NBER WORKING PAPER SERIES

GENETIC ADVERSE SELECTION: EVIDENCE FROM LONG-TERM CARE INSURANCE AND HUNTINGTON DISEASE

Emily Oster
Ira Shoulson
Kimberly Quaid
E. Ray Dorsey

Working Paper 15326
http://www.nber.org/papers/w15326

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2009

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. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Genetic Adverse Selection: Evidence from Long-Term Care Insurance and Huntington Disease
Emily Oster, Ira Shoulson, Kimberly Quaid, and E. Ray Dorsey
NBER Working Paper No. 15326
September 2009
JEL No. D82,I11,I18

ABSTRACT

Individual, personalized genetic information is increasingly available, leading to the possibility of greater adverse selection over time, particularly in individual-payer insurance markets; this selection could impact the viability of these 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. We use these estimates to make predictions about the future of this market as genetic information increases. We argue that even relatively limited increases in genetic information may threaten the viability of private long-term care insurance.

Emily Oster
University of Chicago
Booth School of Business
5807 South Woodlawn Ave
Chicago, IL 60637
and NBER
eoster@uchicago.edu

Ira Shoulson
University of Rochester
School of Medicine and Dentistry
Department of Neurology
601 Elmwood Ave, Box 673
Rochester, New York 14642
Ira.Shoulson@ctcc.rochester.edu

Kimberly Quaid
IUPUI
kquaid@iupui.edu

E. Ray Dorsey
University of Rochester
School of Medicine and Dentistry
Department of Neurology
601 Elmwood Ave, Box 673
Rochester, New York 14642
Ray.Dorsey@ctcc.rochester.edu

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
IUPUI

E. Ray Dorsey
University of Rochester

On Behalf of the Huntington Study Group PHAROS Investigators
Draft: August 31, 2009

Abstract

Individual, personalized genetic information is increasingly available, leading to the possibility of greater adverse selection over time, particularly in individual-payer insurance markets; this selection could impact the viability of these 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. We use these estimates to make predictions about the future of this market as genetic information increases. We argue that even relatively limited increases in genetic information may threaten the viability of private long-term care insurance.

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.[1] This testing, in turn, is likely to increase the degree of private information that individuals have about their mortality and morbidity risks.[2] In this paper we explore the possible impact of this increased information on the markets for long-term care insurance.

*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.

[1]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).

[2]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.[3] Increasingly, the findings in this literature point to the importance of multiple dimensions of preference heterogeneity (Finkelstein and Poterba, 2004; 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 at the moment 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 significantly change the overall correlation between ownership and payouts and have large effects on the viability of this market in the long-term. This issue may well be of policy importance, especially as legislation is increasingly enacted to restrict the ability of insurers to observe individual genetic risk 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 use data on individuals at risk for HD alongside individuals without HD risk to estimate adverse selection in long-term care insurance. We then use our estimates to evaluate the longer term consequences of increased genetic information on the functioning of that market.

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, genetic nature of the disease individuals have significant private information about their disease risk. Those with

[3]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.

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. 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 begin by outlining a simple theory of adverse selection in long-term care insurance, which clearly predicts adverse selection in our setting.[4] We focus on long-term care insurance in this paper, despite the fact that it is a relatively small market, 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 something like 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%.[5] 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 partially observe their genetic

[4] We follow a very 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).

[5] 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.

status[6] and therefore compare individuals at 50% risk to those who know they have the genetic mutation (100% risk) and those that know they do not have the mutation (0% 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 – it is not subject to concerns about selection into testing, for example. However, we are somewhat cautious about this analysis, due to relatively small sample sizes and our need to infer gene test results rather than observing them directly.

Using these data, we find significant evidence of adverse selection. In the general population (the HRS), 10% of individuals own long-term care insurance. In the PHAROS population, 25% of individuals own long-term care insurance. Among those who pursued genetic testing and know they carry the mutation responsible for HD, ownership is close to 50%.[7] The rate of long-term care insurance ownership among those individulas who know they are at 100% risk for developing HD is 30 percentage points higher than those who also pursued genetic test and know they do not carry the genetic mutation for HD.[8] This primary result of the paper can be seen simply in Figure 1. When we look at the timing of ownership around testing, we find limited evidence of increased insurance among tested individuals *before* testing, but a large increase in ownership among carrier individuals after testing, again consistent with HD risk driving adverse selection.

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, much older than individuals in the HD population. We attempt to address this, and other demographic differences, by including extensive controls for demographics (e.g. fixed effects for age, income, education, gender). In general, we find that the effects of HD are not driven by demographics and the magnitude of the HD effect swamps the effect of the 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. As a robustness check, in an appendix we show

[6] In fact, 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.

[7] 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.

[8] Although we can reject equality between individuals who do not carry the genetic mutation and those who have been tested and do not carry the genetic mutation, the latter group do not own significantly less insurance than non-tested, HD-risk individuals. This could be due to stickiness in the ownership of insurance or, more likely, may be driven by an imperfect feature of our data – that we have to infer test results – which we hope will be remedied in later versions of this paper.

estimates of adverse selection based on an alternative dataset (the Consumer Expenditure Survey), which we argue 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.

Armed with these estimates, we then turn to the second aim of the paper: evaluating possible future consequences for the long-term care insurance market. We consider whether the viability of the markets is likely to be threatened by increases in availability of genetic information. Our estimates of adverse selection, plus information on insurance premiums and the chance of needing long-term care, can be used to generate a demand curve for long-term care insurance. Private information shifts this demand curve. We consider the case where a monopolist insurer faces a market with some share of a low type, who have no private information, and some share of a high type, who have private information that they are high risk. We solve for optimal pricing, under a pooling equilibrium, for environments where the share of high type individuals differ.[9] We find that when the share of the high type is small (as is certainly the case with HD), a monopoly insurer can make positive profits selling to both types, since low types are willing to purchase some insurance at actuarially unfair prices. As the share of the high type increases, the market eventually shifts, abruptly, to selling only to the high type. Our calculations suggest that this occurs when the share of the high type is small, around 3%.

This analysis can also be used to calculate consumer surplus under private versus public information. Those calculations suggest that when the insurer is willing to sell to both types, consumer surplus is strictly higher when information is private. However, once the insurer is no longer willing to sell to both types, it would be Pareto-improving to make genetic information available to the insurer: low types would be better off, since they would be able to purchase insurance, and high types would be no worse off, since they are already paying the full, high-type insurance premiums. This suggests that legislation prohibiting insurers from observing genetic information may be a mixed blessing; in the short run, this will increase consumer surplus, but in the long run it may reduce it. Given the difference in the short-run and long-run implications, it seems important to understand the current state of genetic knowledge. A simple summary of available genetic testing suggests that – even if everyone eligible for genetic testing for relevant diseases got it – the share of the "high type" would only be around 0.1%. This indicates that, certainly in the current state and likely for the immediate future, consumer surplus will be higher with private information than without.

In the context of the existing literature on private information and adverse selection, this paper

[9]There is no separating equilibrium in this case, since the high type gets the payout for sure, so the insurer will never be willing to offer them a contract with any "insurance", and they would prefer any contract offered to the low type.

suggests that adverse selection in this market may become much more important over time, as the amount of private genetic information increases. In the current state of the world, with relatively limited private information, variations in preferences seem to outweigh these effects, but this may well not be true even in the relatively near future.

In addition to this contribution, 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. These advantages allow us to make a systematic attempt to calibrate the impact of this information on the viability of insurance markets which has not previously been done and seems the most policy-relevant issue.

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 a very simple theory. Section 4 shows the results, Section 5 analyzes the long-term care market viability, and Section 6 concludes with policy implications.

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 discission, 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 a positive result 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 loads for someone with and without a risk for HD (insurance load in this case is defined as one minus the expected payout for each dollar paid in). We use the methodology from Brown and Finkelstein (2007), along with information about HD disease progression, to calculate these loads. 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. For either type of individual, this is a very good financial investment.[10]

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

[10]To do these calcuations 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 state. 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.

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.

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 is scheduled to conclude 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, like a sibling) 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%). This will be important in the calibration of demand curves in Section 6.

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, where 0 indicated no motor abnormalities and 4 indicated "motor abnormalities that are unequivocal signs of HD ($\geq$ 99% confidence)."

Two questionnaires were administered which covered insurance ownership. The first, which was administered (for most participants) one or two times during the course of the study asked (a) whether the individual had any long-term care or life insurance and (b) if yes, how long they have owned each type. The second questionnaire was intended to be administered at all visits and simply asked whether the individual had made any changes in their insurance since their last visit. Using the information from these two questionnaires together, we coded insurance ownership for as many periods as possible for each individual. In some cases this involved inference (for example, 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 about this inference 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. However, these results were not communicated to either the research participant or to the physician investigator. Because the study has yet to conclude, these results are not yet available for analysis; they are likely to become available eventually. 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. Details of this inference are in Appendix B; of the 91 individuals tested during the sample period, we can infer genetic status for 67 of them.[11] It is important to note that the decision to get tested may be reflective of individual unobservables – for example, individuals who are risk averse may be more likely to get tested and more likely to own long term care insurance – in which case the better comparison is between tested individuals, across test results.

[11] Our inference is much better and more reliable for individuals who carry the genetic mutation for HD than those that do not. We are therefore more confident about our results for those who carry the genetic mutation.

There is also a distinction in our sample between individuals who are currently "asymptomatic" (no symptoms) and those who are "symptomatic" (showing symptoms). To distinguish between these groups, we use the doctor's exam score. Individuals who receive a 4 on this scale show signs of HD consistent with a greater than 99% chance of having the disease. We classify these individuals as symptomatic, as they are typically already showing signs of manifest HD. The symptomatic group is likely to face limited ability to purchase new long-term care insurance, and may well be actually using their long-term care insurance. For this reason, we exclude them from our analysis, although including them does not change our central conclusions.

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. The data from PHAROS is at the individual-year level. Standard errors will therefore be clustered by individual.

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. 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 less education; they are also, on average, 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 C 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 Theoretical Framework and Empirical Strategy

In this section we briefly outline a very simple theoretical framework, and then describe our estimation strategy.

3.1 Theoretical Framework

We assume there are two states that an individual may experience: healthy ("good") or disabled ("bad"). A healthy individual becomes disabled in the next period with probability q. Income differs across states of the world. In the healthy state, consumption is equal to Y_1. In the disabled state consumption, denoted Y_2, is lower due to limited ability to work and high health expenses. Utility over consumption is an increasing and concave function, $u(.)$. We assume the only difference between individuals with HD risk and without is q: individuals who carry the genetic mutation for HD are more likely to become disabled (i.e. have a higher q). Long-term care insurance is purchased for a price γ per unit in the healthy state and pays benefits when disabled; denote units of long-term care insurance as T. We assume

that γ is unrelated to genetic risk; this is consistent with the discussion in Section 2.2.

We analyze the decision to purchase insurance on a period-by-period basis. We assume that if an individual remains healthy in the next period, they get a chance to change their insurance choices, and these choices are unaffected by the current period choices; we denote the continuation utility V. Given these parameters, individuals will choose the value of T to maximize Equation (1) below.

$$EU = u(Y_1 - \gamma T) + qu(Y_2 + T) + (1 - q)V \tag{1}$$

Adverse selection occurs when individuals who look the same to the insurer, and therefore face the same γ, purchase more insurance if they have a higher chance of experiencing the bad state. We expect adverse selection if $\frac{dT}{dq} > 0$. It is straightforward to see that this is the case; the result is summarized in Equation (2) below.

$$\frac{dT}{dq} = \frac{-u'(Y_2 + T)}{\gamma^2 u''(Y_1 - \gamma T) + qu''(Y_2 + T)} > 0 \tag{2}$$

It is very important to note that this result relies crucially on the assumption that the only difference between those with and without the genetic risk is the probability of experiencing the bad state (being disabled). The actual experience in the good and bad states – income, health expenditures – are the same for the two groups. Put differently, learning that you have the genetic mutation for HD changes your expectation of the probability that you will be disabled, which changes the effective price of insurance. However, it does not change the marginal utility of income in the two states, so has no effect on the "insurance" motivation for purchasing insurance.

In the case of long-term care, this seems reasonable. The income before the period of disability is likely to be similar for otherwise comparable individuals with and without the genetic mutation, as are the nursing expenditures. The primary difference between individuals with and without the genetic mutation is the price of insurance: for those with the gene, insurance is a much better financial investment. However, this may be less true for other types of insurance. For example, in the case of life insurance, learning that you have the genetic mutation for HD has a significant impact on your "good" state (alive) consumption, given the expectation of a shorter working life and expensive disability. This means that learning about one's HD risk changes *both* the price of insurance (since death is more likely) and the relative marginal utility of income in the good and bad states, which may work in opposite directions in driving ownership.[12] The fact that the prediction of advese selection is so clear in the long-term care case is one of the major advantages of analyzing these issues in this market.

[12]There are other complications with analyzing life insurance, including the fact that much of it is owned through employers, and becoming disabled often leads to loss of employment. We hope that we will be able to analyze life insurance choices in future work.

3.2 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. Defined Own_i as equal to one of the individual owns insurance, and zero otherwise. Our primary regression equation is below

$$\text{Own}_i = \alpha + \beta(HD_{risk})_i + \mathbf{\Lambda X_i} + \epsilon_i \tag{3}$$

where $\mathbf{X_i}$ is a vector of controls. We interpret the coefficient β as the extent of adverse selection.

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. Finally, define $test_{uncl}$ as equal to one if the individual was tested, but the result cannot be inferred (note that the individual will know their test results; they are unclear, at this point, to the researchers – see Appendix A). We estimate the regression below.

$$\text{Own}_i = \alpha + \beta_1(HD_{risk})_i + \beta_2(test_{pos})_i + \beta_3(test_{neg})_i + \beta_4(test_{uncl})_i + \mathbf{\Lambda X_i} + \epsilon_i \tag{4}$$

Our theory suggests that $\beta_2 > 0$ and $\beta_3 < 0$. Note that β_2 and β_3 in this regression are identified off of variation in risk *within* the HD-risk population. This is helpful in addressing concerns that the results are driven by omitted variable bias. We continue to include data from the general population, however, since this allows us to adjust more precisely for expected variation in ownership by demographics.

In addition, we analyze how behavior is affected by proximity to testing. We define five testing variables: a variable equal to one if the individual reports testing for the first time two visits in the future, a variable equal to one if the individual reports testing for the first time one visit in the future, a variable equal to one if the individual is first tested in the current period, a variable equal to one if the individual reported having been tested last period and is positive and a variable equal to one if the individual reported having been tested last period and is negative. Note that these last two variable are equal to one for all individuals who were tested last period, two periods ago, etc. These variables are all equal to zero for individuals in the HRS. Denote these variables, respectively, as $tested_2$, $tested_1$, $tested_0$, $tested_{pos}$ and $tested_{neg}$. We estimate:

$$\begin{aligned} \text{Own}_i \quad = \quad & \alpha + \delta_1(HD_{risk})_i + \delta_2(tested_2)_i + \delta_3(tested_1)_i \\ & + \delta_4(tested_0)_i + \delta_5(tested_{pos})_i + \delta_6(tested_{neg})_i + \mathbf{\Lambda X_i} + \epsilon_i \end{aligned} \tag{5}$$

We can interpret δ_1 as the adverse selection among un-tested individuals. δ_2 has a straightforward interpretation: this is the increase (or decrease) in insurance ownership, relative to someone who is never tested, among individuals who will be tested for the first time two periods in the future. δ_3 and δ_4 have similar interpretations, but for those who are tested for the first time next period, or for the first time in the current period. δ_5 and δ_6 are the relative ownership for individuals who have been tested *at least* one period in the past, differentiated by test result.

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.[13] Comparing either of these rows to the third row, which shows ownership among untested (roughly 50% risk) individuals in PHAROS, gives us our first result: insurance ownership is significantly higher among at-risk individuals. Depending on the comparison group used, individuals in PHAROS are 15 to 17 percentage points, or about two and a half times, more likely to own long-term care insurance. This difference is strongly significant, as can be seen from the p-values at the bottom of the panel

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 again consistent with the theoretical framework. Individuals who know they carry the genetic mutation are roughly twice as 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 smaller, and not significant. 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

[13] 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.

selection in this market.

Table 3 shows our primary 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 and 4 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 slightly above 13 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.

We can get a visual sense of the magnitude of this adverse selection in Figure 2. To generate this figure we run the regressions in Column 2 of Table 3 and predict long-term care ownership based on demographics and HD risk. The histogram displays predicted ownership levels for the HD-risk population and the general population. It is clear from the graph that, in addition to having different ownership levels on average, there is very little overlap in the distributions. The effect of HD risk overwhelms any observables which determine ownership. Among other things, this may provide comfort that differences in the populations – other than HD risk – are not playing an important role in driving the coefficients.

Column 3 of Table 3 presents evidence on variations across individuals with different risks within the PHAROS population. The coefficient on HD-risk in this column can be interpreted as the additional ownership, relative to matched controls, among individuals who have not been tested. As before, we see higher ownership among this group than the controls. 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.

Column 3 provides some additional evidence of adverse selection. Individuals who know they carry the genetic mutation are 24 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 could be due to "stickiness" in ownership: once individuals own the insurance, it takes them time to get rid of it.[14] An perhaps more likely

[14]It is also possible that this is an issue with our current data – it is more difficult to infer negative test results than positive

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 0.06.

As a further test of adverse selection within the HD population, Column 4 of Table 3 explores the dynamics of purchases around the period of testing by estimating Equation (5) from Section 3. The evidence in Column 4 also points strongly to adverse selection. The coefficients on future testing are positive – suggesting, perhaps, some small increase in purchases prior to testing – but not significant. In contrast, in the periods after testing, individuals who find out they carry the genetic mutation are 47 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. In this case we more strongly reject equality in after-testing ownership for those who test positive and negative. Similar to Column 3, this provides evidence of adverse selection even within the HD-risk population.[15]

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 between 50% and 75% 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 C, 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. 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.

ones. It is our hope that this will be remedied at some point, when genetic testing data are available.

[15]The magnitude of the coefficient on testing positive is larger than in Column 3, since in Column 3 individuals who first test positive in the current period are also included, while in Column 4 the "tested, positive" coefficient is identified off of the period after testing. The larger magnitude suggests that insurance purchases may increase over time, after testing.

Selection Bias

The analysis of variation among tested individuals in Columns 3 and 4 of Table 3 brings up a potentially important issue: are tested individuals comparable to untested individuals? This may be important for two reasons. First, as in Columns 3 and 4 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 will be a selected sample of individuals – the sample is likely to contain individuals with relatively less inclination to be tested, since those who test when they are young will not be able to enroll. To the extent that inclination for testing is correlated with insurance purchases, this will mean this sample does not necessarily give us an accurate picture of the overall population.

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 – they all own insurance, or none of them do – the selection of the PHAROS sample will make little difference in our estimate of the magnitude of adverse selection.

In addition, we can observe directly whether individuals in our sample who choose to undertake predictive testing look, prior to testing, very different than those who do not.[16] 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. To the extent that this coefficient is significantly different from zero, that would point to differences across individuals even prior to testing. The results from this regression suggest zero, or at least very limited, selection. The coefficient on the interaction between HD and future testing is an insignificant 0.03. This should

[16]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.

provide confidence both in the value of this sample for representing the overall HD population, and in the comparability of untested and tested individuals.

5 Calibrating the Future of the long-term Care Insurance Market

The results in the above sections are interesting on their own as a test of whether individuals use available private information in making choices about insurance. We find strong evidence of rational behavior and see individuals making significantly different choices based on the degree of private information they have. In this section we use our estimates of adverse selection to provide some preliminary predictions about the viability of the long-term care insurance market as the share of individuals with private genetic information increases. Further, we will be able to make some preliminary statements about the possible welfare gains and losses from prohibiting or allowing the use of genetic information in insurance screening.

We consider the potential profit for long-term care insurers as the share of individuals with private information – the share which has a genetic test like the one available for HD – increases. Among other things, these calculations can give us an estimate of welfare gain or loss from restricting or allowing firms to require genetic testing. The key contribution of our results from Section 4 to this calculation is mapping out the relationship between insurance pricing and purchase behavior among individuals with different amounts of private information.

It is important to note that the results in this section are inevitably much more speculative than those in previous sections. In order to generate these estimates we make a number of assumptions about the structure of the market, the use of long-term care insurance by individuals with and without HD, and about aspects of the demand curve. Perhaps most importantly, we assume in this section that the insurance purchase behavior among individuals at risk for HD is reflective of how individuals with genetic risks for other diseases with similar mortality or morbidity consequences (like Parkinson disease or Alzheimer disease) will behave. The unique genetic and clinical features of HD may limit its generalizability; nevertheless, we feel that these results are likely to be informative about behavior of individuals with more common diseases.[17]

[17] Taylor et al (2009) give a discussion of similar issues, focusing on the case of Alzheimer disease, although they do not provide a calibration of this type.

5.1 Setup

We consider a monopoly provider of long-term care insurance facing a market of individuals, some of whom have private information.[18] This insurer can offer an insurance product for some premium which pays off if the individual needs long-term care.

For simplicity, we focus on a case with two types of individuals: high types, who constitute a share α of the population and have a very high probability of needing long-term care, and low types (share $1-\alpha$) who have a lower probability of needing care. We consider whether there is a pooling equilibrium which provides positive profits.[19] Since we know that individuals in the general population will purchase actuarially unfair insurance in some cases, it is plausible that for a small enough α, their premiums will cover the high type. How large α can be and still sustain this pooling equilibrium is the primary question of interest here. The crucial input to this calculation is the demand for insurance by each type: how much higher is the demand by the high (expensive) type than the low type? This will be informed by the results in Section 4.

Based on the details of the insurance product (benefit levels, coverage length, etc), insurance premiums and the probability of needing various types of long-term care, we define expected pay in and pay out for each type. Denote the present discounted value of money paid in premiums for type i as π_i and presented discounted value of money paid out in benefits for type i as τ_i. We define insurance *load* as the amount paid out for each dollar paid in: $load_i = \frac{\tau_i - \pi_i}{\pi_i}$. We note that we expect τ_H to be larger than τ_L, since we have defined the high type as having a larger chance of needing long-term care. It is also possible that the pay in (π_i) is lower for the high type, since they pay premiums for a shorter time, although this need not be true. The higher value of τ_H indicates that, for any given insurance product and premium, load will be higher for the high type.

Demand for insurance is a function of insurance load, and we can use the results from Section 4, along with load calculations, to estimate this relationship. Using data from actual insurance contracts and information on transitions between states, we calculate the long-term care insurance load for individuals in the general population, those with a roughly 50% risk for HD and those with 100% risk of HD. We use code based on Brown and Finkelstein (2007) to calculate loads for the general population; for the HD-risk

[18] As we discuss below, for our purposes we could draw similar conclusions using a competitive market with an imposition of zero profits.

[19] We believe there will not be a separating equilibrium in this case. The high types here expect to need long-term care with a probability very close to one. Given this, long-term care "insurance" for them is not really insurance, but rather a valuable financial asset. The insurer will only be willing to insure these high types at a very high price – close to the full price of long-term care – and will need to offer a much lower price-lower coverage plan to the low types. But this latter plan will need to be a more attractive financial asset to entice the low types, and the high types will therefore prefer that, violating the incentive compatibility constraint.

population we use a number of sources (Bolt, 1970; Nance and Saunders, 1996; Walker et al, 1981; Harper, 1996; our PHAROS data) to estimate the need for long-term care among HD-risk individuals. The results in Section 4 tell us the demand (share of individuals purchasing) which correspond to each load value. This gives us three observations which we can use to map demand onto load.[20]

A final important element in this relationship is the load at which demand is equal to zero. We could, in principle, pin this value down using the three points we observe. However, to the extent that there are non-linearities in this relationship (which there appear to be even based on the three primary data points), this may not be accurate. As an alternative, we pick this value – which we call the "zero-purchase load" – to match the premiums and demand that we observe in the market. That is, we identify a value for the zero-purchase load which implies, at least roughly, optimal prices and quantities that match the prices and quantities we observe. This value is -0.4: no one will want to purchase insurance if the payout is less than \$0.60 per \$1 paid in. We will explore the robustness of our results to this assumption. Using this final data point, along with the three primary data points, we estimate a non-parametric demand-load relationship.

This relationship between load and demand is not directly interpretable as a demand curve, since load is not equal to price. However, any price p chosen by the insurer will translate into a load for the high type and a load for the low type and, hence, into quantity demanded. The key is that since the values of τ_i, and possibly π_i, differ across types, the load defined by a single price p will be different for the two types and they will have different demands for the same price. We therefore define a demand curve for each type – $D_H(p)$ and $D_L(p)$.

Given this setup, the insurer revenue and cost functions are straightforward. Revenue (R) is the pay in for each type (π_i), multiplied by the demand from each type ($D_i(p)$) and the share of that type (α or $1-\alpha$). Cost (C) is the sum of the payouts to each type (the τ_is) multiplied by the share of that type consuming the insurance. These are expressed formally in Equations (6) and (7) below.

$$R = \alpha D_H(p)\pi_H + (1-\alpha)D_L(p)\pi_L \tag{6}$$

$$C = \alpha D_H(p)\tau_H + (1-\alpha)D_L(p)\tau_L \tag{7}$$

The monopoly insurer will choose a price p to maximize $R - C$; this price will vary with α.[21] Note that

[20] To do these calcuations for indiviauls at risk for HD 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 state. We use information from the listed sources to generate transition matrices; the matrices we use are available from the authors.

[21] We could also model the insurance market as competitive, and assume a zero profit condition. In this case we would

the insurer always has the option to sell only to the high type. We evaluate profits under different values of α, and report the largest α at which the insurer will continue to sell to both types.

5.2 Results: Demand Curves and Optimal Prices

Throughout this section we focus on a world in which the insurer offers only one insurance plan: the primary insurance plan analyzed by Brown and Finkelstein (2007). This is a policy that pays a maximum benefit of $100 per day, covers home care, assisted living care and nursing care, and lasts for a duration of, at most, 4 years. Based on this plan, and our data on the need for long-term care, we calculate loads and quantity demanded for the two types of individuals under different premiums.

Figure 3 shows the estimated demand curves for long-term care insurance for the two types of individuals. Consistent with our adverse selection results, we find that at any price the demand is larger for high type than low type individuals. At premiums above $1700, no low risk individuals will purchase insurance (actual premiums for this policy, in the data, are around $1190), and the market contains only high-risk individuals. Demand for insurance among high risk individuals is positive for any annual premium under $6500. For prices between $1700 and $6500, the market contains only high risk individuals. For both groups, demand is more elastic at lower prices.

To give some sense of how these demand curves vary under varying assumptions about the zero-purchase load, Appendix Figure 1 shows demand for the high type as this varies (the relative positions of the low type demand curves are similar). The exact shape of the curves varies with the zero-purchase assumption. In addition, not surprisingly, the lower we assume this zero-purchase load is, the higher is the maximum price under which individuals will purchase insurance.

Based on these demand curves we calculate optimal pricing and demand by the two types over a range of values of α. These results are reported in Panel A of Table 5. For a zero value of α, optimal premium is around $1350, close to the actual price for this insurance product. At this point, both high and low types are purchasing insurance – about 7% of the low types, and 35% of the high types, similar to what we see in the data. As α increases, there are some relatively small increases in pricing over some range before the market abruptly changes when the share of the high type reaches above 2.5%. The full dynamics of this can be seen in Appendix Figure 2. Above 2.5% of the high type, the insurer can no longer make non-zero profits by selling to both high and low types. They will still be willing – at a much higher price – to sell insurance to the high types. It's difficult to predict whether or not the insurer would

ask whether there is any price p, under τ_H, which would produce non-negative profits. This analysis leads us to the same conclusions about the highest possible α, since the question is still – in essence – whether there is any surplus available, and the only difference in the market structure is who gets the rents.

continue to operate at this point, with such a small share of the market as possible consumers.

Panel B of Table 5 shows similar calculations for the assumption of -0.6 as the zero-purchase load. The key thing to notice here is that while the exact value of α at which the market shifts to only selling to the high type is larger here – 5% rather than 2.5% – the basic pattern is exactly the same. Indeed, with demand curves that have this basic structure, we would expect the general intuition here to be robust. There will be a point at which the insurer cannot profitably sell to both types, and the market changes structure. As a side note, the fact that with this assumption the premium at $\alpha = 0$ is much higher, and purchases are lower, than what we see in the data suggests that this assumption is less consistent with our data.

5.3 Results: Welfare Under Varying Information Schemes

This analysis also allows us to ask how the level, and distribution, of consumer surplus would change if we allowed insurers to observe individual type – that is, if the type was not private information. This is a policy relevant question: although our survey of current insurance plans suggest they do not require genetic testing, insurers could in principle introduce this requirement. The fear of this has led a number of states to pass laws forbidding the use of genetic test results in insurance pricing, which will prohibit insurers from observing this private genetic information.

Based on the demand curves in Figure 3, and the optimal pricing in Column 2 of Table 5, we are able calculate consumer surplus in the pooling equilibrium for various values of α.[22] The consumer surplus under full revelation is also simple. In this case, the price charged to the low types will be equal to the price when $\alpha = 0$, and the price charged to high types will be equal to what they are charged when they are the only ones in the market.

Figure 4 illustrates the variation in total consumer surplus under the two information schemes as α varies. For small values of α – below the level at which the market switches to selling only to the high type – the consumer surplus is higher under the private information scheme. In addition, this surplus increases as α gets larger. This is due to the fact that, at the optimal prices, the high type individuals get a much, much larger surplus than the low types. As the share of this type increases, the total surplus increases.

This situation is reversed once α increases over the critical value of (in this case) about 2.5%, and the market sells only to the high type when information is private. In this case, the surplus under private information goes only to the high type. The total consumer surplus is strictly larger under full

[22] As is standard, consumer surplus is the area under the demand curve above the price charged.

information revelation, because at that point the insurer could continue to sell to both types, both of whom would receive positive surplus. Importantly, our results suggest that once α is above this critical value, it would be *Pareto*-improving to reveal the private information. The high types would be no worse off, since they are already being charged the full price they would pay under private information, and the low types would be made better off since they would be able to purchase insurance.

This is a strong result. It suggests that above this critical, market-changing value, there are no distributional or equality-based tradeoffs for public versus private information.[23] Public information is at least weakly better for everyone. It is also worth nothing that although the specific critical value is obviously sensitive to the demand curves and the assumptions that go into generating them, the idea that there will be a critical value above which public information is better should not be sensitive to that. For example, in the case where we assume the zero-purchase load is -0.6, the critical value is 5%, but the intuition is identical.

This analysis also makes it clear that in the short term, while the share of high type individuals is below the critical value, whether genetic information is public or private has distributional consequences. Put simply, keeping information private transfers utility from low type to high type individuals. The low types pay higher premiums because of their pooling with the high types, and the high types pay lower premiums because of their pooling with the low types. Similarly, allowing insurers to observe this information transfers utility in the other direction.

Together, this suggests that in the case of long-term care insurance, once the share of individuals with private information gets sufficiently high, laws that keep genetic information private are not optimal. In the short run, such laws may or may not be optimal, depending on how distributional concerns trade off against overall consumer surplus. In a world where the total consumer surplus was the only relevant value, these laws are optimal in the short run. This is, of course, complicated for policy since statues are rarely written to run out when information gets sufficiently advanced, but it is nevertheless the implication of this model.

Existing Genetic Information An important follow-up issue is how much genetic knowledge we currently possess. The evidence above suggest that at relatively small values for α, the insurance market may undergo significant changes. However, the fact that this value of α is small relative to 1 is not the important question. More important is understanding whether it is likely that the actual genetic knowledge will reach this level in the near future. It is important to note first that HD alone is not nearly

[23] By "public" here we indicate that the information is avaialble to the insurer; for obvious reasons making the information available to the public at large might well have negative consequences.

common enough to have significant effects on this market – the share of individuals in the population who carry the HD genetic mutation is only 0.01% – and unlikely to vary much around that. HD alone will not affect the long-term care insurance market.

However, genetic information is increasingly available for other, more common conditions. To get a sense of the magnitude of current information, we focus on genetic knowledge about the diseases that the long-term care insurers are particularly concerned about – namely, those on which insurers screen. We base this on information from long-term care insurance applications. At the start of the application, insurers ask whether the application has any of a set of roughly 10 diseases. If they say yes, they are advised not to continue with the application. Four of these diseases – Huntington, Parkinson, Alzheimer and ALS (Lou Gehrig's Disease) – have at least some genetic basis and the potential for private information. Appendix Table 1 lists, for each of these diseases, the available genetic information, as well as the overall prevalence in the population. In three of the four cases – everything except Parkinson disease – at least some affected individuals could have perfect genetic information prior to symptoms. In the case of Parkinson disease the predictiveness of the genetic testing is positive, but somewhat smaller.

Overall, these data suggest that – at the moment, based on just these diseases – roughly 0.1% of the population could be (with genetic testing) "high types" in the context of the model above. This number is extremely small. Given this, it is not surprising that insurers do not screen extensively on genetics. With even a small cost of genetic testing it would not be sensible to screen given the population shares. Further, it seems clear there are some costs of pricing differently to different groups, since long-term care insurers typically do not price differently based on simple characteristics like sex (Brown and Finkelstein, 2007). On the other hand, the overall shares in the population suggest that significant advances in identifying genes predicting Parkinson and Alzheimer diseases alone could be sufficient to bring the share of high types close to the critical value.

6 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.

Using our estimates of adverse selection in long-term care, embedded in a simple model of the

insurance market, we argue that the viability of private long-term care insurance, at least in its current form, may be threatened by increases in the share of individuals with private genetic information. As these individuals become a larger share of the population, insurers will be unable to make positive profits by offering a single price to all individuals. Barring an ability to observe individual types, private insurance companies may shut down, or sell only to high type individuals. To the extent that this is likely to occur, it may be Pareto-improving to allow insurers to observe genetic information in the state of the world where the share of individuals with private information is sufficiently large. However, a simple calibration suggests that, at the current state of genetic knowledge, total consumer surplus is higher under private information; this suggests that, at least for the immediate future, it may be optimal from a total consumer surplus standpoint to retain systems which keep this information private. This conclusions may be important in driving optimal policy, especially as legislation increasingly limits the ability of insurers to observe or use genetic information in pricing (i.e. United States House of Representatives, 2007).

An interesting and important follow-up question is to what extent are these results likely to apply to other insurance markets. Although it is obviously difficult to draw strong conclusions about this, it seems likely that the basic point that people respond to changes in insurance prices generated by genetic information is portable across contexts. However, other insurance markets may have different institutional or other features which limit (or exacerbate) the adverse selection generated by information. For example, health insurance owned through an employer may be less subject to this type of selection, simply due to more limited individual-specific variation in policies. Exploring the impact of this type of information on other markets seems a fruitful avenue for future work.

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.

Beiser, A, S Seshadri, R Au, and P Wolf, "Unpublished Data from Framingham Heart Study," Technical Report, Departments of Neurology and Biostatistics, Boston University Schools of Medicine and Public Health 2008.

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 Politcal Economy*, 2000, *108* (1), 56–78.

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, Amy Finkelstein Liran 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.

Elbaz, Alexis, James H. Bower, Demetrius M. Maraganore, Shannon K. McDonnell, Brett J. Peterson, J. Eric Ahlskog, Daniel J. Schaid, and Walter A. Rocca, "Risk tables for parkinsonism and Parkinson's disease," *Journal of Clinical Epidemiology*, 2002, *55* (1), 25–31.

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.

Johnston, Clare, Biba Stanton, Martin Turner, Rebecca Gray, Ashley Blunt, David Butt, Mary-Ann Ampong, Christopher Shaw, P. Leigh, and Ammar Al-Chalabi, "Amyotrophic lateral sclerosis in an urban setting," *Journal of Neurology*, December 2006, *253* (12), 1642–1643.

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.

N. I. H. National Institute on Aging, "Alzheimer's Disease Genetics Fact Sheet," nov 2008.

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.

Pankratz, Nathan, William C Nichols, Veronika E Elsaesser, Michael W Pauciulo, Diane K Marek, Cheryl A Halter, Joanne Wojcieszek, Alice Rudolph, Ronald F Pfeiffer, and Tatiana Foroud, "Alpha-synuclein and familial Parkinson's disease," *Movement Disorders: Official Journal of the Movement Disorder Society*, May 2009. PMID: 19412953.

Pasinelli, Piera and Robert H Brown, "Molecular biology of amyotrophic lateral sclerosis: insights from genetics," *Nature Reviews. Neuroscience*, September 2006, *7* (9), 710–723. PMID: 16924260.

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.

Reiman, Eric M et al., "GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers," *Neuron*, June 2007, *54* (5), 713–720. PMID: 17553421.

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.

Tang, M X, Y Stern, K Marder, K Bell, B Gurland, R Lantigua, H Andrews, L Feng, B Tycko, and R Mayeux, "The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics," *JAMA: The Journal of the American Medical Association*, March 1998, *279* (10), 751–755. PMID: 9508150.

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.

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.

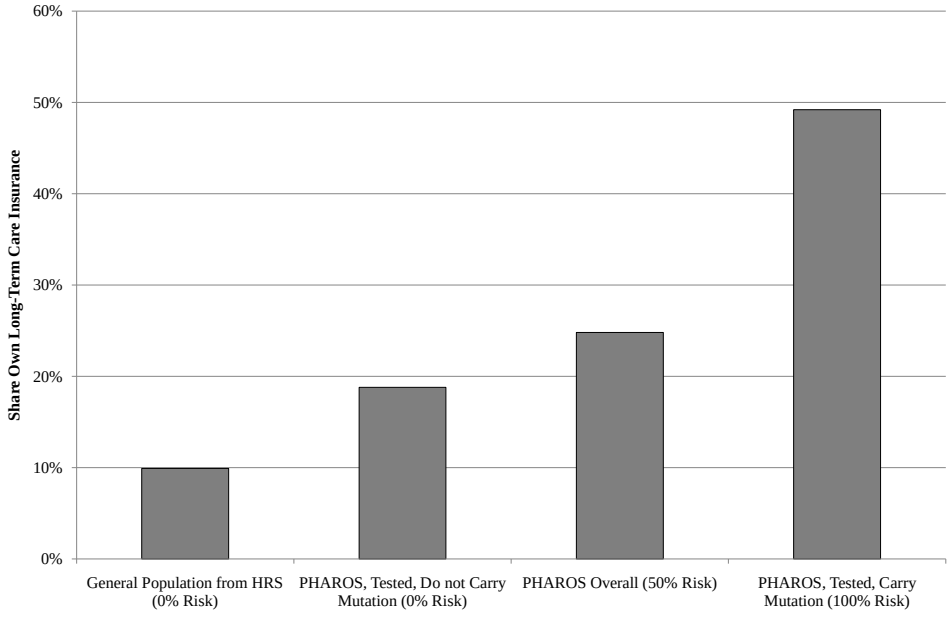
Walker, DA, PS HArper, CEC Wells, A Tyler, K Davies, and RG Newcombe, "Huntington's Chorea in South Wales: A Genetic and Epidemiological Study," *Clinical Genetics*, 1981, *19*, 213–221.

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.

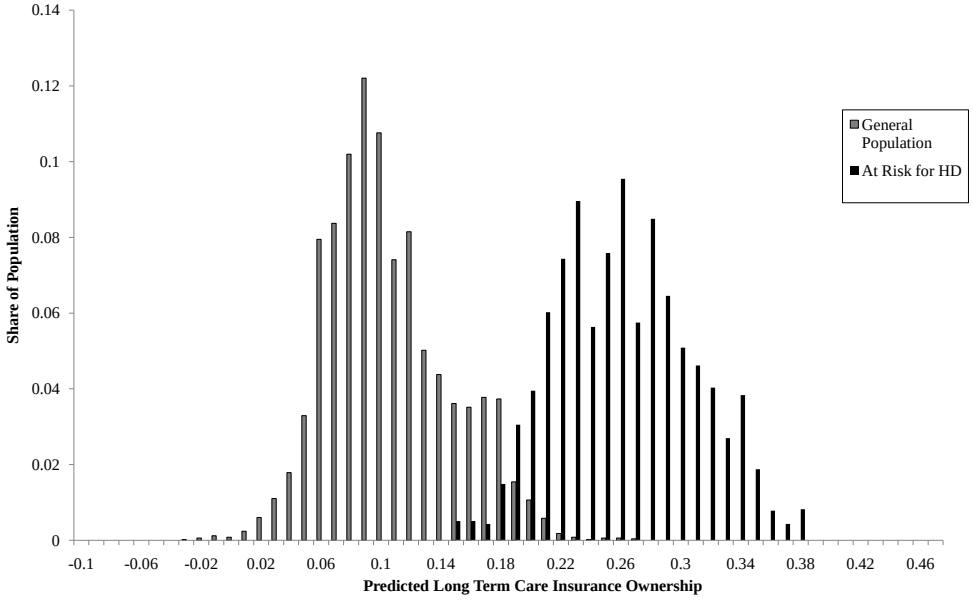
___ **and** ___ , "Genetic Testing For Alzheimers Disease And Its Impact On Insurance Purchasing Behavior," *Health Affairs*, 2005, *24* (2), 483–490.

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



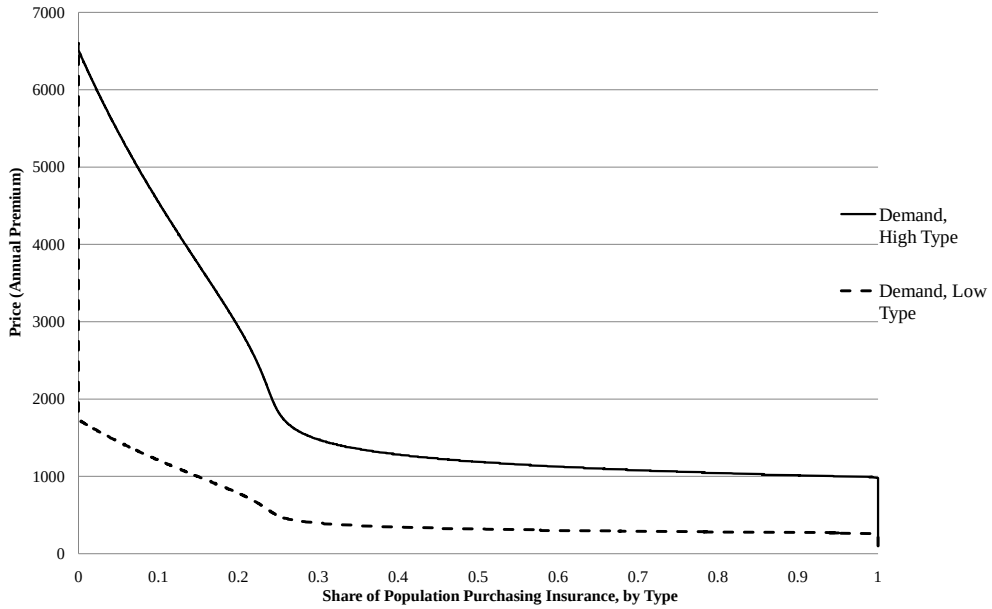
Notes: This shows a simple comparison of mean levels of ownership for individuals in the HRS, and those in PHAROS, by risk status.

Figure 2:
Histogram of Predicted Long-Term Care Insurance Ownership, by HD Risk



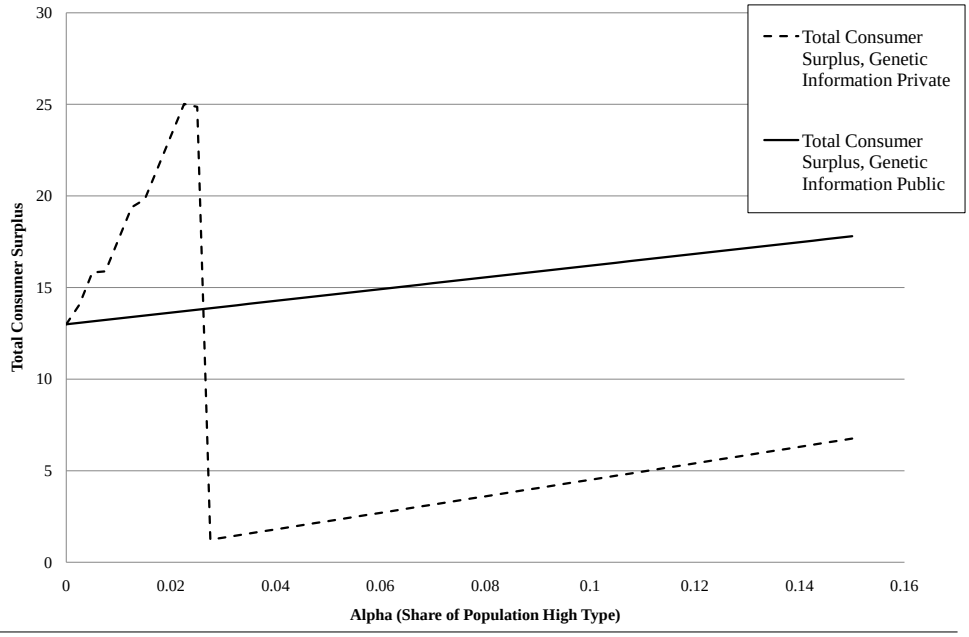
Notes: Histogram of predicted insurance ownership, generated by regressing long term care insurance ownership on observables (marital status, age, children, income , employment and education) and Huntington Disease status (at risk or not) and predicting ownership.l

Figure 3:
Demand Curves for Long Term Care Insurance, By Individual Type



Notes: This figure shows the demand for long term care insurance by a high type (with HD gene) and a low type (no HD gene). Details on how this is graph generated are in Section 5.

Figure 4:
Consumer Surplus under Varying Information Regimes



Notes: This figure shows consumer surplus under private information versus public (i.e. observable to the insurer) information.

Table 1. *Summary Statistics*

Panel A: Individuals at risk for HD (PHAROS Data)			
	Mean	*Standard Deviation*	*Number of Obs.*
Female	0.688	0.463	826
Married	0.708	0.455	685
Children	1.76	1.62	644
Education (yrs)	15.03	2.06	784
Employed	0.867	0.339	657
Age	40.8	7.67	826
Panel B: General Population (HRS Data, Under 66)			
	Mean	*Standard Deviation*	*Number of Obs.*
Female	0.582	0.493	7614
Married	0.768	0.422	7614
Children	3.24	2.02	7614
Education (yrs)	12.89	3.13	7614
Employed	0.828	0.377	6172
Age	58.12	5.25	7614

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.

Table 2. *Insurance Ownership Comparison of Means*

Group	*Share Owning Insurance*	*Sample Size*
HRS, simple mean	9.9%	5980
HRS, weighted by PHAROS age profile	8.2%	5980
PHAROS, untested	24.8%	2436
PHAROS, tested, Carry HD mutation	49.2%	71
PHAROS, tested, do not carry HD mutation	18.8%	32
p-value HRS vs. PHAROS untested		p<.0001
p-value PHAROS untested vs. carry mutation		p<.0001
p-value PHAROS untested vs. do not carry mutation		p=.42
p-value tested carry gene vs. tested do not carry mutation		p=.003

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 we report means for the HRS weighted by the age distribution in PHAROS. The PHAROS data are restricted to asymptomatic individuals to avoid the concern that symptomatic individuals may have already used their long-term care insurance.

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

Dependent Variable: Own long-term Care Insurance				
	(1)	(2)	(3)	(4)
Sample:	*Asymptomatic Individuals*			
Explanatory Variables:				
At Risk for HD	.1319***	.1434***	.1377***	.1340***
	(.019)	(.024)	(.024)	(.024)
Tested, Positive			.2488**	
			(.115)	
Tested, Negative			−.0639	
			(.126)	
Tested, Unclear			−.0351	
			(.118)	
Tested in Two Periods				−.0201
				(.108)
Tested Next Period				.1689
				(.148)
Tested This Period				.1338
				(.143)
Tested Before This Period, Positive				.4706***
				(.18)
Tested Before This Period, Negaive				−.0354
				(.188)
CONTROLS	YES	YES	YES	YES
AGE FIXED EFFECTS	NO	YES	YES	YES
p-value, pos vs. neg			.06	.05
Number of Observations	7315	7315	7292	7167
R^2	.05	.06	.06	.06

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, who are defined as individuals in the HRS or any individuals in PHAROS who are not showing obvious signs of HD. "Testing, Unclear" indicates that the test result could not be inferred from the data, not that the test was unclear to the individual (see Appendix A).

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

Dependent Variable:	*Own long-term Care Ins.*
Explanatory Variables:	
At Risk for HD	.1301***
	(.024)
HD, Ever Tested	.0315
	(.08)
Controls in all columns: income categories, education categories, number of children, married, gender, employed, age fixed effects.	
Number of Observations	7207
R^2	.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.

Table 5. ***Optimal Pricing, Variations in High Type***

Panel A: Primary Results, Zero-Purchase Load = -0.4			
Share of High Type	**Optimal Premium**	**Share of High Type Purchasing**	**Share of Low Type Purchasing**
0.0%	1350	35.2%	6.7%
0.5%	1360	34.5%	6.5%
1.0%	1385	33.3%	6.0%
1.5%	1405	32.3%	5.5%
2.0%	1405	32.3%	5.5%
2.5%	1435	31.3%	4.9%
3.0%	5120	6.7%	0.0%
5.0%	5120	6.7%	0.0%
7.0%	5120	6.7%	0.0%
10.0%	5120	6.7%	0.0%
15.0%	5120	6.7%	0.0%
Panel B: Robustness, Zero-Purchase Load = -0.6			
Share of High Type	**Optimal Premium**	**Share of High Type Purchasing**	**Share of Low Type Purchasing**
0.0%	1665	30.5%	5.1%
0.5%	1695	29.8%	4.9%
1.0%	1695	29.8%	4.9%
1.5%	1750	28.9%	4.5%
2.0%	1775	28.5%	4.3%
2.5%	1790	28.1%	4.1%
3.0%	1830	27.6%	3.9%
5.0%	6305	5.1%	0.0%
7.0%	6305	5.1%	0.0%
10.0%	6305	5.1%	0.0%
15.0%	6305	5.1%	0.0%

Notes: This table shows variation in optimal pricing and share of individuals purchasing as we vary the share of the high type in the population. Results are generated based on the model in Section 6. Panel A shows our primary results; Panel B shows the robustness of these results to variation in our assumption about the load at which purchases are equal to zero.

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 more 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: Inferring Genetic Test Results

This appendix discusses how we infer genetic test results for individuals who report testing in the data. Although genetic tests were performed as part of the study, we do not have access to those data at this point, and must infer test results from other information in our data. An observation in the data is an individual-visit, and visits occur approximately every 9 months to 1 year. At each visit, individuals are asked if they have had genetic testing. At enrollment, no one was tested; over the course of the study, about 10% of individuals eventually get tested. We are concerned only with inferring test results for tested individuals.

There are two pieces of information we use for this inference. The first is the doctor reports on the likelihood they think the individual has HD, which are on a scale from 0 to 4, where 3 indicates "symptoms of HD with greater than 90% confidence" and 4 indicates "symptoms with greater than 99% confidence". The second is self-reported probability of having HD, which is asked at a subset of visits. We observe these variables at visits after testing takes place. We code an individual as having tested positive if at these subsequent visits either (a) a doctor reports they have a 3 or 4 on the rating scale or (b) the individual reports a 100% probability of carrying the genetic mutation. We code an individual as testing negative if they report a 0% probability of carrying the genetic mutation. Some individuals cannot be coded, since they do not fit either of these profiles, typically because they are not asked about their probability of carrying the genetic mutation after testing.

It should be clear from this discussion that our data on positive test results is likely to be more comprehensive than the data on negative test results; it seems likely that a majority of the individuals for whom we cannot infer test results actually tested negative.

Appendix C: 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 polices 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.

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

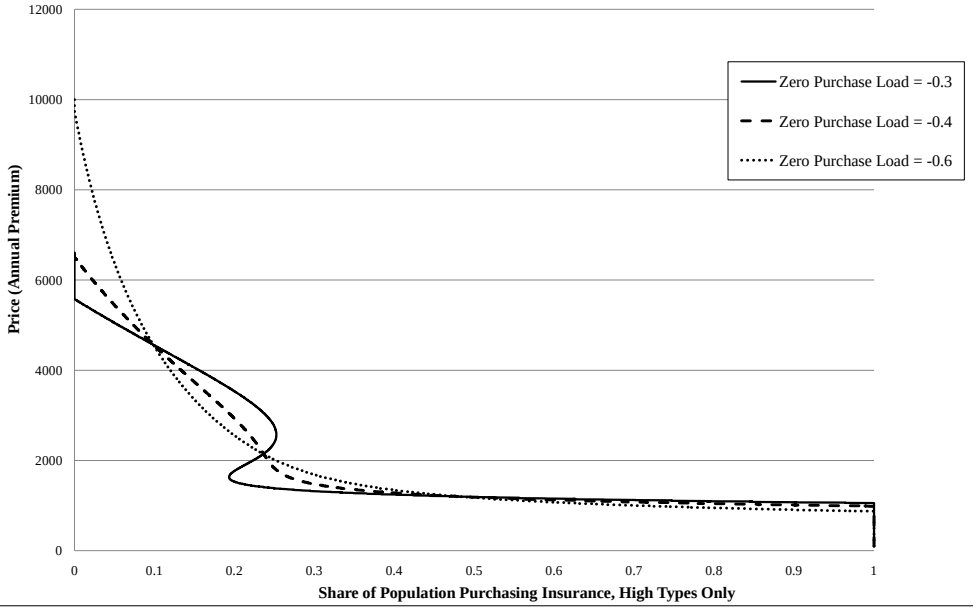
standard errors in parentheses, clustered by individual
* significant at 10%; ** significant at 5%; *** significant at 1%
Notes: This table compares long-term care insurance ownership for individuals at risk for Huntington disease and those in the general population (from the CEX). Controls (in all columns) 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.

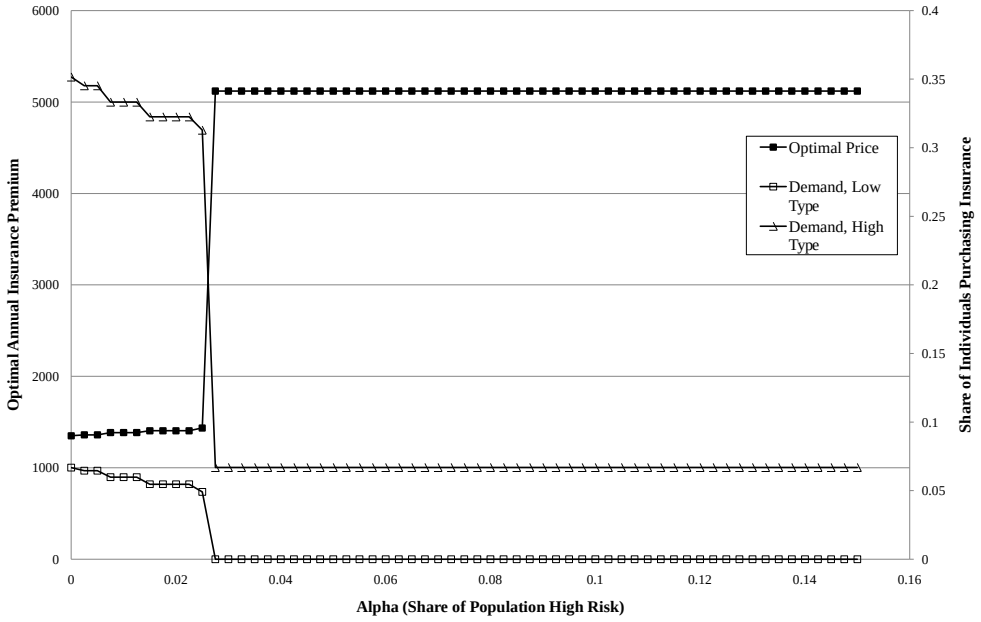
Stand-Alone Appendix Figures and Tables

Appendix Figure 1
Demand Curves, High Type, Robustness to Alternative Assumptions



Notes: This figure shows the demand for long term care insurance by a high type -- an individual who knows they carry the HD gene -- under varying assumptions about the zero-purchase load for insurance. Details are in Section 5.

Appendix Figure 2:
Optimal Insurance Pricing and Demand as High Risk Share Varies



Notes: This figure shows variation in optimal insurance pricing and demand by the two types as the high type share of the population varies.

Appendix Table 1. ***Currently Available Genetic Testing***

Disease	**Prevalence**	**Current Genetic Testing Information**	**Sources**
Huntington disease	0.01%	Perfectly predictive test covers 100% of cases	(Walker, 2007)
Parkinson disease	1.7%	Genetic tests can detect causative mutations in less than 5% of cases, but are not commercially available.	Elbaz et al, 2002; Pankratz et al., 2009
Alzheimer disease	Early Onset: 0.8%; All cases (early and late onset): 11%	Genetic test perfectly predicts early onset for 5% of cases with 100% accuracy, but is not commercially available; a different and widely available genetic test can predict increased risk (2-3 times higher) of late onset	Beiser et al., 2008; Reiman et al., 2007; National Institute on Aging, 2008; Tang et al., 1998
ALS	0.02%	Perfectly predictive test covers 20-25% of inherited cases, which make up 10% of all cases (available); tests theoretically possible for a fraction of other inherited cases, but not performed. No test for the non-inherited 90% of cases.	Johnston et al., 2006; Pasinelli and Brown, 2006.

Notes: This table describes currently available genetic knowledge about diseases which are of primary concern to long-term care insurers. The source column lists only a subset of sources on each topic, typically a summary article.