

Multiple dimensions of private information: evidence from the long-term care insurance market

Amy Finkelstein
Harvard University and NBER

Kathleen McGarry
University of California, Los Angeles and NBER

August 2004

This paper examines empirically the implications of multiple dimensions of private information in insurance markets. Focusing on the long term care insurance market, we demonstrate that two types of people purchase long-term care insurance: individuals with private information that they are high risk and individuals with private information that they have high insurance preferences. Ex-post, the former are higher risk than insurance companies would expect, while the latter are lower risk. As a result, while most standard (uni-dimensional) models of asymmetric information predict a positive correlation between insurance coverage and risk occurrence, these offsetting selection factors result in no such correlation in the long-term care insurance market. Our results demonstrate that insurance markets may suffer from asymmetric information, and its negative efficiency consequences, even if, in aggregate, those with more insurance are not higher risk. They also suggest a straightforward, general test for asymmetric information in insurance markets.

Key Words: asymmetric information; long-term care insurance; adverse selection

JEL classification: D82, G22, I11

We thank Daron Acemoglu, Jeff Brown, Pierre-Andre Chiappori, Raj Chetty, Janet Currie, David Cutler, David de Meza, Seema Jayachandran, Jerry Hausman, Ginger Jin, Larry Katz, Ben Olken, Sarah Reber, Casey Rothschild, Michael Rothschild, Bernard Salanie, Jesse Shapiro, Jonathan Skinner, Jonathan Wright, and numerous seminar participants for helpful comments and discussions. We are especially grateful to Jim Robinson for generously providing us with the actuarial model of long-term care utilization. We thank the NBER and NIA for financial support.

Theoretical research has long emphasized the potential importance of asymmetric information in impairing the efficient operation of insurance markets. This, in turn, has provided the fundamental economic rationale for public intervention in private insurance markets. However, several recent studies of different insurance markets have found no evidence to support the central prediction of many asymmetric information models that those with more insurance should be more likely to experience the insured risk.¹ These findings have challenged the conventional wisdom that asymmetric information is an important phenomenon in insurance markets.²

The standard asymmetric information models on which these studies are based assume that individuals have private information about their risk type alone. This paper provides what is, to our knowledge, the first empirical evidence of an insurance market in which individuals have private information about two characteristics: risk type and risk preferences. Moreover, we show that these multiple forms of private information act in offsetting directions to produce an equilibrium in which – contrary to the predictions of the standard models – those with more insurance are not more likely to experience the insured risk.

Specifically, using data from the Asset and Health Dynamics of the Oldest Old (AHEAD) cohort of the Health and Retirement Survey (HRS), we find that there are two types of individuals who own long-term care insurance: individuals who believe that they are more likely to use a nursing home than the insurance industry would predict (the standard asymmetric information scenario), and individuals who have above-average preferences for insurance conditional on what the insurance company observes about them. We show that the first group does, in fact, have above-average nursing home use, while the second group has below average use. This second group appears to consist of more “cautious” individuals – individuals who are more likely than average to invest in preventive health activities, activities that the insurance company does not observe – as well as wealthier individuals. The net result is that, despite the

¹ See e.g. Chiappori and Salanie 2000, Cardon and Hendel, 2001 and Cawley and Philipson 1999.

² Indeed, even when awarding the 2001 Nobel prize for the pioneering theoretical work on asymmetric information, the Nobel committee noted in its extended citation that empirical evidence of asymmetric information in insurance markets was “ambiguous” (Bank of Sweden, 2001).

presence of asymmetric information about risk type, in equilibrium the insured on average have no higher a risk profile than the uninsured.

Our findings have several important implications for understanding the impact of asymmetric information on insurance markets. First, they demonstrate that asymmetric information may exist even if the insured are not above-average in their risk type; a related implication is that asymmetric information may impair market efficiency without raising the price of insurance above the population-average actuarially fair level. Second, our findings suggest that the widely used test of asymmetric information, based on the presence or absence of a positive correlation between insurance coverage and risk occurrence, can lead to incorrect conclusions. However, they also suggest an alternative approach for testing for asymmetric information in insurance markets that we expect will have wide applicability. Third, our evidence of preference-based selection may provide a unifying explanation for the disparities across different insurance markets in whether the insured are above-average in their risk type. For example, there is evidence that annuitants are longer-lived (i.e. higher risk) than the general population while those with life insurance (which insures the opposite longevity risk) are also longer-lived (McCarthy and Mitchell 2003, Cawley and Philipson 1999, Finkelstein and Poterba 2002, 2004). Preference-based selection may well operate in the opposite direction in these two markets and reconcile the two findings.

The remainder of the paper proceeds as follows. In Section one, we provide some theoretical examples to illustrate how – when individuals differ in terms of their risk preferences as well as their risk type – it is possible to generate any equilibrium correlation between risk type and risk occurrence. We emphasize that such an equilibrium may be inefficient regardless of the observed correlation. As a result, knowledge of the reduced form correlation between insurance coverage and risk occurrence is insufficient for understanding either the efficiency of the market or the structure of information.

Section two provides some background on long-term care expenditure risk and the private long-term care insurance market. Long-term care expenditure risk is among the largest financial risks facing today's elderly, making an investigation of this market's information structure particularly interesting.

Section three contains our empirical findings. In the AHEAD data, we can measure each individual's subjective belief of his probability of entering a nursing home over the next five years, and compare that prediction to his subsequent five-year nursing home utilization. We supplement these data with measures of the insurance companies' information set and risk classification practices; these measures are based on insurance company application forms which reveal the set of individual characteristics observed by the insurance companies, as well as on the industry's actuarial model of nursing home utilization as a function of these observed characteristics. We find that, after controlling for the insurance company's risk-classification, the individual's beliefs about his subsequent nursing home use remain a positive and statistically significant predictor of subsequent nursing home use. This provides direct evidence that individuals in this market have private information about their risk type. Moreover, we find that the individual's private information about his risk type is positively correlated with insurance coverage. However, individuals with long-term care insurance are no more likely to enter a nursing home than those without, even when controlling for the insurance companies' risk classification. This lack of a positive correlation between nursing home utilization and long-term care insurance coverage in the aggregate, combined with private information about risk type, points mechanically to the existence of a second form of unobserved heterogeneity – heterogeneity in preferences – which must offset the risk based selection, obscuring the expected correlation. We conclude Section three with direct evidence of this preference-based selection.

In Section four, we discuss the implications of our findings for testing for asymmetric information in insurance markets. The last section concludes.

Section 1: Theoretical framework

In this section, we use some simple examples to illustrate the intuition behind the two insights that motivate the subsequent empirical work. First, when there is asymmetric information about risk preferences as well as risk type, the market equilibrium may display a positive, negative, or zero

relationship between insurance coverage and risk occurrence. Second, and more importantly, with multiple forms of unobserved heterogeneity, the market may be inefficient, regardless of the reduced form correlation between insurance coverage and risk occurrence. In other words, we emphasize that it is the structure of information and not the reduced form correlation between insurance coverage and risk occurrence that matters for efficiency.

For ease of exposition, we only consider the case of ex-ante private information about (exogenous) risk type (i.e. adverse selection), and ignore any ex-post private information about (endogenous) risk type (i.e. moral hazard). We also limit the discussion to competitive insurance markets. The basic ideas discussed here, however, carry over to these other settings.³

1.1 The standard model: private information about risk type only

Standard theoretical models of asymmetric information consider individuals who differ only in terms of their (privately known) risk type. The classic result – the Rothschild and Stiglitz (1976) separating equilibrium – is shown in Figure 1. There are two types of individuals: high risk individuals have probability π_H of experiencing the financial loss L , and low risk individuals have probability π_L , with $0 < \pi_L < \pi_H < 1$. The proportion of the population who is high risk is commonly known to be λ ; the population actuarially fair (pooling) price, denoted π_p , is therefore $\lambda\pi_H + (1 - \lambda)\pi_L$. The insurance market is competitive. If all individuals are risk averse and there are no transaction costs, the efficient equilibrium is for all individuals to have full insurance; the 45 degree line denotes the set of full insurance policies.

In the equilibrium in Figure 1, the high risk individuals purchase full insurance at their actuarially fair marginal price of π_H . The low risk individuals also pay their own-type actuarially fair marginal price, but are constrained to buy less than their optimal full insurance policy by the incentive compatibility

³ Interested readers should consult Chiappori et al. (forthcoming), de Meza and Webb (2001) and Jullien et al. (2002) for a formal presentation that includes these other cases as well as the one considered here.

constraint of the high risk individuals. The equilibrium is thus inefficient. It exhibits a positive correlation between insurance coverage and risk occurrence.

The separating equilibrium in Figure 1 arises because, as drawn, the low risk individuals prefer to buy the separating policy (which involves less than full insurance at their own-type actuarially fair price and is denoted by α_L) than to purchase any amount of insurance on the population actuarially fair line.

However, it is possible that, instead, the low risk individuals are sufficiently risk averse that they prefer paying the higher pooling price and purchasing more insurance to buying the separating policy α_L . This situation is depicted in Figure 2. In this case, the equilibrium is a pooling equilibrium, with the equilibrium insurance policy γ determined by the maximum amount of insurance that low risk individuals will purchase at the pooling price π_p ; it is shown graphically by the tangency of the low risk indifference curve to the pooled actuarially fair line.⁴

In Figure 2, while there is asymmetric information, there is no correlation between insurance coverage and risk occurrence because high and low risk individuals purchase the same insurance policy. Is the equilibrium therefore efficient? The answer is a resounding “no.” In the pooling equilibrium, the marginal price faced by the low risk individuals (π_p) is distorted above their actuarially fair marginal price (π_L). Because any policy bought by both high risk and low risk individuals is sold at a marginal price above the low-risk actuarially fair price, the maximum amount of insurance that low risk individuals will want to purchase will be less than the optimal (full) insurance. High risk individuals would be willing to buy full insurance at this price, but it is not offered to them because as Figure 2 makes clear, if

⁴ Since the Rothschild and Stiglitz (1976) Nash equilibrium concept does not permit a pooling equilibrium, the equilibrium shown in Figure 2 requires the foresight equilibrium concept of Wilson (1977). In the foresight equilibrium, consumers choose a single insurance policy to maximize expected utility, each policy earns non-negative profits individually, and there is no other set of policies outside of the equilibrium set which, if offered, would earn positive profits in the aggregate and non-negative profits individually, *after the unprofitable policies in the original set have been withdrawn*. The italicized portion of the equilibrium definition indicates the modification of the equilibrium concept used by Rothschild and Stiglitz that permits the existence of a pooling equilibrium rather than non-existence of equilibrium as would be the case in Rothschild and Stiglitz (1976).

more insurance than that available in the equilibrium policy γ were offered at the population-pooling price, it would attract only high risk individuals and therefore lose money.

Thus, even though both high and low risk types are buying the same insurance package, the equilibrium outcome is inefficient for both groups. Indeed, it is precisely because both high risk and low risk individuals pay the same price for the same insurance policy and have different expected costs that it cannot be the case that both groups are paying an actuarially fair price on the margin. The quantity of insurance purchased by at least one group will thus not be first best. Indeed, as depicted in Figure 2, neither group is paying their own-type actuarially fair price on the margin for policy γ , and the insurance coverage for both groups is less than their first best full insurance policy.⁵

In the pooling equilibrium shown in Figure 2, only one type of insurance policy is purchased in equilibrium. In practice, however, in most insurance markets – including the long-term care insurance market on which we focus – multiple policies are purchased in equilibrium (including the null policy). As a result of this unrealistic feature, the one-policy pooling equilibrium has received little attention from the empirical literature devoted to testing for the presence of asymmetric information in insurance markets.

1.2 Heterogeneity in risk preferences as well as risk type

In order to generate an asymmetric information equilibrium with *multiple policies* that does not exhibit a positive correlation between insurance coverage and risk occurrence, individuals must differ on some additional (unobserved) dimension and the price of policies must be marked up above the expected claims cost.⁶ Intuitively, heterogeneity in preferences – such as risk aversion – will ensure that individuals of the same risk type will demand different policies from the same option set. The markup is required because otherwise, high risk individuals will (as long as they are somewhat risk averse) choose either their full insurance policy at their own-type actuarially fair price or the policy with the most insurance

⁵ The pooling equilibrium shown in Figure 2 can only be efficient if low risk individuals are infinitely risk averse in which case, they will buy full insurance even at the pooling price.

⁶ See Chiappori et al. (forthcoming) for a formal proof of this statement in a general setting.

that is available at the population-average pooling price (which is below their own-type actuarially fair price).

An example of such an equilibrium is shown in Figure 3, which makes the two necessary modifications to the example in Figure 2. First, we introduce heterogeneity in preferences: we assume that fraction ω of each risk type is high risk aversion (“worried”), and $1 - \omega$ are low risk aversion (“less worried”).⁷ As a result, there are now four possible types of individuals, defined by their risk type (high or low) and their risk aversion (worried or less worried). Second, as noted above, we assume an (exogenous) markup (θ) on the pricing of the insurance policies that is sufficiently high that the population pooling marginal price ($\pi_p + \theta$) is above the high risk type’s actuarially fair (no-load) marginal price π_H . As a result, even when pooling with low risk individuals, high risk individuals will face a price above their own-type actuarially fair marginal price. They will therefore not demand full insurance. More importantly, the amount of insurance they will demand at this price will vary with their risk aversion, with the more risk averse demanding more insurance. Such an equilibrium is illustrated in Figure 3; both high and low risk, worried individuals (whose indifference curves are given by the dotted lines) pool at the more comprehensive insurance policy on the pooling line, while the high and low risk, less worried individuals (whose indifference curves are given by the solid lines) pool at the less comprehensive insurance policy on the pooling line.

This equilibrium does not result in a positive correlation between insurance coverage and risk type, but yet it is not efficient. In Figure 3, both worried and less worried low risk individuals face a marginal price ($\pi_p + \theta$) above their own-type actuarially fair price ($\pi_L + \theta$); the low risk individuals are therefore underinsured relative to the efficient amount of insurance that they would buy under symmetric

⁷ Note that while we talk about preference heterogeneity here in terms of heterogeneity in risk aversion, in principle the preference heterogeneity could be due to any factor that affects willingness to pay for insurance (other than risk occurrence) and is not observed by the insurance company.

information (which of course is now less than full insurance due to the load).⁸ Again, the inefficiency stems from the fact that individuals with different expected costs pay the same price for insurance. Interestingly, because those with more insurance are not higher risk, the example shown here is one in which asymmetric information has negative efficiency consequences even though it does not raise prices above the population-average actuarially fair price. The inefficiency arises from the fact that, on the *margin*, individuals are not facing their own-type actuarially fair price, even though on *average* prices are actuarially fair for the insured population.

In the above example, we assumed that the fraction of low risk individuals who are “worried” (ω_L) and the fraction of high risk individuals who are worried (ω_H) were the same, and showed that this can result in an equilibrium with no correlation between insurance coverage and risk type. One way to generate a negative correlation between insurance coverage and risk occurrence would be if the proportion of “worried” individuals among the low risk types in Figure 3 is slightly increased, so that $\omega_L > \omega_H$. Similarly, if the proportion of “worried” individuals among the high risk types in Figure 3 is instead slightly increased (so that $\omega_H > \omega_L$), then the equilibrium will display the standard positive correlation between insurance coverage and risk occurrence.

Thus, when the standard asymmetric information model is extended to allow for heterogeneity in risk preference as well as risk type, it is possible that the equilibrium may display any relationship between risk occurrence and insurance coverage and that the equilibrium may be inefficient.⁹ Conversely, the presence of a positive correlation between insurance coverage and risk occurrence is not sufficient to

⁸ High risk types may be over or under-insured relative to the first best, depending on their degree of risk aversion relative to the low risk individuals with whom they are pooling; de Meza and Webb (2001) provide an example in which high risk individuals inefficiently purchase too much insurance.

⁹ Although the examples we have discussed demonstrate inefficient equilibria, this need not be the case. For example, if low risk individuals are substantially more risk averse than high risk individuals and loads are sufficiently high it is possible that the incentive compatibility constraint on high risk individuals will not bind and an efficiently operating market could display a positive (or negative) correlation.

establish the presence of asymmetric information or market inefficiency.¹⁰ Moreover, even if the equilibrium is known to involve private information about risk type, it is not, in general, possible to make conclusions about the relative efficiencies of alternative equilibria based on the observed correlation between insurance coverage and risk occurrence. These different equilibria reflect differences in the correlation between risk type and risk aversion. Welfare comparisons across equilibria in which individuals have different preference structures are notoriously difficult.

In the remainder of this paper, we show that in the case of the long-term care insurance market, these two types of information – private information about risk type and private information about risk preferences – are both present, and act in offsetting directions to produce no correlation between insurance coverage and risk occurrence in the aggregate. Before doing so, it is worth noting that the conditions needed to generate such an equilibrium are – *ex ante* – reasonable ones to expect to encounter in insurance markets. A large empirical literature has documented the presence of heterogeneous consumer preferences not directly observed by the seller for a range of consumer goods, although to our knowledge this paper is the first to provide evidence of such preference heterogeneity in insurance markets. The existence of a marginal load also seems a reasonable characterization of insurance markets (Chiappori et al., forthcoming). These loads may be generated endogenously, if firms have market power. Alternatively, they may come from exogenous factors such as premium taxes or administrative costs.

Section 2: Background on the long-term care insurance market

2.1 Long-term care expenditure risk and private insurance

Annual expenditures on long-term care in the United States totaled \$135 billion in 2004, comprising over 8.5% of total health expenditures, or roughly 1.2 percent of GDP (Congressional Budget Office 2004).

Within the elderly population these expenses are distributed unevenly. For example, Murtaugh et al.

¹⁰ For example, in a perfectly competitive insurance market with symmetric information and a marginal load on policies, if all high risk individuals are “worried” and all low risk individuals are “less worried”, there will be an efficient equilibrium with a positive correlation between coverage and risk.

(1997) estimate that while 60 percent of individuals who reach age 65 will never enter a nursing home, one-fifth of individuals who do enter a nursing home will spend at least five years there. These figures suggest potentially large welfare gains from insurance coverage that reduces this expenditure risk.

However, most of this risk is currently uninsured. One-third of long-term care expenditures for the elderly are paid for out of pocket; double the share paid for out of pocket in the health sector as a whole. This disparity primarily reflects the limited size of the private long-term care insurance market wherein private insurance reimburses only 4 percent of long-term care expenditures, compared to 35 percent of expenditures in the health care sector as a whole (Congressional Budget Office 2004, National Center for Health Statistics 2002).

This limited coverage stems both from the relatively small size of the covered population and gaps in coverage in purchased policies. We estimate that in the 2000 Health and Retirement Study (HRS), only 10 percent of those aged 65 and over had private long-term care insurance coverage. Moreover, policies often have daily and lifetime benefit caps and are typically not indexed for inflation. As a result, Brown and Finkelstein (2004a) estimate that typical private policies pay for only one-third of the expected present discounted value of long-term care costs for a 65 year old.

A variety of theoretical explanations have been proposed for the limited size of the private long-term care insurance market (see Norton, 2000 for a review). Asymmetric information is often raised as one potential explanation, yet there exists very little empirical evidence on its presence in this market.

2.2 Preference heterogeneity and loads in the long-term care insurance market

In Section one we demonstrated that an asymmetric information equilibrium with multiple policies that does not exhibit a positive correlation between insurance coverage and risk occurrence may exist if individuals also differ on an unobserved dimension other than risk type (e.g. preferences) and if the price on policies is marked up above the expected claims cost. Medicaid, the major source of public insurance for long-term care expenditures, is a source of both preference heterogeneity and loads in the private long-term care insurance market.

Currently Medicaid pays for over one-third of long-term care expenditures (Congressional Budget Office, 2004). It is a payer-of-last resort covering long-term care expenditures only after an individual has exhausted nearly all of his financial resources. As a result, it provides substantially more comprehensive insurance for lower-wealth individuals than for higher wealth individuals, making it likely that an individual's wealth – a factor not used in pricing private policies – will be an important source of preference heterogeneity in private insurance demand. Indeed, private insurance coverage rises monotonically with wealth in both our AHEAD sample and in other data (see e.g. HIAA 2000a). We will explore this more formally in Section 3.4 below.

Medicaid also creates substantial marginal loads on private long-term care insurance products by imposing an implicit marginal tax on private policies. This implicit tax stems from the fact that for many individuals a portion of the premiums on a private policy pay for benefits that would have otherwise been covered by Medicaid. Brown and Finkelstein (2004b) document that the implicit tax imposed by Medicaid is substantial.¹¹

Finally, we note that the long-term care insurance market has been essentially unregulated until recently (NAIC 2002a, 2002b, Lewis et al., 2003).¹² This makes it an especially attractive market for studying private information. In a regulated market it is likely to be difficult to infer which aspects of the equilibrium are general and which are brought about by the specific regulations in place.

Section 3: The Structure of Information and Equilibrium in the LTCI Market

To investigate the issue of asymmetric information in the long-term care insurance market, we draw on a rich dataset that contains specific information about individuals' beliefs about their risk type, preferences for preventive health care, insurance choices, ex-post risk occurrence, and a full set of health indicators which allow us to proxy an insurer's risk categorization. Our empirical strategy is to demonstrate first the

¹¹ In addition to the marginal load imposed by Medicaid, premiums are also marked up above expected (gross) claims; Brown and Finkelstein (2004a) estimate that the typical policy purchased pays out, on average, only 82 cents in expected present discounted value (EPDV) benefits for every dollar in EPDV premiums. However it is unclear whether the 18 cent average gross load is a fixed or marginal load, or some combination.

¹² Our data pertain to the period prior to the newly imposed regulations.

existence of private information about risk-type (section 3.1) and to assess the relationship between this private information and insurance coverage (section 3.2). We then show that despite direct evidence of private information about risk type that is positively correlated with insurance coverage, there is no positive correlation in the aggregate between insurance coverage and the use of long-term care (section 3.3). Finally, following the framework developed in Section 1, we present evidence of the offsetting preference based selection which can reconcile these ostensibly contradictory findings (section 3.4).

3.1 Private Information about Risk Type

We use individual-level panel survey data from the Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS) to investigate whether individuals have private information about their chances of using a nursing home.¹³ This cohort, first interviewed in 1993, is representative of the non-institutionalized population born in 1923 or earlier and their spouses. We use the panel nature of the data to track nursing home utilization for the 1995 respondents through the latest currently available interview in 2000.¹⁴ At the beginning of the panel in 1995, the average age of individuals in our sample is 79. Appendix A provides more detail on the data and our sample.

Information on individuals' beliefs about their risk of nursing home utilization comes from responses to the following question asked in the 1995 AHEAD:

“Of course nobody wants to go to a nursing home, but sometimes it becomes necessary. What do you think are the chances that you will move to a nursing home in the next five years?”

Individuals are asked to give a response on a scale from zero to 100, which we rescale to be between 0

¹³ The HRS is a panel survey that follows a nationally representative sample of older individuals with interviews scheduled approximately every other year (with one exception). Respondents are initially selected from the non-institutionalized population but are followed to nursing home should they move to one. Details about the survey and the data itself are available at <http://hrsonline.isr.umich.edu/>

¹⁴ Because the first wave of the survey in 1993 does not provide a reliable measure of long-term care insurance coverage – which will figure prominently in subsequent analyses – our analysis begins with the second (1995) wave. See Appendix A for details.

and 1.¹⁵ We can then follow these individuals over the subsequent five years and compare their predictions to their ex-post 5-year nursing home experience.

An important consideration is whether individuals' reporting of their beliefs contains any meaningful information about their actual beliefs. Two factors are encouraging on this dimension. First, individual predictions appear to be correct on average: the average self-reported probability of nursing home use in a five year period is 18 percent, while 16 percent of the responders enter a nursing home over the five year period. Second, we find that self-reported nursing home entry probabilities co-vary in sensible ways with known risk factors; they are higher for women than for men, and increase monotonically with age and with deteriorating health status. These results are consistent with other work that has found sensible covariance patterns for self-reported mortality probabilities and characteristics such as the individual's age or health status (Hamermesh, 1985, Hurd and McGarry, 2002, Smith et al., 2001).

However, one well-known issue with self-reported probabilities is the phenomenon of "focal" or "categorical" responses wherein respondents give round figures such as 0, 50 or 100 percent. Figure 4 shows a histogram of the responses. Almost 50 percent of respondents report a five-year nursing home entry probability of zero and 14 percent report a 50 percent probability. Hurd and McGarry (1995) and Gan et al. (2003) report a similar preponderance of such categorical responses for self-reported mortality probabilities in the HRS.

It is somewhat unclear how to treat these categorical responses. Our goal is to measure individual beliefs. To the extent that the categorical responses represent the "true" subjective probability of the individual, no adjustment to individuals' statement of their beliefs is warranted. However, the preponderance of focal responses raises the possibility that individual responses convey information about their beliefs of the general nature of their risk (e.g. low, medium, or high) but not about the *scale* of the risk. For example, about 8 percent of individuals who report a zero probability of 5-year nursing home entry have private long-term care insurance; this suggests that an answer of zero may convey a belief that

¹⁵ The question was not asked of the approximately 13 percent of the 1995 respondents for whom the interview was completed by a proxy respondent; this excludes, among others, the most cognitively impaired.

the entry probability is quite low, but not that it is literally zero. In this case, grouping the individual predictions into several categories, rather than including them as a continuous variable, may be more appropriate.

With these considerations in mind, we report results using two alternative measures of the beliefs. First, we use the actual response of the individual, which we refer to as our “continuous measure” of individual beliefs. Second, we use a series of indicator variables for whether the individual reported 0 (49%), 1-49 (30%) or 50-100 (21%). We refer to this as our “categorical measure.” Finally, we note that individuals may not be comfortable reporting probabilistic answers, and may not in fact even think in these terms. If they use probabilistic information in making insurance purchase decisions, but are unable to translate these latent probabilities into numbers when faced with a survey question, the resulting measurement error will lead us to underestimate the extent of the individual’s information.

We estimate the relationship between nursing home utilization and beliefs about nursing home utilization with the following equation:

$$(1) \text{ CARE} = X\beta_1 + \beta_2\text{B} + \varepsilon.$$

We estimate a linear probability model of whether the individual went into a nursing home in the five years between 1995 and 2000 (CARE) on his 1995 self-reported beliefs of this probability (B) and controls for risk classification that would be assigned to the individual by insurance companies (X). As mentioned, the mean of the binary dependent variable is 0.16; probit estimation produces similar results.

The vector X of control variables is crucial for it enables us to investigate whether individuals have residual private information conditional on the insurance company’s rating. Fortunately, in the AHEAD data, we observe extremely rich and detailed information on current health and medical history, as well as other demographics. By examining insurance application forms from five leading long-term care insurance companies we determined which of these characteristics of the individual the insurance companies observe. All collect a limited set of demographic information – age, gender, marital status, and age of spouse – as well as similar and extremely detailed information on current health and on health history. We found only one company that asked applicants to report any financial information

(specifically, whether they had less than \$30,000 in financial assets, presumably to screen for likely Medicaid eligibility).

Essentially all of the information collected by the insurance companies is observable in AHEAD. We also know that firms use relatively little information specific to the individual in pricing policies, despite the absence of regulatory restrictions.¹⁶ Policies are not experience rated. Companies offer age-specific prices with only two or three broad rate classifications within each age based on health information (American Council of Life Insurers, 2001, Weiss 2002, Murtaugh et al. 1995). Premiums do not vary by gender.¹⁷

Given the importance of controlling for the individual's risk classification in the analysis, we experiment with four alternative approaches. First, in the most parsimonious specification we do not include any covariates in estimating equation (1) ("no controls" specification). Second, we control for the individual's age by including a separate indicator variable for each age ("age dummies" specification). Both of these approaches underestimate the amount of categorization done by insurance companies.

Our third approach ("all observables" specification) attempts to control for everything insurance companies observe about the individual. This specification includes not only the age dummies, but also all of the demographic information that insurance companies observe, (gender, marital status and age of spouse), and indicator variables for each of the detailed current health and health history characteristics collected by any insurance companies that we observe in the data. To be conservative, we also include indicator variables for the household's income quartile and asset quartile even though companies do not collect this information. This complete set of extremely rich controls is summarized in Table 1.

By including a separate indicator variable for each health characteristic, the "all observables" specification invokes a much more finely defined categorization of risk than insurance companies actually

¹⁶ According to industry actuaries, insurance companies collect more detailed information than they currently use in risk classification in order to build a detailed claims database for future improvements in actuarial modeling.

¹⁷ This unisex pricing practice is somewhat puzzling, since women use substantially more long-term care than men (Society of Actuaries, 1992). Offsetting this however is the fact that, absent private insurance, women are more likely to end up on Medicaid than men and as a result the net benefits from a private insurance policy are substantially lower for women than men (Brown and Finkelstein, 2004b). Consistent with this, we do not find any differences across gender in insurance coverage.

use. We therefore believe that this specification is likely to overestimate the amount of risk classification done by the insurance company.¹⁸ However, if the insurance company's prediction includes substantial interaction effects among the observable controls, we may misestimate the true relationship between care utilization and insurance coverage by only including these observable controls additively.

To address this limitation, our fourth and final specification substitutes for these linear controls with a single summary measure of the insurance companies' prediction of the individual's 5-year nursing home entry probability. We generated these predictions using the same actuarial model that is employed by much of the long-term care insurance industry; this model and its pedigree are described in detail in Robinson (1996), Robinson (2002), and Brown and Finkelstein (2004a).¹⁹ The predictions depend non-parametrically on the individual's age, gender and membership in one of seven different health states (defined by the number of limitations to instrumental activities of daily living (IADLs), the number of limitations to activities of daily livings (ADLs), and the presence or absence of cognitive impairment); all of this information is available in the AHEAD. This measure provides a parsimonious way of controlling for non-linear (and non-parametric) interactions between the observed characteristics of the individual and care utilization.²⁰

The results from estimating equation (1) are shown in Table 2. Several findings emerge. First, using either the continuous or categorical measure of beliefs, columns (1) and (2) indicate that individual beliefs about the likelihood of entering a nursing home are a statistically significant, positive predictor of subsequent nursing home experience. This provides a complement to studies that have found that individuals have some ability to predict their mortality (e.g. Hurd and McGarry, 1995, 2002; Smith et al.

¹⁸ There are, however, a few characteristics that the insurance companies observe that we cannot measure in the AHEAD. Most are rare health conditions – such as double amputation or unoperated aneurysm—but their omission raises the (we think unlikely) possibility that the “all observables” specification underestimates the amount of risk classification done by the insurance company. To compensate for this omission, we experimented with including as controls *all* of the health measures observed in the AHEAD, including those not observed by the insurance company (e.g. self-reported health status, cataract surgery etc.). We did not find any substantive changes in our results.

¹⁹ We use a version of the model that predicts care utilization for typical individuals in the population and makes no adjustment for potential moral hazard effects of the insurance

²⁰ As an alternative way of dealing with non-linearities in the relationship between observable characteristics and long-term care utilization, we also estimated equation (1) on increasingly homogenous sub-samples of individuals from the perspective of the insurance company (e.g. by age and health conditions). The results were not affected.

2001). The results in column (1) indicate that a 10 percentage point increase in self-reported probability is associated with a 1 percentage point increase in the probability of going into a nursing home. The results in column (2) indicate that individuals who report a subjective probability of five-year nursing home entry of 50 or higher are about 7 percentage points (about 40 percent) more likely to go into a nursing home over the next five years than individuals who report a prediction of 0; individuals who report a prediction of 1 to 49 are no more likely to go into a nursing home than individuals who report 0, but are significantly less likely to go into a nursing home than individuals who report a prediction of 50 or higher.

Perhaps not surprisingly, we also find that the insurance companies' actuarial model generates a prediction that is more highly correlated with subsequent nursing home use than the individual's reported self-assessment. The estimate in column (3) indicates that a 10 percentage point increase in the actuary's prediction of nursing home entry probability is associated with a 5 percentage point increase in nursing home entry probability; by contrast, the estimate in column (1) indicates that the same increase in the individual's prediction is associated with only a 1 percentage point increase in nursing home entry probability. The finding that the actuaries are more accurate, however, is not relevant for the issue of asymmetric information; as long as the individual has residual private information – conditional on the menu of choices offered by the insurance company – asymmetric information can operate as in the theoretical models.

Indeed, most importantly, the results in the remaining columns of Table 2 indicate that the individual has residual private information about his risk type. Regardless of what set of controls for insurance company risk classification or measure of the individual's beliefs is used, these beliefs are positive and statistically significant predictors of subsequent nursing home utilization. These results represent a key finding of the paper: conditional on the insurance company's risk classification, individuals have private information that predicts their subsequent nursing home use.

How large is this private information? Columns 7 and 9 indicate that individuals who report a subjective probability of 50 or higher are 3 percentage points (about 20 percent) more likely to go into a nursing home than individuals who report a lower prediction. Another way to quantify the amount of

private information is to compare the additional predictive power of individual beliefs with that of other determinants of nursing home use. For example, consider the gain in explanatory power from using single-years of age rather than five-year age categories in the vector of explanatory variables in equation (1). In our most detailed specification (“all observables”) the increase in explanatory power from the use of the more finely defined age categories is nearly 12 times larger than the increase in explanatory power associated with adding the individual’s self assessment. Thus, although statistically significant, the predictive value of individual information is substantially less than that of much of the information insurance companies already have.

As noted above, we may be underestimating the amount of individuals’ private information due to measurement error in the reported values. Consistent with this, we find (in results not shown) that if we add to the regression the individual’s response to the same question about beliefs asked in the previous (1993) interview, both the 1993 and 1995 measures of beliefs are significant predictors of nursing home use.

Moreover, it is important to note that even a small amount of private information can have a large effect on the market equilibrium. Suppose, for example, that only the 20 percent of individuals who predict their nursing home risk is 50 to 100 percent buy insurance. The results in Table 2 indicate that such individuals are 3 percentage points more likely to enter a nursing home in the subsequent five years than individuals who gave lower probabilities. Moreover, we also estimate that conditional on going into a nursing home, these high probability individuals are likely to spend 60 more days in the home than other individuals.²¹ If these high probability individuals were the only ones to purchase insurance, the expected number of days in a nursing home (and hence expected premiums in a competitive market) would increase by 150 percent.

This raises an important question: If individuals have a meaningful amount of residual private information about their chances of using a nursing home, why don’t insurance companies attempt to learn

²¹ This is based on Tobit estimates of the relationship between number of nursing home nights over a 5 year period and private information about nursing home risk, controlling for the insurance company’s risk classification. Results (not shown) are statistically significant at the 1 percent level.

more about the individual? An examination of insurance applications indicates that there is clearly some information about the individual that the insurance company could observe but that in practice it does not. This includes, for example, the number, sex, and proximity of the individual's children, the individual's race, religion and education, information on a spouse's health, and whether the individual engaged in each of a variety of potential preventive health measures (described in more detail in Section 3.4). When we add these variables to the "all observables" specification in Table 2, they are jointly significant, but their addition does not affect the magnitude or statistical significance of the coefficient on the individual's prediction. This suggests that feasible collection of additional information about the individual would not correct the problem of asymmetric information vis a vis the consumer's information set, but it would give the company an advantage over competitors that do not collect the information.²²

3.2 Private information and insurance coverage

Given that individuals appear to have private information, we ask next whether this information is correlated with the purchase of long-term care insurance. To do so, we estimate the equation:

$$(2) \text{ LTCINS} = X\delta_1 + \delta_2 B + \mu$$

LTCINS is a binary measure of whether the individual has long-term care insurance in 1995; Appendix A provides more detail on its exact definition. Once again, B and X are, respectively, the subjective probability of entering a nursing home and controls for the risk classification done by the insurance company as measured in the 1995 data.

On average, 11 percent of our sample of non-proxy respondents has long-term care insurance.

Although the AHEAD data do not contain information on the source of the insurance, about 80 percent of private long term care insurance is provided by the individual (non-group) market (HIAA 2000b).

²² We presume that insurance companies do not collect this information because the costs of doing so are high relative to the benefits. In addition, the use of behavioral information in pricing insurance contracts (such as decisions regarding preventive health care investment) could alter the behavioral choices of potential applicants and thus reduce the informative content of these characteristics. Finally, although there are no current restrictions, insurer's may hesitate to price based on demographic characteristics such as race and sex for fear of regulation or court challenges.

Nationwide, the average age of purchase in this market is 67 (HIAA, 2000a).

The results from estimating equation (2) by OLS are presented in Table 3. Across all specifications, individuals who believe that they are higher risk are more likely to have insurance. Results from a probit specification (not shown) are similar.

A substantial fraction of the seemingly uninsured may in fact rely on the public insurance provided by Medicaid, which pays for 35% of all long-term care expenditures (Congressional Budget Office, 2004). To address this issue, we repeat the above analysis, restricting the sample to individuals in the top quartile of the household income or wealth distribution in 1995. Because of the means-testing required for Medicaid eligibility, these high-wealth individuals are less likely to find Medicaid an attractive substitute for private insurance. The results are shown in Table 4. The positive relationship between individuals' beliefs about their nursing home entry risk and their insurance is even stronger among this subset of individuals for whom Medicaid is a poor substitute for private insurance.²³

An interesting question concerns the *timing* of this private information relative to the insurance purchase. In the case of adverse selection, the insured is assumed to have ex-ante superior information to the insurance company about his risk type that influences his insurance purchase decision. In the case of moral hazard, the causality is reversed. Ex-post, insurance coverage lowers the cost of an adverse outcome and thus increases the probability or magnitude of the risk occurrence. The classic explanation is that insurance reduces the individual's incentive to invest in (costly) risk-reducing effort (see e.g. Arnott and Stiglitz, 1988). In the health insurance context, another form of moral hazard may be quantitatively more important: insurance lowers the marginal cost of consuming the insured good (medical care), and may therefore induce additional consumption.

Distinguishing empirically between adverse selection and moral hazard is notoriously difficult and most empirical examinations of asymmetric information do not attempt to do so (Chiappori 2000).

²³ Because insurance companies tend to deny coverage to some unhealthy applicants, we re-estimate equation (2) restricting the sample to individuals whose current health conditions reveal nothing that would be expected to make them ineligible for insurance. The positive correlation between beliefs and insurance coverage was even larger (and still statistically significant) with this healthier subsample.

However, we present two pieces of suggestive evidence that at least some of the private information about risk type that we have detected exists prior to the purchase of insurance. We find that the results of Table 2 that individuals have private information that predicts their nursing home use are unchanged when we restrict the sample to the subset of individuals who do not have private insurance (90% of the total sample) and who would thus not be subject to moral hazard. In addition, analysis of younger cohorts in the HRS (results not reported) shows that beliefs about subsequent nursing home use in 1996 among uninsured individuals aged 60 to 69 were positively and statistically significantly associated with the later acquisition of insurance (which 8 percent of the sample did).²⁴

3.3 Long Term Care Insurance and Long Term Care Use

The results presented thus far provide direct evidence of the existence of private information in the long-term care insurance market. Given the existence of private information about risk type, most standard (uni-dimensional) models of asymmetric information predict a positive correlation between insurance coverage and the use of care, conditional on the insurer's risk classification (Chiappori and Salanie, 2000, Chiappori et al. forthcoming). If this positive correlation prediction is a valid test of asymmetric information, we should see such a correlation in our data. Here, however, we show that despite the direct evidence of asymmetric information in the preceding analysis, we fail to find a positive correlation between insurance coverage and care utilization in the AHEAD data. This ostensible puzzle – that despite the fact that individuals have private information about risk type that is positively correlated with insurance coverage we find no evidence in the aggregate of a positive correlation between insurance coverage and care use – can be explained by the existence of other unobserved preference related characteristics that are positively correlated with insurance demand and negatively correlated with care utilization; Section 3.4 provides direct evidence of these other unobserved factors.

The test of a positive correlation has been broadly applied as a test of asymmetric information across

²⁴ The AHEAD data do not permit us to look at whether private information predicts subsequent insurance purchases because the average initial age in our sample is 79, and hence subsequent insurance purchases are rare.

a wide variety of insurance markets. The results have provided fodder for interesting discussions about the functioning of insurance markets, suggesting the existence of asymmetric information in some markets but not in others. In the case of health insurance markets, there is extensive evidence of a positive correlation between insurance coverage and risk occurrence (see Cutler and Zeckhauser 2000 for a review of the literature) although exceptions do exist (e.g. Cardon and Hendel, 2001.) The evidence from annuity markets also indicates that the insured are at higher risk than the uninsured (Finkelstein and Poterba, 2002, 2004, McCarthy and Mitchell, 2003), but no such correlation is found in life insurance markets (Cawley and Philipson, 1999, McCarthy and Mitchell, 2003). Evidence in the automobile insurance market fails to support a role for asymmetric information (Chiappori and Salanie 2000, Chiappori et al., forthcoming, and Dionne et al., 2001), but again, exceptions do exist (Pueltz and Snow 1994, Cohen, 2001). At the end of the paper, we discuss how the existence of unobserved preference heterogeneity may provide a unifying explanation for these different findings across insurance markets.

Our basic estimating equation directly follows past work (e.g. Finkelstein and Poterba 2004) and is simply a regression of nursing home use on insurance coverage, controlling for risk classification:

$$(3) \quad \text{CARE} = X\beta_1 + \beta_2 \text{LTCINS} + \varepsilon$$

where X again represents a series of covariates designed to control for any risk-categorization of the individual done by the insurance company.²⁵ Note that our coefficient of interest, β_2 does not have a causal interpretation. In a pure moral hazard model with no other unobserved heterogeneity, the coefficient would represent the causal effect of insurance coverage on care utilization. However, in an adverse selection model, the causality is reversed and private information about expected care utilization affects insurance demand.

Our estimates of β_2 for the above equation are reported in Table 5 using linear probability models;

²⁵ Recall that when examining the role of subjective beliefs it was necessary to exclude the 13 percent of the sample for whom interviews were given by proxies. In this section, such an exclusion is not necessary and we therefore use the full sample. When the proxy respondents are included in the sample, the proportion that has any nursing home use between 1995 to 2000 increases from 16 percent to 19 percent and the percent with long-term care insurance falls from 11 percent to 10 percent. However, the regression results are essentially identical if we continue to use the smaller sample instead.

probit results (not reported) are indistinguishable. The dependent variable is a binary measure of whether the individual spent any time in a nursing home between 1995 and 2000; 19 percent of the sample did. The four columns show the results for the four alternative definitions of the control variables (X). The results for the full sample are shown in the top row and indicate that long-term care insurance coverage is negatively associated with utilization. The bottom row shows the results when the sample is restricted to the top quartile of the income or wealth distribution to avoid the potential problem of Medicaid serving as an attractive substitute for those with limited resources; the relationship between insurance coverage and care becomes even more negative.²⁶ To verify that the lack of a positive correlation between insurance coverage and nursing home use is not due to measurement error in insurance coverage, we re-estimated equation (3) using long-term care insurance coverage in 1998 as an instrumental variable for coverage in 1995; the IV coefficient on LTCINS was quite similar to the OLS results and indeed, about half the time, was more negative than the OLS estimates.

To check the robustness of our results we conducted a number of further analyses of the relationship between insurance coverage and care utilization, three of which we note briefly here. First, because we only observe care utilization for a five-year period, and not over the lifetime of the policy, it is possible that the positive correlation property would appear if the data were analyzed over a longer time horizon. We tried several alternative approaches to addressing this issue: we used information on how long the individual has had his policy to restrict the insured individuals in the sample to the two-thirds who had had their policy since at least 1992 (the earliest year for which nursing home utilization data are available) and thus observed 8 years of care utilization data rather than only 5; we limited the sample to the one-third of individuals who died between 1995 and 2000, for whom utilization subsequent to 2000 is not

²⁶ In addition to the means-tested Medicaid program, Medicare – the public health insurance program for the elderly – also reimburses some nursing home expenditures. However, Medicare only pays for skilled (rather than custodial) nursing home stays that directly follow a hospitalization, and will not pay for stays of more than 100 days. The negative relationship between nursing home utilization and long-term care insurance persists if we define nursing home use as a stay beyond 100 days (when Medicare coverage ceases). It also persists if we limit the sample to the 55 percent of individuals who do not go into a hospital between 1995 and 2000 – and whose nursing home stays therefore cannot be covered by Medicare. We also verified that these findings are robust to restricting the sample to individuals who have none of the health conditions that might make them ineligible for private insurance.

possible; and we limited the sample to the top quartile of the age distribution (age 82+), for whom we assume a smaller fraction of the policies' lifetime remains. None of these changes affected the qualitative nature of the results.

Second, it is possible that the positive correlation would manifest itself in other measures of care utilization such as the intensity of care use (i.e. number of days in a nursing home), or home health care use (particularly if long-term care insurance helps the individual avoid institutionalization by substituting for home health care).²⁷ Table 6 therefore reports the results of re-estimating equation (3) using these two other measures of care utilization as the dependent variable. The top row shows the results of a Tobit specification when the dependent variable is the number of nights spent in a nursing home between 1995 and 2000. The conditional (unconditional) mean number of nights is 187 (33). The bottom row shows the results from a linear probability model when the dependent variable (“any care utilization”) measures whether the individual consumed any nursing home *or* and home health care between 1995 and 2000 (mean is 40 percent).²⁸ The results continue to indicate a negative relationship between insurance coverage and care utilization.

Finally, we note that an important limitation to the AHEAD data is that we only know whether the individual has insurance; we do not know details of the insurance policy which affect the quantity of insurance coverage, such as the amount of any deductibles or lifetime limits on coverage. To examine the correlation between the amount of insurance coverage and long-term care utilization, we turn to a proprietary database from a large, private long-term care insurance company. This data set contains information on long-term care utilization for individuals and the exact quantity of insurance they purchased. Furthermore, it allows us to observe directly – and hence stratify on – the risk classification assigned to the individual by the insurance company. Even with this sample we find no substantive or

²⁷ Although over three-quarters of long-term care expenditures are for nursing homes (US Congress, 2000) long-term care insurance policies are increasingly likely to cover home health care expenses (Society of Actuaries 2002, HIAA 2000a).

²⁸ We do look at “any home health care use” directly because home health care utilization is not asked if the individual is residing in a nursing home at the interview date and home health care use is therefore measured with error.

statistical evidence that the amount of insurance is positively related to nursing home use. These data and the results are described in detail in a prior version of this paper (Finkelstein and McGarry, 2003).

3.4 Private Information About Preferences for Insurance

We have found that individuals have private information about their risk type and that this private information is positively correlated with insurance coverage. However, we also found that the insured are no more likely to enter a nursing home than the uninsured. Together, these results imply, mechanically, that there must exist other (unobserved) characteristics of the individual that are positively correlated with insurance coverage but negatively related to use and that thus offset the positive effect of asymmetric information about risk.²⁹

Econometrically, imagine we can measure two aspects of the individual that are unobserved by the insurance company, and related (with noise) to his long-term care utilization. The first of these we have termed beliefs (B) and the second we term preferences for insurance (P). We assume that the insurance company observes neither of these measures. We have already seen that B is positively related to insurance coverage and to care utilization and are looking for a P that has the opposite correlation with insurance coverage and care utilization. Given such a variable P we can estimate the equations:

$$(4) \quad \text{CARE} = Xb_1 + b_3B + b_4P + \varepsilon$$

$$(5) \quad \text{LTCINS} = Xd_1 + d_3B + d_4P + \eta$$

and test the signs of b_4 and d_4 .

It is difficult, if not impossible to measure all of the components of P. By definition, they must be unobserved by the insurance company; many of them are therefore likely to be unobserved by the

²⁹ To verify that absent these unobserved, offsetting preference characteristics, the relationship between insurance coverage and nursing home utilization would be positive and statistically significant, we attempt to single out that portion of the long-term care insurance purchase due solely to private information about risk. To do so we use an instrumental variables procedure to estimate equation (3) with individual beliefs (B) as the instrument for insurance coverage. In this IV specification, the relationship between insurance coverage and nursing home utilization is consistently positive and statistically significant; these results are presented and discussed in more detail in Finkelstein and McGarry (2003).

econometrician as well. We focus primarily on an aspect of individual behavior that has attracted considerable theoretical attention: unobserved heterogeneity in risk aversion. de Meza and Webb (2001) and Jullien et al. (2002) propose that the more risk averse (or more “cautious”) may not only place a higher value on insurance, but may invest more in risk-reducing efforts and thus become endogenously lower risk than less risk averse individuals.

We further exploit the richness of the AHEAD data to construct an innovative measure of the individual’s investment in risk-reducing behavior. We observe, in 1995, whether the individual undertook various gender-appropriate potential preventive health care measures over the last two years. These are: whether the individual had a flu shot, had a blood test for cholesterol, checked her breasts for lumps monthly, had a mammogram or breast x-ray, had a Pap smear, and had a prostate screen.³⁰ The insurance company applications we reviewed did not solicit any of this information.³¹

There is substantial variation in the fraction of gender-appropriate potential preventive activity actually undertaken: the median individual undertakes two-thirds of these activities, but 7% report doing nothing and 30% report engaging in all relevant preventive behaviors. This measure does not appear to primarily reflect whether the individual has seen a doctor over the last two years (over 90% of our elderly sample has) or the type of insurance the individual has (everyone in our sample is covered by Medicare, which covers the expenses from these preventive health measures at least to some extent). Given the extremely rich set of covariates measuring health included in the “all observables” specification (see Table 1), we do not believe these preventive health measures are simply proxies for the individual’s underlying health.

Table 7 reports the results of estimating equations (4) and (5). To conserve space, we report only the results using the continuous measure of individual’s beliefs; the effects are similar if categorical measures

³⁰ If an individual’s risk aversion differs in different aspects of his life this measure is likely particularly well-suited to the issue at hand in that it is likely to be a good proxy for the individual’s risk aversion with respect to health concerns.

³¹ One reason insurance companies may not collect this information is that – if they used it in pricing – it could influence individuals’ choice of preventive behavior and thus distort its usefulness as a proxy for the individuals’ overall level of preventive (or cautious) health behavior.

of beliefs are used instead. The results from the first column within each panel indicate that individuals who undertake a greater fraction of potential preventive health activity (i.e. more cautious individuals) are less likely to enter a nursing home.³² The estimates in the second column within each panel indicate that those who undertake more preventive health activity are also more likely to own insurance. Thus, this variable satisfies the necessary criteria to provide an offset to selection based on private information about risk type.

We use our estimate of equation (5) to decompose insurance coverage (LTCINS) into the component predicted by the preventive health activity ($\hat{\delta}_4 P$, which we denote as LTCINS_HAT_{PREVENT}), the component predicted by individuals' private information about risk type ($\hat{\delta}_3 B$, which we denote as LTCINS_HAT_{RISKTYPE}) and the unexplained component, combined in the residual ($\hat{\eta}$, which we denote by LTCINS_RESID).³³ We then analyze the relationship between long-term care insurance and use of long-term care as in equation (3) but with the LTCINS variable replaced by these three components. Equation (3) is thus rewritten as:

$$(6) \text{CARE} = \beta_1(\text{LTCINS_HAT}_{\text{RISKTYPE}}) + \beta_2(\text{LTCINS_HAT}_{\text{PREVENT}}) + \beta_3(\text{LTCINS_RESID}) + X\beta_4 + \varepsilon$$

Panel A of Table 8 (labeled “Preventative Health Activity”) shows the results of this estimation. As we expected, the effect of LTCINS_HAT_{PREVENT} -- i.e. the component of long-term care insurance predicted by preventive activity – on use is consistently negative, while that of LTCINS_HAT_{RISK} is consistently positive.

Taken as a whole, our empirical findings suggest that private information about risk type and risk preferences produces a market equilibrium in which risk type and insurance coverage are not correlated in

³² The fact that our measure of preventive activity is a statistically significant predictor of nursing home utilization even though we condition on individuals' beliefs about this same nursing home utilization indicates that individuals do not efficiently incorporate all relevant, available information about their characteristics in forming their beliefs about their risk type. This result is consistent with Smith et al.'s (2001) finding that individuals do not efficiently form beliefs about their mortality prospects.

³³ For this specification, we include separate indicator variables for each preventive health activity (and gender) in order to estimate more flexibly the relationship between preventive health activity and insurance coverage. The results are similar if we instead use the fraction of gender-appropriate preventive health activity undertaken.

the aggregate (Table 5), but where higher risk individuals and higher risk averse individuals (who happen to be of lower risk) both purchase more insurance (Table 7).³⁴ Returning to the theoretical framework in Section one, we illustrate such an outcome in Figure 5. Assume there are equal numbers of individuals of two risk types (high (H) and low (L)) and that there are three possible levels of risk aversion (1, 2, 3 in increasing order). High risk individuals are evenly split between risk aversion level 1 and 2 (i.e. denoted H1 and H2), while low risk individuals are evenly split between risk aversion levels 2 and 3 (denoted L2 and L3). In equilibrium, H1 and L2 pool on a less comprehensive policy while H2 and L3 pool on a more comprehensive policy. Because each policy has the same number of high and low risk individuals, there is no correlation between risk type and risk occurrence. However, conditional on a given level of risk aversion, higher risk individuals purchase more insurance (compare L2 to H2); and conditional on a given risk type, more risk averse individuals purchase more insurance (compare L2 to L3 or H1 and H2).

Our findings are consistent with the theoretical models of de Meza and Webb (2001) and Jullien et al. (2002) in which more risk averse individuals are more likely to own insurance and exert more preventive effort, thus becoming endogenously lower risk, but we caution that causality cannot be inferred from our empirical results. While some of the preventative behaviors may themselves forestall or prevent nursing home admissions (e.g. flu shots reduce the risk of pneumonia which is a nontrivial contributor to the need for a nursing home for the elderly) engaging in preventative health activities may simply be correlated with other health investments that themselves cause lower rates of institutionalization. For example, we find that individuals who invest more in our measured preventative health activities are substantially less likely to have a hip fracture, an important contributor to nursing home residence, yet none of the measured activities themselves would be expected to affect bone density or agility. Rather, it is likely that other behaviors (such as a history of exercise and calcium consumption) are positively related to our preventative measures and negatively related to fractures.

Another plausible interpretation is that our measure of preventative health activity is proxying for other

³⁴ We use the term risk aversion here as short hand for preventative activity. Below, we describe alternative interpretations of the preventative activity measure.

unmeasured preference-related characteristics that themselves have a causal effect on nursing home utilization. We find, for example, that individuals who engage in more preventive health activity tend to systematically overestimate their risk probability relative to actual experience. As a result, although beliefs about nursing home risk are positively correlated with nursing home use and preventive activity is negatively correlated with nursing home use, beliefs and preventive activity are themselves positively correlated. This suggests the existence of an unobserved “pessimism” factor that causes individuals to shade upward their beliefs about nursing home risk. It is possible – as conjectured in the theoretical model of Koufopoulos (2003) – that more pessimistic individuals are both more likely to buy insurance and also more likely to invest in risk-reducing efforts, thus endogenously lowering their risk.

Our preventive health measure may also proxy for unmeasured aspects of socio-economic status that affect insurance demand. We noted earlier that individuals with few financial resources are likely to view Medicaid as a good substitute to private insurance coverage. This fact is likely to introduce preference heterogeneity in insurance purchase based on wealth. Insurance companies do not currently collect or use information on wealth in pricing long-term care insurance policies. In an analysis similar to that above for preventive activity, we find that (conditional on the classification done by the insurance company) higher asset individuals are both substantially more likely to have long-term care insurance and substantially less likely to use nursing homes.

These wealthier individuals also appear to be more cautious individuals; our measure of preventive health and individuals’ wealth are positively correlated in the data. Disentangling the relative contribution of two highly correlated variables is, by design, fraught with difficulty, but we nonetheless investigate the correlations. Panel B of Table 8 reports the results from estimating an expanded version of equation (6) wherein long-term care insurance is decomposed into a component predicted by the individual’s wealth ($LTCINS_HAT_{WEALTH}$) as well as $LTCINS_HAT_{PREVENT}$, $LTCINS_HAT_{RISKTYPE}$ and $LTCINS_RESID$. The results indicate that the components of long-term care insurance predicted by preventive health activity (our proxy for risk aversion) and by wealth both remain negatively correlated with subsequent nursing home entry probability. However the coefficient on $LTCINS_HAT_{PREVENT}$ attenuates sharply

relative to the first panel and is less likely to be statistically significant. In the end, we cannot rule out the possibility that our preventive health measure is proxying for unmeasured aspects of socio-economic status which themselves have a causal effect on long-term care utilization. Conversely, wealth itself may proxy risk aversion in that more risk averse individuals save more in order to reduce the possibility of exhausting their resources should they live longer than expected.

Interestingly, other than preventive health activity and wealth, other characteristics of the individual that we can measure and that insurance companies do not observe appear to have the *same* correlation with insurance coverage and risk occurrence. For example, individuals with more children are both less likely to have insurance and less likely to use nursing home care. The same is true for non-whites relative to whites, Hispanics relative to non-Hispanics, and less educated individuals relative to more educated individuals. Thus, while these factors may well be an important component of asymmetric information and knowledge of their values may assist the insurance company in more accurately assessing risk, they do not provide an explanation for the lack of a positive correlation between coverage and care.

Section 4: Implications for testing for asymmetric information

We have found that individuals have private information about their risk type as well as their risk preferences with respect to the long-term care and long-term care insurance, yet the market equilibrium displays no correlation between insurance coverage and risk occurrence. These findings highlight an important shortcoming in the growing body of work challenging the empirical relevance of asymmetric information to insurance markets. This literature has relied on a positive correlation between the amount of insurance and the occurrence of the risk to indicate the existence of asymmetric information (e.g. Chiappori and Salanie 2000, Chiappori et al., forthcoming, Cardon and Hendel 2001, Dionne et al. 2001, Finkelstein and Poterba 2004).³⁵ However, we have demonstrated empirically that asymmetric information can exist even in cases in which the insured are not of above-average risk.

³⁵ An important exception is Cawley and Philipson (1999) who also look directly at the structure of information in the life insurance market and conclude that individuals do not have private information about their mortality risk.

Moreover, our findings suggest an alternative, general approach to testing for asymmetric information in insurance markets which does not require the rich data on subjective beliefs available in the HRS. Rather, we note that, conditional on the information set used by the insurance company, the existence of *any* individual characteristic that is observed by the econometrician, but not by the insurer and that is correlated with both insurance coverage and risk occurrence indicates the presence of asymmetric information. This result is true regardless of the sign of the correlation.

As Chiappori et al. (forthcoming) explain quite nicely, asymmetric information is “relevant” in the sense of affecting equilibrium efficiency if there is a link between the private information of one party to the transaction and the payoff to the other party. In such a case, the consumer’s private information affects the eventual costs of the seller and thus the efficiency of the equilibrium transactions.³⁶ Seen in this light, if individuals have private information about preferences that are in fact correlated with risk and with insurance coverage (such as the risk aversion parameter highlighted here), then there is “relevant” private information, even if the individual is not consciously making the insurance decision in consideration of his beliefs about his risk type.

To see this more formally, assume individual i chooses his insurance purchase (I) to maximize his expected utility as given by the expression $EU = \pi_i u_i(y_i + (1 - m_i)I - L) + (1 - \pi_i)u_i(y_i - m_i I)$ where π_i is the individual’s probability of an accident, u_i is the individual’s sub-utility function (which may vary across individuals due to differences in risk aversion), y_i is the individual’s income, and m_i is the insurance premium (which may also vary across individuals based on individual characteristics that are observed by the insurance company). I is the insurance pay-out in event of accident and L is the financial loss associated with the accident. Maximization of this function yields the familiar first order condition

³⁶ By contrast, private information about individual’s preference for car color in a competitive car market is not relevant to the efficiency of the market equilibrium as these preferences do not affect the seller’s payoff.

$$\frac{u'_i(y_i + (1 - m_i)I - L)}{u'_i(y_i - m_i I)} = \frac{m_i(1 - \pi_i)}{(1 - m_i)\pi_i} \quad (7)$$

Equation (7) indicates that the insurance demand (I) is a function of the price (m_i) the individual faces as well as his risk type (π_i), and preferences (as reflected in u_i and y_i). Any information that the insurer uses about the individual that is correlated with his risk type should be incorporated into the price of the insurance (m_i). If there is “relevant” private information, this means that the individual has information (either in the form of direct private information about his risk type or his preferences) that is correlated with his insurance demand and with his risk type. It does not matter from the perspective of the impact on market equilibrium and efficiency whether the individual acts consciously on information about his accident probability in choosing insurance, or whether preferences for insurance based on such factors as wealth or caution simply happen to be correlated with this accident probability.

One must reject the null hypothesis of symmetric information if there exists some characteristic of the individual that is correlated with both insurance coverage (I) and ex-post risk occurrence (our best measure of risk type π_i) that is unknown (or unused) by the insurer, after conditioning on the information that is used by the insurance company in setting prices. Of course, the test is only one-sided: failure to find such a characteristic might simply reflect a lack of sufficiently rich data, rather than the absence of asymmetric information.

Implementation of the test requires only that the econometrician observe the characteristics of the individual that insurance companies use to assign risk classification as well as some additional information about the individual that the insurance company does not use in pricing but that is correlated with insurance purchase and risk outcomes. These requirements are often satisfied in practice. For example, the government often places legal restrictions on the insurer’s ability to price based on certain characteristics – for example, restrictions on using gender in pricing auto-insurance exist in many states – which may nonetheless be observed in insurance company data or in general surveys. In addition, information that is costly to verify may not be collected by insurance companies but might be collected by

a general survey, where the individual does not have the same incentives to distort his answer. For example, although annuity companies do not collect information on individuals' wealth (despite the known correlation between socio-economic status and mortality), such information is available in many public use surveys. These types of disparities between the information collected and used by the insurance company and that available to the econometrician suggests that this test may find widespread applicability. Thus, while the availability of subjective assessments of individual risk presents an extraordinary opportunity to researchers, tests of asymmetric information need not be limited to studies of such data.

Section 5: Conclusion

In this paper we document the existence of multiple forms of private information in an insurance market and examine the impact of these informational asymmetries on the equilibrium in that market. Our empirical work focuses on the private long-term care insurance market but the ideas we develop have broad applicability.

We begin by using innovative data on self-assessments of nursing home risk, along with elaborate controls for the risk classification used by the insurance company, to demonstrate that private information about risk type does exist and is positively correlated with insurance coverage. We then show that despite this private information, the standard test of asymmetric information fails. We reconcile these findings by presenting evidence of another type of private information: private information about risk tolerance.

Based on these two forms of heterogeneity we identify two types of individuals who purchase long-term care insurance. One group consists of individuals who have private information that they are higher risk than the insurance company would predict and who, ex-post, do in fact turn out to be higher risk. The other group consists of individuals who are risk averse (or have a high preference for insurance). Ex-post, this group is of lower risk than the insurance company would predict based on the factors it observes.

This preference-based selection offsets the selection of high risk persons into the long-term care market,

and as a result, we find no evidence in the market equilibrium of a positive correlation between individuals' insurance coverage and the occurrence of the insured risk.

Our findings highlight the potential pitfalls of attempting to infer the presence or absence of asymmetric information simply from the reduced form correlation between insurance coverage and risk occurrence. When individuals have private information about risk preferences as well as risk type, an inefficient asymmetric information equilibrium in an insurance market can look very different from the standard theoretical case with heterogeneity in risk type alone.

As always, the existence of asymmetric information per se is not sufficient to draw specific policy recommendations (see e.g. Crocker and Snow 1985). Relatedly, another unanswered question and an important avenue for further work is the importance of asymmetric information in contributing to the extremely limited size of the private long-term care insurance market.

For insurance markets more generally, our evidence of offsetting, preference-based selection in the long-term care insurance market suggests a potential unifying explanation for the apparent differences across insurance markets in whether the insured are above-average in their risk type. For example, in several different countries, evidence of a positive correlation between insurance coverage and risk occurrence exists for annuity markets but not for life insurance (Cawley and Philipson 1999, McCarthy and Mitchell 2003, Finkelstein and Poterba 2002, 2004). These markets are fundamentally related in that annuities insure against living "too long" while life insurance insures the opposite risk of dying "too soon". A possible explanation for the apparent selection differences between these markets is that preference-based selection operates in the opposite direction in these two markets. Characteristics of the individual that the insurance company does not observe, such as their level of caution or their wealth, may be positively correlated with demand for both annuities and life insurance, but *negatively* correlated with the life insurance risk of dying and *positively* correlated with the annuity risk of living. Such an explanation would be consistent with either symmetric or asymmetric information about risk type. We pose this potential explanation as an interesting one to pursue in future work.

REFERENCES

- American Council of Life Insurers. 2001. Life Insurers Fact Book, 2001. Washington, DC
- Arnott, R and Joseph Stiglitz. 1988. "The basic analytics of moral hazard." *Scandinavian Journal of Economics*, 90: 383-413.
- Bank of Sweden, 2001. "Press Release: The 2001 Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel, Advanced Information." Available at <http://www.nobel.se/economics/laureates/2001/eoadv.pdf>
- Brown, Jeffrey and Amy Finkelstein. 2004a. "High prices, limited benefits: new evidence from long-term care insurance." Unpublished mimeo.
- Brown, Jeffrey and Amy Finkelstein. 2004b. "The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market." Unpublished mimeo.
- Cawley, John and Tomas Philipson. 1999. "An empirical examination of information barriers to trade in insurance." *American Economic Review* Vol 89, no 4 pp.827-846.
- Cardon, James and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics* 32, 408-427.
- Chiappori, Pierre-Andre. 2000. Econometric Models of Insurance under Asymmetric Information. In Georges Dionne, ed. Handbook of Insurance Economics. London: Kluwer.
- Chiappori, Pierre-Andre and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108, 56-78.
- Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie. forthcoming. *Rand Journal of Economics*. "Asymmetric Information in Insurance: General Testable Implications."
- Cohen, Alma. 2001. "Asymmetric Information and Learning in the Automobile Insurance Market." Mimeo, Harvard University, Department of Economics.
- Cohen, Marc. Forthcoming. "Private Long-Term Care Insurance." *Journal of Aging and Health*.
- Congressional Budget Office. 2004. "Financing Long-Term Care for the Elderly." April.
- Crocker, Keith and Arthur Snow. 1985. "The efficiency of competitive equilibria in insurance markets with asymmetric information." *Journal of Public Economics* Volume 26 (2): 207 – 219.
- Cutler, David and Richard Zeckhauser. 2000. The Anatomy of Health Insurance, in The Handbook of Health Economics, Culyer AJ and Newhouse JP (eds)
- de Meza, David and Webb, David. 2001. "Advantageous selection in insurance markets." *Rand Journal of Economics*. Summer.
- Dionne, Georges, Christian Gourieroux and Charles Vanasse. 2001. "Testing for evidence of adverse selection in the automobile insurance market: a comment." *Journal of Political Economy* 109(2): 444-

451.

Finkelstein, Amy and Kathleen McGarry. 2003. "Private information and its effect on market equilibrium: new evidence from long-term care insurance." NBER Working Paper 9957.

Finkelstein, Amy and James Poterba. 2002. "Selection Effects in the Market for Individual Annuities: New Evidence from the United Kingdom." *Economic Journal* 112, 28-50.

Finkelstein, Amy and James Poterba. 2004. "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market." *Journal of Political Economy*.

Gan, Li, Michael Hurd and Daniel McFadden. 2003. "Individual subjective survival curves." NBER Working Paper 9480.

Hamermesh, Daniel. 1985. "Expectations, Life Expectancy, and Economic Behavior." *Quarterly Journal of Economics* Vol 100, Issue 2, pp.389-408.

Hurd, Michael D. and Kathleen McGarry, 1995. "Evaluations of the Subjective Probability of Survival in the Health and Retirement Survey," *Journal of Human Resources*, 30 : s268-s292.

Hurd, Michael D. and Kathleen McGarry, 2002. "The Predictive Validity of Subjective Probabilities of Survival" *Economic Journal*, 112 (482) : 966-985.

Health Insurance Association of America (HIAA). 2000a. "Who buys LTC insurance in 2000?" Washington DC.

Health Insurance Association of American (HIAA). 2000b. "LTC insurance in 1997 – 1998." Washington DC.

Jullien, Bruno, Bernard Salanie, and Francois Salanie. 2002. "Screening risk-averse agents under moral hazard." Unpublished mimeo.

Koufopoulos, Kostas. 2003. "Asymmetric Information, Heterogeneity in Risk Perceptions, and Insurance: An Explanation to a Puzzle." Unpublished mimeo.

Lewis, Stephanie, John Wilkin and Mark Merlis. "Regulation of Private Long-term Care Insurance: Implementation Experience and Key Issues." Prepared for The Kaiser Family Foundation, March 2003. Available at <http://www.kff.org/content/2003/6073/>.

McCarthy, David and Olivia Mitchell. 2003. "International Adverse Selection in Life Insurance and Annuities." NBER Working Paper 9975.

Mehta, Kala M., Kristine Yaffe, Kenneth Langa, Laura Sands, Mary Whooley, Kenneth Covinsky (2002). "Additive Effects of Cognitive Function and Depressive Symptoms on Mortality in Older Community Living Adults." mimeo, UCSF.

Mellor, Jennifer. 2001. "Long-term care and nursing home coverage: are adult children substitutes for insurance policies?" *Journal of Health Economics* 20 pp.527-547.

Murtaugh, Christopher, Peter Kemper and Brenda Spillman. 1995. Risky Business: Long-term care insurance underwriting. *Inquiry* 32:371-284.

- Murtaugh, Christopher, Peter Kemper, Brenda Spillman, and Barbara Carlso. 1997. "Amount, distribution, and timing of lifetime nursing home use." *Medical Care* (35): 3 pp. 204-218.
- National Association of Insurance Counselors (NAIC). 2002a. "Executive Summary of Amendments to the Long-Term Care Insurance Model Regulation." January.
- National Association of Insurance Counselors (NAIC) 2002b. "State Survey of Long-Term Care Insurance Rating Practices and Consumer Disclosure Amendments. September
- National Center for Health Statistics, 2002. *Health, United States, 2002 With Chartbook on Trends in the Health of Americans*. Hyattsville, Maryland: 2002.
- Norton, Edward. 2000. "Long-term Care." In AJ Culyer and JP Newhouse (eds) Handbook of Health Economics, vol 1, ch 17. Elsevier Science.
- Puelitz, Robert and Arthur Snow. 1994. "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market." *Journal of Political Economy*, 102: 236-257.
- Robinson, James. 1996. "A Long-Term Care Status Transition Model," presented at the 1996 Bowles Symposium on the Old-Age Crisis: Actuarial Opportunities. Georgia State University, September.
- Robinson, James. 2002. "A long-term care status transition model." Mimeo, University of Wisconsin, Madison.
- Rothschild, Michael and Joseph Stiglitz. 1976. "An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90: 629-649.
- Sloan, Frank and Edward Norton. 1997. "Adverse selection, bequests, crowding out, and private demand for insurance: evidence from the long-term care insurance market." *Journal of Risk and Uncertainty* 15: 201-219.
- Smith, V Kerry, Donald Taylor and Frank Sloan. 2001. "Longevity expectations and death: can people predict their own demise?" *American Economic Review* September pp.1126-1134.
- Society of Actuaries. 1992. "Report of the long-term-care experience committee: 1985 national nursing home survey utilization data." *Transactions of the Society of Actuaries*, 1988-89-90 Reports. Pp.101-164.
- Society of Actuaries. 2002. "Long-term care experience committee InterCompany Study 1984-1999." Available at http://www.soa.org/sections/intercompany_study.pdf.
- United States Congress, Committee on Ways and Means (2000). Green Book 2000: Background Material on Programs Under the Jurisdiction of the Committee on Ways and Means, Washington DC: GPO.
- Weiss Ratings Inc. 2002. "Weiss Ratings' Consumer Guide to LTC Insurance." Palm Beach Gardens, Fl.
- Wilson, Charles. 1977. A model of Insurance Markets with Incomplete Information. *Journal of Economic Theory* 16, pp.167-207.

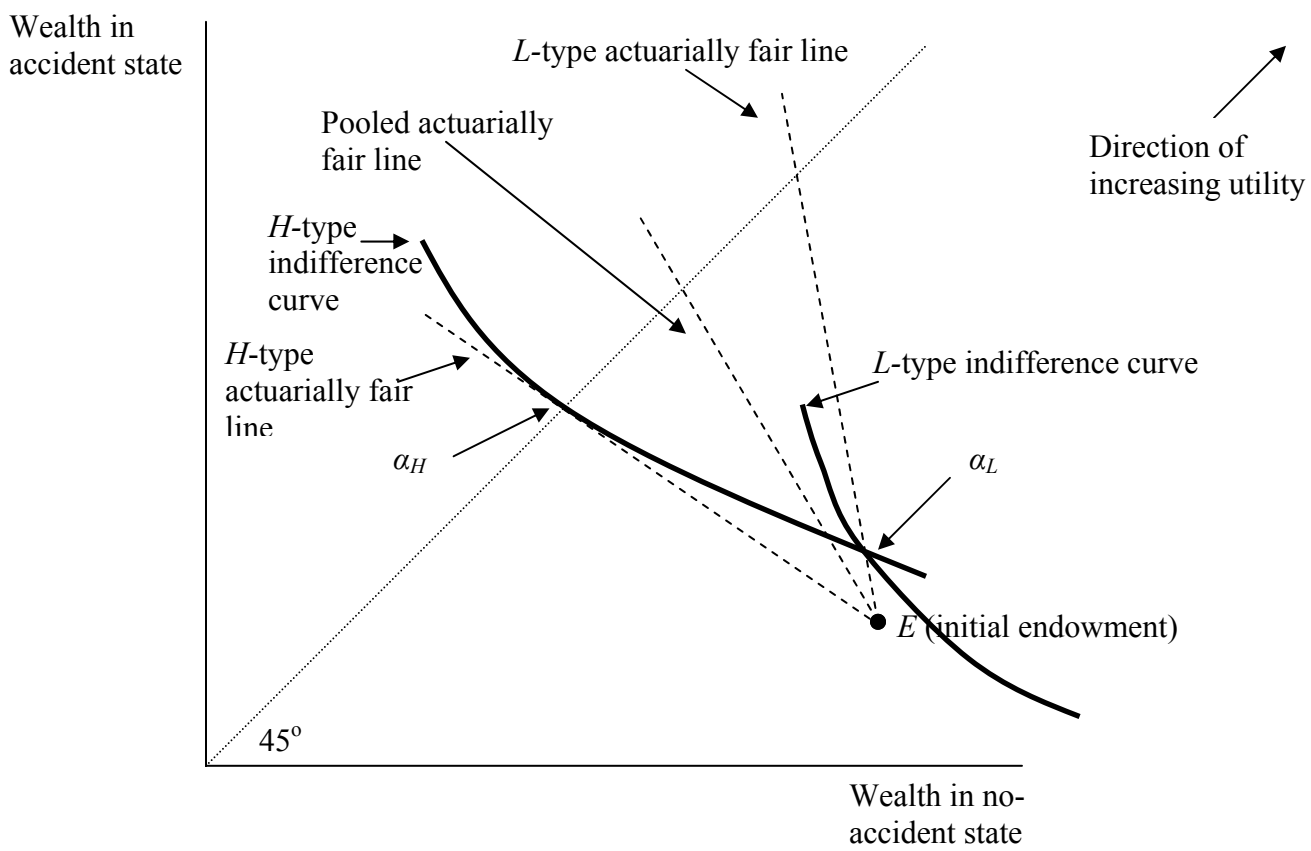


Figure 1: Separating Equilibrium

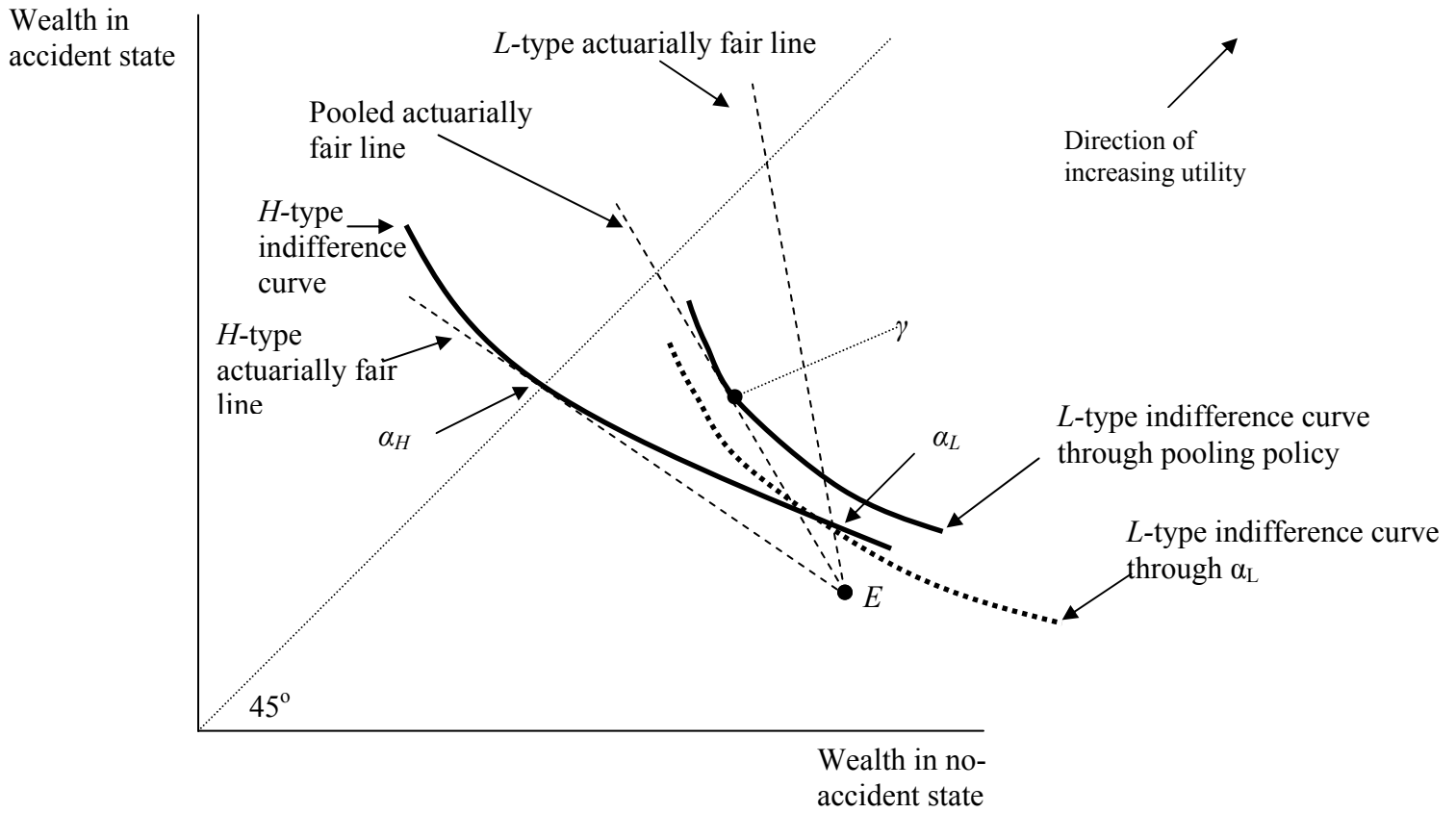


Figure 2: Pooling Equilibrium

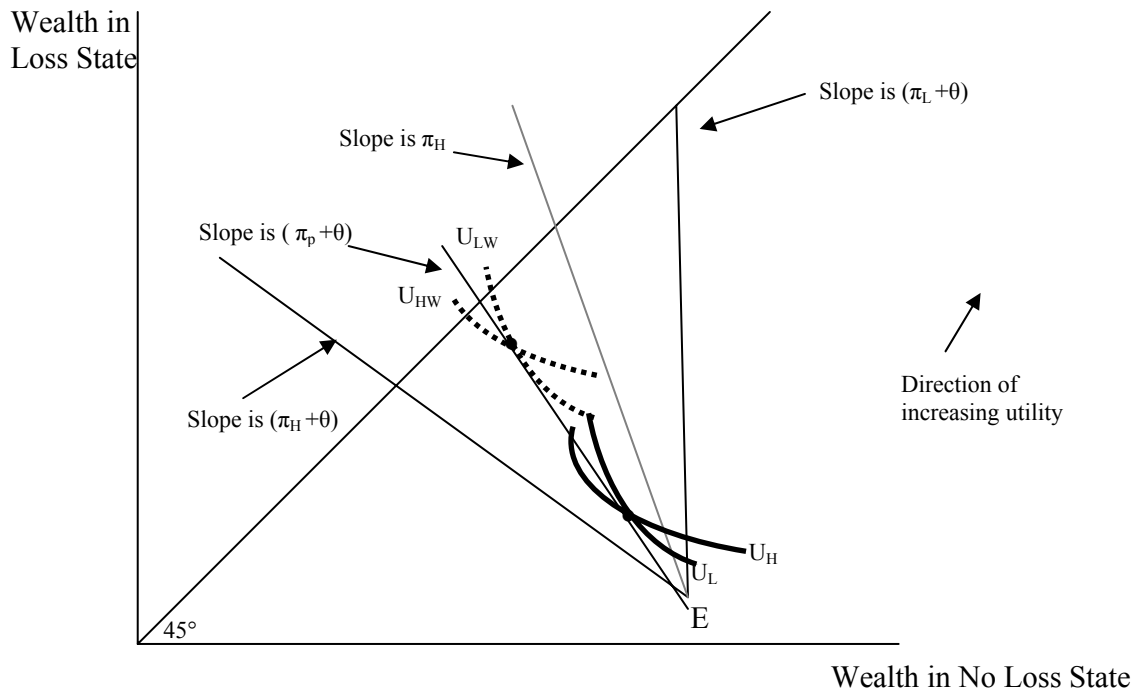
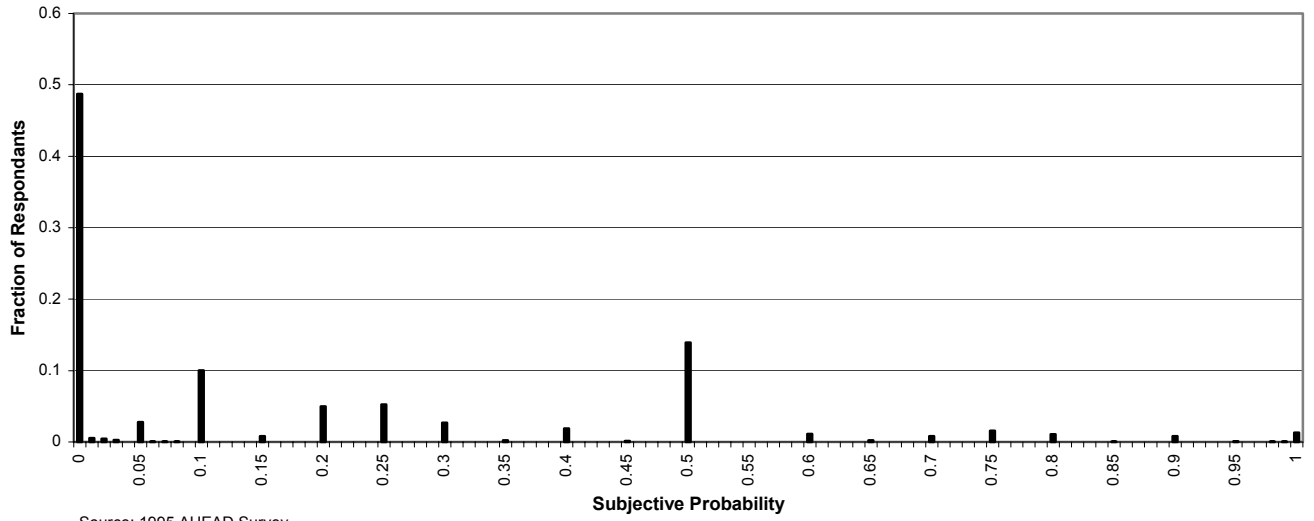


Figure 3: Asymmetric information with multiple policies and no positive correlation

Figure 4: Distribution of Subjective Probability of Entering Nursing Home Within 5 Years



Source: 1995 AHEAD Survey

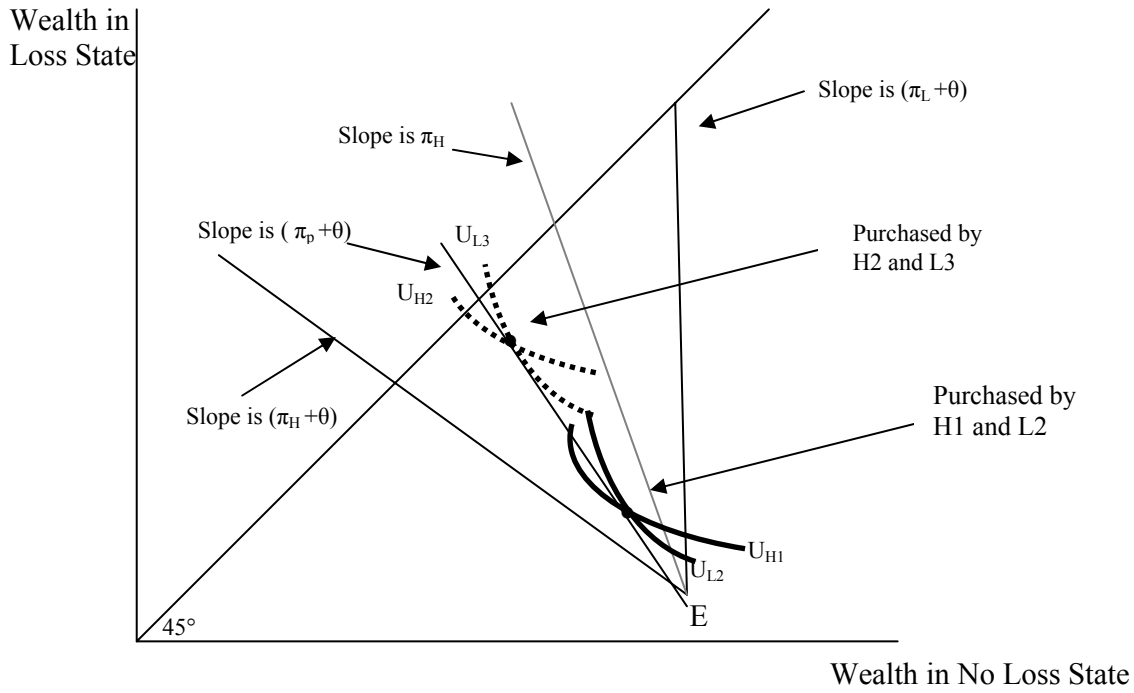


Figure 5: Replicating the empirical features of the long-term care insurance market equilibrium

Table 1: Means of Control Variables Included in “All Observables” Specification

Variable	Whole Sample	Insured	Uninsured
<i>Demographics (1995)</i>			
Age	78.6	77.4	78.8
Female	0.63	0.61	0.64
Married	0.54	0.60	0.54
Spouse’s age (if married)	73.8	73.3	73.9
Household Assets (median)	138,000	218,000	130,100
Household Income (median)	18,000	25,000	17,000
<i>Current Health (1995):</i>			
ADL limitation: bathing	0.11	0.08	0.12
ADL limitation: eating	0.05	0.04	0.05
ADL limitation: dressing	0.13	0.09	0.14
ADL limitation: toileting	0.08	0.07	0.08
ADL limitation: walking	0.10	0.06	0.10
Incontinence	0.22	0.25	0.21
Cognitively impaired	0.03	0.02	0.03
Use wheelchair	0.03	0.03	0.03
Use walker	0.07	0.03	0.08
Use crutches	0.003	0.002	0.004
Use cane	0.13	0.09	0.13
Use oxygen	0.01	0.008	0.01
Regularly use prescription drugs	0.79	0.82	0.78
IADL limitation: grocery shopping	0.15	0.11	0.15
IADL limitation: managing medication	0.05	0.03	0.05
Low BMI	0.10	0.09	0.10
High BMI	0.13	0.09	0.14
Currently smoke	0.08	0.09	0.08
<i>Health History (1995 and before):</i>			
Home Health Care Use	0.17	0.13	0.17
Nursing Home Use	0.02	0.03	0.02
Depression	0.19	0.14	0.20
Drinking Problem	0.03	0.04	0.03
Diabetes	0.14	0.12	0.14
Diabetes treated with insulin	0.05	0.04	0.05
Kidney Failure Assoc w. Diabetes	0.02	0.01	0.02
Stroke	0.12	0.09	0.12
Heart condition	0.34	0.32	0.34
Medication for heart problem	0.22	0.20	0.22
Heart Attack	0.09	0.08	0.09
Congestive Heart Failure	0.04	0.04	0.04
High Blood Pressure	0.54	0.59	0.53
Hip fracture	0.05	0.05	0.05
Lung Disease	0.12	0.13	0.12
Cancer	0.16	0.17	0.16
Psychiatric problems	0.15	0.16	0.14
Arthritis	0.54	0.49	0.54
Injury from falling	0.15	0.15	0.15

Note: All means are weighted. See Appendix A for our construction of cognitive impairment, depression, drinking problem, and household assets.

Table 2
Individuals' predictions of nursing home entry

	Control Variables								
	No Controls			Age Dummies		"All observables"		Insurance Company Prediction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual's Prediction									
<i>Continuous measure</i>	0.097*** (0.024)			0.073*** (0.023)		0.041* (0.022)		0.044** (0.022)	
<i>Categorical measure</i>									
Predicts 0 (omitted)		--			--		--		--
Predicts 1 to 49		0.063*** (0.011)			-0.004 (0.011)		0.003 (0.011)		-0.004 (0.011)
Predicts 50 to 100		0.067*** (0.013)			0.047*** (0.015)		0.032** (0.014)		0.033** (0.015)
Actuarial Prediction			0.507*** (0.028)					0.501*** (0.029)	0.498*** (0.029)
R ²	0.004	0.001	0.099	0.104	0.104	0.169	0.169	0.100	0.100
N	5,072	5,072	5,072	5,072	5,072	4,960	4,960	5,072	5,072

Note: Reported coefficients are from linear estimation of equation (1). Dependent variable is whether individual enters nursing home over subsequent five years; mean is 0.16. "Continuous measure" of individual's prediction uses the individual's reported prediction, rescaled to range from 0 to 1. When categorical measure of beliefs is used instead, the omitted category is "individual predicts 0". The column headings describe the additional covariates included in the regression. Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 3
The Relationship between Insurance Coverage and Individuals Beliefs about Risk Type

Control Variables								
Coefficient	No Controls		Age Dummies		"All observables"		Insurance Company Prediction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual's Prediction								
<i>Continuous measure</i>	0.090*** (0.020)		0.095*** (0.020)		0.095*** (0.020)		0.103** (0.020)	
<i>Categorical measure</i>								
Predicts 0 (omitted)		--		--		--		--
Predicts 1 to 49		0.063*** (0.011)		0.059*** (0.011)		0.048*** (0.011)		0.059*** (0.011)
Predicts 50 to 100		0.067*** (0.013)		0.070*** (0.013)		0.069** (0.013)		0.074*** (0.013)
Actuarial Prediction							-0.122*** (0.017)	-0.114*** (0.017)
R ²	0.005	0.011	0.013	0.018	0.050	0.052	0.013	0.017
N	5,072	5,072	5,072	5,072	4,960	4,960	5,072	5,072

Note: Reported coefficients are from linear estimation of equation (2). Dependent variable is whether the individual has LTC insurance; mean is 0.11. See notes to table 2 for more details.

Table 4: The Relationship between Insurance Coverage and Individuals Beliefs about Risk Type
(Sample restricted to top quartile of household income or wealth distribution in 1995)

Control Variables								
Coefficient	No Controls		Age Dummies		“All observables”		Insurance Company Prediction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual’s Prediction								
<i>Continuous measure</i>	0.157*** (0.042)		0.163*** (0.042)		0.158*** (0.043)		0.168*** (0.042)	
<i>Categorical measure</i>								
Predicts 0 (omitted)								
Predicts 1 to 49	0.083*** (0.020)		0.081*** (0.020)		0.077*** (0.020)		0.081*** (0.020)	
Predicts 50 to 100	0.099*** (0.026)		0.100*** (0.026)		0.098*** (0.027)		0.105*** (0.026)	
Actuarial Prediction							-0.139*** (0.039) -0.129*** (0.040)	
R ²	0.009	0.015	0.023	0.027	0.062	0.066	0.014	0.019
N	1,911	1,911	1,911	1,911	1,887	1,887	1,911	1,911

Note: Sample consists of those in the top quartile of the distribution of income or wealth. Reported coefficients are from linear estimation of equation (2). Dependent variable is whether the individual has LTC insurance; mean is 0.11. See notes to table 2 for more details.

Table 5
The Relationship Between Long-term Care Insurance and Nursing Home Entry

Sample	No Controls (1)	Controls for Age Dummies (2)	Controls for “all observables” (3)	Controls for insurance company prediction (4)
Entire Sample	-0.045*** (0.016) [N=6,280]	-0.016 (0.015) [N=6,280]	-0.001 (0.015) [N=6,083]	-0.012 (0.015) [N=6,275]
Top quartile of wealth or income distribution (i.e. those for whom Medicaid is not as good a substitute)	-0.045** (0.018) [N=2,161]	-0.032* (0.018) [N=2,161]	-0.028 (0.017) [N=2,123]	-0.026 (0.017) [N=2,161]

Notes: Each cell reports the coefficient on LTCINS from estimating equation (3) on the dependent variable “any nursing home use” and the set of control variables defined in the column heading. LTCINS is measured in 1995. The dependent variable covers the time period 1995 – 2000. The mean of the dependent variable is 0.19. The column headings describe the set of control variables used. See text and Appendix for detailed description of these covariates. All estimates are from linear probability models. Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 6:
Long-Term Care Insurance and Other Measures of Long-Term Care Utilization

Dependent Variable	No Controls (1)	Controls for Age Dummies (2)	Controls for “all observables” (3)	Controls for insurance company prediction (4)
Number of nights in nursing home (mean = 33)	-71.589*** (25.774) [N=6,189]	-28.853 (25.099) [N=6,189]	-15.673 (25.241) [N=5,998]	-30.067 (24.762) [N=6,181]
Any long-term care utilization (mean = 0.40)	-0.074 (0.021) [N=6,506]	-0.045** (0.020) [N=6,506]	-0.020 (0.018) [N=6,299]	-0.026 (0.019) [N=6,501]

Notes: Each cell reports the coefficient on LTCINS from estimating equation (3) on a specific dependent variable and definition of the set of control variables. All dependent variables cover the time period 1995 – 2000. The column headings describe the set of control variables used. See text and Appendix for detailed description of these covariates. Reported coefficients are from a Tobit model (top row) and a linear probability model (bottom row). Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 7: Preference-based Selection: Preventive Health Activity

	No Controls		Age Dummies		“All Observables”		Insurance Company Prediction	
	NH ENTRY	LTCINS	NH ENTRY	LTCINS	NH ENTRY	LTCINS	NH ENTRY	LTCINS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Preventive Activity	-0.111*** (0.019)	0.064*** (0.016)	-0.036** (0.018)	0.051*** (0.016)	-0.019 (0.019)	0.013 (0.017)	-0.055*** (0.018)	0.051*** (0.016)
Individual Prediction	0.102*** (0.024)	0.087*** (0.020)	0.072*** (0.023)	0.091*** (0.120)	0.040* (0.022)	0.095*** (0.020)	0.048** (0.023)	0.099*** (0.020)
N	5,010	5,010	5,010	5,010	4,900	4,900	5,010	5,010

Note: The top row of column headings describe the covariates included in the regression. The second row of column headings identifies the dependent variable. The dependent variable is either whether the individual went into a nursing home between 1995 and 2000 (odd columns) or whether the individual had private long-term care insurance in 1995 (even columns). All estimates are by OLS. “Preventive activity” measures the fraction of gender-appropriate preventive health activity undertaken by the individual. All regressions also include a control for sex because the fraction of potential preventive activity undertaken may vary with sex simply because the number of potential preventive activities is 3 for men and 5 for women. “Individual prediction” measures the individual’s continuous prediction. Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 8: Preference Based Selection: Preventive Health Activity and Wealth

	PANEL A: Preventive Health Activity				PANEL B: Wealth and Preventive Health Activity			
	No Controls	Age Dummies	“All observables”	Insurance Company Prediction	No Controls	Age Dummies	“All observables”	Insurance Company Prediction
LTCINS_HAT _{PREVENT}	-1.433*** (0.266)	-0.708*** (0.244)	-0.434 (0.349)	-0.771*** (0.251)	-0.819*** (0.244)	-0.488** (0.246)	-0.363 (0.350)	-0.305 (0.233)
LTCINS_HAT _{WEALTH}	-----	-----	-----	-----	-0.964*** (0.111)	-0.600*** (0.110)	-0.180 (0.111)	-0.454*** (0.116)
LTCINS_HAT _{RISKTYPE}	1.222*** (0.270)	0.781*** (0.246)	0.405* (0.232)	0.460*** (0.228)	0.929*** (0.243)	0.700*** (0.228)	0.425* (0.233)	0.421* (0.219)
LTCINS_RESID	-0.040*** (0.015)	-0.022 (0.015)	-0.007 (0.015)	-0.013 (0.015)	-0.022 (0.015)	-0.011 (0.015)	-0.009 (0.015)	-0.008 (0.015)
N	5,010	5,010	4,900	5,010	5,063	5,063	5,063	5,063

Note: Each column reports the results from estimating equation (6); the dependent variable is an indicator variable for whether the individual went into a nursing home between 1995 and 2000; all estimates are by OLS. See text for explanation of the construction of the right hand side variables. Although income and assets are typically included in the “all observables” specification as control variables, they are excluded from this specification in Panel B since they are used instead to generate LTCINS_HAT_{WEALTH}. Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Appendix A: The AHEAD Sample and Variable Definitions.

Sample definition: Our sample is drawn from the original Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS). This consists of a representative sample of individuals born in 1923 or earlier, and their spouses.³⁷ The AHEAD respondents were interviewed in 1993, 1995, 1998 and 2000. For reasons discussed below, we restrict our analysis to data from 1995 to 2000. We also exclude the 3 percent of original respondents who were in a nursing home in 1995. Non-death (i.e. “real”) attrition is just over 4 percent from 1995 and 2000. All of our estimates from the AHEAD data are weighted using the 1995 household weights.

Measuring long-term care insurance: We measure individuals’ insurance coverage in 1995, the first wave for which reliable information is available. Our indicator variable LTCINS is coded 1 if the individual answers yes to the following question:

R15: Aside from the government programs, do you now have any insurance which specifically pays any part of long-term care, such as, personal or medical care in the home or in a nursing home?

Although a few papers have used answers to questions about long-term care insurance in the 1993 wave (see e.g. Norton and Sloan 1997 or Mellor 2001) we are uncomfortable with this measure. In that year the survey asked specifically about a variety of types of health insurance and then asked if the respondent had any (other) type of insurance:

R6. Do you have any (other) type of health insurance coverage?

R7. What kind of coverage do you have? It is basic health insurance, a supplement to Medicare (MEDIGAP) or to other health insurance, long-term care insurance, or what?

The question thus does not specifically target long-term care insurance coverage. It yields an estimated coverage rate of just over 2 percent, substantially below what other analyses have indicated for this time period. By contrast, the reported coverage rate using the 1995 measure (10 percent) matches other existing estimates (see e.g. Cohen, forthcoming and citations therein). Our concern about the accuracy of the 1993 long-term care insurance measure was corroborated in email correspondence with David Weir, Assistant Director of HRS (April 2002).

Construction of some of the health measures collected by insurance companies

Cognition: We follow Mehta et al. (2002) who work specifically with AHEAD and define an individual as cognitively impaired if he has a score of 8 or less (out of 35) on the Telephone Interview for Cognitive Status (TICS). For proxy interviews, cognitive impairment is based on assessments offered by the proxy.

Depression: We again follow Meta et al. (2002) and define depression as a score of 3 or greater (out of 8) on the CES-D8. This measure is not available for proxy respondents. For the 9% of the sample who were interviewed by proxy, we set the depression measure to zero and include an indicator for proxy interview.

Drinking problem: We define a drinking problem as 3 or more drinks per day.

Assets. Household assets are defined as total bequeathable assets (including housing wealth but not Social Security or Defined Benefit pension wealth) less debts.

³⁷ A younger cohort, born in the years 1931-1941, was interviewed for the companion HRS survey. We use the AHEAD cohort because the HRS cohort was not asked to report their subjective probability of entering a nursing home (a key variable in Section 3) until later waves.