of adverse selection.<sup>10</sup>

In addition to this analysis, we will run several analyses in which the coefficients are identified off of variation within the PHAROS sample. Define  $test_{pos}$  as equal to one if an individual in PHAROS has been tested, and their test result was positive, and zero otherwise. Define  $test_{neg}$  in parallel, but with a negative test results. The omitted category in this regression is untested individuals, and coefficients are interpreted relative to that group. We note that for a small number of individuals we know they are tested but do not know their results; we exclude these individuals from this regression. This has no impact on the results. These regressions we limit to individuals in the PHAROS sample. We estimate the regression below.

$$Own_i = \alpha + \beta_1(test_{pos})_i + \beta_2(test_{neg})_i + \mathbf{A}\mathbf{X}_i + \epsilon_i \quad (2)$$

Our theory suggests that  $\beta_1 > 0$  and  $\beta_2 < 0$ . Note that  $\beta_1$  and  $\beta_2$  in this regression are identified off of variation in risk *within* the HD-risk population.

In addition, we supplement this with two additional analysis. First, we estimate Equation (2) excluding people who receive a score of 2 on the doctor exam described in the data section. Our standard

---

<sup>10</sup>We define adverse selection as occurring if individuals with private information that they will need insurance purchase more of it. This is consistent with the large theoretical literature on genetic adverse selection (e.g. Hoy and Polborn, 2002; Hoy and Witt, 2007; Viswanathan et al, 2007; Pauly et al, 2003; MacMinn and Brockett, 2004; Strohmenger and Wambach, 2000; Subramanian et al, 1999; Hoy, Orsi, Eisinger and Moatti, 2003; Doherty and Thistle, 1996; Hoel and Iverson, 2002).

regressions excludes all individuals who are clearly symptomatic, but do include those with some symptoms which a doctor reports “may be” evidence of HD. If this is correlated with ownership, or with ability to purchase insurance, it could influence our results. Excluding these individuals is therefore a useful robustness check.

Second, we analyze how behavior is affected by proximity to testing. To do this, we define two additional variables: a variable equal to one if the individual reports testing for the first time two visits in the future and a variable equal to one if the individual reports testing for the first time one visit in the future. We include these variables along with the variables defined above; the coefficients will give us a sense of whether there is any increase in ownership in the period leading up to testing. Again, this analysis is limited to individuals in PHAROS.

Before moving to the results, we have a final note on weighting. As discussed, an observation in PHAROS is an individual-year. The goal is to capture long-term care ownership in the “average” person-year in the data; given that people change over time, using the data at the individual-year level seems more sensible than, say, using a measure of whether people ever own insurance. However, this means that some individuals will be over-represented in the data (for example, if they completed more visits or we observe more insurance ownership data on them). To address this, we weight by the inverse of the number of times each individual is observed. In addition, in regressions standard errors are clustered by individual.

## 4 Results: Adverse Selection in Long-term Care Insurance

We begin by showing, in Table 2, basic comparisons of long-term care insurance ownership across groups with different risk levels. In this table, and throughout the paper, we limit the data to asymptomatic individuals, as discussed in Section 2 (these results can also be seen graphically, without statistical tests, in Figure 1). The first two rows show average ownership among the not-at-risk population in the HRS. The first row shows the simple mean level of ownership; the second row addresses the concern that the HRS and the PHAROS data are not well-matched in terms of age by showing ownership among HRS individuals weighted to replicate the PHAROS population.<sup>11</sup>

In the third row, we show ownership among untested (roughly 50% risk) individuals in PHAROS. As noted, an observation in PHAROS is an individual-year; these data are weighted as described above.

---

<sup>11</sup>To do this weighting, we divide the sample into four age categories – 25-34, 35-44, 45-54 and 55-64 and calculate insurance ownership among each group in the HRS. We then calculate the share of individuals in each group in PHAROS, and weight the average ownership by group in the HRS using the age distribution from PHAROS.

Comparing this row to either the first or second row gives us our first result: insurance ownership is significantly higher among at-risk individuals. Depending on the comparison group used, individuals in PHAROS are 17 to 19 percentage points, or about two and a half times, more likely to own long-term care insurance. At the bottom of the panel we reported p-values from a weighted t-test; the difference between the groups is strongly significant.

The fourth and fifth rows show ownership among other risk categories within PHAROS: those who have been tested and know they carry the genetic mutation (100% risk) and those that have been tested and do not carry the mutation (0% risk). Both of these samples – particularly the negative test group – are small, but the patterns are consistent with what we would expect. Individuals who know they carry the genetic mutation are 17 percentage points more likely to own long-term care insurance as at-risk individuals, and this difference is significant. Those who were at risk but found out they do not carry the genetic mutation are slightly less likely, although this difference is very small. Perhaps most important, the difference between individuals who tested positive and those who tested negative is large and statistically significant. These results are all consistent with our expectation of adverse selection in this market.

Finally, in the fifth and sixth rows of Panel A we look at insurance ownership more than one period after testing. The basic results in rows 3 and 4 include individuals who may have been tested right before the last visit; if people respond to a positive test results by increasing their long term care insurance after testing, it may take more time for this to appear in the data. What we see is consistent with this, at least on the positive side. Ownership levels at visits after the first tested visit are even higher (at 53%) for those who test positive; there is some reduction (to 22.9%) for those who test negative, but this is extremely minimal. It is important to note that the sample sizes in this table are quite small. This is especially true when we consider that these sample sizes represent person-years; the count of unique individuals is as low as 12 or 13 individuals in some of the tested (negative) samples; in the tested (positive) samples we have around 30 unique individuals. This suggests the results should be interpreted with significant caution. It is also the reason that we emphasize the comparison of the PHAROS sample overall to the HRS, where the sample sizes are both much larger.

In Panel B of Table 2 we explore variation in the tested sample over time in a bit more depth. In particular, we would like to compare individuals before and after testing, by test results. The goal is to compare individuals who should have similar expectations about their probability of HD, although different realizations. To do this, we restrict the sample to individuals with effectively no motor symptoms of HD and look at them before and after testing. As noted above, all of our analyses exclude

people with obvious symptoms of HD; however, we do include individuals who show minor symptoms of HD which might be noticeable to the individual or doctor, but would not be noticed by an insurer. Because individuals might respond to these symptoms, limiting to people with no symptoms lets us compare people with identical expectations.

The first two rows show average ownership rates among these individuals when they are untested; although the sample sizes are small, the levels of ownership are very similar. Rows 3 and 4 in Panel B then show ownership rates among these individuals after testing, again restricting to individuals with no motor symptoms. We see that in this state, after testing, the insurance ownership is dramatically different in the two groups, suggesting significant purchases by those who tested positive, and some reduction by those who test negative. We should note that the ownership among the positive group here is lower than the overall ownership among positive individuals, since we are restricting the sample here to people who appear completely healthy. These individuals will have further opportunities to ramp up their ownership over time as they get closer to being sick.

Table 3 shows our estimates of adverse selection adjusting for demographics. Columns 1-2 of this table focus on individuals at risk for HD versus the general population; Columns 3-5 estimate the effect of variation in risk level within the HD population. Column 1 estimates the effect of HD risk on insurance ownership adjusting for some demographics (income, education, employment, children and marital status) but *not including* controls for age. As we noted in the data section, since the HRS sample is older, and long-term care insurance is more common among older people, this should be a lower bound on the estimate of adverse selection. The effect is about 15 percentage points, similar to what we saw in Table 2. In Column 2 we add in fixed effects for age: consistent with our intuition about the effect of this control, adding it in increases the coefficient on HD risk. It should be noted, though, that this increase is small: differences in ownership across age groups in the HRS are not enormous, so this adjustment is relatively minor. In both columns the effect is highly significant.

Column 3 of Table 3 presents evidence on variations across individuals with different risks within the PHAROS population. Our expectation of adverse selection indicates that the coefficient on “tested positive” should be positive, since these individuals know they will develop HD. Similarly, the coefficient on “tested negative” should be negative: these individuals are just like the general population in terms of risk.

The evidence is supportive of this expectation. Individuals who know they carry the genetic mutation are 22 percentage points more likely to own insurance than those who are at risk but have not been tested; this is significant. Individuals with a negative test result have a lower ownership level, but

this is not significant. This second result reflects what we saw in the summary statistics above, and could reflect a number of things. At a most basic level, this may reflect issues with our data – it is more difficult to infer negative test results than positive ones. It may also reflect some “stickiness” in ownership: once individuals own the insurance, it takes them time to get rid of it.<sup>12</sup> A final, more likely, possibility is that individuals who choose to get tested are, on average, more likely to own insurance (they are more cautious people, more prepared, etc). If this is correct, then the better test for adverse selection is whether individuals who test and find out they carry the genetic mutation have significantly higher ownership than those who test and do not carry the genetic mutation; the p-value for this test is shown at the bottom of Column 3 and we find the difference is significant, with a p-value of 0.02.

Column 4 replicates the regression from Column 3, but excludes individuals with a doctor score of 2, who are showing motor symptoms which “may be” a sign of HD. Our results are, if anything, stronger. Finally, Column 5 of Table 3 explores the dynamics of purchases around the period of testing. We see some statistically weak evidence that purchases of insurance increase in the period before testing (the point estimate is positive and large, but not significant). The effects are then very strong in the periods after testing: individuals who find out they carry the genetic mutation are 38 percentage points more likely to own insurance, whereas those who test negative own slightly less insurance than at risk individuals, although this is not significant.<sup>13</sup>

Overall, based on Tables 2 and 3, we see about one-quarter of individuals who are at 50% risk for HD own long-term care insurance, and around 50% of those who know they carry the genetic mutation. This is in comparison to ownership rates of roughly 10% on average in the HRS and, as can be seen in Appendix B, even lower rates in the Consumer Expenditure Survey. Although this clearly points to significant adverse selection, there is a lingering question of why these results are not even larger. Based on the discussion in Section 2, it seems likely that nearly all individuals who carry the HD genetic mutation will need some long-term care, and they should know this. Why is long-term care insurance ownership not universal in this population?

One possibility is Medicaid crowd-out: the PHAROS sample is relatively poor, and may believe they are likely to have their long-term care covered by Medicaid, and therefore do not need insurance. Further, given that they are poor, the private nursing options they have available may not be much better than what is covered by Medicaid, which increases the relative value of Medicaid coverage.

---

<sup>12</sup>This may be particularly likely since insurance payments are often front-loaded, so once an individual pays in for a while, the insurance becomes a better deal even if they find out they have only the general population risk of needing long term care.

<sup>13</sup>The magnitude of the coefficient on testing positive is larger than in Column 4, since this restricts to individuals who are observed for a longer period, and this sub-sample has larger results.

Another possibility is that individuals believe they would not be able to get insurance, due to their genetic risk. Although our analysis of insurance contracts suggests this is probably not correct, concerns about genetic discrimination are high in this population. Unfortunately, we do not have sufficient information in these data to fully tease out these explanations, or others which might be operating.

### **Selection Bias**

One important issue raised by the analysis above is whether tested individuals are comparable to untested individuals. This may be important for two reasons. First, as in Columns 3-5 above, we would like to make statements about differences in ownership between untested individuals (approximately 50% risk) and tested individuals who know they carry the genetic mutation (100% risk). However, to the extent that tested individuals are different in their insurance ownership for reasons other than testing, this will not be a reasonable comparison. It will still be reasonable to compare those who test and find out they are carriers (100% risk) to those who test and find out they are not (0% risk), but not to make the further comparison with the untested individuals.

The second issue relates to the use of the PHAROS sample overall as representative of at HD-risk individuals. To be enrolled in PHAROS, individuals must not have been tested at the time of enrollment. This means that they are selected to contain individuals who are less interested in testing, since those who test when they are young will not be able to enroll. The first thing to note – related largely to the second concern – is that predictive testing for HD is quite rare. Only about 5% of individuals who are at risk choose to get tested (Walker, 2007). This means that, in terms of magnitudes, there is a limit to how much this can influence our results. Even if individuals who choose to get tested are very different, the possible impact on our results is minimal.

In addition, we can observe directly whether individuals in our sample who choose to undertake predictive testing look, prior to testing, different than those who do not.<sup>14</sup> Column 1 of Table 4 shows a regression of long-term care insurance ownership among untested individuals on an indicator for HD risk, and then HD risk interacted with whether the individual chooses to be tested in the future. Column 2 shows the same regression limited to the PHAROS sample, which drops the HD risk variable. The results suggest at most very limited selection. The coefficient on the interaction between HD and future testing ranges from -0.01 to -0.03 and is not significant. This should provide confidence both in the value of this sample for representing the overall HD population, and in the comparability of untested and tested

---

<sup>14</sup>This regression excludes tested individuals after they have been tested, and also excludes data for tested individuals from the period immediately prior to testing. This is done because of the concern – consistent in terms of direction with Column 4 of Table 3 that individuals may increase their insurance ownership right before testing as a precaution.

individuals.

## 5 Discussion and Conclusion

In this paper we use data on individuals at risk for a high-morbidity disease to estimate adverse selection in long-term care insurance purchases. We find strong evidence of adverse selection. Individuals with a 50% risk of carrying the HD genetic mutation are about two and a half times as likely to own long-term care insurance, relative to matched controls, and those who know they carry the genetic mutation (100% risk) are about five times as likely as matched controls.

These results suggest that there is potential for significant adverse selection in insurance when people have large amounts of private information; it seems likely that the share of people in this group will increase over time as genetic knowledge improves. Huntington disease is a rare disease, but it shares features with a number of more common diseases like Parkinson and Alzheimer disease. In the longer term, as genetic testing improves, it is possible that this private information could threaten the viability of private long term care insurance. This is especially true as governments (both state and federal) restrict the ability of insurers to observe genetic information (e.g. United States House of Representatives, 2007).

In addition to implications for the private market, the results in this paper relate to a provision of the recent health care bill, the “Community Living Assistance Services and Support Act” (CLASS Act) (United States Senate, 2010). This provision of the bill requires the government to offer a long-term care insurance product which would provide support for expenditures on nursing care or other long-term care needs. The provisions of the bill are fairly simple: (1) participants pay in monthly premiums for at least 5 years before the insurance vests, (2) estimated benefits are \$50 per day with no limits on payout period and no lifetime maximum, (3) no underwriting is permitted, other than on age at enrollment and (4) premiums must be set so the program breaks even (i.e. no taxpayer funds will be used). This last provision puts a significant restriction on pricing: the office in charge of the program must set premiums to balanced expected payments.

Adverse selection is clearly a concern given these pricing restrictions. If those who purchase insurance are higher risk, this will result in a need to charge higher premiums, which could limit participation by low risk individuals. In contrast to options available to private insurers, the very simple nature of the insurance product eliminates the ability to screen out high risk types with product variations. Our results suggest that increasing genetic information may exacerbate this potential for

adverse selection. We find that individuals who know they carry the genetic mutation for HD have insurance ownership rates 4 to 5 times higher than the general population, and have costs which are likely to be at least 5 or 6 times as high. While HD is rare, this finding raises the possibility that increasing genetic information about more common diseases like Alzheimer disease could raise expected costs and premiums of the CLASS program if such individuals exhibit purchasing behaviors which echo those of HD individuals. At the same time, there is clearly value in providing an insurance product like this which protect individuals who have the misfortune to inherit bad genetic risks; this may argue for some type of “health status” insurance, as in Cochrane (2009).

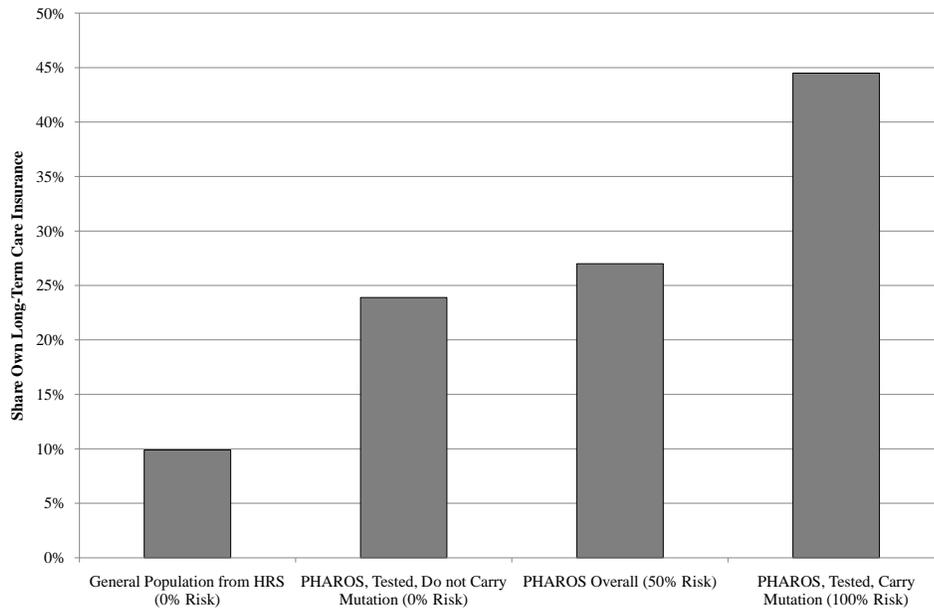
## References

- Aktan-Collan, Katja, Ari Haukkala, and Helena Kaariainen**, “Life and Health Insurance Behaviour of Individuals Having Undergone a Predictive Genetic Testing Programme for Hereditary Non-Polyposis Colorectal Cancer,” *Community Genetics*, 2001, 4 (4), 219–224.
- Armstrong, Katrina et al.**, “Life Insurance and Breast Cancer Risk Assessment: Adverse Selection, Genetic Testing Decisions, and Discrimination,” *American Journal of Medical Genetics Part A*, 2003, 120A (3), 359–364.
- Bolt, Jean**, “Huntington’s Chorea in the West of Scotland,” *British Journal of Psychiatry*, 1970, 116 (532), 259–270.
- Brown, Jeffrey and Amy Finkelstein**, “Why Is the Market for Long Term Care Insurance so Small,” *Journal of Public Economics*, 2007, 91 (10), 1967–1991.
- and —, “The Private Market for Long-Term Care Insurance in the U.S.: A Review of the Evidence,” *Journal of Risk and Insurance*, 2009.
- Burton, Paul et al.**, “Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls,” *Nature*, 2007, 447 (7145), 661–678.
- Cawley, John and Tomas Philipson**, “An Empirical Examination of Information Barriers to Trade in Insurance,” *American Economic Review*, 1999, 89 (4), 827–846.
- Chiappori, Pierre-Andre and Bernard Salanie**, “Testing for Asymmetric Information in Insurance Markets,” *Journal of Political Economy*, 2000, 108 (1), 56–78.
- Cochrane, John**, “Health-Status Insurance: How Markets Can Provide Health Security,” *Policy Analysis*, 2009, (633).
- Cohen, Alma and Liran Einav**, “Estimating Risk Preferences from Deductible Choice,” *American Economic Review*, 2007, 97 (3), 745–788.
- Cutler, David, Amy Finkelstein, and Kathleen McGarry**, “Preference Heterogeneity in Insurance Markets: Explaining a Puzzle,” *American Economic Review Papers and Proceedings*, 2008, 98 (2), 157–162.
- De Meza, David and David Webb**, “Advantageous selection in insurance markets,” *RAND Journal of Economics*, 2001, 32, 249–262.
- Doherty, Neil and Paul Thistle**, “Adverse Selection with Endogenous Information in Insurance Markets,” *Journal of Public Economics*, 1996, 63 (1), 83–102.
- Einav, Liran, Amy Finkelstein, and Paul Schrimpf**, “Optimal Mandates and The Welfare Cost of Asymmetric Information: Evidence from The U.K. Annuity Market,” *NBER Working Paper No. 13228*, 2009.
- Fang, Hanming, Michael Keane, and Dan Silverman**, “Sources of Advantageous Selection: Evidence from the Medigap Insurance Market,” *Journal of Political Economy*, 2008, 116 (2), 303–350.
- Finkelstein, Amy and James Poterba**, “Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market,” *Journal of Political Economy*, 2004, 112 (1), 183–208.

- **and Kathleen McGarry**, “Multiple Dimensions of Private Information: Evidence From the Long-Term Care Insurance Market,” *American Economic Review*, 2006, 96 (4), 938–958.
- Harper, Peter**, *Huntington’s Disease*, WB Saunders, 1996.
- Hoel, Michael and Tor Iversen**, “Genetic Testing When There is a Mix of Compulsory and Voluntary Health Insurance,” *Journal of Health Economics*, 2002, 21 (2), 253–270.
- Hoy, Michael and Julia Witt**, “Welfare Effects of Banning Genetic Information in the Life Insurance Market: The Case of BRCA1/2 Genes,” *Journal of Risk and Insurance*, 2007, 74 (3), 523–546.
- **and Michael Polborn**, “The Value of Genetic Information in the Life Insurance Market,” *Journal of Public Economics*, 2002, 78 (3), 235–252.
- **, Fabienne Orsi, Francois Eisinger, and Jean Paul Moatti**, “The Impact of Genetic Testing on Healthcare Insurance,” *Geneva Papers on Risk and Insurance*, 2003, 28 (2), 203–221.
- Huntington Study Group PHAROS Investigators**, “At risk for Huntington disease: The PHAROS Prospective Huntington At Risk Observational Study cohort enrolled,” *Archives of Neurology*, 2006, 63, 991–996.
- MacMinn, Richard and Patrick Brockett**, “Genetic Testing and Adverse Selection,” *Illinois State University Working Paper*, 2004.
- McCarthy, David and Olivia Mitchell**, “International Adverse Selection in Life Insurance and Annuities,” *NBER Working Paper 9975*, 2003.
- Meyer, Bruce and James Sullivan**, “Further Results on Measuring the Well-Being of the Poor Using Income and Consumption,” *NBER Working Paper No. 13413*, 2007.
- Meyers, RH**, “Huntington’s Disease Genetics,” *NeuroRx*, 2004, 1, 255–262.
- Nance, MA and G Saunders**, “Characteristics of Individuals with Huntington Disease in Long-Term Care,” *Movement Disorders*, 1996, 11 (5), 542–548.
- Oster, Emily, E. Ray Dorsey, Jan Bausch, Aileen Shinaman, Elise Kayson, David Oakes, Ira Shoulson, and Kimberly Quaid**, “Fear of health insurance loss among individuals at risk for Huntington disease,” *American Journal of Medical Genetics, Part A*, 2008, 146 (16), 2070–2077.
- Pauly, Mark, Kate Withers, Krupa Viswanathan, Jean Lemaire, John Hershey, Katrina Armstrong, and David Asch**, “Price Elasticity of Demand for Term Life Insurance and Adverse Selection,” *NBER Working Paper No. 9925*, 2003.
- Smart, Michael**, “Competitive insurance markets with two unobservables,” *International Economic Review*, 2000, 41 (1), 153–169.
- Strohmenger, Rainer and Achim Wambach**, “Adverse Selection and Categorical Discrimination in the Health Insurance Markets: The Effects of Genetic Tests,” *Journal of Health Economics*, 2000, 19 (2), 197–218.
- Subramanian, Krupa, Jean Lemaire, John Hershey, Mark Pauly, Katrina Armstrong, and David Asch**, “Estimating Adverse Selection Costs from Genetic Testing for Breast and Ovarian Cancer: The Case of Life Insurance,” *Journal of Risk and Insurance*, 1999, 66 (4), 531–550.

- Taylor, Donald, Robert Cook-Deegan, Susan Hiraki, J. Scott Roberts, Dan Blzer, and Robert Green**, “Will APOE risk assessment for Alzheimers disease affect private long term care insurance coverage rates?,” *Duke University Working Paper*, 2009.
- United States House of Representatives**, “Genetic Information Non-Discrimination Act,” Federal Law 2007.
- United States Senate**, “The Patient Protection and Affordable Care Act,” Federal Law 2010.
- Viswanathan, Krupa, Jean Lemaire, Kate Withers, Katrina Armstrong, Agnieszka Baumritter, John Hershey, Mark Pauly, and David Asch**, “Adverse Selection in Term Life Insurance Purchasing Due to the BRCA1/2 Genetic Test and Elastic Demand,” *Journal of Risk and Insurance*, 2007, 74 (1), 65–86.
- Walker, DA, PS HARper, CEC Wells, A Tyler, K Davies, and RG Newcombe**, “Huntington’s Chorea in South Wales: A Genetic and Epidemiological Study,” *Clinical Genetics*, 1981, 19, 213–221.
- Walker, Francis**, “Huntington’s Disease,” *Lancet*, 2007, 359, 218–228.
- Zick, Cathleen et al.**, “Genetic Testing, Adverse Selection, and the Demand for Life Insurance,” *American Journal of Medical Genetics Part A*, 2000, 93 (1), 29–30.
- **and** — , “Genetic Testing For Alzheimers Disease And Its Impact On Insurance Purchasing Behavior,” *Health Affairs*, 2005, 24 (2), 483–490.

**Figure 1:**  
**Long-Term Care Ownership, by HD Risk**



Notes: This shows a simple comparison of mean levels of ownership for individuals in the HRS, and those in PHAROS, by risk status.

**Table 1. Summary Statistics**

<b>Panel A: Individuals at risk for HD (PHAROS Data)</b>			
	<i>Mean</i>	<i>Standard Deviation</i>	<i>Number of Obs.</i>
Female	0.688	0.463	838
Married	0.709	0.454	691
Children	1.77	1.63	650
Education (yrs)	14.94	2.15	799
Employed	0.862	0.344	663
Age	40.8	7.65	838

<b>Panel B: General Population (HRS Data, Under 66)</b>			
	<i>Mean</i>	<i>Standard Deviation</i>	<i>Number of Obs.</i>
Female	0.578	0.493	6164
Married	0.777	0.416	6164
Children	3.06	1.87	6164
Education (yrs)	13.93	2.20	6153
Employed	0.840	0.366	5141
Age	57.90	5.31	6164

Notes: This table shows summary statistics for the two samples used in the data. The Health and Retirement Survey (HRS) is from 2000; the PHAROS data is collected over the period from 1999-present. In the analysis, we use multiple observations per individual from the PHAROS data, but the summary statistics here include only one observation per individual. The HRS is limited to individuals 65 or younger with at least a high school degree (this increases the demographic overlap with PHAROS).

Table 2. Insurance Ownership Comparison of Means

Panel A: Basic Ownership Levels		
<i>Group</i>	<i>Share Owning Insurance</i>	<i>Sample Size</i>
HRS, simple mean	9.9%	5980
HRS, weighted by PHAROS age profile	8.2%	5980
PHAROS, untested	27.0%	2468
PHAROS, tested, Carry HD mutation	44.5%	73
PHAROS, tested, do not carry HD mutation	23.9%	42
<b>Tested, Longer After Testing</b>		
PHAROS, Carry HD mutation	53.3%	46
PHAROS, do not carry HD Mutation	22.9%	24
<i>p-value HRS vs. PHAROS untested</i>		p<.001
<i>p-value PHAROS untested vs. carry mutation</i>		p=.011
<i>p-value PHAROS untested vs. do not carry mutation</i>		p=.735
<i>p-value tested carry gene vs. tested do not carry mutation</i>		p=.068
<b>Longer After Testing:</b> <i>p-value tested carry gene vs. tested do not carry mutation</i>		p=.013
Panel B: Tested Individuals with No Motor Symptoms, Over Time		
<b>Before Testing, No Motor Symptoms</b>		
PHAROS, Carry HD mutation	18.1%	16
PHAROS, do not carry HD Mutation	19.7%	24
<b>After Testing, No Motor Symptoms</b>		
PHAROS, Carry HD mutation	33.2%	33
PHAROS, do not carry HD Mutation	13.5%	34
<i>p-value, Before Testing</i>		p=.504
<i>p-value, After Testing</i>		p=.052

Notes: This table shows simple means of ownership levels by risk status for long-term care insurance, and t-tests of differences between groups. In the last row of Panel A we report means for the HRS weighted by the age distribution in PHAROS. The PHAROS data are restricted to asymptomatic individuals. An observation in the PHAROS data is an individual-year. Averages are weighted by the inverse of the number of times each individual appears in the data; p-values come from weighted t-tests. The “Longer After Testing” data shows ownership among tested individuals at visits after the first visit they report testing; this allows for more time to purchase after knowing test result. Panel B includes only individuals in the sample who ever have no motor symptoms (motor score < 5). The “Before Testing” data shows average ownership among untested individuals who eventually test, by eventual test status, restricted to time periods in which they show no motor symptoms. The “After Testing” shows average ownership among tested individuals who have no motor symptoms.

**Table 3. Adverse Selection in Long-term Care Insurance**

<i>Dependent Variable: Own long-term Care Insurance</i>					
	(1)	(2)	(3)	(4)	(5)
<i>Sample:</i>	<i>PHAROS and HRS</i>		<i>PHAROS Only</i>		
<i>PHAROS Restrictions:</i>	<i>Doctor Score &lt; 3</i>		<i>Doctor Score &lt; 3</i>	<i>Doctor Score &lt; 2</i>	<i>Doctor Score &lt; 3</i>
Explanatory Variables:					
At Risk for HD	.1494*** (.019)	.1603*** (.02)			
Tested, Positive			.2203** (.103)	.2356** (.111)	.3817** (.16)
Tested, Negative			-.0874 (.094)	-.0984 (.087)	-.0465 (.205)
Tested Next Period					.1943 (.132)
Tested in Two Periods					-.0163 (.097)
Demographic Controls	YES	YES	YES	YES	YES
Age Fixed Effects	NO	YES	YES	YES	YES
<i>p-value, pos vs. neg</i>			.026	.016	.098
Number of Observations	7356	7356	2340	2114	2246
R <sup>2</sup>	.03	.04	.07	.07	.07

standard errors in parentheses, clustered by individual

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table compares long-term care insurance ownership for individuals at risk for Huntington disease and those in the general population (from the HRS). Controls (in all columns) are: income categories, education categories, number of children, married, gender and employment status. All columns limit the sample to asymptomatic individuals from PHAROS. Columns 1 and 2 include the HRS individuals; Columns 3-5 include only PHAROS (and therefore do not include the control for HD Risk since all individuals in the sample are at risk). Column 4 excludes individuals with a doctor score of 2, which indicates “possible” symptoms of HD. An observation in PHAROS is an individual-year and regressions are weighted by the inverse of the number of times an individual is observed in the data.

**Table 4. Sample Selection: Insurance Ownership Among Ever Tested Versus Not**

<i>Dependent Variable:</i>	<i>Own long-term Care Ins.</i>	
<i>Sample:</i>	<i>PHAROS and HRS</i>	<i>PHAROS Only</i>
Explanatory Variables:		
At Risk for HD	.1501*** (.021)	
HD, Ever Tested	-.0327 (.071)	-.0108 (.072)
Controls in all columns: income categories, education categories, number of children, married, gender, employed, age fixed effects.		
Number of Observations	7241	2259
R <sup>2</sup>	.03	.06
standard errors in parentheses, clustered by individual		
* significant at 10%; ** significant at 5%; *** significant at 1%		
Notes: This table estimates the difference in adverse selection among individuals who were tested at all during the sample versus those who were never tested. The sample is limited to people who have not yet been tested. The “Ever Tested” variable is therefore equal to one if the individual will be tested in the future. An observation in PHAROS is an individual-year and regressions are weighted by the inverse of the number of times an individual is observed in the data.		

## Appendix A: Inferring Insurance Ownership

This appendix discusses how we infer insurance ownership among individuals in the PHAROS data. There are two PHAROS surveys which cover insurance. The first, the “Life Decision Survey”, was administered (typically) once or twice to each individual during the study. In this survey individuals were asked, for each type of insurance, whether they currently had insurance. Based on this question alone – whether the individual has insurance – it would be possible to run our analysis. However, in this case we would have only one or two observations per individual (i.e. only one or two years). Having multiple observations over time is helpful in identifying variation within the HD sample and, for example, exploring decision making in the period around testing.

We use two pieces of data to identify changes over time. First, in the Life Decision Survey individuals were also asked how many years they had held insurance for. Second, a separate survey, called the “Insurance and Employment Survey”, was administered at a larger number of visits. In this survey individuals were asked (among other things) if they had made any changes to their insurance in the last year. Using these two pieces of information together, we inferred ownership for years that were not covered by the primary Life Decision Survey question.

The procedure was as follows. For all internally consistent points (i.e. none of the observations for a given individual gave information that could not logically be true given some other observation for that individual):

1. If there was data indicating the number of years the individual had insurance, the former years were filled in as having insurance, and the year prior to having the insurance (if it was in the observation period) was marked as not having insurance.
2. If no data on the number of years the individual had insurance was available in any observation, or if values remained unknown, then all points where the state of insurance (having or not having) and the presence of change (i.e. is this state different than last year or next year) were known had the state imputed backwards or forwards a period, respectively.

For all points that were inconsistent, meaning that at least one of the observations gave information that was logically inconsistent with other data provided:

1. For two or three inconsistent observations (there were never more than that) if one of them was corroborated by other data points, then that information was extrapolated in the manner described above.
2. For two or three inconsistent observations, if none of them could be corroborated by other data points, then precedence was given to the earliest observation, its information was extrapolated, and all other points were filled in.

We note that number of inconsistent observations was fairly small, and the results we report are robust to excluding these.

## Appendix B: Alternative Data Sources for non-HD Population

Our primary data source for not-at-risk individuals is the Health and Retirement Survey. As we note above, in many ways these data are well suited as a comparison: the structure of the information on long-term care insurance is similar to what we have in the PHAROS data and similar controls are available. The main downside of these data is the mismatch in terms of age. Given the HRS sample frame, the sample is on average much older than the PHAROS sample. As we note, there is actually full overlap in ages available, since the HRS does sample some younger individuals, but there is relatively little data available on younger individuals. On average, we expect this to bias our results downward, especially since long-term care insurance ownership is much more common among older people.

In this appendix we use another dataset– the Consumer Expenditure Survey (CEX) – to estimate our main results. This has the advantage of covering a larger age range, and the overall age profile is similar to the PHAROS data. The main downside is that the data on insurance ownership is much less comparable (this is described in further detail below), and we have significant concerns about under-reporting of insurance coverage (Meyer and Sullivan, 2007). Nevertheless, we will see from the analyses below that the results are quite similar with these datasets and, as we expect, if anything our results understate the extent of adverse selection.

There are two questions in the CEX which can be used to infer long-term care insurance ownership. First, households are asked to report their expenditures on long-term care insurance. Second, they are later asked to

report the total number of policies that they pay for. From these two sources, we construct a measure of the number of long-term care policies held by the household. There are at least two issues with these data. First, we observe counts of the number of policies held by the household overall, not measures of individual ownership. In some cases, allocation is easy – if there are two policies in a household with two people, it seems most likely that one is held by each. In other cases, allocation is difficult – if there is only one policy in a two person household, or two policies in a three adult, multi-generational household. Second, the CEX is known to have issues with under-reporting; in recent years comparison to the national accounts suggests only about 55% of expenditures are reported (Meyer and Sullivan, 2007). This could lead us to overstate adverse selection.

To address the first of these issues, we simply allocate fractions of policies if necessary. That is, we take the number of policies owned by the household, divide it by the number of adults in the household, and assume each adult owns that fraction of a policy. If there are two adults and two policies, each has one. If there are two adults and one policy, each have one half. This is equivalent to assigning policies randomly if there are more individuals than policies. In cases where there are more policies than individuals (this is very unusual) we code only one policy per individual, to match how this is reported in the PHAROS data. We address the second issue only after observing the results. In particular, after we observe the actual adverse selection in the data, we can ask how large it would be if ownership in the CEX was twice as large.

The table below shows estimates of adverse selection in long-term care insurance, using the CEX rather than the HRS as a comparison. This is comparable to Table 3 in the paper; we leave out the latter two columns, since the coefficients on testing positive and negative are identified off of variations within the PHAROS population and are therefore unchanged by the change in comparison set.

<i>Dependent Variable: Own long-term Care Insurance</i>		
	(1)	(2)
<i>Sample:</i>	<i>Asymptomatic Individuals</i>	
Explanatory Variables:		
At Risk for HD	.1796*** (.018)	.177*** (.018)
CONTROLS	YES	YES
AGE FIXED EFFECTS	NO	YES
Number of Observations	11395	11395
R <sup>2</sup>	.13	.13

standard errors in parentheses, clustered by individual

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table compares long-term care insurance ownership for individuals at risk for Huntington disease and those in the general population (from the CEX). Controls (in all columns) are: income categories, education categories, number of children, married, gender and employment status. All columns limit the sample to asymptomatic individuals, who are defined as individuals in the CEX or any individuals in PHAROS who are not showing obvious signs of HD.

Relative to the primary results, in Table 3, the comparison with the CEX suggests slightly higher adverse selection (18% versus 13%). However, a simple adjustment for under-reporting brings the results almost completely into line. The average rate of long-term care ownership in the CEX is 3.8%. Assume only 50% of expenditures are reported and that this translates to 50% of individuals who own long-term care not reporting having a policy. Ownership is therefore around 7.6%, the coefficient would be 14.1%, very close to what we see in Table 3.

Regardless of whether we adjust the coefficients or not, however, the message is similar. If anything, our estimates using the HRS understate the degree of adverse selection in long-term care insurance, although not by very much.