Preliminary and incomplete.
Do not cite or circulate without permission.

Asymmetric information in insurance markets: evidence and explanations from the long-term care insurance market

Amy Finkelstein
Harvard University and NBER

Kathleen McGarry
UCLA and NBER

March 2003

Asymmetric information has often been viewed as a key factor hindering the efficient operation of insurance markets. Yet several recent empirical analyses have concluded that asymmetric information may not exist in insurance markets as diverse as the automobile and life insurance markets. The conclusions of these studies are based on a widely used test of asymmetric information, specifically that its presence would imply a positive relationship between insurance coverage and risk occurrence. In this paper, we show empirically that despite the lack of a positive relationship, asymmetric information may still impair market functioning. We analyze the market for long term care insurance and find no evidence that individuals with more long-term care insurance are more likely to use nursing home care. However, we also find direct evidence of asymmetric information: controlling for the information set of the insurance company, individuals have residual private information about their risk type and this private information is positively correlated with insurance coverage. Further, we show that the lack of a positive relationship between insurance coverage and care utilization in equilibrium – *despite* asymmetric information about risk type – is attributable to other unobserved characteristics of the individual that are positively related to coverage and negatively related to care utilization. For example, we find that more cautious individuals are both more likely to have insurance and less likely to enter a nursing home. Such preference-based selection can offset the positive correlation between insurance coverage and care utilization that asymmetric information about risk type would otherwise produce. It cannot, however, offset the negative efficiency consequences of this asymmetric information for insurance coverage.

We thank Daron Acemoglu, Jeff Brown, Pierre-Andre Chiappori, Raj Chetty, Seema Jayachandran, Ben Olken, Sarah Reber, Jesse Shapiro, and Jonathan Wright for helpful comments and discussions. We are especially grateful to the employees of the insurance company who generously provided us with the proprietary data used in this paper and answered many questions about it, and to Jim Robinson for generously providing us with the actuarial model of long-term care utilization.

Theoretical research has long emphasized the potential importance of asymmetric information in impairing the functioning of insurance markets. Its empirical relevance, however, remains the subject of considerable debate. Several recent studies of the automobile, health, and life insurance markets have concluded that asymmetric information may not exist in these insurance markets (e.g. Chiappori and Salanie, 2000; Cardon and Hendel, 2001; and Cawley and Philipson, 1999). Specifically, contrary to the predictions of many moral hazard and adverse selection models, there appears to be no evidence that individuals with more of these types of insurance are more likely to experience the insured risk. These findings appear to challenge the conventional wisdom among economists and policymakers that asymmetric information is likely to create problems in insurance markets.[1]

These findings also raise an important but, as yet largely unanswered, question: *why* do we fail to find evidence of asymmetric information in some insurance markets? Alternative explanations for the same observed equilibrium relationship between insurance coverage and risk occurrence can have very different implications for the structure of information, for market efficiency, and potentially for public policy. Understanding the underlying structure of the insurance market that produces this equilibrium is therefore critical. However, to our knowledge, there exists no systematic empirical investigation of this issue.

We investigate this issue in the context of the private long-term care insurance market in the United States. In addition to providing an interesting setting to study asymmetric information, the long-term care insurance market is of substantial interest in its own right. Long-term care expenditures represent one of the largest uninsured financial risks faced by the elderly in the United States. As the baby boomers age and medical costs continue to rise, the implications of the lack of long-term care insurance for the welfare of both the elderly and their children, will only become more pronounced. One potential explanation for the private market's limited size is the presence of adverse selection or moral hazard. To date, however,

[1] 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).

there exists very little empirical evidence on the presence of asymmetric information in this market, let alone evidence of whether it is an important factor in limiting the market's size.

We begin by following the existing literature and examine whether there is positive correlation between the quantity of insurance purchased and the occurrence of the risk which, in this case, is the individual's ex-post use of a nursing home. We analyze data from three different sources: aggregate actuarial tables produced by the Society of Actuaries, proprietary micro data from a large private long-term care insurance company, and individual-level panel data from the Asset and Health Dynamics Among the Oldest Old (AHEAD) survey. In none of the data sets do we find evidence that those with more long-term care insurance end up using more nursing home care—in anything, we find suggestive evidence of the reverse.

Similar evidence of a lack of a positive correlation between insurance and risk occurrence in other insurance markets has been interpreted as evidence against the presence of adverse selection and moral hazard in these markets (see e.g. the papers cited in the opening paragraph). Indeed, one potential explanation for these findings is that individuals do not have private information about their risk type. If this were true and information was indeed symmetric, the market equilibrium for insurance against the risky event could be first best. However, recent theoretical work by de Meza and Webb (2001), Jullien et al. (2002) and Chiappori et al. (2002) suggests an alternative explanation: the insurance company may lack information not only about the individual's risk type, but also about preference-related characteristics which have the *opposite* correlation with insurance coverage and with risk occurrence (e.g. risk aversion). These other "private factors" may offset the positive correlation between insurance coverage and risk occurrence that asymmetric information about risk type would otherwise produce. But, as we explain in more detail below, this offsetting effect does not correct the market inefficiencies produced by asymmetric information. In equilibrium, even if the insured are not above-average in their risk type, coverage levels will not be optimal.

To explore which of these competing explanations is responsible for the equilibrium in the private long-term care insurance market, we directly examine whether individuals have private information about their risk type. To do so, we draw on several complementary and rich data sources to construct measures of the individual's beliefs about his risk type, the insurance company's beliefs about his risk type, and the insurance company's information set. Our results indicate that, after controlling for the risk-classification of the individual done by the insurance company, the individual's beliefs about his subsequent nursing home use remain positively and statistically significantly correlated with this subsequent use. This result provides direct evidence of asymmetric information in the private long-term care insurance market: the individual *does* have residual private information about his risk type. Moreover, we find that this private information is positively correlated with whether the individual has insurance coverage.

We then demonstrate that the existence of unobserved heterogeneity not only in risk type but also in preferences can reconcile the direct evidence that individuals have private information about their risk type with the fact that, in equilibrium, there is no positive relationship between insurance coverage and care utilization. We also provide direct evidence of the existence and nature of these other unobserved, preference-related characteristics. For example, consistent with the theoretical models of de Meza and Webb (2001) and Jullien et al. (2002), we find that more cautious individuals (a characteristic not observed by the insurance companies) are both more likely to own long-term care insurance and less likely to end up using long-term care. We discuss below why insurance companies may not collect such information.

Thus in the case of the long-term care insurance market, there appears to be asymmetric information about risk type *despite* the lack of a positive correlation between insurance coverage and care utilization. Asymmetric information may therefore be an important contributor to the limited size of the private long-term care insurance market. However, the lack of a positive correlation between insurance coverage and care utilization does suggest that asymmetric information about risk type should not raise the *price* of long-term care insurance above the population-average actuarially fair price; from the viewpoint of the insurance company, the insured do not have above-average risk characteristics.

The rest of the paper is structured as follows. Section one provides some brief background on long-term care and the private long-term care insurance market. Section two provides the conceptual framework for the paper and outlines the empirical approach. The next three sections present, respectively, the three main empirical findings. Section three documents the lack of a positive correlation between long-term care insurance coverage and nursing home care utilization. Section four provides evidence that individuals have private information about their risk type. Section five investigates the relationship between characteristics of the individual unobserved by the insurance company and the individual's insurance coverage and care utilization. Section six presents suggestive evidence of the existence of insurance rationing in the private long-term care insurance market. The final section summarizes our findings and discusses their implications for other insurance markets.

1. Background on long-term care and long-term care insurance

At almost $100 billion in the year 2000, long-term care expenditures in the United States represent almost 10% of total health expenditures *for all ages*, and about 1% of GDP. There is substantial variation among the elderly in their long-term care utilization (see e.g. Dick et al. (1994) and Kemper and Murtaugh (1991)), suggesting potentially large welfare gains from insurance coverage. For example, Dick et al. (1994) estimate that about two-thirds of individuals who reach age 65 will never enter a nursing home, one-quarter of women who do enter a nursing home will spend at least three years there.

Most of this risk is uninsured, making long-term care expenditures one of the largest uninsured financial risks facing the elderly (and potentially their children). Only about 10 percent of those aged 60 and over had private long-term care insurance coverage in 2000.[2] Moreover, most private insurance policies provide very limited insurance benefits. In particular, they tend to specify a maximum amount that will be reimbursed per day in covered care that is considerably below average *current* daily nursing home costs and is not scheduled to rise with medical cost inflation (Cutler (1996); Brown and Finkelstein (2003)). Public insurance is also quite limited. Medicare, the public health insurance program for the elderly, covers only a very restricted set of long-term care services. And Medicaid, the public health insurance program for the indigent, requires a deductible of almost one's entire wealth before it will cover

[2] Authors' calculation based on 2000 Health and Retirement Survey. This is consistent with other estimates (e.g. Cohen, forthcoming).

long-term care expenses. As a result, less than 1 percent of long-term care expenditures for the elderly in 2000 were paid for by private insurance contracts and 40 percent were paid for out of pocket. (US Congress, 2000). By contrast, in the health sector as a whole, private insurance paid for 35 percent of expenditures, and only 17 percent were paid for out of pocket (National Center for Health Statistics, 2002).

Currently most buyers of long-term care insurance are of retirement age or older. In 2000, the average age of buyers in the individual market was 67, and over one-fifth of the buyers were 75 or older at the time of purchase (HIAA 2000a). Premiums, which tend to be a constant nominal amount paid on a monthly or annual basis, increase linearly with the maximum benefit the policy allows the individual to receive per day in covered care.[3] Coverage rates are roughly comparable for men and women but increase substantially with asset levels, probably due to the means-tested nature of the public Medicaid insurance (HIAA, 2000a). About 80% of private insurance is provided by the individual (non-group) market, with the remaining share sold through employer-sponsored plans or life insurance (HIAA 2000b). Presumably to reduce moral hazard, insurance policies specify a set of standard health-related criteria ("benefit triggers") that must be satisfied before an individual is eligible to receive benefits for covered care (Wiener et al., 2000).

Despite the absence of regulatory restrictions, firms use relatively little information in pricing policies. Policies are not experienced rated, and premiums tend to vary only with age and several broad health categories (ACLI, 2001; Weiss 2002). Most notably, they do not vary with gender, despite known long-term care utilization differences by gender (Society of Actuaries, 1992). This is something of a puzzle.[4] One potential explanation is that – even though a given policy covers only a single life – policy ownership is highly correlated among couples; for example, although only 10% of the elderly in the 1995 AHEAD data have private long-term care insurance coverage, over 60% of the spouses of individuals in

[3] Pricing data are from Weiss Ratings Inc and from the insurance company whose proprietary data we analyze in Section 3.2.
[4] The puzzle of why insurance companies do not use many observable and relevant characteristics in pricing insurance exists in other insurance markets as well, such as annuities (Finkelstein and Poterba, 2000).

these data with long-term care insurance also have this insurance.

A variety of theoretical explanations have been proposed for the limited size of the private long-term care insurance market (see e.g. Norton, 2000 for a review). Asymmetric information is one potential explanation, yet there exists very little empirical evidence on its presence in this market. Consistent with moral hazard, Garber and MaCurdy (1993) present evidence of an increase in nursing home discharges when the Medicare nursing home benefit is exhausted. The widespread use of deductibles in long-term care insurance policies (Brown and Finkelstein, 2003) is also suggestive of asymmetric information in this market.[5]

2. Theoretical background

2.1 The "positive correlation" prediction

The standard empirical test for asymmetric information is to test for a correlation between the amount of insurance coverage and the ex-post occurrence of the (potentially) insured risk. A wide variety of asymmetric information models predict that – in markets in which observationally identical individuals are offered a choice from a menu of insurance contracts – coverage and the probability of the risky event being realized will be positively correlated (Chiappori and Salanie, 2000; Chiappori et al., 2002). Throughout, the paper we will refer to this test as the "positive correlation" prediction Of course, the prediction – and hence a valid empirical test – applies among individuals who would be treated symmetrically by the insurance company (i.e. placed in the same risk category and offered the same menu of insurance contracts, defined over price and benefit characteristics).[6]

The "positive correlation" can arise from either adverse selection or moral hazard, both of which result in a market that is inefficient relative to the first best. The two types of market failures do, however, imply different mechanisms through which this positive correlation arises. In the case of adverse

---

[5] Sloan and Norton (1997) examine the relationship between insurance coverage and individuals' beliefs about their nursing home risk using an early wave of the HRS. However, this early wave does not provide a reliable measure of long-term care insurance coverage (see Appendix A).

[6] See Dionne et al. (2001) for a more detailed discussion of this point.

selection, the insured is assumed to have superior information to the insurance company about his risk type. Because individuals who appear to the insurance company to be observationally equivalent face the same menu of insurance options, and because the marginal utility of insurance at a given price is increasing in risk type, those with private information that they are high risk will select contracts with more insurance than those with private information that they are low risk (Rothschild and Stiglitz, 1976). In the case of moral hazard, the causality is reversed: coverage by insurance 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.

Empirically, the positive correlation property appears to exist in some insurance markets but not others. In health insurance, Cutler (2002) reviews an extensive empirical literature that finds evidence in support of this prediction, although he notes the existence of exceptions, such as Cardon and Hendel (2001). There is also evidence from annuity markets that the insured are higher risk (Finkelstein and Poterba 2000, 2002), but no such evidence in life insurance markets (Cawley and Philipson, 1999). In the automobile insurance market, the empirical evidence is mixed. Chiappori and Salanie (2000) and Dionne et al. (2001) fail to reject the null hypothesis of no correlation; but Pueltz and Snow (1994) and Cohen (2001) find support for the positive correlation prediction.

2.2 What does a lack of positive correlation imply for the information structure and for market efficiency?

The explanations that have been offered to explain a lack of a positive correlation between insurance coverage and risk occurrence can be broadly classified into two classes, with different implications for the underlying structure of information and for market efficiency.

The first class of explanations challenges the central assumption of adverse selection models that the individual has information about his risk type (i.e. expected loss experience) that is unknown to the insurance company. Given the vast amount of information that insurance companies can, and do, collect about potential customers, information about the individual's risk type may be symmetric.[7] With symmetric information, all else equal, insurance coverage should not vary systematically with expected loss experience and the equilibrium insurance coverage against the underlying financial risk can be first best.[8] However, a market with symmetric information about risk type will not provide insurance against the *classification* risk of being a high-risk type (Hirshliefer, 1971, Crocker and Snow, 2000).

A symmetric information story, however, is not enough to explain the apparent lack of moral hazard that may also be implied by the failure to find the positive correlation property. It may be that moral hazard does not exist in particular insurance markets. For example, in the case of long-term care insurance, the unappealing nature of nursing homes, and the use of health-related criteria that must be satisfied before care will be reimbursed, may be sufficient to dampen any potential moral hazard effects.

The second class of explanations for a lack of a positive correlation between insurance coverage and risk occurrence offers an explanation in a market in which both asymmetric information about risk type and moral hazard may be present. This class of explanations argues that risk type is not the only source of unobserved heterogeneity. Individuals may also differ on unobserved preference-based characteristics – such as risk aversion – that are correlated with the demand for insurance coverage and risk occurrence. We refer to this as "preference-based" selection to distinguish it from traditional adverse (or "risk based") selection based directly on the individual's private information about his risk type.

If unobserved preferences are positively correlated with insurance demand and negatively correlated with risk occurrence, then there may be no positive correlation between insurance coverage and risk occurrence in equilibrium, even in the presence of asymmetric information about risk type. For example, if more risk averse individuals – who value insurance more – are also lower risk, the correlation between

[7] It is also possible that the insurance company, with its access to sophisticated actuarial methods, may have *superior* information about the individual's risk type. In this case, the equilibrium may involve a *negative* correlation between risk type and amount of insurance coverage. (Villeneuve 2000, Villeneuve forthcoming).
[8] An important caveat, is that if the information is costly for the insurance company to collect, the private market may collect *too much* information relative to what is socially efficient (Crocker and Snow (1987)).

insurance coverage and risk occurrence may be zero or even negative.[9] The positive correlation prediction of asymmetric information models therefore arises *only after conditioning on all of the risk classification done by the insurance company* and *all determinants of insurance-buying behavior other than the individual's risk type.*

The exact nature of the equilibrium with multiple unobserved characteristics can be complex, in part because the single crossing property may no longer hold. The equilibrium will, however, be inefficient relative to the first best, and will generally result in underinsurance. In equilibria that pool different risk types (see e.g. Smart (2000) and Wambach (2000)), no risk type faces his actuarially fair price on the margin and all individuals receive less than full insurance; it would therefore be Pareto improving if each type could buy additional insurance at his risk-type-specific actuarially fair price.[10] Equilibria that separate different risk types involve the standard inefficiency that at least one type is constrained from receiving his optimal insurance coverage. These equilibria need not yield a positive correlation between insurance coverage and risk occurrence since, in the presence of multiple unobservables, it may be optimal for an insurance company with market power to constrain the less risk averse (but higher risk) to receive less coverage than the higher risk averse (but lower risk) (see e.g. Chiappori et al. 2002). In this case, the first best outcome could be achieved if all information about the individual were publicly available and the monopolist could thus perfectly price discriminate.

It is less clear, however whether such first-best outcomes – or even Pareto improvements over the existing equilibrium – are feasible given the information structure. In the above models, feasible Pareto

---

[9] The more risk averse may be exogenously assumed to be of lower risk, or their lower risk may be determined endogenously in models of both adverse selection and moral hazard in which the more risk averse investment more in risk-reducing effort (Jullien et al. (2002), or de Meza and Webb (2001)).
[10] Of course, a single pooling equilibrium – and the resultant lack of a correlation between insurance coverage and risk occurrence – may also exist in asymmetric information models with only unobserved risk type (see e.g. Wilson, 1977). However, the positive correlation prediction applies to market in which observationally equivalent individuals choose *different* contracts. For the positive correlation property not to exist in a model with asymmetric information on multiple characteristics *and* multiple contracts requires a wedge between price and expected loss. Jullien et al. (2002) and Chiappori et al. (2002) develop monopoly models with these features; de Meza and Webb (2001) develop a competitive model with administrative costs (i.e. loading) that also exhibits these features. The equilbria in these models may exhibit a positive, negative, or zero correlation between insurance coverage and risk occurrence (see e.g. Jullien et al, 2002).

improvements are not readily apparent; in other words, the equilibrium may be second best (or constrained) Pareto efficient. However, in the de Meza and Webb (2001) model in which the more risk averse are also lower risk (and thus the positive correlation property need not apply), their additional assumption of an administrative load on the insurance policy produces an equilibrium that may exhibit overinsurance. Government taxation of insurance may thus drive out individuals whose valuation of insurance is less than the (net of administrative costs) costs of providing this insurance and thus produce a Pareto improvement.

Evidence of a lack of a positive correlation between insurance coverage and risk occurrence is therefore insufficient for making inference about either the structure of information or the efficiency of the market. Similar limitations apply to interpreting evidence of a positive correlation between insurance coverage and risk occurrence as evidence of an (inefficient) asymmetric information equilibrium. The positive correlation may reflect the presence of asymmetric information about risk type; alternatively, this relationship may reflect an (efficient) symmetric information equilibrium in a competitive insurance market with a marginal load on insurance, in which the lower risk are also less risk averse and therefore choose (optimally) to buy less insurance in equilibrium. Direct information about the information structure is therefore required in order to test for asymmetric information.

2.3 Overview of empirical approach

In this section, we briefly sketch the set of empirical tests used in the paper to distinguish among these alternative models of the insurance market. We assume throughout that individuals differ on a set of characteristics (X) that insurance companies use to classify individuals into risk categories, and that these characteristics are all observable to the econometrician. As the preceding discussion demonstrated, the positive correlation prediction is motivated by a model in which, conditional on X, the unobserved characteristics of the individual (which we denote by T for the individual's unobserved type) relate only to his probability of risk occurrence; any (unobserved) preferences are identical across individuals.

If T were observable to the econometrician, we could estimate:

$$(1)\ \text{CARE} = Xb_1 + \text{b}_2\text{LTCINS} + b_3\text{T} + \varepsilon$$

$$(2)\ \text{LTCINS} = Xd_1 + d_3T + \eta$$

CARE is a dependent variable that measures the amount of long-term care used by the individual (i.e. the occurrence of the risk). LTCINS is a measure of the individual's long-term care insurance coverage. By definition, $b_3 > 0$ (higher risk individuals use more care). If moral hazard is present, we expect $b_2 > 0$.

Now assume that T is unobserved by the econometrician and thus omitted from estimation of equation (1). We therefore can only estimate:

$$(1a)\qquad \text{CARE} = X\beta_1 + \beta_2\text{LTCINS} + \varepsilon'$$

Equation (1a) suffers from the classic omitted variables bias problem, with the sign of the bias on $\hat{\beta}_2$ equal to the sign of $b_3 * d_3$. If the only unobserved characteristic of the individual is his risk type, adverse selection models predict $d_3 > 0$ (those with private information that they are high risk will select more insurance in equilibrium), hence $\hat{\beta}_2$ will be an upward biased estimate of the moral hazard parameter b_2 in equation (1). The "positive correlation" prediction examined in the existing empirical literature is therefore that if *either* adverse selection *or* moral hazard (or both) is present in the market, we should find $\hat{\beta}_2 > 0$ in equation (1a). In Section 3 we estimate various versions of equation (1a) but find no evidence that $\hat{\beta}_2$ is positive.

As discussed above, there are two potential explanations for this finding: 1) individuals do not have private information about their risk type and there is no moral hazard in this market, or 2) individuals do have private information about their risk type and there are also unobserved preferences for insurance that are correlated with their risk probabilities. In the latter case, it will not necessarily be true that those with higher risk buy more insurance *even in the presence of private information about risk probabilities.* In other words, d_3 in equation (2) is not necessarily positive, and can be negative or zero. As a result the sign of the bias on $\hat{\beta}_2$ as an estimator of b_2 is ambiguous and, consequently, $\hat{\beta}_2$ need not be positive.

To distinguish between these two alternative hypotheses, in Section 4 we investigate directly whether individuals, conditional on the classification done by the insurance company (X), have any residual ability to predict their risk type. To do so, we take advantage of an innovative question in the AHEAD that asks each individual his beliefs (B) about his chance of entering a nursing home in the next five years.

We therefore estimate:

(3) $\text{CARE} = X\beta_1 + \beta_3 \text{B} + \varepsilon''$

We find direct evidence of asymmetric information: $\beta_3 > 0$.[11] In Section 5, we show empirically that the existence of preference-based selection can reconcile the direct evidence of asymmetric information with the lack of a positive correlation between insurance coverage and long-term care utilization

3. Is there a positive correlation between LTC insurance and LTC utilization?

Long-term care encompasses both institutional care (i.e. care in a nursing home) as well as home health care. Institutional care accounts for three-quarters of total long term care expenditures (US Congress, 2000). Until quite recently, policies tended to cover exclusively nursing home care.[12] As a result of this, and data limitations discussed below, we focus our empirical tests on the relationship between insurance coverage and nursing home utilization. We comment briefly on the limited evidence of the relationship between insurance coverage and home health care utilization.

Our measure of care utilization is usually a binary measure of whether the individual consumes *any* nursing home care; where data permit, we also test for a positive correlation between insurance coverage and the *intensity* of care utilization (i.e. length of stay). We make two basic types of comparisons. First, we compare care utilization *among* insured individuals with different amounts of insurance using a proprietary database on insurance purchases and subsequent claims experiences of a large, private long-

---

[11] We do note control for LTCINS in estimating equation (3) – as might be suggested by equation (1) – because to the extent that private information about risk type is correlated with insurance coverage due to adverse selection, we would be controlling away part of the individual's information.

[12] The long-term care insurance market began in the early 1980s with policies that tended to cover nursing home care only. Even in 1990, two-thirds of policies sold covered only nursing homes. By 2000, however, over three-quarters of new policies covered both home care and nursing home care (AARP 2002, SOA 2002, HIAA 2000a).

term care insurance company. Second, we compare care utilization for the insured population with that of the general population using individual-level panel survey data from the ADEAD.

Before embarking on this formal analysis, however, we begin by presenting some simple, but illustrative, comparisons using actuarial data from the Society of Actuaries (SOA).

3.1 Actuarial Data from the Society of Actuaries (SOA)

The SOA publishes tables of nursing home admission rates for the general population and for about 60% of the privately-insured population (SOA, 2000, 2002). Figure 1 plots the ratio of nursing home admission rates for insured individuals whose policies have no deductible to admission rates for the general population, by age and by sex.[13] The positive correlation property predicts that the insured-to-population ratio of admission rates should be larger than one. Figure 1 is not consistent with this prediction. We observe similar admission rates at younger ages and much *lower* nursing home admission rates for the insured relative to the population at older ages.

Figures 2 and 3 compare nursing home admission rates (by age) among insured individuals with different benefit periods and different maximum daily benefits, respectively.[14] The benefit period denotes the total number of days that benefits can be received during the lifetime of the policy; the maximum daily benefit denotes the maximum amount of incurred care expenditures that the policy will reimburse per day in care. Increases in either aspect of the policy increase the total amount of insurance in the contract. However, the figures do not indicate a positive correlation between the amount of insurance in the contract and nursing home admissions; nursing home admission rates are non-monotonic in the daily benefit amount (Figure 3) and are in fact *decreasing* in the benefit period (Figure 2).

---

[13] We limit the insured data to the approximately 12% of the exposure that reflects the experience of policies with no deductible because the insurance company data only record an admission that results in a claim; admissions that do not stay longer than the policy's deductible period will therefore be missed in the insurance company data. Of course, admissions that do not result in a claim because of failure to meet the health-related benefit triggers will also be missed. The omission of these "uncovered" admissions" will tend to understate the nursing home utilization of the insured relative to the population, although the impact is likely to be small; only about 6 to 10 percent of nursing home admissions in the population are for individuals who fail to meet these benefit triggers (Brown and Finkelstein, 2003).

[14] Both figures show results for the most common deductible (20 days), which represents about one-third of the exposure in the data; results for other common deductibles are similar (not shown).

Although suggestive, the results in Figures 1 through 3 do not represent a formal test of the positive correlation prediction. Most importantly, they do not condition on the risk classification of the individuals done by the insurance companies. In addition, the comparison of utilization rates between the insured and the general population further suffers from the fact that many “uninsured” individuals may in fact be able to collect public Medicaid insurance should they end up in a nursing home. Our formal analysis of micro data in the next two sub sections is designed to address these issues.

3.2 Proprietary policyholder data from a large private insurance company

3.2.1 Data and empirical framework

We have data on the complete set of individual (non-group) private long-term care insurance policies sold by a large U.S. private long-term care insurance company from January 1, 1997 through December 31, 2001. The company is among the top-five companies in this market (which combined account for almost two-thirds of premiums (LIMRA ,2001)) and sold about 150,000 policies during the period covered by the data. We observe a complete description of the features of each policy, as well as the information needed to classify the individuals into the “observable” risk categories created by the company. As is typical of the industry as a whole, this risk classification depends on the individual’s age at the time of issue, whether the individual is rated preferred, standard, or substandard based on detailed health information, and when the policy was issued; we observe all three of these features. We also observe a complete description of all claims incurred through December 31, 2001.

Although these data come from a single company, they appear comparable to the broader market on a variety of dimensions. These include the average age at purchase, the gender-mix of the purchasers, the average daily benefit, and the average length of the benefit period.[15] In addition, this particular company has experienced similar growth rates in policy sales to the industry as a whole over the last five years (LIMRA 2001). However, on several other dimensions the company appears quite different from industry averages. Almost all of the policies sold cover both home care and nursing home care, whereas

[15] Table 1 provides the statistics for the company data. HIAA (2000a) provides comparable industry-wide statistics.

in the industry as a whole only about three-quarters of recent policies do (HIAA 2000a). In addition, the policies tend to have larger deductibles than industry averages and are more likely to specify maximum daily benefits that escalate over time in nominal terms rather than maximum daily benefits that remain constant in nominal terms. Lacking more detailed data on the industry, it is difficult to know whether in fact each of the large companies has its own idiosyncrasies or whether the other large companies look more like the industry averages.

We estimate hazard models of nursing home care utilization for individuals who purchase different types of policies. The positive correlation property implies that those with more generous coverage are more likely to use a nursing home. However, we only observe care utilization for which a claim is paid. Therefore, to be able to analyze the effect of the deductible period on nursing home admissions, we restrict the sample to the 94% of policies that have a deductible of 100 days or less (and were issued at least 100 days before the end of the sample period) and define a "failure" in our hazard model as having at least 100 continuous days of nursing home care.[16] The results are similar if we restrict the sample to the almost 90% of policies with a 100-day deductible.

The average failure rate (0.3 percent) is quite low, but is consistent with market-wide and population statistics on nursing home care utilization (SOA 1992, 2002). Due to this low failure rate, the sample size appears insufficient to analyze the relationship between policy characteristics and length of stay beyond 100 days, or the occurrence of multiple stays of at least 100 days in length. We therefore focus on whether or not an individual had any stay lasting 100 days or more.[17]

We let $\lambda(t, x_i, \beta, \lambda_0)$ denote the hazard function, the probability that a policyholder with personal and policy characteristics x_i enters their 100th day of continuous nursing home care *t* periods after purchasing the policy, conditional on not having done so prior to *t*. We use the standard proportional

---

[16] The deductible must be used up anew for each care episode.

[17] Conditional on entering a nursing home, stays of more than 100 days are quite common. For example, published estimates suggest that between 40 and 55% of 65 year olds who enter a nursing home will spend more than 1 year (365 days) there (Dick et al, 1994, Kemper and Murtaugh, 1991, and Murtaugh et al. 1997).

hazard model which assumes that $\lambda(t, x_i, \beta, \lambda_0)$ can be decomposed into a baseline hazard $\lambda_0(t)$ and a proportional "shift factor" $\exp(x_i'\beta)$ as follows:

(4) $$\lambda(t, x_i, \beta, \lambda_0) = \exp(x_i'\beta)\lambda_0(t).$$

We estimate a semi-parametric Cox proportional hazard model to avoid making any parametric assumptions about the baseline hazard $\lambda_0(t)$.

The hazard model framework is particularly well-suited to handling the extensive right-censoring in the data. Censoring (exiting the sample for reasons other than failure) occurs either because the sample period ends or because the policy terminates due to death or to failure to pay premiums. Slightly less than 10 percent of our policies terminate; this is comparable to industry-wide termination rates (SOA 2002).[18]

Because the positive correlation prediction is conditional on the individual's risk classification, we include a set of covariates designed to capture this risk classification. These consist of indicator variables for issue year, rating category (standard, preferred or substandard), and issue age (which we divide into five roughly equal size bins that are less than 60, 60-64, 65-69, 70-74, and 75+). This information almost uniquely identifies the price a policyholder would be charged for any given plan.[19]

A second set of covariates controls for four aspects of the policy that affect the quantity of insurance in the policy and that can therefore be used to test for a positive correlation between the amount of insurance coverage and nursing home utilization.[20] These are: (1) the deductible, (2) the maximum daily benefit, (3) the total number of days for which benefits may be received in the lifetime of the policy ("benefit period"), and (4) how the nominal maximum benefit increases over time ("benefit escalation"). The positive correlation property predicts that the hazard rate should be increasing in the benefit amount,

[18] Treating terminated policies as censored at the date of termination is equivalent to a competing risks framework in which the two risks (termination and failure) are assumed independent. This may not be an appropriate assumption. We ascertained that our results were not sensitive to instead keeping terminated policies in the "at risk" sample through the end of the observed sample period (or even beyond).
[19] Technically, price may vary with issue age more finely than in five-year intervals. Sample size considerations suggest the use of coarser issue-age categories.
[20] For completeness, a third set of covariates controls for the remaining features of the policy for which asymmetric information theory makes no strong prediction about their sign. These are described in the notes to Table 2.

the benefit period and the amount of benefit escalation, all of which increase the amount of insurance in the contract; similarly, the hazard should be decreasing in the size of the deductible, which reduces the amount of insurance in the contract.

We measure the deductible with indicator variables for 20-day, 60-day and 100-day deductibles. We measure the daily benefit amount using three roughly-equal sized indicator varies for less than \$100, \$100, and more than \$100 per day.[21] In measuring the benefit period, we create a series of indicator variables that take account of two factors. First, we distinguish among policies with benefit periods of 1-4 years, 5+ years (but finite), and unlimited. Second, among policies with finite benefit periods, we further distinguish policies that reset the allowable benefit period to the original benefit period if the individual has had 180 continuous days since the last day of receiving covered care; this reset option effectively extends the benefit period. Finally, we use indicator variables for the four possible benefit escalation options: constant nominal benefits, benefits escalate at 5 percent of the original benefit per year ("simple" escalation), benefits escalate at 5 per year ("compound" escalation), and benefits are increased by the greater of 5% compounded annually over 3 years or CPI-growth over the last3 years at the option of the policy holder ("indexed").

Table 1 provides summary statistics on the main individual and policy characteristics examined in the analysis. We do not control for the premium because we have controlled for all of the characteristics of the individual and the policy that determine it. We also do not control for gender because it is not used in determining the pricing of contracts.

<u>3.2.2 Results</u>

Table 2 reports the results from estimating equation (4). The first column shows results for the entire sample. Since some of these policies have been in effect for only a short time, in the second column we also report results for the subset of policies issued in 1997 or 1998, all of which have had at least three

---

[21] We do not include the daily benefit for home health care separately because the correlation between the two daily benefit is very high (0.92). We do include an indicator for whether the policy pays lower home health care benefits than nursing home benefits. Twenty percent of policies pay the same amount for both types of care, while the remainder cap home health care costs at either 50 or 80 percent of the nursing home limit.

years of exposure. The results are substantially unchanged.[22]

The results in the top half of the table show the estimated coefficients on several covariates that reflect the insurance company's risk-categorization of the individual. As expected, the hazard rate increases monotonically with the individual's issue age and assessed risk category.

The lower portion of the table reports the coefficients on covariates for which the positive correlation property makes predictions. There is little evidence in support of these predictions. The coefficients on the benefit escalation and benefit period variables all have the opposite sign from what is predicted by the positive correlation property. The coefficients on the deductible and daily benefit variables are positive as predicted (those with shorter deductible periods and higher daily benefits are more likely to use services) but are not statistically different from zero. More importantly, their magnitudes suggest that any effect is quantitatively unimportant. For example, the change in hazard rate associated with a 20-day deductible compared to a 100-day deductible (which is the largest right-signed coefficient) is not only statistically insignificant but only about half the magnitude of the change in the hazard rate associated with being rated standard risk instead of high risk; it is considerably smaller in magnitude than the change in hazard associated with any 5-year increase in issue age.

One potential concern with these findings is that our inclusion of a series of indicator variables for the individual's age and rating category may produce misleading estimates of the relationship between features of the contract and nursing home utilization if this relationship differs for individuals in different risk categories. We therefore estimated a more flexibly specified version of equation (4) in which we included fixed effects for each risk class; we defined the risk class either by the individual's issue age category and rating category, or by the issue age category, rating category and issue year. We also re-estimated the hazard model restricting our sample to an increasingly homogenous population with respect to the insurance company's risk categories. For example, we limited the sample to the two-thirds to three-quarters who are rated standard risk, and we also tried further restricting the sample to the approximately

[22] The results are similarly unchanged if we limited the sample to individuals who are 75 and older at the time of purchasing the policy, and for whom we therefore observe a greater fraction of the policies' actual lifetime.

one-fifth of the original sample who are rated standard risk *and* purchased the policy between ages 65 and 69 (the largest age/rating combination in the data). We found (in results not shown) that the coefficients on the policy characteristic variables in these alternative specifications were, if anything, less consistent with the predictions of the positive correlation property than those shown in Table 2.

3.3 Evidence from the individual panel data in the AHEAD

The proprietary insurance company data provide detailed information on the relationship between the amount of insurance and subsequent claims. However, they contain no comparative information on the experience of those without private insurance. Our final data set provides this comparison.

3.3.1 Data and empirical framework

We use the Asset and Health Dynamics (AHEAD) cohort from the Health and Retirement Study (HRS). This sample, first interviewed in 1993, was designed to be representative of the non-institutionalized individuals born in 1923 or earlier and their spouses. Because the first wave of the survey does not provide a reliable measure of long-term care insurance coverage, our analyses begin with the second interview in 1995. We use the panel nature of the data to track these individuals' nursing home utilization through the latest currently available interview in 2000. Appendix A provides more detail on the sample and variable definitions used.

The basic estimating equation is exactly that shown in equation (1a). We regress a measure of the individual's long-term care utilization from 1995 through 2000 (CARE) on whether he has long-term care insurance coverage in 1995 (LTCINS); 10% of the sample has long-term care insurance coverage. We include as controls a series of covariates (X) designed to control for any categorization of the individual done by the insurance company.

We use two different measures for the dependent variable CARE. The first is a binary measure of whether the individual spent any time in a nursing home in the five years between 1995 and 2000 (mean is 0.19). The second measure is the total number of nights that the individual spent in a nursing home in this period (mean is 33 nights)[23].

As discussed, the correct empirical test requires controlling for the “categorization” of the individual done by the insurance companies. This categorization depends both on the characteristics of the individual observed by the insurance companies and on how finely they make distinctions among individuals based on the observed information. To determine the characteristics observed by the insurance company, we collected insurance applications from five leading long-term care insurance companies.[24] All companies collect a limited set of demographic information: age, gender, marital status, and age of spouse. They all collect similar and extremely detailed information on current health and on health history. The only noticeable difference across companies is that we found only one company that asked applicants to report any financial information (specifically, whether they had less than $30,000 in financial assets).

Despite the very detailed medical information collected, companies, as discussed, offer age-specific prices with only two or three broad rate classifications within each age based on the medical information (ACLI 2001, Weiss 2002). This same practice was in place in the late 1980s, when individuals in our data may have been purchasing insurance (Kemper et al. 1995).[25] Therefore, while the very rich information in the AHEAD about the individual’s current health and medical history allows us to measure essentially all of the information observed by the insurance companies, we do not actually observe the individual’s risk classification directly, as we do in the proprietary data.

Given the importance of controlling for the individual’s risk classification in the analysis, we investigate the sensitivity of our findings to four alternative approaches. First, we do not include any covariates (X) in estimating equation (1a) (“no controls” specification). Second, we control for the one measure that insurance companies all appear to use in risk classification – the individual’s age – by

---

[23] For those who spent any time in a nursing home over this period, the average amount of time was 187 nights.
[24] John Hancock, UNUM, CNA, TIAA-CREF and CALPERs.
[25] Conversations with actuaries suggest that 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.

including a separate indicator variable for each age ("age dummies" specification).[26] Both of these approaches are likely to underestimate the amount of categorization done by insurance companies.

As a likely overestimate of the risk classification done by the insurance company, our third approach ("all observables" specification) tries to control for everything the insurance companies observe about the individual, even though we know that most of it is not used in pricing. This specification therefore includes not only the age dummies, but also all of the demographic information that insurance companies observe, (gender, marital status and age of spouse, which we enter linearly), and indicator variables for each of the detailed current health and health history characteristics collected by any insurance companies that we can measure in the data. To be conservative, we also include indicator variables for the household's income quartile and asset quartile, even though it appears that most companies do not collect this information. This complete set of controls is summarized in Table 3 and described in more detail in Appendix A.[27]

The "all observables" specification invokes a much more finely defined categorization of risk than insurance companies actually do. However, by merely including each observed characteristic as an additive control we may misestimate the true relationship between insurance coverage and care utilization if there are substantial interaction effects among these controls. Our final alternative specification therefore substitutes these linear controls with a single summary measure of the insurance companies' prediction about each individual in the AHEAD's chance of entering a nursing home in the next five years. These predictions were generated from an actuarial model of nursing home use that is widely used in the long-term care insurance industry.[28] They 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

---

[26]At the tails of the age distribution we use catch-all indicators for "younger than 70" and "older than 90".
[27] Despite the very detailed health information in the AHEAD, there are a few rare health conditions that are not measured by the AHEAD but that insurance companies sometimes collect (e.g. "unoperated aneurysm," "double amputation"). We therefore experimented with including additional health measures observable in the AHEAD and not in the insurance applications and did not find any notable changes in our results.
[28] We use a version of the model that predicts care utilization for typical individuals in the population. For more details on the model and its pedigree, see Brown and Finkelstein (2003), Robinson (1996), or Robinson (2002).

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.[29]

3.3.2 Results

Table 4 describes the results of estimating equation (1a) for these four alternative definitions of the control variables (X). When the dependent variable is the binary measure of any nursing home use, we report results from OLS estimation of equation (1a); probit estimation produces similar results. When the dependent variable is the cumulative number of nights spent in a nursing home since 1995, we report estimates from a Tobit model; a linear model produces similar results, as does a linear estimate of the number of nights spent in a nursing home among those who report any nursing home stays.

The results are not supportive of a positive correlation between long-term care insurance coverage and long-term care utilization. Instead, in all but one specification, long-term care insurance coverage is *negatively* associated with long-term care utilization; in the one case in which the relationship is positive, it is substantively and statistically indistinguishable from zero. Across all specifications, we can reject a higher probability of nursing home utilization for the insured relative to the uninsured of more than 3 percentage points with 95 percent confidence.

We subjected these results to a number of specification tests. First, we ascertained that these results are robust to adding the insurance company's prediction as an additional control variable in the "all observables" specification. Second, as an alternative way to deal with potential non-linearities in the relationship between observe characteristics and care utilization, we re-estimated equation (1a) restricting the sample to individuals who are more homogenous with respect to their risk category. In particular, we

[29] When the dependent variable is the number of nights spent in a nursing home over the subsequent five years, we would like to control for the actuarial estimate of the individual's expected number of nights in a nursing home over the next five years, not the probability that the individual will enter a nursing home in the next five years, as we currently do. The actuarial mode allows us to generate this prediction regarding expected number of nights in a nursing home and we plan on using this in future versions.

restricted the sample to the most common two-year age range (ages 73 and 74 which constitute about 15% of the sample) and we further restricted the sample to the healthiest individuals (those with no ADLs, IADLs, or cognitive impairment) within this two-year age range (12% of the total sample). The relationship between insurance coverage and long-term care utilization was negative with all of these sample restrictions. Third, to address the possibility of attenuation bias due to measurement error in long-term care insurance coverage, we ascertained that the results were not affected by instrumenting for long-term care insurance coverage reported in 1995 with long-term care insurance coverage reported in 1998, the next available measure. Fourth, because we do not observe care consumption 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 experimented with several approaches to try to gauge the sensitivity of our findings to this possibility. For example, 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 and we defined the dependent variable based on care utilization from 1992 through 2000 (thus observing 8 years of information rather than only 5).[30] We also tried limiting the sample to the one-third of individuals who died between 1995 and 2000, and for whom further care consumption was therefore not possible. None of these alternative specifications affected the qualitative nature of the results.

A final – and potentially most important –set of specification checks concerned the sample of individuals analyzed. A concern with the results in Table 4 is that a substantial fraction of the seemingly uninsured may in fact rely on the public insurance provided by Medicaid, which pays for 40% of all nursing home expenditures (US Congress, 2000). Since Medicaid coverage requires a deductible of almost all of one's assets (AARP 2000), Medicaid is a more attractive substitute for private insurance for lower-wealth individuals, and may therefore induce selection based on wealth levels. Moreover, some of those without private insurance may experience moral hazard effects from public insurance, which could bias downward our estimate of any moral hazard effects of private insurance. To account for these

[30] 1992 is the furthest back that nursing home care utilization is available since the first interview wave, in 1993, asked about nursing home use in the preceding 12 months.

possibilities, we repeat the regressions shown in Table 4, restricting the sample to individuals who are in the top quartile of the household income or wealth distribution in 1995, and who are therefore least likely to find Medicaid an attractive substitute for private insurance. The top panel of Table 5 indicates that the relationship between insurance coverage and care utilization appears *more negative* when the sample is restricted to these individuals. Indeed, across all specifications, we can now reject a higher probability of nursing home utilization for the insured relative to the uninsured of more than 0.7 percentage points with 95 percent confidence.

Another concern with the sample definition is that many individuals in our sample would have been denied insurance coverage because of a health condition, had they applied. The application forms from long-term care insurance companies explicitly discourage individuals with certain health conditions from applying; Figure 4 provides illustration using the application forms of two of the top-10 long-term care insurance companies.[31] We therefore repeat the regressions from Table 4 for a restricted sample of individuals who are unlikely to be classified as ineligible; Appendix A describes the criteria for constructing this sample. The middle panel of Table 5 shows the results based on restricting the sample to the eligible population. The bottom panel shows the results when the sample is restricted to those who are *both* eligible *and* unlikely to view Medicaid as a good substitute for private insurance. The coefficients on long-term care insurance remain consistently negative across specifications.

3.4 Insurance coverage and home health care utilization

The analysis thus far has focused exclusively on the relationship between long-term care insurance coverage and nursing home care utilization. However, increasingly, long-term care insurance policies also cover home health care. We therefore examine – where the data permit – the relationship between insurance coverage and other measures of care utilization (results not shown).

Only two of our data sets allow some measure of long-term care utilization other than nursing home

---

[31] In Section 6, we discuss how this denial practice may reflect that unraveling of the asymmetric information equilibrium for (observably) high risk individuals, and how preference-based selection may increase the likelihood of such unraveling.

utilization. In the AHEAD, we can measure whether the individual consumed any nursing home *or* any home health care between 1995 and 2000 (40% of the same did). We cannot measure home health care use separately since, in follow-up interviews, individuals who are currently in a nursing home are not asked if they had used home health care since the previous interview.

We re-estimate equation (1a) using as the dependent variable whether the individual used *any long-term care*. The relationship between insurance coverage and *any care utilization* is slightly <u>more negative</u> than the relationship between insurance coverage and *any nursing home utilization*. This finding persists if we restrict the insured sample to the two-thirds whose policies provide some home health care benefits. This suggests that the relationship between insurance coverage and home care is also likely to be negative. Indeed, when we restrict our sample to the approximately 80% of individuals who do not use nursing home care – and for whom we therefore observe an accurate measure of home health care utilization – the relationship between insurance coverage and home care utilization appears more negative than the relationship between insurance coverage and nursing home care shown in Tables 4 and 5.

In the proprietary data, we directly observe home health care utilization. We therefore re-estimate equation (4) with the failure defined as the 100^{th} consecutive day of home health care; this failure rate is 0.2%. Here, the statistically significant coefficients on the deductible and maximum daily benefit amount variables are of the sign predicted by the positive correlation property. There is also some (weaker) evidence of the positive correlation property on the benefit period variables. The coefficients on the benefit escalation variables, however, tend to be wrong signed. If we define the failure as the 100^{th} consecutive day of *any care*, there is even weaker evidence of the positive correlation property – existing mainly on the coefficient on the daily benefit amount – which reflects the fact that the results for nursing home failure tend to be the opposite of what is predicted by the positive correlation property.

The evidence from the AHEAD and the proprietary data therefore point to opposite conclusions, and together are inconclusive about the relationship between insurance coverage and home health care utilization. We suspect that if there is a positive relationship between insurance coverage and home health care utilization, this most likely reflects a larger role of moral hazard for care that is not in a nursing home; home care may provide some net consumption value and therefore be more susceptible to moral hazard than the less desirable nursing home care. Because policies that cover both types of care offer similar benefits for each type of care, we do not think a differential selection explanation is compelling.

Another possible interpretation of the weak evidence of a positive correlation between insurance coverage and home health care utilization is that long-term care insurance induces greater use of home health services, which in turn substitute for the (less desirable) nursing home care and explains the lack of a positive correlation between insurance coverage and nursing home care. We consider this interpretation unlikely, however. We ascertained that the negative relationship between the quantity of insurance coverage and nursing home care in each data set persists when we restrict the insured sample to those with no home health care coverage.[32] This is consistent with quasi-experimental evidence that home health care does not appear to substitute for nursing home care (Kemper (1988), McKnight (2002)).

4. The structure of information in the private long-term care insurance market

The previous section indicated that there appears to be no positive correlation between insurance coverage and nursing home utilization, raising the possibility that asymmetric information may not exist in the private long-term care insurance market. In this section, we test directly for asymmetric information by examining whether individuals have any private information about their risk type, conditional on the risk classification done by the insurance company.

4.1 Data

Information on individual 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 and 1. The question was not asked of the approximately 13 percent of the 1995 respondents for whom the

[32] In the SOA data, this requires no restrictions since essentially all of the exposure meets this criterion (SOA, 2002).

interview was completed by a proxy respondent; this excludes, among others, the most cognitively impaired.[33] The results in Tables 4 and 5 are robust to this sample restriction.

An important consideration is whether individuals' reporting of their beliefs contains any meaningful information about their actual beliefs. Several factors are encouraging on this dimension. First, individuals' predictions appear right on average: the average self-reported probability was 18 percent, and 16 percent of the responders enter a nursing home over the next five years.[34] Second, we find that self-reported nursing home entry probabilities co-vary in consistent 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 quite evident in our data. This is the problem of "focal responses" wherein respondents give round figures such as 0, 50 or 100 percent (Hurd and McGarry, 1995; Gan et al. 2003). Figure 5 shows a histogram of the responses; almost 50% of responders reported a probability of zero, 14 percent report a 50 percent probability, and about 1 percent reported a probability of 100 percent.

It is somewhat unclear how to treat the issue of focal responses – or other forms of reporting error more generally – in individual responses. Our goal is to measure individual beliefs. To the extent that the focal responses represent the "true" subjective probability of the individual, correcting for what mistakenly appears to be measurement error will tend to overstate the predictive power of the individual. However, if the responses capture with error the individual's true subjective probability, this will produce attenuation bias in our estimation of individuals' private information. A final possibility is that the

[33] The response rate among those asked is high (95%).
[34] The accuracy of the average prediction holds for both men and women. We find some evidence that those with insurance and those in better health tend to overestimate their risk. It is unclear, however, to what extent the apparent underestimation of risk by those in poor health is driven by the tendency of individuals to give focal responses, particularly 0 or 0.5.

preponderance of zero-responses may reflect the fact that individuals are not comfortable thinking in terms of probabilities. To the extent that they also do not think about their risk probabilities in purchasing long-term care insurance, this is not a problem for our analysis. However, if somehow they latently use probabilistic information without realizing it, this will lead to an underestimate of the extent of private information.

Given these issues, we adopt three alternative approaches to measuring the individual's beliefs. First, we use the unadulterated response of the individual for the individual's beliefs. Second, we add an indicator variable for whether the individual reported zero. Third, to try to correct for any classical measurement error in the self-reported probabilities, we instrument for individuals' self-reported probability in 1995 using their answer to the same question in 1993.[35] To the extent that there is idiosyncratic, classical measurement error, lagged self-reported beliefs may be a useful instrument. Of course, if there is a large individual-specific measurement error component to self-reported beliefs (such as a misunderstanding of probability), lagged beliefs will be a less useful instrument.[36]

4.2 Results

The estimating equation was shown in equation (3). We estimate a linear probability model of whether the individual went into a nursing home in the five years between 1995 and 2000 on his 1995 self-reported beliefs of this probability (B) and controls for the risk classification done by the insurance company (X). The results are shown in Table 6. We report results separately for the three alternative measures of the individual's beliefs described above, and the four alternative definitions of risk classification used in Tables 4 and 5.

Two basic findings emerge across all three measures of the individual's beliefs. First, as shown in the first three columns, individual beliefs about the likelihood of entering a nursing home are a significant predictor of subsequent nursing home experience. This provides a complement to studies that have found

---

[35] The first stage coefficient from a regression of 1995 prediction on 1993 prediction is 0.34; t stat = 15.9; $R^2 = 0.09$.
[36] There does not seem to be a large individual-specific component to giving a focal response. Although 64 percent (71 percent) of respondents give a focal response of 0, 50 or 100 in 1995 (1993), the correlation between giving a focal response in 1993 and giving a focal response in 1995 is only 0.23

that individuals have some ability to predict their mortality risk (see e.g. Hurd and McGarry, 1995, 2002; Smith et al. 2001). The strength of this predictive power, however, varies substantially with the measure of beliefs. For example, column 1 indicates that a 10 percentage point increase in self-reported probability is associated with only a 1 percentage point increase in the probability of going into a nursing home (a coefficient of 1 would indicate perfect predictive ability). By contrast, the estimates using the instrumented measure of the individual's prediction (column 3) suggest substantially greater predictive power for the individual; a 10 percentage point increase in self-reported probability is now associated with a 4.6 percentage point increase in the probability of going into a nursing home.

As discussed, it is unclear whether instrumenting to correct for measurement error is appropriate, given that our desire is to best measure the individual's true beliefs rather than the true probability. If instrumenting is appropriate, however, its substantial impact raises the possibility that the availability of better instruments might produce evidence of even better predictive power on the part of the individual.[37] We investigated this issue by experimenting with alternative sets of instruments for 1995 beliefs about nursing home entry; we used as instruments various combinations of the individual's beliefs in 1995 about the probability that he will need financial help from his children, that he will die within five years, or that his medical expenses will exceed his savings, as well as the original instrument based on 1993 beliefs about nursing home entry. In general, the results reported in column (3) lie within the range of estimates of the predictive power of individuals' beliefs obtained through the use of alternative sets of instruments.

Second, a comparison of the results in the left hand panel (no controls) with those in the other three panels indicates that, although the risk categorization done by the insurance company reduces the predictive power of the individual's information about his risk type in about half, the individual still has residual private information about his risk type *conditional on the risk class that the insurance company*

---

[37] Another interpretation is that – given the large number of focal responses – self-reported probability may be closer to a categorical variable than a continuous variable, in which case the IV estimator may be biased upward (Kane et al. 1999). We plan on investigating alternative approaches to dealing with potential non-classical measurement error in a next draft.

assigns to the individual. This result holds for all definitions of the individual's beliefs and all definitions of the risk classification done by the insurance company. In results not reported, we ascertained that these results were not sensitive to a number of alternative specifications, such as adding the actuarial prediction to the "all observables" specification. We also verified that restricting the sample to individuals who are more homogeneous with respect to their risk category (i.e. similar ages and/or similar health status), to individuals likely to be eligible for insurance, or to individuals in the top quartile of the income or asset distribution did not affect the qualitative results in Table 6; the coefficient on the individual's beliefs was not sensitive in magnitude but was sometimes no longer statistically significant in these smaller samples.

These findings provide direct evidence of the presence of asymmetric information in the private long-term care insurance market. More specifically, they provide direct evidence of the assumption of adverse selection models that individuals have private information about their risk type. An alternative interpretation of the results in Table 6 might be that individuals and insurance companies initially have symmetric information but that individuals, in reporting their beliefs, anticipate the moral hazard effects of insurance. However, this is not corroborated by the data; the finding of residual private information on the part of the individual is unaffected by restricting the sample to the 90% of uninsured individuals. Interestingly, we do not find evidence that the insured are better predictors of their utilization than the uninsured. This suggests that individuals do not appear to update their beliefs about their risk type based on the price offered by the insurance company.[38]

The results in the last panel of Table 6 that control for the insurance company prediction provide some additional insight about the structure of information in this market. Interestingly, the results suggest that it may be the insurance company who is the better predictor, presumably because its superior

[38] We also investigated whether predictive power varies systematically across other observable groups. More educated and older individuals tend to be better predictors; there is weak evidence that women may be better predictors than men. We found no evidence that eligible individuals are better predictors than ineligible individuals, or have greater residual private information.

forecasting ability can compensate for its coarser information set.[39] In addition, the fact that when both the individual and the actuarial prediction are included on the right hand side, each remains statistically significant, but the predictive content of the self-reported probability attenuates substantially, suggests that the insurance company and the individual form their predictions based, to some extent, on complementary rather than overlapping information. Consistent with this, the correlation between the two predictions is only about 0.1. As a result, *despite* the fact that the insurance company may be better able to predict the individual's risk type than the individual, the individual appears to still have residual private information.[40]

If individuals have residual private information about their chances of using a nursing home, this raises the interesting question of why insurance companies do not collect additional information about the individual. Some of the information – such as how much the individual dislikes nursing homes or how likely their spouse or children are to take care of them in the home – may simply not be feasible for the insurance company to observe. However, there is some information about the individual that the insurance company could in principle observe but that in practice it does not. These include, for example, additional health conditions measured in the AHEAD data but not by the insurance company as well as measures of the individual's race, religion, education, spouse's health, the number, gender, and proximity of the individual's children, and indicator variables for whether the individual engaged in each of a variety of potential preventive health measures (described in more detail in Section 5). We found that the inclusion of these variables as additional controls in the "all observables" specification of Table 6 does

[39] Measurement error in individual's beliefs is of particular concern when comparing the predictive power of individual's beliefs to that of the actuary, because the latter is based on aggregate data and therefore presumably has less measurement error. However, even when we instrument for individuals' beliefs (column 12), the individual is not a better predictor than the actuary.

[40] In results not reported, we also find that the predictive power of the full actuarial model is almost identical to that of a restricted model in which actuarial predictions are constrained to be the same across gender. The residual private information of the individual, conditional on the actuarial prediction, is also similar whether we control for a gender-specific or unisex actuarial prediction. This may help explain why insurance companies do not use gender in pricing insurance.

not affect the magnitude or statistical significance of the individual's residual private information. However, an F-test on the added variables indicates that they are jointly statistically significant.

These results therefore suggest that feasible collection of additional information about the individual would not help the insurance company vis a vis the consumer's information set, but it would give it an advantage over competitor insurance companies that do not collect the information. Insurance companies may not collect this information because the costs of doing so are high relative to the benefits, or because they are able to sustain a collusive equilibrium in which they agree not to collect this (costly) information. In addition, to the extent that some of the characteristics the insurance company might collect reflect the outcome of behavioral choices (such as decisions regarding preventive health care investment like flu shots and mammograms), insurance companies' use of these characteristics in pricing insurance contracts could affect individuals' behavioral choices and thus reduce the informative content of these characteristics.

5. Unobserved characteristics, insurance coverage, and care utilization

In this section, we investigate how to reconcile the direct evidence of asymmetric information in Section 4 with the evidence in Section 3 that there is no positive correlation between long-term care insurance coverage and nursing home utilization.

5.1 The relationship between individual beliefs and insurance coverage

First, we show that individuals who believe that they are higher risk are more likely to purchase insurance. Specifically, we estimate:

(5) $\mathrm{LTCINS} = X\delta_1 + \delta_3 \mathrm{B} + \mu$

where B is once again the individual's beliefs about his chances of going into a nursing home and X is a series of controls for the risk classification done by the insurance company. The results are shown in Table 7 and indicate $\delta_3 > 0$. Individuals who believe that they are higher risk are, controlling for the risk categorization done by the insurance company, more likely to purchase insurance. We consider the most natural interpretation of this result to be that it is evidence of risk-based selection. However, we cannot rule out the possibility that the causality goes the other way and that the result reflects individuals' rational incorporation of the moral hazard effects of insurance into their perceptions of their chances of using care.

The combined evidence from Tables 6 and 7 thus indicates that individuals have private information about their risk type *and* that this private information is positively correlated with insurance coverage. Therefore, to explain the lack of a positive correlation between insurance coverage and care utilization, there must be some other unobserved characteristic of the individual (which we denote by P for the individual's unobserved preferences) that has the opposite correlation with insurance coverage and care utilization.

To demonstrate this, we decompose LTCINS from estimating equation (5) into a component explained by the individual's residual private information about his risk type $(\hat{\delta}_3 B)$, which we refer to as RISKTYPE_HAT, and the residual (RESID_HAT) from equation (5) which is the unexplained portion of the individual's long-term care insurance coverage. We then estimate:

(6) $\text{CARE} = X\beta_1 + \lambda_1 \text{RISKTYPE\_HAT} + \lambda_2 \text{RESID\_HAT} + \varepsilon$

The results are shown in Table 8.[41] For comparison, we also report the estimated coefficient on LTCINSURANCE from estimating equation (1a) in which LTCINS is included directly on the right hand side of equation (6), rather than it being decomposed into its different components.[42] The coefficient on RISKTYPE_HAT is positive and statistically significant in all specifications; the portion of insurance coverage that is correlated with individuals' private information about risk type is positively correlated with the probability of using a nursing home. However, the coefficient on RESID_HAT is always negative; the portion of insurance coverage that is not explained either by the individual's private information about risk type or by the categorization done by the insurance company is negatively

[41] Because the individual beliefs are about the five-year entry probability, we only report results where "nursing home entry" is the dependent variable; results for length of nursing home stay, however, are quite similar.
[42] The results from estimating equation (1a) that are reported in Table 8 differ very slightly from those reported in Table 4 due to differences in sample size (the sample is now restricted to the approximately 87% who are asked to assess their nursing home risk).

correlated with nursing home use. In other words, there are characteristics of the individual not controlled for in equation (6) that have the *opposite* correlation with insurance coverage and care utilization. We ascertained that these results are not substantively sensitive to restricting the sample to eligible individuals or to individuals in the upper quartile of the income or asset distribution, although the coefficient on RISKTYPE_HAT is often not statistically significant in these smaller samples.

5.2 Evidence of preference-based selection

Finally, we explore directly what these other unobserved factors – that have the opposite correlation with risk occurrence and preferences for insurance – might be. Imagine we can measure two aspects of the individual that are unobserved by the insurance company, and related (with noise) to the individual's (unobserved) type T: his beliefs about his type (B) and some aspect of his preferences (P) that the insurance company does not observe. We have 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.[43] We therefore estimate:

$$(7) \qquad \text{CARE} = Xb_1 + \text{b}_2\text{LTCINS} + b_3\text{B} + b_4\text{P} + \varepsilon$$

$$(8) \qquad \text{LTCINS} = Xd_1 + d_3B + d_4P + \eta$$

and look for variables to measure P that produce the opposite sign on b_4 and d_4.

It is difficult, if not impossible, to measure all (or even most) of these unobserved characteristics. By definition, they must be unobserved by the insurance company; many of them are therefore likely to be unobserved by the econometrician as well. We focus on implementing a test for one particular type of preference-based selection that has attracted considerable theoretical attention. De Meza and Webb (2001) and Jullien et al (2002) propose models in which individuals have private information about their

[43] The fact that P is correlated with care utilization but not incorporated into the individual's beliefs about his care utilization (B) indicates that the individual does not use information about P in forming his beliefs. Note that this implies that the individual is inefficient in forming his beliefs; it does not imply that his beliefs must exhibit any systematic biases or mistakes. Consistent with this view, Smith et al. (2001) find that individual's self-reported *mortality* probability in the HRS does not reflect all of the private information individuals have about their survival prospects, even though it is a good predictor of actual mortality and is updated in sensible ways based on experienced health shocks.

risk aversion. The risk averse (or "cautious") individuals choose both higher levels of coverage and greater preventive effort; this greater preventive effort by those with higher coverage can reduce or reverse the prediction that those with more coverage should have more accidents.

The AHEAD data provide a nice 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. There is substantial variation in the fraction of gender-appropriate potential preventive activity actually undertaken: the median individual does two-thirds of these activities, but 7% report doing nothing and 30% report doing all.

Of course, this measure may reflect characteristics of the individual other than his level of caution. In particular, it might be substantially affected by whether the individual has been to a doctor or what his insurance will cover. Fortunately, over 99% of our sample is covered by Medicare, which reimburses for all of these preventive health measures. In addition, over 90% of the sample had a doctor's visit since the previous interview; the results are not sensitive to limiting the sample to those who had been to a doctor.

Table 9 reports the results. Each panel reports the results separately for a different set of controls for the insurance company risk classification.[44] The first column of each panel indicates that individuals who undertake a greater fraction of potential preventive health activity (i.e. more cautious individuals) are more likely to own insurance. The second column of each panel indicates that those who undertake more preventive health activity are also less likely to subsequently go into a nursing home.[45] Thus, consistent with the theoretical models of de Meza and Webb (2001) and Jullien et al (2002), more cautious

---

[44] We also include a control for gender since the fraction of potential preventive activity undertaken may vary with gender simply because the *number* of potential preventive activities is 3 for men and 5 for women. The results reported in Table 9 are not sensitive to whether we include this control.

[45] Ex ante, it was conceivable that this relationship might go the other way, as individuals who engage in more preventive health activity might be expected to live longer, and longevity is positively correlated with nursing home utilization. In practice, any longevity effect appears to be offset by an improved health effect. For example, we find that individuals who invest more in preventive health activity are also less likely to have had a fractured hip (which can be an important contributor to nursing home residence).

individuals are both more likely to own insurance and less likely to experience the insured risk.

We used the results from estimating the insurance coverage equation (8) to decompose insurance coverage into a component predicted by the preventive health activity (PREVENT_HAT), a component predicted by individuals' private information about risk type (RISKTYPE_HAT) and the residual (RESID_HAT).[46] The third column of each panel of Table 9 shows the results of estimating the care utilization equation (6) with these three different components of insurance coverage on the right hand side. Not surprisingly, given the previous results in Table 9, PREVENT_HAT is always negative; the variation in insurance coverage that is positively correlated with preventive health activity is negatively correlated with long-term care utilization.

We briefly explored whether we could identify other forms of preference-based selection that would tend to reduce the positive correlation between insurance coverage and risk occurrence that risk-based selection would otherwise produce. Of course, such an exploration is limited to characteristics that the insurance company does not observe (or does not use in risk classification). Many of these factors have the *same* correlation with insurance coverage and risk occurrence and would thus tend to reinforce a positive correlation. For example, individuals with less schooling or more children are both less likely to have insurance and less likely to use nursing home care. Similarly, nonwhites, Hispanics, and Catholics are each less likely to have insurance and less likely to use care, while Jews are both more likely to have insurance and more likely to use care.

However, we did find evidence that higher asset-individuals are substantially more likely to have insurance and less likely to use nursing homes.[47] Preference-based selection along wealth – which may be driven in large part by the crowd out of private insurance demand by Medicaid for lower-wealth individuals (see e.g. Brown and Finkelstein, 2003) – may therefore also contribute to the lack of a positive

---

[46] For this specification, we include separate indicator variables for each preventive health activity (and gender) in order to more flexibly estimate 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.

[47] Since we found only one insurance company who collected information on income and assets (and they only inquired whether financial assets were less than $30,000), it seems reasonably to look for potential preference-based selection along this dimension.

correlation between insurance coverage and care utilization. It is important to note, however, that even if we add all of these unobserved factors to the right hand side of equation (1a) in which we estimate the relationship between care utilization and insurance coverage, we still do not recover a positive relationship between insurance coverage and care utilization. This indicates that, not surprisingly, we are not able to fully measure all of the unobserved characteristics of the individual that have the opposite correlation with insurance coverage and with risk type.

6. Insurance rationing with risk-based and preference-based selection

As discussed above, private information about risk type can produce inefficient restrictions on the amount of insurance purchased by one or more risk types in equilibrium, even if preference-based selection undoes any positive correlation between insurance coverage and risk occurrence. The direct evidence of asymmetric information presented in this paper therefore suggests that asymmetric information may contribute to the limited size of the private long-term care insurance market. In this section, we briefly discuss how insurance appears to be rationed in this market along two dimensions, both of which may reflect the effects of asymmetric information. Of course, more work is needed to establish the extent to which this rationing is due to asymmetric information, as opposed to other factors, and we regard this as an important direction for further work.

First, policies tend to insure only a very limited fraction of long-term care expenditure risk. In addition to deductibles and limits to the total number of days an individual can receive benefits, most policies specify a maximum daily benefit that is substantially below the *current* daily cost of nursing homes (see Cutler 1996 and Brown and Finkelstein, 2003). Brown and Finkelstein (2003) present evidence that individuals who would not find the purchase of the currently-available limited-coverage policies welfare-enhancing might want to purchase policies with more comprehensive coverage. The fact that such contracts are not offered may reflect the effects of asymmetric information in this market, although other explanations for the structure of these contracts are possible (Cutler, 1996).

Second, many (observably) high-risk applicants would be denied insurance coverage at any price. This practice is non-trivial, and persists despite the lack of any regulatory restrictions on the level of prices or on actuarially-based pricing differentials. For example, Weiss (2002) estimates that 15% of non-group long-term care insurance *applications* are denied, while Murtaugh et al. (1995) estimate that up to one-quarter of 65 year olds and one-third of 75 year olds would be denied insurance based on their observable health characteristics if they applied. Such denials may reflect an adverse selection equilibrium in which there is no interior pricing solution for certain (observably) high risk groups (i.e. "unraveling").[48] Of course, other explanations for denials are possible, such as a lack of much residual variation in care utilization – and hence much insurance value – among the very high risk, or binding wealth constraints for the observably high risk.

Nevertheless, it is interesting to note that preference-based selection that tends to offset a positive correlation between insurance coverage and risk occurrence that would otherwise be produced by private information about risk may make it more likely for the market to unravel for a given (observable) risk class. Figure 6 illustrates this phenomenon graphically in a simple model with private information about risk type and linear pricing (e.g. Stiglitz and Weiss, 1981, Cutler 2002). C(p) denotes the expected claims over all those with risk type p or higher. If individuals differ only in terms of their unobserved risk type, the value of insurance at a given price (V_0(p)) is rising with risk type (p). The equilibrium shown in Figure 6 exhibits an interior pricing solution in which all those of risk type p* or higher buy insurance at price C(p*). Now imagine that we make the higher risk (unobservably) less risk averse. As we decrease the risk aversion of the higher risk individuals, the V(p) curve flattens from V_0(p) to V_1(p), and the market may unravel: as drawn, there is no longer any $\hat{p}$ at which the value of insurance (V_1(p)) at all $p > \hat{p}$ exceeds the expected claims for all insured individuals (C($\hat{p}$)).

7. Conclusion

A growing body of empirical work has begun to question the empirical relevance of theoretical models of asymmetric information to insurance markets. In several different insurance markets, recent

[48] Asymmetric information can cause a market to unravel completely, resulting in no trade in equilibrium (Akerlof, 1970). Cutler and Reber (1998) present evidence of such unraveling in a group private health insurance market.

papers have found no evidence of a positive correlation between the amount of insurance and the occurrence of the risk. In this paper, we show empirically that asymmetric information may exist *even if* the insured are not above-average in their risk type.

We explore these issues in the context of the private long-term care insurance market in the United States. We do not find a positive correlation between individuals' insurance coverage and their consumption of nursing home care in any of three complementary data sources. However, using information about individuals' assessment of their nursing home risk, we find direct evidence of asymmetric information. Even after conditioning on the information set of the insurance company, the individual's beliefs about his risk type are positively correlated with both insurance coverage and subsequent care utilization.

The lack of a positive correlation between insurance coverage and care utilization – *despite* the presence of private information about risk type – is explained by the existence of another type of private information: individuals have private information not only about their risk type but also about preference-related characteristics that have the opposite correlation with insurance coverage and risk occurrence. For example, we find evidence that more "cautious" individuals – as measured by their investment in preventive health measures, which is not observed by the insurance company – are both more likely to have long-term care insurance and less likely to use nursing home care.

Such preference-based selection can offset the tendency of asymmetric information about risk type only to produce a positive correlation between insurance coverage and risk occurrence. As a result, the insured risk pool is not "adversely-selected"; its risk is not above-average relative to the population. This suggests that the presence of private information about risk type in the long-term care insurance market does not raise prices above their population-average actuarially fair price. However, as discussed above, preference-based selection cannot offset the negative efficiency consequences of private information about risk type. The quantity of insurance is often restricted relative to the first best (see e.g. Chiappori et al. 2002), although models with overinsurance are also possible (de Meza and Webb, 2001). Further work is needed to try to quantify the insurance rationing that asymmetric information may produce, and its contribution to the extremely limited nature of the private long-term care insurance market.

It is particularly striking that preference-based selection appears to offset the positive correlation that asymmetric information would otherwise produce between insurance coverage and risk occurrence in the long-term care insurance market, given the extensive evidence of such a positive correlation in the market for acute medical care insurance (see e.g. Cutler (2002) for a review). It may be that preference-based selection plays more of a role in the long-term care insurance market, and or that moral hazard may play less of a role, than in the market for acute medical care insurance. Additionally, it seems likely that private information about risk type – and hence risk-based selection – may be greater in the market for acute medical care, where the individual may be continually learning more about his risk, than in the market for long-term care where the risk occurs less frequently and later in life.

The results in this paper suggest that preference-based selection may help explain the apparent differences across insurance markets in the presence or absence of a positive correlation between insurance coverage and risk occurrence. For example, there is evidence of such a positive correlation in annuities (Finkelstein and Poterba, 2000) but not in life insurance (Cawley and Philipson, 1999) which insures the (opposite) longevity risk to that insured by annuities. One potential explanation is that there are characteristics of the individual that the insurance company does not observe – such as their level of caution or their wealth – that are positively correlated with preferences for both types of insurance but are *negatively* correlated with the life insurance risk of dying and *positively* correlated with the annuity risk of living. We regard this as an interesting direction for further work.

REFERENCES

AARP. 2000. "Medicaid Financial Eligibility for Older People: State variations in access to home and community-based waiver and nursing home services." April.

AARP. 2002. "Long-Term Care Insurance: An Assessment of States' Capacity to Review and Regulate Rates." February, MS # 2002-02.

ACLI. 2001. Life Insurers Fact Book, 2001. Washington, DC

Akerlof, George. 1970. "The Market for "Lemons": Quality, Uncertainty and The Market Mechanism". *Quarterly Journal of Economics* (74): 488-500.

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/ecoadv.pdf

Brown, Jeffrey and Amy Finkelstein. 2003. "Why is the market for private long-term care insurance so small?" 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.

Cutler, David and Sarah Reber. 1998. "Paying for health insurance: the tradeoff between competition and adverse selection." *Quarterly Journal of Economics* 113 (2): pp.433-466.

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. 2002. "Asymmetric Information in Insurance: General Testable Implications." Unpublished mimeo.

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: A look Ahead." *Journal of Aging and Health.*

Crocker, Keith and Arthur Snow. 1986. "The efficiency effects of categorical discrimination in insurance markets." *Journal of Political Economy* Vol 4, No 2, pp.321-344.

Crocker, Keith and Arthur Snow. 2000."The theory of risk classification." In G Dionne (ed) Handbook of Insurance, Kluwer Academic Publishers.

Cutler, David. 1996. "Why Don't Markets Insure Long-Term Risk?" Mimeo.

Cutler, David. 2002. "Health Care and the Public Sector." In A. Auerbach and M. Feldstein, eds., Handbook of Public Economics, Volume 4 (Amsterdam: North Holland), forthcoming.

De Meza, David and Webb, David. 2001. "Advantageous selection in insurance markets." *Rand Journal of Economics*. Summer.

Dick, Andrew, Alan Garber and Thomas MaCurdy. 1994. Forecasting Nursing Home Utilization of Elderly Americans. In David Wise (ed) Studies in the Economics of Aging.

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.

Ettner, Susan. 1993. "Do elderly Medicaid patients experience reduced access to nursing home care?" *Journal of Health Economics* 12(3): 259-280.

Finkelstein, Amy and James Poterba, 2000. "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market." NBER Working Paper 8045, Cambridge, MA.

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.

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

Garber, Alan and Thomas MaCurdy. 1993. Nursing home discharges and exhaustion of Medicare benefits." *Journal of the American Statistical Association* 88(423): 727-736.

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.

Hirshleifer, J. 1971. "The private and social value of information and the reward to inventive activity." *American Economic Review*, 61, 561-574.

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

Kemper, Peter and Christopher Murtaugh. 1991. Lifetime use of nursing home care." *New England Journal of Medicine* 324: 595-600.

Kemper, Peter. 1988. "The evaluation of the national long term care demonstration: overview of findings." *Health Services Research* 23(1): 162-174.

Lakdawalla, Darius and Tomas Philipson. 2002. "The rise of old-age longevity and the market for long-term care." *American Economic Review* Vol 92, Number 1 pp. 295-306.

LIMRA. 2001. "Individual Medicare Supplement and Long-Term Care Insurance."

McKnight, Robin. 2002. "Home Care Reimbursement, Long-term Care Utilization, and Health Outcomes." MIT mimeo.

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 Center for Health Statistics, 2002. H*ealth, 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.

Pueltz, 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.

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.

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.

Smart, Michael. 2000. "Competitive insurance markets with two unobservables." *International Economic Review*, Vol 41, No 1. pp. 153-169.

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. 1996. "Long-term care insurance valuation methods; society of actuaries long-term care insurance valuation methods task force." *Transactions of the Society of Actuaries*, 1995. vol 47. pp.599-773.

Society of Actuaries. 2000. "Long-term care experience committee InterCompany Study 1984-1993." Available at http://www.soa.org/sections/ltc_experience.html.

Society of Actuaries. 2002. "Long-term care experience committee InterCompany Study 1984-1999." Available at http://www.soa.org/sections/intercompany_study.pdf.

Stiglitz, Joseph and Andrew Weiss. 1981. “Credit Rationing in Markets with Imperfect Information.” *American Economic Review* 71(3): 393-410.

US Bureau of the Census, 2000. Statistical Abstract of the United States 2000. (120th Edition). Washington DC.

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.

Villeneuve, Bertrand. 2000. “The consequences for a monopolistic insurance firm of evaluating risk better than customers: the adverse selection hypothesis reversed.” *Geneva Papers on Risk and Insurance Theory* 25: 65-79.

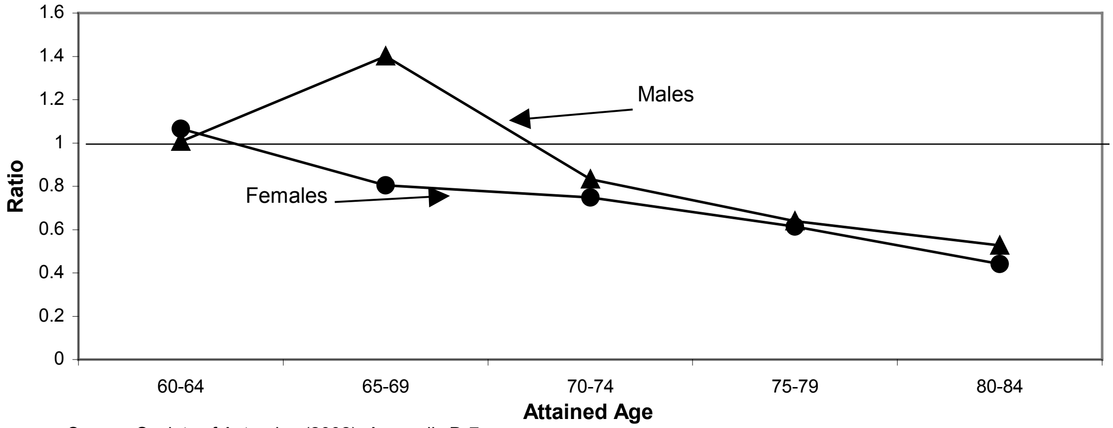
Villeneuve, Bertrand. Forthcoming. “Competition between Insurers with Superior Information” *European Economic Review.*

Wambach, Achim. 2000. “Introducing Heterogeneity in the Rothschild-Stiglitz Model.” *Journal of Risk and Insurance* Vol 67, No 4 pp.113.

Wiener, Joshua, Jane Tilly and Susan Goldenson. 2000. “Federal and state initiatives to jump start the market for private long-term care insurance.” *The ElderLaw Journal* 8(1): 57-99.

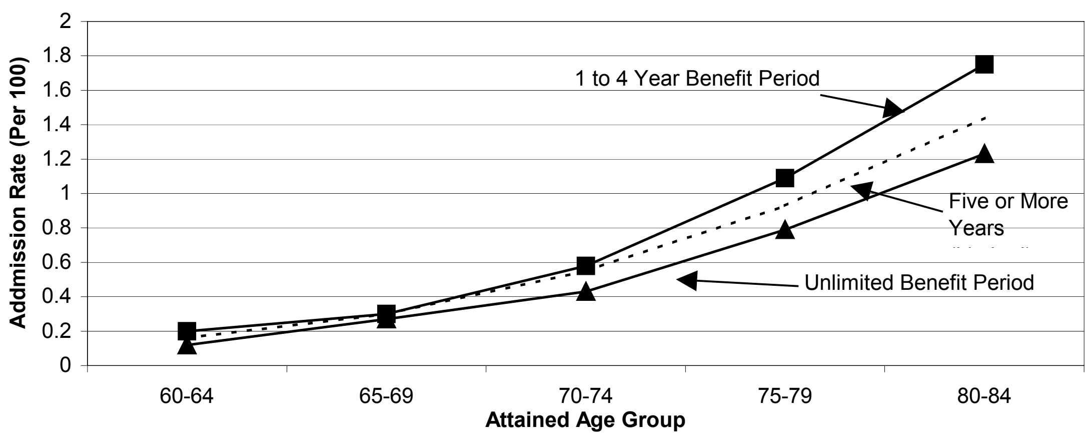
Weiss Ratings Inc. 2002. “Weiss Ratings’ Consumer Guide to Long-Term Care Insurance.” Palm Beach Gardens, Fl.

Figure 1: Ratio of Insured to Population Nursing Home Admission Rate



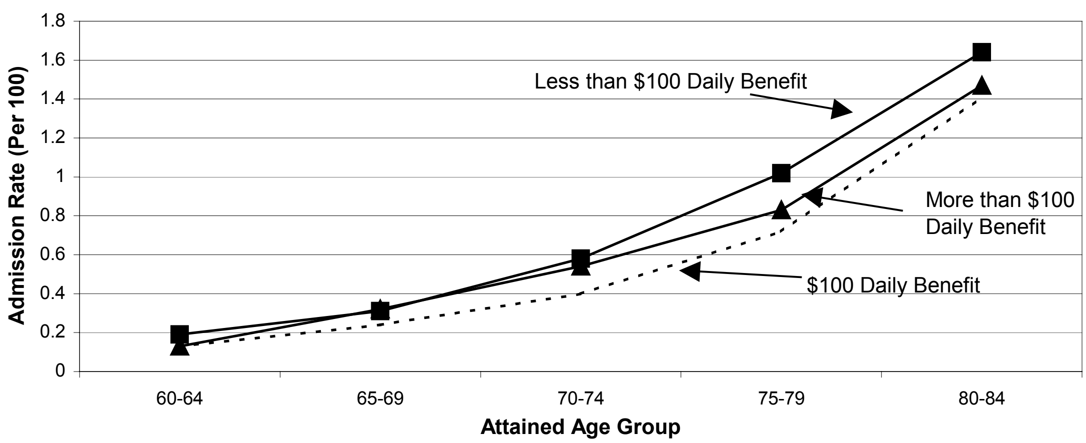
Source: Society of Actuaries (2002), Appendix D-7
Note: Insured estimates based on policies with no deductible

Figure 2: Variation in Nursing Home Admission Rate by Benefit Period



Source: Society of Actuaries, Table D-5
Note: Data reported for policies with 20-day deductible

Figure 3: Variation in Nursing Home Admission Rate by Daily Benefit



Source: Society of Actuaries, Table D-6
Note: Data reported are for policies with 20-day deductible

Table 1: Summary statistics for proprietary insurance company data.

	Policies Issued 1997-2001	Policies Issued 1997 or 1998
FAILURE RATE	0.3%	0.6%
CATEGORIZATION OF INDIVIDUALS		
Median Issue Date	September 1, 1999	February 8, 1998
Average Issue Age	64.4	65.3
% Rated Low Risk	29	16
% Rated Standard Risk	66	79
% Rated High Risk	5	5
CHARACTERISTICS OF POLICY		
Deductible		
% with 20-day deductible	5	5
% with 60-day deductible	8	8
% with 100-day deductible	87	87
Maximum Daily Benefit		
Average nursing home daily benefit (in $)	119	113
Average home health care daily benefit (in $)	112	103
% With Home Care Benefit Less than Nursing Home Benefit	19	26
Benefit Period		
% of policies with Unlimited Benefit Period	18	18
Average Benefit Period for Policies w/ limited benefit period	4.3	4.2
% of Policies with Limited Benefit Period that allow extension	10	13
Benefit Escalation		
% with no benefit escalation	2	----
% with 5% "simple" benefit escalation	30	29
% with 5% compound benefit escalation	28	18
% with "indexed" escalation	40	53
N	144,798	49,887

Table reports means. Percentages all rounded to nearest whole number. Dashed line indicates less than 1 percent. 60% of policy sales are to women; we do not report this in the above table since it is not a characteristic used by the insurance company to categorize individuals.

Table 2: Hazard rate of receiving nursing home care for 100th consecutive day

Sign predicted by the "positive correlation" property	Covariates in Regression	Policies Issued 1997 - 2001	Policies Issued 1997 or 1998
	Issue Age Categories (reference category = less than 60)		
	Issue Age 60-64	1.199***	1.039**
		(0.423)	(0.505)
	Issue Age 65-69	1.729***	1.798***
		(0.423)	(0.475)
	Issue Age 70-74	2.944***	2.928***
		(0.400)	(0.469)
	Issue Age 75+	4.010***	3.913***
		(0.403)	(0.473)
	Price categories (reference category = rated high risk)		
	Rated low risk	-1.100***	-0.964***
		(0.259)	(0.322)
	Rated standard risk	-0.535***	-0.562***
		(0.175)	(0.200)
	Deductible (Reference category = 100-day deductible)		
Positive	60-day deductible	0.024	-0.030
		(0.208)	(0.252)
Positive (and larger than 60-day deductible)	20-day deductible	0.233	0.312
		(0.238)	(0.268)
	Daily Benefit (Reference Category = < $100)		
Positive	Daily Benefit = $100	0.095	-0.007
		(0.127)	(0.141)
Positive (and larger than $100 category)	Daily Benefit > $100	0.240*	0.143
		(0.134)	(0.151)
	Benefit Period (Reference Category = 1-4 years w/ no extension possible)		
Positive	1-4 years w/ ext. possible	-0.306	-0.509**
		(0.207)	(0.254)
Positive	5+ years w/ no extension	-0.391**	-0.543***
		(0.162)	(0.193)
Positive (and larger than 5+ years w/ no extension)	5+ years w/ ext. possible	-0.160	-0.257
		(0.343)	(0.389)
Positive (and larger than 5+ years with extension)	Unlimited Benefit Period	0.168	0.075
		(0.153)	(0.175)
	Escalation of Benefits (Reference Category = 5% compound)		
Negative	No escalation of daily benefit	0.438	------
		(0.399)	
Negative (but larger than "no escalation")	5% "simple" escalation of daily benefit	0.213	0.270
		(0.244)	(0.302)
Larger than "no escalation"	"Index option" for escalation of daily benefit	0.102	0.254
		(0.236)	(0.288)
	Failure Rate	0.3%	0.6%
	N	144,798	49,888

Note: Table reports exponentiated coefficients from estimation of Cox proportional hazard model. Variables that are included in regression but not shown in table are: indicator variables for issue year, whether the policy is tax qualified, whether the home health care benefits are lower than (rather than equal to) the nursing home benefits, frequency of policy premium payments, and an indicator variable for a "shared care" rider benefit (which makes the spouse eligible for the policy benefits if the individual dies within a specified time period after policy issue).

Table 3: Summary Statistics for care utilization, insurance coverage, and covariates in the AHEAD

	Whole		Top quartile by inc or assets	
	Whole	Eligibles	Whole	Eligibles
Dependent Variables:				
Any NH Utilization (1995 – 2000)	0.187	0.113	0.130	0.089
Total # of nights in NH (1995 – 2000)	32.7	16.23	17.39	11.06
Key Independent Variable:				
Long-term care insurance coverage (1995)	0.104	0.109	0.155	0.155
Control Variables				
Demographics (1995)				
Age	77.5	76.4	76.2	75.5
Female	0.63	58.3	55.5	50.1
Married	0.54	0.60	0.74	0.77
Spousal Age if Married	73.8	73.5	73.6	73.7
Household Assets (median)	138,400	159,000	420,750	422,000
Household Income (median)	18,000	20,000	35,000	35,000
Current health (1995)				
ADL limitation: bathing	0.11	0	0.07	0
ADL limitation: eating	0.05	0	0.03	0
ADL limitation: dressing	0.13	0	0.09	0
ADL limitation: toileting	0.08	0	0.06	0
ADL limitation: walking	0.10	0	0.06	0
Incontinence	0.22	0	0.22	0
Cognitively impaired	0.03	0	0.02	0
Use wheelchair	0.03	0	0.02	0
Use walker	0.07	0	0.04	0
Use crutches	0.003	0	0.0009	0
Use Cane	0.13	0	0.09	0
Use oxygen	0.01	0	0.008	0
Regularly use prescription drugs	0.79	0.73	0.78	0.74
IADL limitation: grocery shopping	0.15	0.03	0.09	0.02
IADL limitation: managing medication	0.05	0.007	0.03	0.003
Low BMI	0.10	0.09	0.09	0.09
High BMI	0.13	0.11	0.10	0.09
Currently smoke	0.08	0.08	0.07	0.07
Health History (1995 and before)				
Home Health Care Use	0.17	0.07	0.11	0.06
Nursing Home Use	0.02	0.008	0.02	0.006
Depression	0.21	0.12	0.15	0.09
Drinking Problem	0.03	0.04	0.03	0.04
Diabetes	0.14	0.12	0.11	0.10
Diabetes treated with insulin	0.05	0.03	0.03	0.02
Kidney Failure Assoc w. Diabetes	0.02	0.005	0.007	0.002
Stroke	0.12	0.07	0.09	0.06
Heart condition	0.34	0.30	0.30	0.27
Medication for heart problem	0.22	0.19	0.19	0.18
Heart Attack	0.09	0.07	0.07	0.07
Congestive Heart Failure	0.04	0.02	0.03	0.02
High Blood Pressure	0.54	0.49	0.49	0.46
Hip fracture	0.05	0.03	0.04	0.02
Lung Disease	0.12	0.9	0.10	0.08
Cancer	0.16	0.14	0.17	0.15
Psychiatric problems	0.15	0.11	0.12	0.09
Arthritis	0.54	0.44	0.47	0.39
Injury from falling	0.15	0.10	0.14	0.10

Note: All means are weighted. See Appendix A for our construction of cognitive impairment, depression, drinking problem, household assets and BMI. All of the listed control variables are used in the "all observables" specification.

Table 4: Long-term care insurance coverage and long-term care utilization in the AHEAD

Dependent Variable	No Controls (1)	Controls for Age Dummies (2)	Controls for "all observables" (3)	Controls for insurance company prediction (4)
Any nursing home utilization	-0.045*** (0.016) [N=6,277]	-0.014 (0.015) [N=6,276]	0.0005 (0.015) [N=6,080]	-0.012 (0.015) [N=6,273]
Number of nights spent in nursing home	-71.44*** (25.79) [N=6,186]	-25.32 (25.11) [N=6,185]	-12.19 (25.29) [N=5,995]	-25.82 (24.76) [N=6,182]

Notes: Each cell reports the coefficient on LTCINS from estimating equation (1a) on a specific dependent variable and definition of the set of control variables. The column headings describe the set of control variables used. See text and Appendix for detailed description of these covariates. Coefficients in the first row are from a linear probability model. Coefficients in the second row are from a Tobit model. Heteroskedacticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 5: LTC insurance coverage and LTC utilization in the AHEAD; RESTRICTED SAMPLES

Dependent Variable	No Controls (1)	Controls for Age Dummies (2)	Controls for "all observables" (3)	Controls for insurance company prediction (4)
Panel 1: Restricted to Individuals for whom Medicaid is not a Close Substitute for Private Insurance (top quartile of income or asset distribution)				
Any nursing home utilization	-0.046*** (0.017) [N=2,159]	-0.027* (0.016) [N=2,158]	-0.031* (0.017) [N=2,158]	-0.027 (0.017) [N=2,121]
Number of nights spent in nursing home	-91.18*** (35.39) [N=2,138]	-61.90* (34.27) [N=2,137]	-62.86* (33.12) [N=2,101]	-66.81** (34.18) [N=2,137]
Panel 2: Restricted to Eligible Individuals				
Any nursing home utilization	-0.013 (0.017) [N=3,600]	-0.007 (0.016) [N=3,599]	-0.0004 (0.017) [N=3,526]	-0.004 (0.017) [N=3,599]
Number of nights spent in nursing home	-37.16 (35.83) [N=3,563]	-30.23 (35.66) [N=3,562]	-28.84 (36.64) [N=3,491]	-17.08 (35.35) [N=3,562]
Panel 3: Restricted to Eligible Individuals for whom Medicaid is not a good substitute				
Any nursing home utilization	-0.026 (0.019) [N=1,362]	-0.030 (0.019) [N=1,361]	-0.032* (0.019) [N=1,343]	-0.026 (0.019) [N=1,361]
Number of nights spent in nursing home	-108.75** (50.90) [N=1,354]	-109.69** (50.21) [N=1,353]	-109.69** (48.21) [N=1,336]	-103.08** (50.52) [N=1,353]

Notes: See notes to Table 4.

UnumProvident Application for Individual Long-Term Care Insurance (p.1)

Part 2: Insurability Profile		
❐ Yes ❐ No	2a.	Do you use mechanical devices such as: a wheelchair, walker, quad cane, crutches, hospital bed, dialysis machine, oxygen, or stairlift?
❐ Yes ❐ No	2b.	Are you cognitively impaired or do you currently need or receive help in doing any of the following: bathing; eating; dressing; toileting; transferring; or maintaining continence?

If you answered "Yes" to either of these questions in Part 2, DO NOT SUBMIT THIS APPLICATION.

John Hancock Application for Individual Long-Term Care Insurance (p. 2)

Part 3 -- Should You Submit This Application?			
3a.	❑ Yes	❑ No	Do you currently use mechanical devices, such as: a wheelchair, walker, crutches, hospital bed, dialysis machine, oxygen, or stairlift?
3b.	❑ Yes	❑ No	Do you currently need or receive help in doing any of the following: bathing; eating; dressing; toileting; transferring from bed to chair; maintaining continence?
3c.			Do you currently have, or have you ever had a diagnosis for or symptoms of:
	❑ Yes	❑ No	1. Alzheimer's disease, dementia, or organic brain syndrome?
	❑ Yes	❑ No	2. multiple sclerosis, muscular dystrophy, ALS (Lou Gehrig's Disease) or Parkinson's Disease?
3d.	❑ Yes	❑ No	Have you been diagnosed or treated by a member of the medical profession for AIDS or AIDS related complex?
STOP HERE!			**If you answered "YES" to any questions in part 3, you may not wish to submit this application as it may not be favorably considered. If not, please continue.**

Table 6: Individuals' predictions of nursing home entry

	No Controls			Controls for Age Dummies			Controls for "All observables"			Controls for Insurance Company Prediction			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Individual Prediction	0.097*** (0.024)	0.153*** (0.032)	0.458*** (0.087)	0.069*** (0.022)	0.092*** (0.031)	0.303*** (0.082)	0.037* (0.022)	0.036 (0.030)	0.214** (0.083)	0.044** (0.022)	0.055* (0.030)	0.289*** (0.085)	
Individual Predicts 0		0.039*** (0.014)			0.016 (0.014)			-0.0007 (0.014)			0.008 (0.014)		
Actuarial Prediction										0.505*** (0.029)	0.504*** (0.029)	0.470*** (0.033)	0.511*** (0.029)
IV?	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
R^2	0.004	0.006	------	0.11	0.11	-----	0.17	0.17	-----	0.10	0.10	----	0.098
N	5,146	5,146	4,549	5,146	5,146	4,549	5,032	5,032	4,455	5,146	5,146	4,459	5,146

Note: Reported coefficients are from linear estimation of equation (3). Dependent variable is whether individual enters nursing home over subsequent five years. In the IV specification, the 1995 self-reported probability is instrumented for using the 1993 self-reported probability. The column headings describe the additional covariates included in the regression; see text and Appendix A for more details.

Table 7: The relationship between insurance coverage and individual beliefs about risk type

	No Controls			Controls for Age Dummies			Controls for "all observables"			Controls for Insurance Company Prediction		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Individual Prediction	0.077*** (0.019)	-0.008 (0.027)	0.275*** (0.070)	0.082*** (0.019)	0.003 (0.027)	0.307*** (0.071)	0.084*** (0.019)	0.026 (0.028)	0.300*** (0.073)	0.091*** (0.019)	0.016 (0.027)	0.331*** (0.072)
Individual Predicts 0		-0.060*** (0.013)			-0.055*** (0.013)			-0.040*** (0.014)			-0.052*** (0.013)	
Actuarial Prediction										-0.128*** (0.019)	-0.119*** (0.017)	-0.149*** (0.021)
IV?	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.004	0.008	-----	0.01	0.016	-----	0.05	0.05	-----	0.02	0.02	----
N	5,233	5,233	4,621	5,233	5,233	4,621	5,118	5,118	4,525	5,233	5,233	4,621

Note: Reported coefficients are from linear estimation of equation (5). Dependent variable is whether individual has long-term care insurance. Instrument for 1995 self-reported probability is 1993 self-reported probability. The column headings describe the additional covariates included in the regression; see text and Appendix A for more details.

Table 8: Decomposing the relationship between care utilization and insurance

	No Controls		Controls for Age Dummies		Controls for "all observables"		Controls for insurance company prediction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LTCINSURANCE	-0.046***		-0.021		-0.006		-0.017	
	(0.015)		(0.015)		(0.015)		(0.015)	
RISKTYPE_HAT		1.274***		0.863***		0.454*		0.443***
		(0.307)		(0.276)		(0.266)		(0.220)
RESID_HAT		-0.051***		-0.025*		-0.008		-0.020
		(0.015)		(0.015)		(0.015)		(0.015)
N	5,070	5,070	5,070	5,070	4,958	4,958	5,070	5,070

Notes: All estimates are by OLS. The dependent variable is "any nursing home entry". The column headings describe the additional covariates included in the regression; see text and Appendix A for more details. Columns (1), (3) and (5) report the coefficient on LTCINSURANCE from estimating equation (1a). Columns (2), (4) and (6) report the coefficients on RISKTYPE_HAT and RESID_HAT (which are predicted based on the results from estimating the insurance coverage equation (5)) from estimating equation (6). These reflect, respectively, the portion of insurance coverage that is explained by the individual's beliefs about his risk type, and the portion of insurance coverage that is unexplained by either these beliefs or by the risk classification done by the insurance company. Heteroskedacticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively. See text for more details.

Table 9: Preference-based selection

	(1) LTCINS	(2) NH ENTRY	(3) NH ENTRY	(4) LTCINS	(5) NH ENTRY	(6) NH ENTRY	(7) LTCINS	(8) NH ENTRY	(9) NH ENTRY	(10) LTCINS	(11) NH ENTRY	(12) NH ENTRY
Preventive Activity	0.064*** (0.016)	-0.112*** (0.019)		0.050*** (0.016)	-0.036** (0.018)		0.011 (0.020)	-0.017 (0.019)		0.051*** (0.016)	-0.056*** (0.018)	
Individual Prediction	0.087*** (0.019)	0.104*** (0.024)		0.091*** (0.020)	0.072*** (0.023)		0.094*** (0.020)	0.039* (0.022)		0.099*** (0.020)	0.050** (0.023)	
PREVENT _HAT			-1.383*** (0.222)			-0.662*** (0.244)			-0.400 (0341)			-0.714*** (0.246)
RISKTYPE _HAT			1.371*** (0.316)			0.923*** (0.286)			0.457* (0.267)			0.559*** (0.259)
RESID _HAT			-0.042*** (0.015)			-0.022 (0.015)			-0.007 (0.015)			-0.016 (0.015)
Risk Classification	None	None	None	Age dummies	Age dummies	Age dummies	All observables	All observables	All observables	Insurance company prediction	Insurance Company prediction	Insurance Company prediction
N	5,070	5,070	5,070	5,070	5,070	5,070	4,958	4,958	4,958	5,070	5,070	5,070

Note: All estimates are by OLS on the binary dependent variable given in the top row. "Preventive activity" measures the fraction of gender-appropriate preventive health activity undertaken by the individual. All regressions include a control for gender. The "risk classification controls" row describes the additional covariates included in the regression. PREVENT_HAT, RISKTYPE_HAT and RESID_HAT are generated based on the results from estimating the insurance coverage equation (8) and reflect the portion of insurance coverage that is explained, respectively, by "Preventive activity", the individual's beliefs about his risk type, and the portion unexplained by either preventive activity, individual beliefs, or the risk classification controls. Heteroskedacticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively. See text for more details.

Figure 5: Sample distribution of self-reported probability of entering a nursing home in next five years

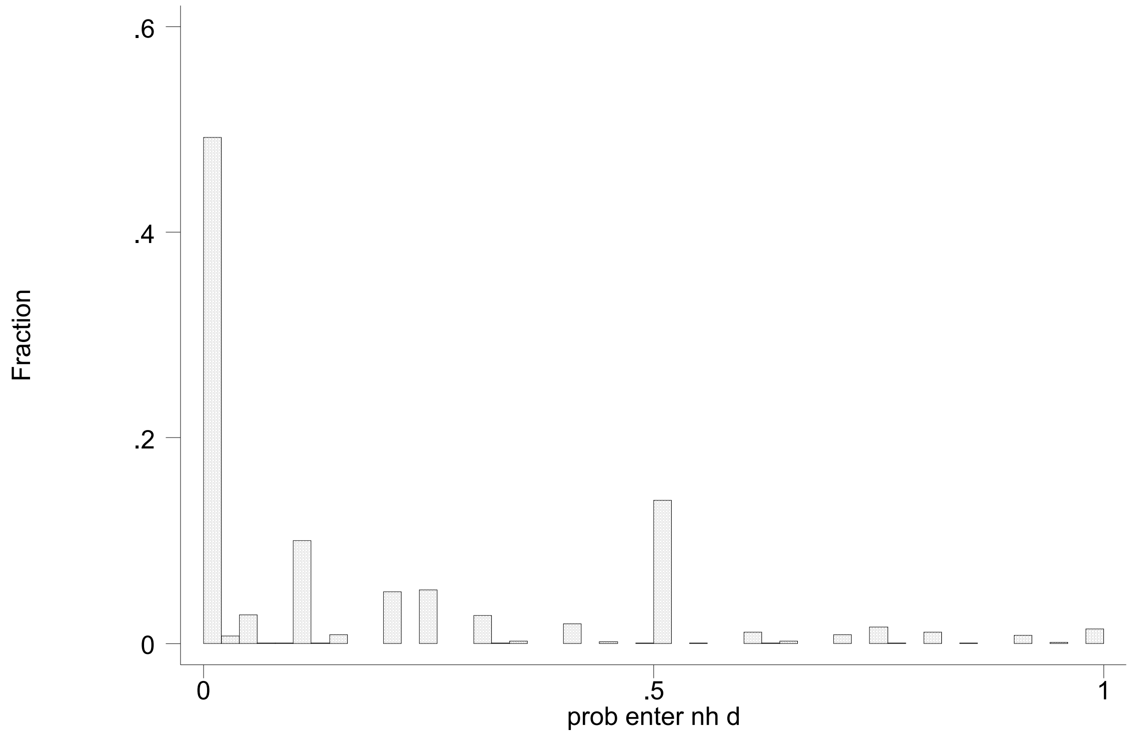
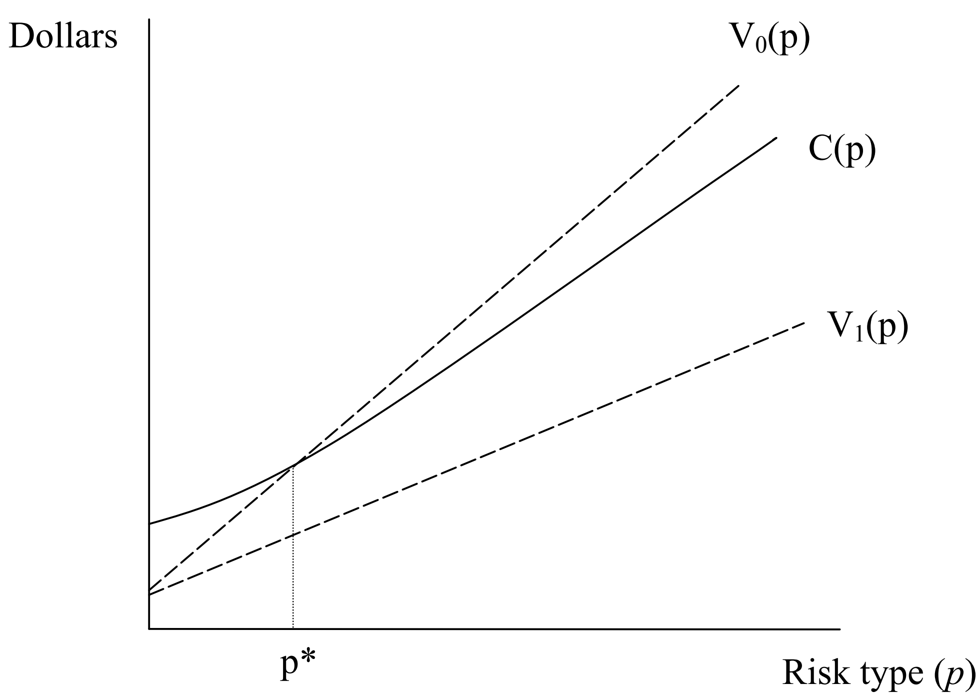


Figure 6: Unraveling



Appendix A: Detailed information on 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. This original AHEAD cohort consists of individuals born in 1923 or earlier and their spouses; when appropriately weighted, it is representative of the non-institutional population of this age group.[49] To increase sample size, we include observations on the sample members spouses even if they are outside this age range (1.5% of the sample), but we exclude 50 spouses who were younger than 60 at the 1995 interview. We also exclude the 3 percent of original respondents who were in a nursing home in 1995. The results are not sensitive to any of these inclusions or exclusions.

The AHEAD respondents were interviewed in 1993, 1995, 1998 and 2000. We restrict our analysis to data from 1995 to 2000, omitting information in 1993, because the question of long-term care insurance in that year was poorly worded and we do not believe reflects true insurance coverage (see below). Non-death attrition (i.e. "real" attrition) from our sample is just over 4 percent from 1995 and 2000. Those who attrited from the sample have significantly lower income and wealth, on average, and are slightly less likely to have long term care insurance, although the difference is not statistically different from zero.

Measuring care utilization: Our analyses use the panel nature of the data to track individual care utilization through the 2000 interview wave, which is the latest currently available wave of data. We use responses to questions about care utilization since the last interview obtained in the 1998 and 2000 surveys. Specifically, at each interview we know whether the individual is currently in a nursing home or has been in a nursing home since the last interview and if so, for how many nights. If the individual is not in a nursing home, they are also asked whether they are currently receiving home care or have since the last interview**.**

Sample weights: All of the means and the regression results reported in the paper from the AHEAD data are weighted using the 1995 household weights. The use of household weights rather than respondent weights allows us to include out-of-age-range spouses The regression results are not sensitive to using respondent-level weights instead or to running un-weighted regressions.

Key independent variable: Long-term care insurance. As we noted above, 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 relying on 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?

---

[49] 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 (the key variable for the analysis in Section 4) until later waves.

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 (see e.g. Cohen, forthcoming and citations therein). Our concern about the accuracy of long-term care insurance coverage measurement in the 1993 AHEAD was corroborated by the staff of the HRS (email correspondence with David Weir, Assistant Director of HRS, April 2002).

With the 1995 question, the reported coverage rate in the 1995 wave was 10 percent. This estimate roughly matches that of Cohen (forthcoming) who estimates – based on industry survey data – that 3.5 to 4.0 million Americans have private long-term care insurance. (Although people of any age may hold long term care insurance, it tends to be held by the elderly (HIAA 2000a)). By means of comparison there are about 35 million individuals aged 65 and over in the United States (U.S. Bureau of the Census, 2000).

The long-term care insurance question was altered again in 1998 to define long-term care insurance as a policy covering *stays of a year or more,* in order to distinguish long-term care policies from policies that cover short stays related to acute care. The mean coverage rate did not change and we have found our results to be robust to the use of the 1998 measure in lieu of the 1995 measure.

A note on other covariates

Cognition: Cognitive functioning is an important factor in considering potential nursing home use, and our understanding is that insurance companies pay a great deal of attention to assessing cognitive limitations. Insurance companies ask directly about cognitive impairment and use other techniques such as interviews to assess cognition. (One company asks for a hand-written statement.) Fortunately AHEAD provides numerous measures of cognition allowing for a rich measure. We follow Mehta et al. (2002) who work specifically with AHEAD and use a modified version of the Telephone Interview for Cognitive Status (TICS) score. A score of 8 or less on a scale of 35 is used as our definition of cognitive impairment. The questions that are used in the TICS include the respondent's ability to report the day and date, count backwards from 20, count backwards from 100 by 7, define a set of commonly used words, and remember a list of words (immediate and delayed recall). For proxy respondents, cognitive ability was based on assessments offered by the proxy.

Depression: Our measure of depression also follows that used in Mehta et al. (2002) and is based on the CES-D8. We use scores of 3 (out of 8) or greater as an indicator of depression. The CES-D8 questions ask if the respondent considers himself depressed, whether he feels that everything he does is an effort, if he has trouble sleeping, feels happy, lonely, sad, and enjoys life. (The scaling of "happy questions" is inverted in summing the responses.) Based on this measures, 20 percent of our sample is categorized as depressed. This measure is not available for proxy respondents. An indicator of a proxy interview is included in the regressions and the depression measure is set to zero.

Alcohol uses: Although many insurance companies query respondents about drinking, we could find no commonly accepted survey measure of a drinking problem. We define 3 or more drinks per day as a drinking problem.

BMI. Insurance companies collect information on individuals' height and weight. We used this to construct a measure of body mass index (BMI) defined as weight in kilograms divided by height in meters squared. We include controls for extreme BMI (above 30 or below 20) as an indicator of poor health. A BMI of 30 or more is considered obese, and a BMI of 18.5 or less is considered underweight according to the Centers for Disease Control and Prevention (http://www.cdc.gov/nccdphp/dnpa/bmi/bmi-adult.htm). The results are not sensitive to instead including height and weight linearly in the regression in place of our categorical measure of BMI.

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

Missing values: Some of our regressions include a very rich set of covariates, particularly for current health and medical history. A few of these variables are missing for relatively large fractions of the sample. In order to retain observations with missing values in our regression analyses we use dummy variables to indicate a missing value on a variable, and set the variable itself equal to zero. Similarly, some questions were not asked of proxy respondents. We thus include a dummy variable indicating that the interview was conducted by proxy (9% were), and set the value of the unasked variable to zero. The results are not sensitive to dropping any observations with missing values from the sample.

<u>Restriction of sample to individuals who are not likely to view Medicaid as a good substitute</u>
In addition to having low income, Medicaid coverage essentially requires the spend-down of nearly all assets.[50] We therefore define a restricted sub-sample that consists only of those individuals who are unlikely to qualify for Medicaid. We experimented with several approaches to restricting the sample. In the estimates we report in the paper, we select as Medicaid-ineligible only those respondents who are in the top quartile of either the income or asset distribution. This restriction eliminates two-thirds of the original sample. Our measure is similar in spirit to Cutler and Gruber's (1996) measure of "conditional coverage" by Medicaid among the non-elderly.

<u>Restriction of sample to "eligibles".</u> We identified current insurance company denial practices using information from long-term care insurance applications as well as underwriting guides from the insurance company. This information of course reflects current (2002) denial practices, while we are analyzing insurance coverage in 1995. We therefore investigate the consistency of practices over time. Murtaugh et al. (1995) collected information on long-term care insurance denial and pricing practices in the late 1980s using sources similar to those we employ in 2002. Their description suggests that the basic practice has not changed much of time. This pattern was confirmed in conversations with actuaries and we are therefore comfortable with our definition.

The measure of eligibility used in the paper uses denial criteria that are common across the current applications of several major insurance companies as well as the older applications described in Murtaugh et al. (1995). The three criteria are: limitations with respect to activities of daily livings (bathing, eating, dressing, toileting, walking, and maintaining continence), use of mechanical devices (wheelchair, walker, crutches, quad cane, oxygen) or cognitive impairment. There are a few cases where some companies employ additional tests. We do not use these additional criteria because any item that is limited to a single firm (or to a handful of firms) may be non-binding; an individual who is denied based on this uncommon parameter can simply apply to another company that does not impose the same restrictions. We experimented with stricter definitions of eligibility in which criteria used to deny individuals in several firms were also used to exclude them from the eligible sample; the results were not sensitive to this alternative approach.

Based on this algorithm, we classify 40% of the sample as ineligible for long-term care insurance. Our algorithm is likely to overly restrict the eligible sample because it is based on *current* health conditions and individuals may have applied at a younger age when they would likely have been in better health. Consistent with this strict selection, those ineligible, based on our definition, have only a slighter lower long-term care insurance coverage rate than the general population (9.6 vs. 10.4 percent). However, even in the population, ineligibility is a non-trivial phenomenon. Weiss (2002) reports that approximately 15% of non-group long-term care insurance *applications* are denied. Murtaugh et al. (1995) estimate that between 12 and 23% of 65-year olds would have their applications denied if they applied for insurance and that between 20 and 31% of 75-year olds would be denied.

[50] There is a substantial asset allowance for a non-institutionalize spouse that is excluded from the determination of Medicaid eligibility (AARP, 2000).