Re-examining the test for asymmetric information in insurance markets: New evidence from long-term care insurance

Amy Finkelstein
Harvard University and NBER

Kathleen McGarry
University of California, Los Angeles and NBER

July 2003

This paper examines the standard test for asymmetric information in insurance markets: that its presence will result in a positive correlation between insurance coverage and risk occurrence. We show empirically that while there is no evidence of this positive correlation in the long-term care insurance market, asymmetric information still exists. We use individuals' subjective assessments of the chance they will enter a nursing home, together with the insurance companies' own assessment, to show that individuals do have private information about their risk type. Moreover, this private information is positively correlated with insurance coverage. We reconcile this direct evidence of asymmetric information with the lack of a positive correlation between insurance coverage and risk occurrence by demonstrating the existence of other unobserved characteristics that are positively related to coverage and negatively related to risk occurrence. Specifically, we find that more cautious individuals are both more likely to have long-term care insurance and less likely to enter a nursing home. Our results demonstrate that insurance markets may suffer from asymmetric information, and its negative efficiency consequences, even if those with more insurance are not higher risk. They also suggest an alternative approach to testing for asymmetric information in insurance markets.

We thank Daron Acemoglu, Jeff Brown, Pierre-Andre Chiappori, Raj Chetty, David de Meza, Seema Jayachandran, Jerry Hausman, Larry Katz, Ben Olken, Sarah Reber, Bernard Salanie, Jesse Shapiro, Jonathan Wright, and seminar participants at Berkeley, Stanford, San Diego and the University of Chicago 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. We thank the NBER and NIA for financial support.

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). These studies are all based on the same widely used test of asymmetric information: they test for whether there is a positive correlation between insurance coverage and risk occurrence. Contrary to the predictions of many moral hazard and adverse selection models, these papers find no evidence that individuals with more insurance are more likely to experience the insured risk. These findings have challenged the previous conventional wisdom that asymmetric information is an important phenomenon in insurance markets.[1]

In this paper, we show empirically that asymmetric information may exist in an insurance market even when the expected positive correlation fails to materialize. Individuals may have private information not only about their risk type but also about preference-related characteristics (such as risk aversion). If these unobserved preference-related characteristics have the *opposite* correlation with insurance coverage and with risk occurrence, they may offset the positive correlation between insurance coverage and risk occurrence that private information about risk type would otherwise produce. Thus, rather than indicating symmetric information, the lack of a positive correlation between insurance coverage and risk occurrence may indicate that there exist multiple forms of private information, acting in different directions.

In contrast to a symmetric information explanation for a lack of a positive correlation between insurance and risk occurrence, an explanation based on multiple forms of private information indicates that the market equilibrium is not efficient. Distinguishing between these two explanations for the same observed equilibrium is therefore critical to an understanding of the underlying structure of the insurance

[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).

market. It is also a necessary first step in evaluating potential public policy interventions. Despite its importance, to our knowledge, there has been no systematic empirical exploration of this issue. We provide one here.

While the ideas advanced in this paper are applicable to a wide variety of insurance markets, we focus our empirical work on the private long-term care insurance market in the United States. In addition to providing an interesting setting for studying asymmetric information, this market is of substantial importance in its own right. Long-term care expenditures currently represent one of the largest uninsured financial risks faced by the elderly in the United States. As the baby boomer generation ages and medical costs continue to rise, the nature of the long-term care insurance market will have profound implications for the well-being of both the elderly and their children. The limited size of the long-term care insurance market is well-known, but not well understood. Adverse selection and moral hazard may play a role, yet there exists little empirical evidence on their existence in this market.

We begin by following the existing literature and examine whether there is a positive correlation between the amount of insurance individuals have and the occurrence of the risk (in this case, the individual's ex-post use of a nursing home). We analyze data from two complementary sources: proprietary micro data from a large private long-term care insurance company, and individual-level panel data from the Asset and Health Dynamics of the Oldest Old (AHEAD) cohort of the Health and Retirement Survey (HRS). We find no evidence that, after controlling for the risk classification done by the insurance company, those with more long-term care insurance end up using more nursing home care. If anything, we find suggestive evidence that they use less nursing home care.

To distinguish whether this equilibrium reflects the presence of symmetric or of asymmetric information, we directly examine whether individuals have private information about their risk type. In the AHEAD data, we can measure each individual's subjective belief of his probability of entering a nursing home over the next five years, and compare that prediction to his subsequent five-year nursing home utilization. We supplement these data with measures of the insurance companies' information set and risk classification practices; these measures are based on insurance company application forms which reveal the set of individual characteristics observed by the insurance companies, and on the industry's actuarial model of nursing home utilization as a function of these observed characteristics.

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 provides direct evidence of asymmetric information in the private long-term care insurance market: Individuals have information about the likelihood of risk occurrence that the insurance company does not capture. Moreover, we find that the individual's private information about his risk type is positively correlated with whether the individual has insurance coverage. Most importantly, we demonstrate that the variation in long-term care insurance coverage explained by individuals' private information about their risk type is positively and statistically significantly associated with subsequent nursing home utilization.

These results – together with the fact that, overall, nursing home utilization and long-term care insurance coverage are not positively correlated – indicate the existence of unobserved heterogeneity not only in risk type but also in preferences that has the opposite correlation with insurance coverage and nursing home utilization. We provide direct evidence of the existence and nature of these other unobserved, preference-related characteristics. 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.

The rest of the paper is structured as follows. Section one describes the standard empirical test for whether insurance coverage and risk occurrence are positively correlated and discusses different possible explanations for a lack of a positive correlation; we emphasize their different implications for the structure of information and for market efficiency. Section two provides some brief background on the private long-term care insurance market. The next three sections present 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 and that differences in insurance coverage due to this private information are positively and statistically significantly related to subsequent care utilization. The combined evidence from Sections three and four suggests that other unobserved factors have the opposite correlation with insurance coverage and care utilization and thus mask the role of adverse selection and moral hazard. Section five provides direct evidence of the existence and nature of these other unobserved factors. The final section summarizes our findings and discusses their implications for testing for asymmetric information in other insurance markets.

1. Theoretical background

1.1 The "positive correlation" prediction

A wide variety of asymmetric information models predict that there will be a positive correlation between the amount of insurance and the probability of the risky event occurring (Chiappori and Salanie, 2000; Chiappori et al., 2002). As a result, the standard test for asymmetric information used in the empirical literature is to test for a correlation between the amount of insurance coverage and the ex-post occurrence of the (potentially) insured risk. Throughout the paper we will refer to this test as the "positive correlation" prediction. Of course, this prediction applies only among individuals who would be treated symmetrically by the insurance company (i.e. placed in the same risk category and offered the same set of insurance contract options). Our discussion of the "positive correlation" property therefore always refers to the correlation conditional on the risk classification of the individual done by the insurance company, although we will not always state this qualification explicitly. In the empirical work below, we will take great care to appropriately condition on this risk classification.

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 mechanism by which the positive correlation arises differs however, for the two phenomena. In the case of adverse selection, the insured is assumed to have ex-ante 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, 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 (see e.g. Rothschild and Stiglitz, 1976). In the case of moral hazard, the causality is reversed and the informational asymmetry occurs ex-post: 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. Cutler (2002) reviews an extensive empirical literature that finds evidence of the positive correlation property in health insurance, although exceptions exist (e.g. 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.

1.2 Implications for the structure of information and market efficiency

There are two broad classes of possible explanations for a lack of a positive correlation between insurance coverage and risk occurrence. One argues that there is symmetric information and no moral hazard, while the other argues that there is asymmetric information about *both* risk type and preferences, effects that offset each other. The central motivation for our empirical analysis is that these two alternative explanations have very different implications for the structure of information and for market efficiency. Here, we provide some simple intuition for why the positive correlation may break down under either of these scenarios, and what the efficiency implications are. Interested readers should consult Chiappori et al. (2002), Jullien et al. (2002), and de Meza and Webb (2001) for more formal discussions.[2]

Consider first a model in which information is symmetric and there is no moral hazard. Given the vast amount of information that insurance companies can, and do, collect about potential customers, it may be unreasonable to assume that the individual has private information; indeed, the insurance company, with its sophisticated actuarial methods, might even have *superior* information about the individual's risk type.[3] It is also possible that moral hazard may 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. If individuals have no private information and there is no moral hazard, insurance coverage need not be positively correlated with risk occurrence.[4] In addition, the equilibrium insurance coverage will be first best, because with symmetric information there is no reason for anyone's insurance purchases to be distorted.

An alternative explanation for a lack of a positive correlation between insurance coverage and risk occurrence is that – unlike in the standard models of asymmetric information – risk type may not be the only source of unobserved heterogeneity. Individuals may also differ with respect to unobserved preferences – such as risk aversion – that are correlated with both the demand for insurance coverage and with risk occurrence. We refer to this as "preference-based" selection to distinguish it from traditional adverse (or "risk based") selection based on the individual's private information about his risk type. If unobserved preferences are positively correlated with insurance demand and negatively correlated with

[2] Chiappori et al. (2002) show not only that asymmetric information may exist in the absence of a positive correlation (which is the possibility we focus on here), but also that asymmetric information may *not* exist even *with* a positive correlation.

[3] In this case, the equilibrium may exhibit a *negative* correlation between insurance coverage and risk occurrence (Villeneuve 2000, Villeneuve forthcoming).

[4] If there are no marginal production costs, there will be no correlation between insurance coverage and risk occurrence. With symmetric information but marginal production costs, any correlation is possible. Consider, for example, a constant marginal production cost on an insurance policy in a competitive market. Individuals will no longer purchase full insurance (as they would with no marginal production costs), and the amount of insurance they purchase will vary with their risk aversion; the correlation between risk type and amount of insurance coverage will depend on the correlation between risk type and risk aversion.

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, more risk averse individuals value insurance more; if they are also lower risk – perhaps because they investment more in risk-reducing effort – the correlation between insurance coverage and risk occurrence may be positive, zero or even negative (Jullien et al., 2002, de Meza and Webb, 2001).[5]

In the presence of multiple sources of private information, the equilibrium will not be efficient, even if the positive correlation property does not obtain. The intuition for this can be seen clearly in a simply monopoly model of insurance provision. Assume there is no moral hazard and that the first best allocation is for all individuals to be fully insured. If information is completely symmetric, the monopolist can engage in perfect price discrimination and therefore the equilibrium will involve all individuals having full insurance (i.e. the first best). Now assume, instead, that individuals have private information about risk type and about risk aversion. Assume further that risk aversion is decreasing in risk type and it is decreasing sufficiently rapidly that, at a given price, the value of insurance is *decreasing* in risk type. In other words, we have reversed the standard single crossing property. The equilibrium may therefore be the opposite of the Stiglitz (1977) equilibrium: the monopolist may design a set of non-linear contracts to induce the *lower risk* (but higher insurance valuation) individuals to self-select into more comprehensive contracts, and the higher risk (lower insurance valuation) individuals to self-select into less than full insurance. There will therefore be a negative correlation between risk type and insurance coverage. However, just as in the standard asymmetric information model in which risk type is the only source of private information, the resulting equilibrium is inefficient: one type (although now it is the higher risk) is constrained to have sub-optimal insurance coverage. Other models with unobserved risk type and risk aversion can generate the opposite inefficiency, with one type receiving over-insurance relative to the first best (de Meza and Webb, 2001); in their model, not only is the private market equilibrium inefficient, but there is scope for Pareto improvements through government intervention.

---

[5] Chiappori et al. (2002) show that for preference-based selection to result in no positive correlation requires either imperfect competition in the insurance market or some marginal production cost or load.

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

At almost $100 billion a year in 2000, long-term care expenditures in the United States comprise 7.5% 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; for example, Dick et al. (1994) estimate that while 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. This suggests potentially large welfare gains from insurance coverage that reduces this expenditure risk.

However, most of this substantial expenditure risk is uninsured. Over 40 percent of long-term care expenditures for the elderly are paid for out of pocket, compared to only 17 percent of the elderly's expenditures in the health sector as a whole. This disparity partly reflects the limited public insurance coverage for long-term care. Medicare, the public health insurance program for the elderly, covers only a very restricted set of long-term care services. Medicaid, the public health insurance program for the indigent, is available only to elderly individuals with little or no wealth. The extremely limited nature of private long-term care insurance coverage is also an important factor.

Private insurance covers only about 5 percent of the elderly's long-term care expenditures, compared to 35 percent of the elderly's overall health expenditures (US Congress, 2000, National Center for Health Statistics, 2002). Only about 10 percent of those aged 65 and over had private long-term care insurance coverage in 2000.[6] Moreover, most private insurance policies insure only a limited fraction of long-term care expenditures. For example, Brown and Finkelstein (2003) estimate that typical private long-term care insurance contracts pay for less than 50 percent of the expected present discounted value of long-term care costs for a 65 year old.

About 80 percent 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). In 2000, the

[6] Authors' calculation based on 2000 Health and Retirement Survey.

average age of purchase of long-term care insurance in the individual market was 67 (HIAA 2000a). Coverage rates are roughly comparable for men and women but increase substantially with asset levels, probably due, at least in part, to the means-tested nature of the Medicaid program, (HIAA, 2000a). Presumably to reduce moral hazard, insurance policies specify a set of standard health-related criteria 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 with several broad health categories (ACLI, 2001; Weiss 2002). Most notably, premiums do not vary with sex, despite known differences by sex in long-term care utilization (Society of Actuaries, 1992). One potential explanation for this ostensibly puzzling practice 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 insured individuals also have this insurance. Premiums tend to be a constant nominal amount paid on a monthly or annual basis.

A variety of theoretical explanations have been proposed for the limited size of the private long-term care insurance market (see Norton, 2000 for a review). Asymmetric information is 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.

3. Is there a positive correlation between LTC coverage and care use?

Long-term care includes both care in a nursing home and home health care. Nursing homes are substantially more expensive (MetLife 2002), and account for over three-quarters of long-term care expenditures (US Congress, 2000). Moreover, until quite recently, policies tended to cover only nursing home care.[7] Because of these patterns, and data limitations discussed below, our empirical analysis focuses primarily on the relationship between insurance coverage and nursing home utilization.

Figure 1 – based on aggregate data from the Society of Actuaries (SOA, 2002) – shows the ratio of nursing home admission rates for insured individuals to admission rates for the general population, by age and by sex.[8] The positive correlation property predicts that the insured-to-population ratio of admission rates should be larger than one. The pattern displayed in 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. The SOA (2002) also provides data on the relative nursing home admission rates, by age, among insured individuals with varying amounts of insurance. Again there is no evidence of a positive correlation between the amount of insurance and nursing home admission rates; indeed, for some policy features that increase the amount of insurance, there is a *negative* correlation between admission rates and increases in the amount of insurance.

Although suggestive, the SOA data do not provide 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, 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 remainder of this section is designed to address these issues.

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 the insurance purchases and subsequent claims experiences of customers in a large, private long-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 AHEAD.

---

[7] When the market began in the early 1980s, most policies covered nursing homes 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).

[8] We limit the insured data to the approximately 12% of the exposure that reflects the experience of policies with no deductible. The insured admission rate only records admissions that result in a claim; therefore the data will underestimate the admission rates for insured individuals whose policies have a deductible. Estimated admission rates for the entire insured population are in fact substantially lower than for insured individuals with no deductible.

3.1 Proprietary policyholder data from a large private insurance company

3.1.1 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)). Although the data come from a single company, they appear comparable to the broader market in terms of the age and gender-mix of purchasers and the product mix of policies sold. In addition, the company experienced similar growth rates in policy sales over the last five years to the industry as a whole (LIMRA 2001).

We observe a complete description of the features of each policy. Crucially, we also observe the risk classification of the individual done by the insurance company; it follows typical industry practices. The company varies the premium based on the individual's age at the time of policy issue, the date that the policy is issued, and whether the individual is rated preferred, standard, or substandard based on detailed health information; we observe the individual's age at purchase, issue date, and rating category, although we do not observe the underlying health information on which the rating category is determined. We also observe a complete description of all claims incurred through December 31, 2001.

To test the positive correlation prediction, we examine the relationship between the characteristics of the policy that affect the quantity of insurance provided and nursing home utilization. In contrast to the comparison using the SOA data, here we can condition on the individual's risk classification. However, because we still only observe care utilization if it results in a claim, we will miss stays that do not exhaust the deductible (which must be satisfied anew for each care episode). Therefore, we define a "failure" in our hazard model as having at least 100 continuous days of nursing home care and 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). Conditional on entering a nursing home, stays of more than 100 days are quite common (Dick et al, 1994, Kemper and Murtaugh, 1991, and Murtaugh et al. 1997). The average failure rate in our sample, 0.3 percent, is quite low, but is consistent with market-wide and population statistics on nursing home utilization (SOA 1992, 2002).[9]

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 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:

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

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).[10]

We include a set of covariates to control for the insurance company's risk classification of the individuals. 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+).[11]

The primary covariates of interest measure the four main aspects of the policy that affect the quantity of insurance in the policy. These are: (1) the deductible, (2) the total number of days for which benefits may be received in the lifetime of the policy ("benefit period"), (3) the maximum amount of incurred

[9] This low failure rate prohibits an analysis of 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.
[10] 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. It is not obvious that this assumption is appropriate. We therefore tested the robustness of our results to instead maintaining the terminated policies in the "at risk" sample after policy termination. The results were not substantively affected.
[11] Including separate indicator variables for each age rather than five-year intervals does not affect the coefficients of interest. We adopt the coarser set of controls as our main specification simply for ease of presentation.

nursing home care expenditures that the policy will reimburse per day in care ("maximum daily benefit"), and (4) how the nominal maximum daily benefit increases over time after purchase of the policy ("benefit escalation"). The positive correlation property predicts that the hazard rate should be increasing in the benefit amount, 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 maximum daily benefit amount using three roughly-equal sized indicators for less than $100, $100, and more than $100 per day. 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 out of care since the last day of receiving benefits; this reset option effectively extends the benefit period. Finally, we use indicator variables for the four possible benefit escalation options. In order of increasing benefit levels these are: constant nominal benefits, benefits escalate at 5 percent of the original benefit per year ("simple" escalation) and benefits escalate at 5 percent per year ("compound" escalation). The fourth option – benefits are increased by the greater of 5% compounded annually over 3 years or CPI-growth over the last 3 years at the option of the policy holder ("indexed") – represents higher benefits than no escalation. For completeness, we also control for the remaining policy features as described in the notes to Table 2.

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 sex because it is not used in determining the pricing of contracts.

3.1.2 Results: Table 2 reports the results from estimating equation (1). We show results for the entire sample of policies. Because some of these policies have been in effect for only a short time, we also report results for the subset of policies issued in 1997 or 1998, all of which have had at least three years of exposure. The results look similar if we instead limit the sample to individuals who are 75 and older at the time of purchase, and for whom we therefore observe a greater fraction of the policies' actual lifetime (results not shown).

The results in the top portion 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 with the assessed risk category. For example, individuals who are rated standard risk have about a 55 percent lower baseline hazard rate of entering a nursing home for at least 100 days than individuals who are rated high risk.

The lower portion of the table reports the coefficients on covariates for which the positive correlation property makes predictions; these predictions are summarized in the right-most column. There is little evidence in support of these predictions. The coefficients on the benefit escalation and benefit period variables tend to have the *opposite* sign from what is predicted by the positive correlation property. The coefficients on the deductible and daily benefit variables tend to be positive as predicted (those with shorter deductible periods and higher daily benefits are more likely to use services) but not statistically different from zero. Moreover, 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 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 additive controls for the individual's risk classification may produce misleading estimates of the relationship between features of the contract and nursing home utilization if there are important interaction effects among the various determinants of the individual's risk classification and nursing home utilization. We therefore estimated a more flexibly specified version of equation (1) in which we included fixed effects for each risk class, defined by the interactions of the individual's issue age, 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 classification. 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.

3.2 Evidence from individual panel data in the AHEAD

3.2.1 Data and empirical framework: 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. Such information is available in the Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS). This sample, first interviewed in 1993, is 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, at which point the average age of individuals in our sample is 79. We use the panel nature of the data to track nursing home utilization for the 1995 respondents through the latest currently available interview in 2000. Appendix A provides more detail on the sample and variable definitions.

The basic estimating equation is:

$$\text{(2)} \qquad \text{CARE} = X\beta_1 + \beta_2 \text{LTCINS} + \varepsilon$$

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 such coverage. We include as controls a series of covariates (X) designed to control for any risk-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; 19 percent of the sample did. The second is the total number of nights that the individual spent in a nursing home in this period. On average, individuals spent 33 nights in a nursing home; conditional on entering a nursing home, the mean is 187 nights.

As discussed, the correct empirical test requires controlling for the risk classification of the individual done by insurance companies. In the proprietary insurance company data, we directly observed this risk classification. In the AHEAD data we do not. However, we do observe extremely rich and detailed information on current health and medical history, as well as other demographics. We determined what characteristics of the individual the insurance companies observe by examining insurance application forms from five leading long-term care insurance companies. All collect a limited set of demographic information – age, gender, marital status, and age of spouse – as well as similar and extremely detailed information on current health and on health history. We found only one company that asked applicants to report any financial information (specifically, whether they had less than $30,000 in financial assets).

We can observe in the data essentially all of the information collected by the insurance companies. We also know that companies offer age-specific prices with only two or three broad rate classifications within each age based on health information (ACLI 2001, Weiss 2002, Kemper et al. 1995).[12] However, we do not know the algorithm mapping the observable characteristics into the rate classifications. Given the importance of controlling for the individual's risk classification in the analysis, we therefore experiment with four alternative approaches. First, we do not include any covariates (X) in estimating equation (2) ("no controls" specification). Second, we control for the individual's age by including a separate indicator variable for each age ("age dummies" specification). Both of these approaches are likely to underestimate the amount of categorization done by insurance companies.

Therefore our third approach ("all observables" specification) tries to control for everything the insurance companies observe about the individual. This specification includes not only the age dummies, but also all of the demographic information that insurance companies observe, (gender, marital status and age of spouse, which we enter linearly), and indicator variables for each of the detailed current health and

---

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

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.

By including a separate indicator variable for each health characteristic, the "all observables" specification invokes a much more finely defined categorization of risk than insurance companies actually use. We therefore believe that this specification is likely to overestimate the amount of risk classification done by the insurance company. However, there are two potential issues with this approach. First, there are a few characteristics that the insurance companies observe that we cannot measure in the AHEAD. Most are rare health conditions – such as double amputation or unoperated aneurysm. To try to compensate for this, we added all of the additional health measures observed in the AHEAD and not by the insurance company to the analysis (e.g. self-reported health status, cataract surgery etc.). We did not find any substantive changes in our results. In addition, because many of the characteristics we cannot measure (e.g. handwriting samples) are only collected by insurance companies for observably high risk individuals, we verified that the results were not sensitive to restricting the sample to the observably most healthy individuals.

Second, by merely including each observed characteristic as an additive control in the "all observables" specification, 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. We generated these predictions using the same actuarial model that is employed by much of the long-term care insurance industry; this model and its pedigree are described in detail in Robinson (1996), Robinson (2002), and Brown and Finkelstein (2003). We use a version of the model that predicts care utilization for typical individuals in the population and makes no adjustment for potential moral hazard effects of the insurance. The predictions depend non-parametrically on the individual's age, gender and membership in one of seven different health states (defined by the number of limitations to instrumental activities of daily living (IADLs), the number of limitations to activities of daily livings (ADLs), and the presence or absence of cognitive impairment); all of this information is available in the AHEAD. This measure provides a parsimonious way of controlling for non-linear (and non-parametric) interactions between the observed characteristics of the individual and nursing home utilization.[13]

3.2.2 Results: The top panel of Table 4 describes the results of estimating equation (2) 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 (2); 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 both for the whole sample and when limited to those who report any time in a nursing home.

The results are not supportive of a positive correlation between long-term care insurance coverage and long-term care utilization. In all specifications, long-term care insurance coverage is *negatively* associated with long-term care utilization. Across all specifications, we can reject a higher probability of nursing home utilization for the insured relative to the uninsured of more than 2.8 percentage points with 95 percent confidence.

A potential problem with this analysis 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). To address this issue, we repeat the regressions shown in the top panel of Table 4, restricting the sample to those individuals who are least likely to find Medicaid an attractive substitute for private insurance. Specifically, because Medicaid coverage requires a deductible of almost all of one's assets (AARP 2000) and is therefore a more attractive substitute for lower-wealth individuals, we restrict the sample to individuals in the top quartile of the household income or wealth distribution in

[13] As an alternative way of dealing with non-linearities in the relationship between observable characteristics and long-term are utilization, we also estimated equation (2) on increasingly homogenous sub-samples of individuals from the perspective of the insurance company (e.g. by age and health conditions). The results were not affected.

1995. The bottom panel of Table 4 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 0.6 percentage points with 95 percent confidence.

In results not reported, we ascertained that the results in Table 4 were robust to a number of other alternative specifications. Two in particular are worth noting. First, insurance companies tend to deny some observably unhealthy individuals private long-term care insurance coverage; for example, Weiss (2002) estimates that 15% of non-group long-term care insurance applications are denied. We therefore re-estimated equation (2) restricting the sample to individuals who have none of the health conditions that tend to provoke denials; Appendix A provides details of our classification approach. The coefficient on long-term care insurance remains consistently negative, even if the sample is further restricted to those for whom Medicaid is also not a close substitute for private insurance.[14]

Second, because we only observe care utilization for a five-year period, and not over the lifetime of the policy, it is possible that the positive correlation property would appear if the data were analyzed over a longer time horizon. We tried several alternative approaches to addressing this issue, none of which affected the qualitative nature of the results. 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 (the earliest year for which nursing home utilization data are available) and thus observed 8 years of care utilization data rather than only 5. We also tried limiting the sample to the one-third of individuals who died between 1995 and 2000, for whom utilization subsequent to 2000 is not possible.

Finally, it is worth noting that our analysis thus far has focused exclusively on nursing home utilization. Increasingly, long-term care insurance policies also cover some home health care, although, it

[14] We verified that the results presented in the remainder of the paper were also not substantively affected by limiting the sample to the top quartile of the income or wealth distribution, or to individuals unlikely to be denied insurance, although in some specifications the standard errors increased so that the results in these smaller samples were no longer statistically significant.

is a small component (about one-quarter) of long-term care expenditures (US Congress, 2000). 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 sample did, compared to 19% for nursing home use alone). When we re-estimate equation (2) using as a dependent variable whether the individual used *any long-term care*, we find a relationship between insurance coverage and *any long-term care utilization* that is slightly more negative than the relationship between insurance coverage and *any nursing home utilization* (results not shown). 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 negative. However, in the proprietary insurance company data discussed in the previous subsection, we find some weak evidence, for policies that also cover home health care, of a positive correlation between the amount of insurance in the contract and the hazard rate of entering the 100th consecutive day of home care (results not shown).

In contrast to the results for nursing home use, there may therefore be some weak empirical support for a positive relationship between insurance coverage and home care use. We suspect that any such evidence reflects the fact that home care – unlike nursing home care – may provide some direct consumption value, and therefore be more susceptible to moral hazard than the less desirable nursing home care. It is unlikely that our failure to find a positive correlation between insurance coverage and nursing home use is due to insurance-induced substitution of home health care use for nursing home use. For we find that the negative relationship between insurance coverage and nursing home care in both data sets persists when we restrict the insured sample to those with no home health care coverage. This is consistent with studies based on quasi-experimental evidence that conclude that home health care does not serve as a substantial substitute for nursing home care. (Kemper, 1988; McKnight, 2002).

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

The preceding section indicates no evidence of a positive correlation between insurance coverage and nursing home utilization. As discussed, this result is consistent with there being symmetric information in the market, and similar evidence in other insurance markets has been interpreted as such. However, the result is also consistent with individuals' having private information about both their risk type and their preferences. We therefore test directly for asymmetric information by examining whether individuals have private information about their risk type and whether this information is related to insurance coverage.

4.1 Measuring individual's beliefs about their risk of nursing home utilization

Information on beliefs about the individual's 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 interview was completed by a proxy respondent; this excludes, among others, the most cognitively impaired; the results in Table 4 are robust to this sample restriction.

An important consideration is whether individuals' reporting of their beliefs contains any meaningful information about their actual beliefs. Two factors are encouraging on this dimension. First, individuals' predictions appear right on average: the average self-reported probability of nursing home use in a five year period was 18 percent, while 16 percent of the responders actually did enter a nursing home over the five year period.[15] Second, we find that self-reported nursing home entry probabilities co-vary in sensible ways with known risk factors; they are higher for women than for men, and increase monotonically with age and with deteriorating health status. These results are consistent with other work that has found

[15] 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 50.

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" or "categorical" responses wherein respondents give round figures such as 0, 50 or 100 percent. Figure 2 shows a histogram of the responses; almost 50 percent of respondents report a five-year nursing home entry probability of zero; 14 percent report a 50 percent probability, and about 1 percent report a probability of 100 percent. Hurd and McGarry (1995) and Gan et al. (2003) report a similar preponderance of such categorical responses for self-reported mortality probabilities.

It is somewhat unclear how to treat these categorical responses. Our goal is to measure individual beliefs. To the extent that the categorical responses represent the "true" subjective probability of the individual, no adjustment to individuals' statement of their beliefs seems warranted. However, the preponderance of categorical responses raises the possibility that individual responses convey information about their beliefs of the general nature of their risk (e.g. low, medium, or high) but not about the *scale* of the risk. For example, about 8 percent of individuals who report a zero probability of 5-year nursing home entry have private long-term care insurance; this suggests that an answer of zero may convey a belief that the entry probability is quite low, but not that it is literally zero. In this case, grouping the individual predictions into several categories, rather than including them as a continuous variable, may be more appropriate.

With these considerations in mind, we report results use two alternative measures of the individual's beliefs. First, we use the actual response of the individual, which we refer to as our "continuous measure" of individual beliefs. Second, we use a series of indicator variables for whether the individual reported 0 (49%), 1-49 (30%) or 50-100 (21%). We chose these break points to create categories of roughly equal size. We refer to this as our "categorical measure."

Finally, we note that individuals may not be comfortable reporting probabilistic answers, and may not in fact even think in these terms. If they use probabilistic information in making insurance purchase decisions, but are unable to translate these latent probabilities into numbers when faced with a survey question, the resulting measurement error will lead us to underestimate the extent of private information.

4.2 Do individuals have private information about the likelihood of nursing home utilization?

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

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

We estimate a linear probability model of whether the individual went into a nursing home in the five years between 1995 and 2000 (CARE) on his 1995 self-reported beliefs of this probability (B) and controls for the insurance companies' risk classification (X). We do not control for LTCINS in equation (3) because if private information about risk type is correlated with insurance coverage due to adverse selection, controlling for LTCINS would control away part of the individual's information. In practice, our results are not affected by this choice.

The results are shown in Table 5. Two main findings emerge. First, using either the continuous or categorical measure of beliefs, columns (1) and (2) indicate that individual beliefs about the likelihood of entering a nursing home are a statistically significant, positive predictor of subsequent nursing home experience. This provides a complement to studies that have found that individuals have some ability to predict their mortality (e.g. Hurd and McGarry, 1995, 2002; Smith et al. 2001). The results in column 1 indicate that a 10 percentage point increase in self-reported probability is associated with a 1 percentage point increase in the probability of going into a nursing home. The results in column 2 indicate that individuals who report a prediction of 50 or higher are 6 percentage points (about 40 percent) more likely to go into a nursing home than individuals who report a prediction of 0; individuals who report a prediction of 1 to 49 are no more likely to go into a nursing home than individuals who report 0, but are significantly less likely to go into a nursing home than individuals who report a prediction of 50 or higher.

Second, and most importantly, the results in the remaining columns indicate that the individual still has residual private information about his risk type *conditional on the risk class that the insurance* *company assigns to the individual*. No matter what set of controls for risk classification or measure of the individual's beliefs is used, the individual's beliefs remain a positive, and statistically significant predictor of subsequent nursing home utilization. Because of concerns about possible measurement error in individual beliefs, we also tried adding to the regression the individual's response to the same question about beliefs asked in the previous interview (1993). We found that this additional measure of beliefs – as well as the 1995 measure – is always a statistically significant predictor of nursing home use (results not shown). This suggests that measurement error in individual beliefs is indeed an issue and that the results in Table 5 are likely underestimates of individuals' private information.

These findings thus provide direct evidence of the presence of asymmetric information in the private long-term care insurance market. In particular, the results are supportive of the assumption of adverse selection models that individuals have private information about their risk type prior to the purchase of insurance. An alternative interpretation of the results 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 results look similar if we restrict the sample to the 90% of individuals without private insurance.[16]

If individuals have residual private information about their chances of using a nursing home, why don't insurance companies attempt to collect additional information about the individual? There is clearly some information about the individual that the insurance company could in principle observe but that in practice it does not. This includes, 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, sex, and proximity of the individual's children, and whether the individual engaged in each of a variety of potential preventive health measures (described in more detail in Section 5). When we add these variables to the "all observables" specification in Table 5, they are jointly

[16] 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. 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.

significant, but their addition does not affect the magnitude or statistical significance of the coefficient on the individual's prediction. This suggests that feasible collection of additional information about the individual would not correct the problem of asymmetric information vis a vis the consumer's information set, but it would give the company an advantage over competitors that do not collect the information.[17]

4.3 The relationship between private information about risk and insurance coverage

Table 5 demonstrates that individuals have private information about their risk of entering a nursing home. We next examine the role of this private information in affecting the purchase of long-term care insurance. We estimate the equation:

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

Once again, LTCINS is a binary measure of insurance coverage, and B and X are, respectively, the subjective probability of entering a nursing home and controls for the risk classification done by the insurance company. The results are presented in Table 6. Across all specifications, individuals who believe that they are higher risk are more likely to purchase insurance.

Table 6 raises the interesting question of whether the private information occurs ex-ante (as it would in the case of adverse selection) or ex-post (as it would in the case of moral hazard). In other words, do individuals have ex-ante private information that influences subsequent purchase decisions, or do individuals who purchase insurance incorporate into their beliefs the effect that the insurance will have on their risk reducing behavior or their consumption of care? The AHEAD data do not permit us to distinguish between these alternatives because the average initial age in our sample is 79, and hence subsequent insurance purchases are rare. However, evidence from younger cohorts in the Health and Retirement Survey suggests that individuals have private information about risk type *prior* to purchasing insurance, and that such information influences subsequent insurance purchases. Specifically, we found

[17] We presume that insurance companies do not collect this information because the costs of doing so are high relative to the benefits. In addition, the use of behavioral information in pricing insurance contracts (such as decisions regarding preventive health care investment) could alter the behavioral choices of potential applicants and thus reduce the informative content of these characteristics.

(in results not reported) that beliefs about subsequent nursing home use in 1996 among *uninsured* individuals aged 60 to 69 (average age of 63) were positively and statistically significantly associated with *acquiring* insurance by 2000 (which 8 percent of the sample did); the magnitude of the coefficient on beliefs as a predictor of subsequent insurance coverage was about two-thirds the magnitude of the contemporaneous correlation between beliefs and insurance coverage shown in Table 6.

4.4 The effect of private information on the market equilibrium

The combined evidence from Tables 5 and 6 indicates that individuals have private information about their risk type *and* that this private information is positively correlated with insurance coverage. In order to reconcile these findings with the results in Table 4 – which indicate that the insured are no more likely to enter a nursing home than the uninsured – there must, mechanically, be some other unobserved characteristic of the individual that has the opposite correlation with insurance coverage and care utilization. Therefore, to isolate the relationship between care utilization and the component of insurance coverage that is explained by individual beliefs about risk, we re-estimate equation (2) using the individual's subjective assessment of nursing home risk as an instrument for insurance coverage. It is important to emphasize that we are not using individual beliefs as an instrument in the usual sense (to recover a causal parameter on insurance coverage) but rather, we employ this technique to obtain the reduced form correlation between care utilization and the variation in insurance coverage explained by individuals' beliefs about their risk type. The IV estimates thus provide a partial equilibrium answer to the question of what the equilibrium relationship between insurance coverage and nursing home use would look like absent offsetting preference-based selection effects (i.e. if private information about risk were the sole systematic determinant of insurance coverage). Moreover, a comparison of the IV and OLS estimates of equation (2) allows us to gauge whether offsetting preference-based selection effects are substantively and statistically significant.

Table 7 reports the OLS and IV estimates of equation (2). Not surprisingly, given the results in Tables 5 and 6, the IV estimates consistently indicate a positive correlation between care utilization and the portion of long-term care insurance explained by individuals' beliefs about their risk type. These estimated effects are significantly different from zero in nearly all specifications. If we include individuals' belief in 1993 as an additional instrumental variable along with the 1995 beliefs, the IV estimates are then statistically significant at at least the 10 percent level in all specifications (results not shown).

The results in Table 7 indicate that the market for long-term care insurance would be substantially more adversely selected if individual beliefs about risk type were the sole determinant of long-term care insurance coverage. The IV estimates suggest that, holding prices constant, the insured would be at least 20 percentage points more likely to enter a nursing home than individuals without insurance. Of course, such a change in the risk composition of the insured would also affect the pricing of insurance. One way to gauge how insurance prices would change in the absence of preference-based selection is to compare OLS and IV estimates of the relationship between total nights spent in a nursing home and insurance coverage. In results not reported here, we find that the OLS estimates suggest that the insured spent an (insignificant) 5 to 14 days *less* in a nursing home over a five year period than the uninsured, while the IV estimates (using beliefs about risk type as instruments) suggest that the insured spend a (significant) 40 to 212 days *more* in a nursing home than the uninsured (median estimate is 69). The national average daily cost of a nursing home in 2002 was $143 (MetLife, 2002). Therefore, if we assume the OLS estimate is 0 and take the median IV estimate, this suggests that in the absence of preference based selection, the expected (non-discounted) nursing home expenditures of an insurance policy would rise by almost $10,000 over a five year period. Presumably the general equilibrium effects of these price changes on selection in the long-term care insurance market (and hence further effects on prices) would result in a market that, in the absence of offsetting preference-based selection, would be even more adversely selected than our partial equilibrium estimates in Table 7 imply.

To verify that these unobserved factors are statistically important we conduct a Hausman test of the difference between the IV and OLS estimates; this provides a test of the null hypothesis that in the OLS estimates in equation (2), LTCINS is not correlated with ε. The results of the test, reported in table 7, indicate that, for seven out of the eight IV estimates reported, we can reject with at least 10 percent confidence the null hypothesis that the OLS estimates of equation (2) are consistent. This suggests the presence of statistically significant factors not controlled for in equation (2) that have the *opposite* correlation with insurance coverage and care utilization. We now turn to a direct examination of what these offsetting preference-based factors might be.[18]

5. Direct evidence of preference-based selection

We exploit the richness of the AHEAD data to identify characteristics of the individual that are unobserved by the insurance company and that have the opposite correlation with risk occurrence and preferences for insurance. Econometrically, 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.[19] We therefore estimate the equations:

$$(5) \qquad \mathrm{CARE} = Xb_1 + \mathrm{b}_2\mathrm{LTCINS} + b_3\mathrm{B} + b_4\mathrm{P} + \varepsilon$$

$$(6) \qquad \mathrm{LTCINS} = Xd_1 + d_3B + d_4P + \eta$$

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 the components of P. By definition,

[18] One possibility is that differences between the OLS and IV results in Table 7 stem primarily from classical measurement error in our long-term care insurance variable. However, we believe this to be unlikely. We re-estimated equation (2) using long-term care insurance coverage in 1998 as an instrumental variable for coverage in 1995. This is more typical of the type of instrument selected to deal with classical measurement error. When doing so we found IV results that were only slightly different from the OLS results (never varying by more than 1.6 percentage points) and about half the time are *more negative* than the OLS estimates.

[19] 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. This implies that the individual is inefficient in forming his beliefs in that he fails to use all available information; it does not imply that his beliefs must exhibit any systematic biases or mistakes. Consistent with this, Smith et al. (2001) find that individuals do not use all of the available information efficiently in formulating beliefs about their mortality prospects, although their responses are not biased on average.

they must be unobserved by the insurance company; many of them are therefore likely to be unobserved by the econometrician as well. We therefore focus on a candidate that has attracted considerable theoretical attention: selection based on risk aversion. De Meza and Webb (2001) and Jullien et al. (2002) propose that the risk averse (or more "cautious") may not only place a higher value on insurance, but may invest more in preventive effort and thus end up lower risk than less risk averse individuals.

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. The insurance company applications we reviewed did not solicit any of this information.

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 engaging in all relevant preventive behaviors. This measure does not appear to primarily reflect whether the individual has seen a doctor over the last two years (over 90% of our elderly sample has) or the type of insurance the individual has; Medicare (which covers 99% of our sample) reimburses for all of these preventive health measures.

Table 8 reports the results of estimating equations (5) and (6). To conserve space, we only report results using the continuous measure of individual's beliefs; the results look similar if categorical measures of beliefs are used instead. The results from the first column within each panel indicate that individuals who undertake a greater fraction of potential preventive health activity (i.e. more cautious individuals) are more likely to own insurance. The results from the second column within each panel indicate that those who undertake more preventive health activity are also less likely subsequently to go into a nursing home.

We used the results from estimating the insurance coverage equation (6) to decompose insurance coverage into a component predicted by the preventive health activity ($\hat{\delta}_4 P$, which we denote as PREVENT_HAT), a component predicted by individuals' private information about risk type ($\hat{\delta}_3 B$, which we denote as RISKTYPE_HAT) and the residual ($\hat{\eta}$, which we denote by RESID_HAT).[20] The third column of each panel of Table 8 shows the results of estimating the care utilization equation (5) with these three different components of insurance coverage on the right hand side. 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.

Thus, consistent with the theoretical models of de Meza and Webb (2001) and Jullien et al (2002), more cautious individuals are both more likely to own insurance and less likely to experience the insured risk. It is not clear, however, whether our results reflect the type of causal relationship between risk aversion and risk occurrence posited by these theoretical papers. It is possible that there is a causal relationship; for example, flu shots reduce the risk of pneumonia which is a nontrivial contributor to the need for a nursing home for the elderly. Alternatively, these preventative measures may be correlated with other investments that themselves cause lower rates of institutionalization. For example, we find that individuals who invest more in our measured preventive health activities are substantially less likely to have a hip fracture, another important contributor to nursing home residence. This is unlikely to reflect a causal effect of any of our measured preventive health activities – flu shots and mammograms are not thought to reduce the probability of a hip fracture – but may reflect a causal effect of other unmeasured preventive health activities (such as greater exercise and calcium consumption in earlier years) and a positive correlation across preventative activities.

Another possibility is that our preventive health measures proxy for unmeasured aspects of socio-economic status. Indeed, we find a strong positive correlation between our measure of preventive health and individuals' wealth, which most insurance companies do not observe. Moreover, we find that higher

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

asset individuals are both substantially more likely to have long-term care insurance (which may reflect the crowd-out of Medicaid on insurance demand among the less wealthy) and substantially less likely to use nursing homes. Although the results in Table 8 are robust to adding controls for the individual's asset quartile to the three specifications that do not already have them, we cannot rule out the possibility that our preventive health measure is proxying for unmeasured aspects of socio-economic status.

Interestingly, other than preventive health activity and wealth, other characteristics of the individual that we can measure and that the insurance companies do not observe appear to have the *same* correlation with insurance coverage and risk occurrence. They are therefore not potential contributors to offsetting, preference-based selection. 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. Relatedly, if we re-estimate the relationship between care utilization and insurance coverage in equation (2) adding as controls all of the factors that we observe and that the insurance company does not, 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. 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 papers have found no evidence of a positive correlation between the amount of insurance and the occurrence of the risk, and have concluded that asymmetric information may therefore not exist in these markets. 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 find no evidence of a positive correlation between individuals' insurance coverage and their consumption of nursing home care in several complementary data sources. However, using information about individuals' assessments of their nursing home risk, we find direct evidence of asymmetric information. After conditioning on the information set of the insurance company, the individual's beliefs about his risk type are positively and statistically significantly correlated with both subsequent care utilization and insurance coverage.

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 – 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 positive correlation between insurance coverage and risk occurrence that asymmetric information about risk type would tend to produce. Indeed, we demonstrate that the private long-term care insurance market would be substantially adversely selected (i.e. those with insurance would be substantially higher risk than those without insurance) in the absence of preference-based selection. The fact that the long-term care insurance market is not adversely selected indicates that, due to the offsetting preference-based selection, private information about risk type in the long-term care insurance market does not raise prices above their population-average actuarially fair price. However, a market with such an information structure will not provide an efficient amount of insurance coverage relative to the first best. An unanswered question – and an importance avenue for further work – is whether asymmetric information in the long-term care insurance market is an important contributor to the extremely limited size of this private insurance market.

The results in this paper suggest two other interesting directions for further work. First, the evidence of offsetting, preference-based selection in the long-term care insurance market suggests a potential unifying explanation for the apparent differences across insurance markets in whether the insured are above-average in their risk type. For example, there is evidence of a positive correlation between insurance coverage and risk occurrence in annuities (Finkelstein and Poterba, 2000, 2002) but not in life insurance (Cawley and Philipson, 1999), which insures the (opposite) longevity risk to that insured by annuities. It may be that preference-based selection operates in the opposite direction in these two markets. Characteristics of the individual that the insurance company does not observe – such as their level of caution or their wealth – may be positively correlated with demand for both annuities and life insurance, but *negatively* correlated with the life insurance risk of dying and *positively* correlated with the annuity risk of living.

Second, the results suggest an alternative, general approach to testing for asymmetric information in insurance markets. Conditional on the information set of the insurance company, the existence of an individual characteristic that is correlated with both insurance coverage and risk occurrence indicates the presence of asymmetric information. This is true regardless of the sign of the correlation. Nor does this test require the type of information about individuals' subjective assessments of their risk occurrence used in this paper. The test only requires that the econometrician observe more information about the individual than the insurance company observes, which may often be the case. Information that is costly to verify may not be collected by insurance companies – since individuals would have an incentive to lie if it were used in pricing – but might be collected by a general survey, where such incentives do not operate. For example, although annuity companies do not collect information on individuals' wealth (despite the known correlation between socio-economic status and mortality), such information is available in many public use surveys. This type of disparity between the incentives for truthfully reporting information to insurance companies and to data surveyors suggests that the test may find widespread application.

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

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.

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.*

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.

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.

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.

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.

Merlis, Mark. 2003. “Private long-term care insurance: who should buy it and what should they buy?” Kaiser Family Foundation, March. http://www.kff.org/content/2003/6072/

MetLife Mature Market Institute. 2002. “MetLife Market Survey of Nursing Home and Home Care Costs 2002.”

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.

Robinson, James. 1996. “A Long-Term Care Status Transition Model,” presented at the 1996 Bowles Symposium on the Old-Age Crisis: Actuarial Opportunities. Georgia State University, September.

Robinson, James. 2002. “A long-term care status transition model.” Mimeo, University of Wisconsin, Madison.

Rothschild, Michael and Joseph Stiglitz. 1976. “An Essay on the Economics of Imperfect Information.” *Quarterly Journal of Economics* 90: 629-649.

Sloan, Frank and Edward Norton. 1997. “Adverse selection, bequests, crowding out, and private demand for insurance: evidence from the long-term care insurance market.” *Journal of Risk and Uncertainty* 15: 201-219.

Smith, V Kerry, Donald Taylor and Frank Sloan. 2001. “Longevity expectations and death: can people predict their own demise?” *American Economic Review* September pp.1126-1134.

Society of Actuaries. 1992. “Report of the long-term-care experience committee: 1985 national nursing home survey utilization data.” *Transactions of the Society of Actuaries*, 1988-89-90 Reports. Pp.101-164.

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

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.*

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

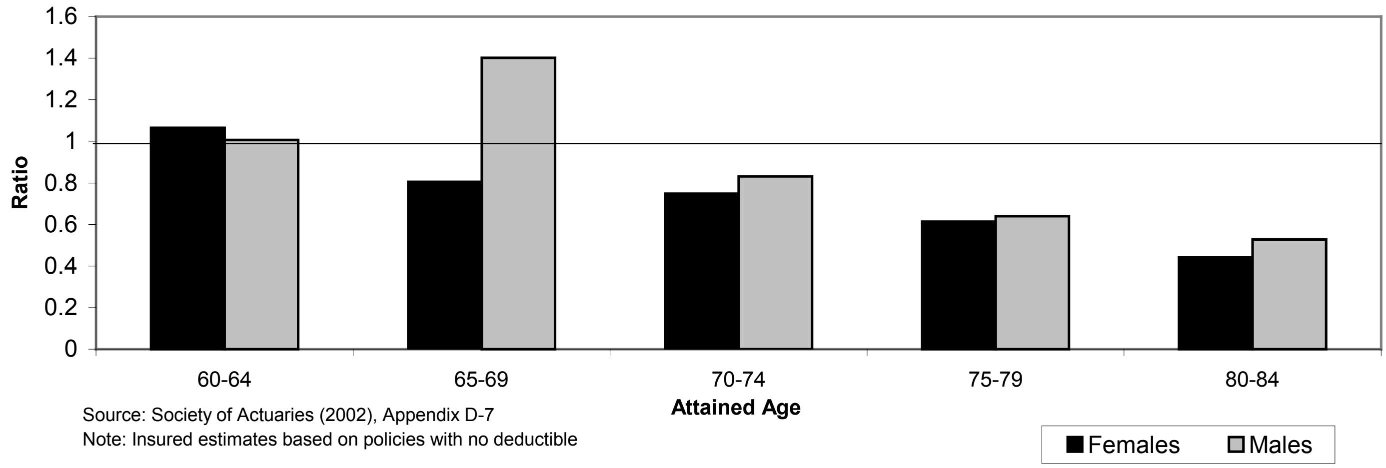

Figure 2: Distribution of the Subjective Probability of Entering a Nursing Home within Five Years

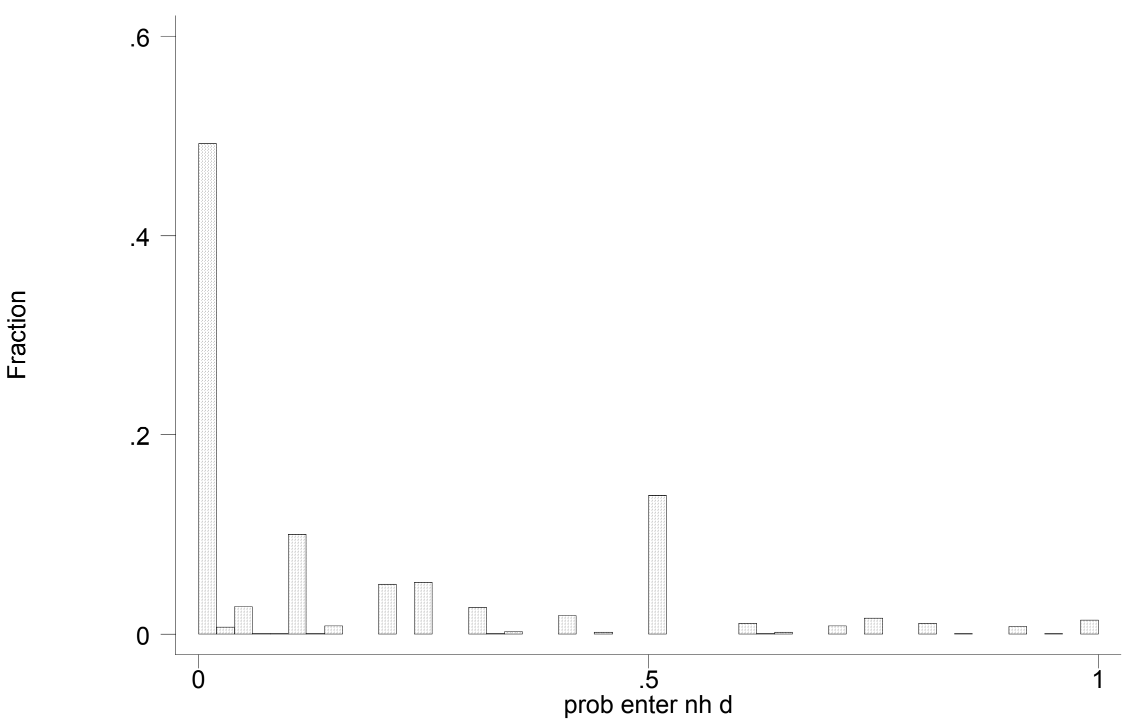

Table 1
Summary Statistics for Proprietary Insurance Company Data

	Policies Issued 1997-2001	Policies Issued 1997 or 1998
Failure rate	0.3	0.6
Risk Classification of Individual:		
Median issue date	September 1, 1999	February 8, 1998
Average issue age	64.4	65.3
Percent rated low risk	29	16
Percent rated standard risk	66	79
Percent rated high risk	5	5
Policy Characteristics:		
Deductible		
Percent with 20-day deductible	5	5
Percent with 60-day deductible	8	8
Percent 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
Percent With Home Care Benefit < than Nursing Home Benefit	19	26
Benefit Period		
Percent of policies with Unlimited Benefit Period	18	18
Average Benefit Period for Policies w/ limited benefit period	4.3	4.2
Percent with Limited Benefit Period that allow extension	10	13
Benefit Escalation		
Percent with no benefit escalation	2	----
Percent with 5% "simple" benefit escalation	30	29
Percent with 5% compound benefit escalation	28	18
Percent with "indexed" escalation	40	53
Number of observations	144,798	49,887

"Failure rate" denotes the percentage of the sample who experience at least 100 continuous day of nursing home care during the sample period. Dashed line indicates less than 1 percent. 60% of policy sales are to women; we do not report this in the above table because it is not a characteristic used by the insurance company to categorize individuals.

Table 2
Hazard of Receiving Nursing Home Care for 100th Consecutive Day

Covariates in Regression	Policies Issued 1997 - 2001		Policies Issued 1997 or 1998		Prediction w/ "positive correlation"
	Coeff	Std err	Coeff	Std err	
Issue Age Category:					
Age < 60 (omitted)	--	--	--	--	
Age 60-64	1.199***	(0.423)	1.039**	(0.505)	
Age 65-69	1.729***	(0.423)	1.798***	(0.475)	
Age 70-74	2.944***	(0.400)	2.928***	(0.469)	
Age 75+	4.010***	(0.403)	3.913***	(0.473)	
Rating Category:					
High risk (omitted)	--	--	--	--	
Rated low risk	-1.100***	(0.259)	-0.964***	(0.322)	
Rated standard risk	-0.535***	(0.175)	-0.562***	(0.200)	
Deductible:					
100-day deductible (omitted)	--	--	--	--	
60-day deductible	0.024	(0.208)	-0.030	(0.252)	Positive
20-day deductible	0.233	(0.238)	0.312	(0.268)	Positive; > 60-day
Daily Benefit:					
Daily benefit ≤ $100 (omitted)	--	--	--	--	
Daily Benefit = $100	0.095	(0.127)	-0.007	(0.141)	Positive
Daily Benefit > $100	0.240*	(0.134)	0.143	(0.151)	Positive; > $100
Benefit Period:					
1-4 years, extension (omitted)	--	--	--	--	
1-4 years, possible extension	-0.306	(0.207)	-0.509**	(0.254)	Positive
5+ years, no extension	-0.391**	(0.162)	-0.543	(0.193)	Positive
5+ years, possible extension	-0.160	(0.343)	-0.257	(0.389)	Positive; > 5+ no ext
Unlimited	0.168	(0.153)	0.075	(0.175	Positive; > 5+ w/ ext
Escalation of Benefits:					
5% compound (omitted)	--	--	--	--	
No escalation	0.438	(0.399)	----		Negative
5% "simple" escalation	0.213	(0.244)	0/270	(0.302)	Negative; > no esc
"Index option" for escalation	0.102	(0.236)	0.254	(0.288)	> no escalation
Failure Rate	0.3%		0.6%		
N	144,798		49,888		

Note: Estimation from a Cox proportional hazard model. Also included are: indicators of issue year, whether the policy is tax qualified, frequency of policy premium payments, whether the policy has 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), and whether the home health care benefits are lower than (rather than equal to) the nursing home benefits. We do not include the home health care daily benefit separately because the correlation between daily benefit for home health care and daily benefit for nursing home is 0.92.

Table 3: Means of Variables Used in the Analysis

Variable	Whole Sample	Insured	Uninsured
Any NH Utilization (1995 – 2000)	0.187	0.146	0.191
Total # of nights in NH (1995 – 2000)	32.7	20.67	33.9
Long-term care insurance coverage (1995)	0.103	1.00	0.00
Demographics (1995)			
Age	78.6	77.4	78.8
Female	0.63	0.61	0.64
Married	0.54	0.60	0.54
Spouse's age (if married)	73.8	73.3	73.9
Household Assets (median)	138,000	218,000	130,100
Household Income (median)	18,000	25,000	17,000
Current Health (1995):			
ADL limitation: bathing	0.11	0.08	0.12
ADL limitation: eating	0.05	0.04	0.05
ADL limitation: dressing	0.13	0.09	0.14
ADL limitation: toileting	0.08	0.07	0.08
ADL limitation: walking	0.10	0.06	0.10
Incontinence	0.22	0.25	0.21
Cognitively impaired	0.03	0.02	0.03
Use wheelchair	0.03	0.03	0.03
Use walker	0.07	0.03	0.08
Use crutches	0.003	0.002	0.004
Use cane	0.13	0.09	0.13
Use oxygen	0.01	0.008	0.01
Regularly use prescription drugs	0.79	0.82	0.78
IADL limitation: grocery shopping	0.15	0.11	0.15
IADL limitation: managing medication	0.05	0.03	0.05
Low BMI	0.10	0.09	0.10
High BMI	0.13	0.09	0.14
Currently smoke	0.08	0.09	0.08
Health History (1995 and before):			
Home Health Care Use	0.17	0.13	0.17
Nursing Home Use	0.02	0.03	0.02
Depression	0.19	0.14	0.20
Drinking Problem	0.03	0.04	0.03
Diabetes	0.14	0.12	0.14
Diabetes treated with insulin	0.05	0.04	0.05
Kidney Failure Assoc w. Diabetes	0.02	0.01	0.02
Stroke	0.12	0.09	0.12
Heart condition	0.34	0.32	0.34
Medication for heart problem	0.22	0.20	0.22
Heart Attack	0.09	0.08	0.09
Congestive Heart Failure	0.04	0.04	0.04
High Blood Pressure	0.54	0.59	0.53
Hip fracture	0.05	0.05	0.05
Lung Disease	0.12	0.13	0.12
Cancer	0.16	0.17	0.16
Psychiatric problems	0.15	0.16	0.14
Arthritis	0.54	0.49	0.54
Injury from falling	0.15	0.15	0.15

Note: All means are weighted. See Appendix A for our construction of cognitive impairment, depression, drinking problem, 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 Utilization in the AHEAD Data

Dependent Variable	No Controls (1)	Controls for Age Dummies (2)	Controls for "all observables" (3)	Controls for insurance company prediction (4)
Entire Sample				
Any nursing home utilization	-0.045*** (0.016) [N=6,280]	-0.016 (0.015) [N=6,280]	-0.001 (0.015) [N=6,083]	-0.012 (0.015) [N=6,275]
Number of nights spent in nursing home	-71.589*** (25.774) [N=6,189]	-28.853 (25.099) [N=6,189]	-15.673 (25.241) [N=5,998]	-30.067 (24.762) [N=6,181]
Restricted to those for whom Medicaid is not a close substitute for private insurance				
Any nursing home utilization	-0.045** (0.018) [N=2,161]	-0.032* (0.018) [N=2,161]	-0.028 (0.017) [N=2,123]	-0.026 (0.017) [N=2,161]
Number of nights spent in nursing home	-91.087*** (35.391) [N=2,140]	-64.487* (34.347) [N=2,140]	-64.626* (33.063)* [N=2,103]	-70.576** (34.183) [N=2,140]

Notes: Each cell reports the coefficient on LTCINS from estimating equation (2) 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. When the dependent variable is "any nursing home utilization", reported coefficients are from a linear probability model. When the dependent variable is "number of nights spent in nursing home", the reported coefficients 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. Restricted sample ("those for whom Medicaid is not a close substitute") consists of those in the top quartile of the distribution of income or wealth.

Table 5
Individuals' predictions of nursing home entry

	Control Variables								
	No Controls		Age Dummies		"All observables"		Insurance Company Prediction		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual's Prediction									
Continuous measure	0.097*** (0.024)		0.073*** (0.023)		0.041* (0.022)		0.044** (0.022)		
Categorical measure									
Predicts 0 (omitted)		--		--		--		--	
Predicts 1 to 49		-0.019 (0.012)		-0.004 (0.011)		0.003 (0.011)		-0.004 (0.011)	
Predicts 50 to 100		0.062*** (0.015)		0.047*** (0.015)		0.032** (0.014)		0.033** (0.015)	
Actuarial Prediction							0.501*** (0.029)	0.498*** (0.029)	0.507*** (0.028)
R^2	0.004	0.007	0.104	0.104	0.169	0.169	0.100	0.100	0.099
N	5,072	5,072	5,072	5,072	4,960	4,960	5,072	5,072	5,072

Note: Reported coefficients are from linear estimation of equation (3). Dependent variable is whether individual enters nursing home over subsequent five years. "Continuous measure" of individual's prediction uses the individual's reported prediction, rescaled to range from 0 to 1. When categorical measure of beliefs is used instead, the omitted category is "individual predicts 0". The column headings describe the additional covariates included in the regression. Heteroskedacticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 6
The Relationship between Insurance Coverage and Individuals Beliefs about Risk Type

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

Note: Reported coefficients are from linear estimation of equation (4). Dependent variable is whether the individual has LTC insurance. See notes to table 5 for more details.

Table 7
Decomposing the Relationship between Utilization of Nursing Home Care and Insurance

	Control Variables											
	No Controls			Age Dummies			"All observables"			Insurance Company Prediction		
	OLS	IV	IV	OLS	IV	IV	OLS	IV	IV	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
LTCINS	-0.046*** (0.015)	1.079*** (0.356)	0.262 (0.176)	-0.023	0.767*** (0.288)	0.324* (0.172)	-0.007 (0.015)	0.435* (0.251)	0.335* (0.190)	-0.017 (0.014)	0.428* (0.233)	0.231 (0.164)
Hausman test p-value	------	<0.0001	0.07	------	0.001	0.04	-----	0.06	0.06	-------	0.04	0.12
Instruments	-------	Continuous Beliefs	Categorical Beliefs	-------	Continuous Beliefs	Continuous Beliefs	-------	Continuous Beliefs	Continuous Beliefs	-------	Continuous Beliefs	Categorical Beliefs
N	5,072	5,072	5,072	5,072	5,072	5,072	4,960	4,960	4,960	5,072	5,072	5,072

Notes: Reported coefficients are from estimates are equation (2). The dependent variable is "any nursing home entry". The column headings describe the additional covariates included in the regression as well as whether the estimation is OLS or IV. Heteroskedacticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Table 8: Preference-based Selection

	Control Variables											
	No Controls			Age Dummies			"All Observables"			Insurance Company Prediction		
	LTCINS	NH ENTRY	NH ENTRY	LTCINS	NH ENTRY	NH ENTRY	LTCINS	NH ENTRY	NH ENTRY	LTCINS	NH ENTRY	NH ENTRY
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Preventive Activity	0.064*** (0.016)	-0.111*** (0.019)		0.051*** (0.016)	-0.036** (0.018)		0.013 (0.017)	-0.019 (0.019)		0.051*** (0.016)	-0.055*** (0.018)	
Individual Prediction	0.087*** (0.020)	0.102*** (0.024)		0.091*** (0.120)	0.072*** (0.023)		0.095*** (0.020)	0.040* (0.022)		0.099*** (0.020)	0.048** (0.023)	
PREVENT_HAT			-1.433*** (0.226)			-0.708*** (0.244)			-0.434 (0.349)			-0.771*** (0.251)
RISKTYPE_HAT			1.122*** (0.270)			0.781*** (0.246)			0.405* (0.232)			0.460** (0.228)
RESID_HAT			-0.040*** (0.015)			-0.022 (0.015)			-0.007 (0.015)			-0.013 (0.015)
N	5,010	5,010	5,010	5,010	5,010	5,010	4,900	4,900	4,900	5,010	5,010	5,010

Note: All estimates are from a linear probability model. The column headings describe the additional covariates included in the regression and, below that, the dependent variable. Columns (1), (4), (7), and (10) report the results from estimating equation (6). Columns (2), (5), (8), and (11) report the results from estimating equation (5). "Preventive activity" measures the fraction of gender-appropriate preventive health activity undertaken by the individual. All regressions also include a control for sex because the fraction of potential preventive activity undertaken may vary with sex simply because the *number* of potential preventive activities is 3 for men and 5 for women. "Individual prediction" measures the individual's continuous prediction. PREVENT_HAT, RISKTYPE_HAT and RESID_HAT are generated based on the results from estimating the insurance coverage equation (6) and reflect the variation in insurance coverage that is explained, respectively, by "Preventive activity", the individual's beliefs about his risk type, and the variation 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.

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.[21] To increase sample size, we include observations on the spouses of sample members 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 who have a zero respondent-level weight. The regression results are not sensitive to the choice of weights.

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?

[21] 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, since there are about 35 million individuals aged 65 and over in the United States (U.S. Bureau of the Census, 2000). (Although people of any age may hold long term care insurance, it tends to be held by the elderly (HIAA 2000a)).

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.

<u>A note on other covariates</u>

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 interviews, 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.[22] 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, 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. This is of course a substantial overestimate of the percentage of individuals who would be ineligible for insurance since we are excluding individuals – whose average age is 79 – if they *currently* have any health conditions that might result in ineligibility. However, they may not have had these conditions at the younger age at which they purchased insurance. Indeed, our "ineligible" individuals are only about 20 percent less likely to have private insurance than our eligible individuals. We choose an overly restrictive definition of eligibility, however, because we are more concerned with avoiding including ineligibles in our eligible sample than vice versa.

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