

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

Amy Finkelstein and Kathleen McGarry^{*}

Abstract: We demonstrate the existence of multiple dimensions of private information in the long-term care insurance market. Two types of people purchase insurance: individuals with private information that they are high risk and individuals with private information that they have high insurance preferences. Ex-post, the former are higher risk than insurance companies expect, while the latter are lower risk. In aggregate, those with more insurance are not higher risk. Our results demonstrate that insurance markets may suffer from asymmetric information even absent a positive correlation between insurance coverage and risk occurrence. The results also suggest a general test for asymmetric information. (JEL D88, G22, I11)

Theoretical research has long emphasized the potential importance of asymmetric information in impairing the efficient operation of insurance markets. However, several recent studies in different insurance markets have found no evidence to support the central prediction of many asymmetric information models that those with more insurance should be more likely to experience the insured risk.¹

In this paper, we use a new method to test for the presence of asymmetric information in the long-term care insurance market in the United States. We use individuals' subjective assessments of the chance they will enter a nursing home to show that, conditional on the insurance companies' own assessment of the individuals' risk type, individuals have residual private information that predicts their eventual risk. Moreover, this residual private information is also

positively correlated with insurance coverage. Combined, these two findings provide direct evidence of asymmetric evidence in the long-term care insurance market.

Yet despite this result, the same data provide no evidence of a positive correlation between individuals' insurance coverage and their risk experience. A resolution of this apparent puzzle may be that individuals have private information about a second determinant of insurance purchase, their preferences for insurance coverage, as well as private information about their risk type. If individuals with private information that they have strong tastes for insurance are lower risk than the insurance company would predict, private information about risk type and private information about insurance preferences can operate in offsetting directions to produce an equilibrium in which those with more insurance are not more likely to experience the insured risk.

We provide some suggestive evidence of the nature of this offsetting private information about preferences in the long-term care insurance market. Specifically, we show that wealthier individuals and individuals who exhibit more cautious behavior – as measured either by their investment in preventive health care or by seat belt use – are both more likely to have long-term care insurance coverage and less likely to use long-term care; neither of these characteristics is used by insurance companies in pricing long term care insurance.

Our findings have several important implications for understanding the impact of asymmetric information on insurance markets. They demonstrate that the equilibrium in insurance markets with multiple types of correlated unobserved heterogeneity may look very different from what standard uni-dimensional models of asymmetric information about risk type alone would predict. Because unobserved preference heterogeneity can offset the positive correlation between insurance coverage and risk occurrence that private information about risk type alone would

produce, these findings also suggest that the widely used “positive correlation” test for asymmetric information can lead to incorrect conclusions.

This insight suggests an alternative approach for testing for asymmetric information in insurance markets that we expect will have wide applicability. Conditional on the information set used by the insurance company, the existence of *any* individual characteristic that is observed by the econometrician, but is not used by the insurer, and is correlated with both insurance coverage and risk occurrence, indicates the presence of asymmetric information that affects the payoffs to the parties to the transaction.

Our evidence of preference-based selection may also provide a unifying explanation for the disparities across different insurance markets in whether the insured appear to be above-average in their risk type. For example, there is evidence from several different countries, that annuitants are longer-lived (i.e. higher risk) than the general population while those with life insurance (which insures the opposite longevity risk) are also longer-lived (i.e. lower risk) relative to the general population (Cawley and Philipson 1999; Amy Finkelstein and James Poterba 2002, 2004; David McCarthy and Olivia S. Mitchell 2003). A possible explanation for the apparent selection differences between these markets is that preference-based selection may operate in the opposite direction for annuities and for life insurance. 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 forms of insurance, but *negatively* correlated with the life insurance risk of dying and *positively* correlated with the annuity risk of living. Such an explanation would be consistent with the existence of private information about risk type in both markets.

The remainder of the paper proceeds as follows. In Section I, we provide some conceptual background on the empirical predictions made by models of insurance markets with asymmetric information as well as institutional background on the private long-term care insurance market. Section II describes our data and outlines our empirical approach. Section III presents our findings. The final section concludes.

I. Background

A. Theoretical effect of asymmetric information on market equilibrium

Standard models of asymmetric information consider individuals who differ only in terms of their (privately known) risk type. A robust prediction of these models is that in equilibrium there will be a positive correlation between the amount of insurance and the occurrence of the risky event (Chiappori and Salanie, 2000; Dionne et al., 2001, Chiappori et al., forthcoming). This observation has motivated the standard test for asymmetric information used in the literature, which is to test for a positive correlation between the amount of insurance coverage and the ex-post occurrence of the (potentially) insured risk.²

This “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. In the case of adverse selection, the insured is assumed to have ex-ante private information about his risk type relative to what the insurance company knows; those with private information that they are high risk select contracts with more insurance than those with private information that they are low risk (see e.g. Michael Rothschild and Joseph Stiglitz, 1976). In the case of moral hazard, the causality is reversed and the informational asymmetry occurs ex-post: insurance coverage lowers the cost of an adverse outcome and thus increases the probability or magnitude of the risk occurrence (see e.g. Richard J. Arnott and Stiglitz 1988).

Recent theoretical work has enriched these standard asymmetric information models to allow for private information about preferences as well as about risk type. With these multiple forms of private information, a positive correlation between insurance coverage and risk occurrence may be neither a necessary nor sufficient condition for the presence of asymmetric information about risk type (Michael Smart 2000, Achim Wambach 2000, David de Meza and David C. Webb 2001, Bruno Jullien et al. 2002, Chiappori et al. forthcoming). For example, if unobserved preferences are positively correlated with insurance demand and negatively correlated with risk occurrence, then the correlation between insurance coverage and risk occurrence in equilibrium can be negative, despite the presence of asymmetric information about risk type.³

Just as with a single dimension of private information, an equilibrium with multiple forms of private information is unlikely to be efficient relative to the first best. This result applies even absent a positive correlation between insurance coverage and risk occurrence. Intuitively, consider the case in which private information about risk type and risk preferences offset each other to produce an equilibrium with no correlation between insurance coverage and risk occurrence. In this case one could well imagine two different groups of individuals purchasing a given insurance policy: low risk, high risk aversion individuals, and high risk, low risk aversion individuals. If these groups pay the same price for the insurance policy, but have different expected costs, both cannot pay an actuarially fair price, and the quantity of insurance purchased by at least one group will therefore not be first best. However, whether the equilibrium is inefficient relative to the second-best, or whether there is scope for Pareto improvement for government intervention, varies across models, just as in the case of uni-dimensional models of private information (Keith J. Crocker and Arthur Snow, 1985).⁴

B. Long-term care expenditure risk and private insurance

We focus on the long-term care insurance market for two main reasons. First, long-term care expenditure risk is among the largest financial risks faced by today's elderly, making an investigation of this market of particular interest. Annual expenditures on long-term care in the United States totaled \$135 billion in 2004, comprising over 8.5 percent of total health expenditures, or roughly 1.2 percent of GDP (Congressional Budget Office 2004). These expenses are distributed unevenly. For example, Christopher M. Murtaugh et al. (1997) estimate that while 60 percent of individuals who reach age 65 will never enter a nursing home, one-fifth of those who do will spend at least five years in the institution.

These figures suggest that there exist potentially large welfare gains from insurance to reduce this expenditure risk. Yet most of the risk is uninsured. Only about 10 percent of individuals aged 65 and over have private long-term care insurance, and most of the private policies provide only limited coverage (HIAA 2000a; Jeffrey R. Brown and Finkelstein, 2004a; Marc Cohen, 2003). As a result, only 4 percent of expenditures are reimbursed by private insurers while one-third are paid for out of pocket (Congressional Budget Office 2004, National Center for Health Statistics 2002).

Second, the long-term care insurance market is unusual among insurance markets in the United States in that during the period of our analysis, there was essentially no direct regulation of the prices charged or policies offered (NAIC 2002a, 2002b, Stephanie Lewis et al., 2003). In a regulated market it is more difficult to infer which aspects of the equilibrium are inherent to the market itself and which stem from regulatory constraints.⁵

The average age of purchase for long-term care insurance is 67 (HIAA, 2000a). Once purchased, policies are guaranteed renewable for the lifetime of the individual at a pre-specified constant nominal premium.

II. Empirical approach

To investigate the nature of private information in the long-term care insurance market, we draw on a rich dataset that contains the respondents' own assessments of their nursing home risk as well as data on ex-post risk occurrence, insurance coverage, precautionary activities, and a full set of health and demographic indicators that are sufficiently detailed to allow us to proxy the insurers' risk categorization. Our empirical strategy proceeds in three simple steps. First, we demonstrate that individuals have private information about their risk type and that this private information is positively correlated with insurance coverage. Second, we show that despite this private information, the equilibrium does not exhibit the positive correlation between insurance coverage and the use of long-term care predicted by uni-dimensional models of asymmetric information. These two sets of findings point mechanically to the existence of a second form of unobserved heterogeneity – heterogeneity in preferences – which offsets the risk based selection and obscures the expected positive correlation between insurance coverage and risk occurrence. In the third step of the analysis, we present direct evidence of the form of this offsetting preference based selection.

A. Data

Our individual-level survey data are from the Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS). At baseline this cohort is representative of the non-institutionalized population born in 1923 or earlier and their spouses. Appendix A provides more detail on the data and our sample. The average age of our respondents when we observe them in 1995 is 78, and 11 percent have long-term care insurance. These respondents are followed over time, allowing us to observe actual nursing home use from 1995 to 2000. Sixteen percent of our initial community-based sample enters a nursing home at some point during this five year period.

Central to our analysis is a measure of individual beliefs about nursing home use. We take advantage of a question in AHEAD that directly asks respondents about their perceived likelihood of nursing home use. The specific question we use from the 1995 survey is:

“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 of zero to 100, which we rescale to be between 0 and 1. Their responses may reflect information about their health status and thus their chance of needing nursing care. They may also reflect their alternatives to nursing care should they need it, such as the willingness of a spouse or child to take care of them, or their insurance coverage.

On several dimensions the subjective responses to this question appear reasonable. Individual predictions appear to be correct on average; the average self-reported probability of nursing home use over the five-year period is 18 percent, while 16 percent of the responders actually enter a nursing home over the five-year period. We also find that self-reported nursing home entry probabilities co-vary in sensible ways with known risk factors; they are higher for women than for men, and increase monotonically with age and with deteriorating health status. These results are consistent with other work that has found sensible covariance patterns for self-reported mortality probabilities and characteristics such as the individual’s age or health status (Daniel Hamermesh, 1985; V. Kerry Smith et al., 2001; and Michael D. Hurd and Kathleen McGarry, 2002).

One well-known drawback with self-reported probabilities, however, is the propensity of respondents to report round figures such as 0, 50 or 100 percent (Hurd and McGarry, 1995; Li Gan et al., 2005). The preponderance of focal responses suggests that individuals may not be

comfortable reporting probabilistic answers, and may not in fact even think in these terms. If individuals use probabilistic information in making insurance purchase decisions, but are unable to translate these latent probabilities into numbers when faced with a survey question, our results are likely to produce underestimates of the extent of the individual's information.⁶

We supplement the AHEAD data with information obtained directly from insurance companies about the information they collect from applicants and their risk classification practices. We draw on both the application forms used by several large long-term care insurance companies which reveal the set of individual characteristics the companies observe, and on the industry's actuarial model of nursing home utilization which itself is a function of these observed characteristics.

B. Econometric approach

The first step of our analysis is to examine the relationship between an individual's beliefs about his subsequent nursing home utilization on the one hand and his actual subsequent nursing home utilization or his current long-term care insurance holdings on the other. We therefore estimate the following two probits:

$$(1) \text{ Prob}(\text{CARE} = 1) = \Phi(X\beta_1 + \beta_2 B)$$

$$(2) \text{ Prob}(\text{LTCINS} = 1) = \Phi(X\delta_1 + \delta_2 B)$$

CARE is a binary variable for whether the individual went into a nursing home in the five years between 1995 and 2000. LTCINS is a binary variable for whether the individual has long-term care insurance in 1995. The coefficient of interest in each equation is that on B, the individual's self-reported beliefs (measured in 1995) of his probability of entering a nursing home between 1995 and 2000. X is a vector of covariates to control for the risk classification that would be assigned to the individual by insurance companies in 1995.⁷

The second step of our analysis is to implement the standard positive correlation test for

asymmetric information. To do so we employ two approaches used in the literature (the results are quite similar across the two). In one approach, we follow Chiappori and Salanie (2000) and estimate a bivariate probit of insurance coverage and risk occurrence, conditional on the risk classification variables (X). This is equivalent to estimating the probits in (1) and (2) simultaneously, with the beliefs variable (B) omitted from the regression. The key variable of interest is the correlation between the error terms (ρ). A uni-dimensional model of asymmetric information predicts that residuals will be positively correlated ($\rho > 0$) and an inability to reject the null hypothesis that $\rho = 0$ constitutes a failure to reject the null of symmetric information.⁸

Our second approach follows the work of Finkelstein and Poterba (2004) and estimates a probit model with nursing home use a function of insurance coverage, controlling for risk classification:

$$(3) \quad \text{Prob}(\text{CARE} = 1) = \Phi(X\beta_1 + \beta_2 \text{LTCINS})$$

The positive correlation prediction is that $\beta_2 > 0$. Note that β_2 does not have a causal interpretation. In a pure moral hazard model, the coefficient would represent the causal effect of insurance coverage on care utilization. However, in an adverse selection model, the causality is reversed and private information about expected care utilization affects insurance demand.

In the final step in our empirical work, we provide several examples of additional individual characteristics that are not used in the insurer's risk classification, but that when added on the right hand sides of equations (1) and (2), have opposite signed relationships with CARE and with LTCINS.

C. Controlling for risk classification

Any analysis of private information, whether the approach we undertake in the first part of

our analysis or the standard positive correlation test in our second step, requires that we condition on the risk classification of the individual by insurance companies (X). This conditioning allows us to test for the existence of *residual* private information, which is the economically meaningful question.

Using insurance applications from five leading long-term care insurance companies we determined what individual characteristics insurance companies observe. All of them 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. This same information is observable in AHEAD, which collects extremely rich and detailed information on current health and medical history, as well as standard demographic data. We can therefore accurately replicate the insurer’s information set.

Our preferred approach is to control for the insurance companies’ actuarial prediction of the individual’s risk type, because it is this measure that is used to generate the price the individual is charged. We generated the insurance companies’ prediction of the probability the individual will go into a nursing home over a five year period using the same actuarial model that is employed by many of the firms in the industry; the model and its pedigree are described in detail in James Robinson (1996), Robinson (2002), and Brown and Finkelstein (2004a).⁹ The predictions depend non-parametrically on the individual’s age, sex, 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 living (ADLs), and the presence or absence of cognitive impairments). As noted, 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 used by the actuaries in

pricing insurance and care utilization.¹⁰

One potential issue with this approach is that we are implicitly assuming that none of the information that is collected by the insurance company, but omitted from the actuarial model, is used in pricing the policies. This assumption appears broadly consistent with insurance company practice. Companies offer age-specific prices with only two or three broad health-based rate classifications within each age (Murtaugh et al. 1995; American Council of Life Insurers, 2001; Weiss 2002,) and do not further adjust premiums over time if the characteristics of the individual changes.¹¹ However, we cannot rule out the possibility that this information is somehow used by the insurance company in determining which of the two or three broad health-based rate classifications to assign the individual to, or in whether the company agrees to insure the individual in the first place. Therefore, we also present results from an alternative approach, which we term the “application information” specification. In this specification, we attempt to control for everything insurance companies observe about the individual. We include a full set of single year of age dummies, all of the demographic information that insurance companies collect in their applications (sex, marital status and age of spouse), and over 35 indicator variables for each of the detailed current health and health history characteristics collected by any insurance company.¹² To be conservative, we also include indicator variables for the household’s income quartile and asset quartile, even though we found only one company that collected any financial information.¹³ Finally, to allow for possible non-linearities among these various characteristics, we also include a complete set of two-way and three-way interactions between age, sex, and the health variables that are included in the actuarial model (ADLs, IADLs and cognitive impairment). The “application information” specification thus invokes a more finely defined categorization of risk than insurance companies likely use in pricing.¹⁴

III. Results

A. Private information about risk type and its relation to insurance coverage

Table 1 reports the marginal effects from probit estimation of equation (1) of the relationship between beliefs and subsequent nursing home use. The results in column (1) indicate that individual beliefs about the likelihood of entering a nursing home are a statistically significant, positive predictor of subsequent nursing home experience. Our finding of a relationship between risk perception and actual risk of nursing home use complements similar findings for mortality risk (e.g. Smith et al., 2001; Hurd and McGarry, 2002).

The results in column (2) indicate that, perhaps not surprisingly, the insurance companies' prediction is more highly correlated with subsequent nursing home use than is the individual's. Comparing columns 1 and 2, we see that a 10 percentage point increase in the self-reported probability of nursing home use is associated with a 0.9 percentage point increase in eventual use, whereas a 10 percent increase in the insurer's prediction is associated with a 4 percentage point increase in eventual use. This is not, however, the relevant metric for testing for asymmetric information; as long as the individual has residual private information – conditional on the menu of choices offered by the insurance company – asymmetric information can operate as in the theoretical models.¹⁵

Indeed, most importantly, columns (3) and (4) indicate that, controlling for the insurer's risk classification, individual beliefs remain a positive and statistically significant predictor of subsequent nursing home use. Individuals thus appear to have residual private information about their risk type.¹⁶

Table 2 reports the results from estimating the relationship between beliefs and insurance coverage in equation (2). The results indicate that individuals who believe that they are of higher

risk are also more likely to have insurance. By contrast, the insurance company's prediction of the individual's risk is negatively related to insurance coverage; this is consistent with the results below that, in fact, on average risk type and insurance coverage are negatively correlated. It also supports our use of the insurance company prediction as a proxy for insurance pricing; conditional on the individual's risk assessment, a higher insurance company prediction implies a higher price relative to the individual's perception of an actuarially fair price, and therefore reduces the probability of purchase.

Taken together, the results in Tables 1 and 2 indicate that individuals have residual private information that predicts their risk type and is positively correlated with insurance ownership. This provides direct evidence of asymmetric information. It does not, however, allow us to distinguish between ex-ante private information (adverse selection) and ex-post private information (moral hazard). Other empirical evidence suggests that demand for nursing home use is relatively price inelastic (Grabowski and Gruber 2005); we therefore suspect – but can not directly corroborate – that at least some of what we are detecting reflects ex-ante private information.¹⁷

We now turn to an examination of what the results from the standard positive correlation test would suggest about asymmetric information in this market. This test also does not distinguish between adverse selection and moral hazard.

B. Long Term Care Insurance and Long Term Care Use

The standard test for residual asymmetric information, based on a positive correlation between insurance coverage and risk occurrence conditional on insurance company risk classification, has been applied across a variety of insurance markets with differing results. In the case of health insurance, David Cutler and Richard Zeckhauser (2000) review an extensive

literature that tends to find evidence of this positive correlation. The positive correlation also appears in annuity markets (Finkelstein and Poterba, 2002, 2004; McCarthy and Mitchell, 2003). Several papers, however, find no evidence of a positive correlation in life insurance markets (Cawley and Philipson, 1999; McCarthy and Mitchell, 2003) or in automobile insurance markets (Chiappori and Salanie, 2000; Georges Dionne et al., 2001; and Chiappori et al., forthcoming).

Table 3 shows the results of this standard test in the long-term care insurance market. The top row shows the correlation of the residuals from a bivariate probit of long-term care insurance and nursing home use, as in Chiappori and Salanie (2000). The bottom row shows the marginal effect from probit estimation of nursing home use on long-term care insurance (equation 3), as in Finkelstein and Poterba (2004). Both approaches yield the same findings. With no controls for the insurers' information set, the relationship between coverage and risk occurrence is negative and statistically significant. This finding is consistent with other aggregate data on relative rates of nursing home use for the insured and general population (Society of Actuaries, 2002). When we control for the insurance companies' risk classification in columns 2 and 3, as required for by the positive correlation test, we are unable to reject the null hypothesis of zero correlation.

Proper implementation of the positive correlation test requires that we examine insurance demand among individuals who face the same set of possible insurance contracts. The controls for risk classification – and hence the premium the insurer would offer the individual – represent an important component of this choice set. However, they may not fully capture all determinants of the individual's choice set. We therefore use two approaches to verify that our failure to reject the null hypothesis of symmetric information with the positive correlation test is not an artifact of our inability to control completely for the options offered to the individuals.

First, we implement the positive correlation test in a more homogenous sub-sample of

individuals who are likely to face the same option set. We identified two additional characteristics of the individual that are particularly likely to affect his effective choice set: his wealth, and the presence of certain adverse health conditions.¹⁸ The public long-term care insurance provided by Medicaid offers a substitute for private insurance; indeed, Brown and Finkelstein (2004b) estimate that Medicaid may have substantial crowd-out effects on demand for private long-term care insurance. However, because Medicaid requires that individuals exhaust virtually all of their financial assets before it will cover long term care expenses, Medicaid is a substantially better option for lower wealth individuals. In addition, individuals with some observably very poor health conditions are often denied insurance coverage, at least by the large long-term care insurance companies (Murtaugh et al., 1995; Weiss, 2002).¹⁹ For such individuals, simply controlling for the price they would face, if they were to be offered coverage, may produce misleading results. We therefore re-implement the positive correlation test limiting our analysis to the 20 percent of individuals who are in the top quartile of the wealth distribution and who have none of the health conditions that could trigger a denial.²⁰ Consistent with the better option set available to the healthier and wealthier subsample, 17 percent of our selected sample have long-term care insurance, compared to only 11 percent of the total sample. Table 4 shows the results from implementing the positive correlation test on this much more homogenous subsample. There is again no evidence of a positive correlation between coverage and nursing home use. Indeed, there is now statistically significant evidence of a *negative* correlation.²¹

Second, to identify more precisely the options that individuals face and the policies they choose, much of the existing literature testing for asymmetric information has relied on proprietary data provided by insurance companies themselves. These data typically include the

individual's risk classification as assigned by the company, details about the policy he has selected from the menu of available policies, and his ex-post risk experience (see e.g. Chiappori and Salanie, 2000 or Finkelstein and Poterba 2004). To implement this alternative approach, we obtained proprietary data from a large long-term care insurance company that contain all of the information used in similar analyses, including the individual's choice from the menu of contracts, his exact risk classification, and his ex-post risk experience.²² This analysis complements the analysis with the AHEAD data because it allows us to examine the relationship between the *quantity* of insurance coverage (conditional on having insurance) and risk occurrence. Using these data, we once again fail to reject the null hypothesis of no positive correlation. The data and results from this exercise are described more fully in Appendix B.

We also undertook numerous additional tests of robustness in the AHEAD data, many of which we present in detail in the working paper version (Finkelstein and McGarry, 2003). For example, we verified that the positive correlation does not manifest itself in other measures of care utilization such as the intensity of care use (i.e. number of nights in a nursing home), or home health care use. We also verified that the positive correlation does not emerge if the relationship between insurance coverage and risk occurrence is analyzed over a longer time horizon than the five-year period studied here. Finally, although policies once purchased are guaranteed renewable for life, some individuals stop paying their premiums and thereby forfeit some or all of their potential future nursing home benefits; we therefore verified that the results are unaffected by excluding from the sample the 10 percent of insured individuals in 1995 who subsequently report having dropped their insurance coverage.²³

The results presented in Tables 1 and 2 point to the presence of asymmetric information, even though the results in Tables 3 and 4 indicate that the standard positive correlation test is

unable to reject the null of symmetric information. This suggests that our beliefs-based test may be a more discerning test for asymmetric information than the standard positive correlation test. Moreover, as noted previously, because our beliefs measure is a highly imperfect proxy for an individual's private information, our findings in Tables 1 and 2 likely understate the amount of private information in this market.

Nonetheless, a natural question is whether the private information we detected in Tables 1 and 2 is sufficiently large that we should have expected to obtain a positive correlation between insurance coverage and long-term care use in the absence of any offsetting selection. A simple decomposition suggests that this is indeed the case. Specifically, we used the results from estimating equation (2) to decompose long-term care insurance coverage into the portion explained by individual beliefs and a residual that is not explained by these beliefs. We then re-estimated the nursing home utilization equation (3) substituting these two components of long term care insurance for the LTCINS variable. In doing so we found a positive and statistically significant relationship between nursing home utilization and the portion of insurance coverage predicted by individual beliefs, with a coefficient of around 0.5 when controls for risk classification are included. This suggests that the relationship between nursing home utilization and insurance coverage would indeed be positive in the absence of any offsetting selection effects. The next section presents more direct evidence of the existence of these offsetting selection effects.

C. Private Information About Preferences for Insurance

Our finding that insured individuals are no more likely to enter a nursing home than individuals without insurance might ostensibly seem at odds with our finding that individuals have private information about their risk type that is positively correlated with insurance

coverage. However, a natural reconciliation of these two sets of results is the existence of other (unobserved) characteristics of the individual that are positively correlated with insurance coverage but negatively related to insurance use. These factors must be items that are omitted from the pricing formulas used by insurance companies and that have the opposite correlation with insurance coverage and care utilization. We now turn to an empirical examination of what some of these offsetting preference-based factors might be.

It is worth noting at the outset that if individuals' self-reported beliefs are a sufficient statistic for their private information and if individuals efficiently incorporate all available information when forming their beliefs, then conditional on these beliefs other characteristics of the individual should have no predictive power in explaining subsequent utilization. In practice, however, our belief measures are likely to be an imperfect proxy for individual's beliefs. Moreover, evidence from Smith et al.'s (2001) study of belief formation about mortality prospects suggests that in fact individuals do not efficiently incorporate all available information in forming and updating their beliefs.

We focus on two potential dimensions of offsetting preference-based selection – wealth and cautiousness – which have previously attracted theoretical attention (de Meza and Webb 2001, Bruno Jullien et al., 2002). Neither of these factors is used by insurance companies in pricing long-term care insurance.

As we noted earlier, Medicaid offers a substantially better substitute for private insurance for lower wealth individuals. This makes it likely that an individual's wealth – a factor not used in pricing private policies – may be an important source of preference heterogeneity in private insurance demand. The top panel of Table 5 reports the results of adding indicator variables for the individual's financial net wealth quartile to equations (1) and (2). Individuals in higher

wealth quartiles are *less* likely to go into a nursing home (odd columns) but *more* likely to have long-term care insurance (even columns).

We measure an individual's cautiousness by his investment in risk reducing activities. We observe, in 1995, whether the individual undertook various gender-appropriate preventive health care measures in the previous two years. These activities 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 median individual undertakes two-thirds of gender-relevant activities; 7 percent report doing nothing and 30 percent report engaging in all relevant activities. The insurance company applications we reviewed did not solicit any of this information.

Individuals who invest more in such preventive activities may also be more risk averse and thus place a higher value on insurance (de Meza and Webb, 2001). However, more cautious individuals are not necessarily more risk averse; preventive activity affects the mean as well as the variance of the risk distribution (see e.g. Dionne and Eeckhoudt, 1985 and Jullien et al., 1999). Thus the sign of the correlation between cautious behavior and insurance coverage is an empirical question. The results in Table 5 panel B indicate that individuals who undertake a greater fraction of potential preventive health activities (i.e. more cautious individuals) are in fact more likely to own insurance; they are also less likely to enter a nursing home.²⁴

One interpretation of the negative relationship between preventive health behaviors and nursing home use is that the preventive behaviors endogenously lower the individual's risk type by forestalling or preventing nursing home admissions. For example, flu shots reduce the risk of pneumonia which is a nontrivial contributor to nursing home use among the elderly. However, we also 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 use) yet none of the measured activities themselves would be expected to affect bone density or agility. We therefore suspect that these preventive health activities are correlated with other health investments that themselves cause lower rates of institutionalization. It is also possible that these preventive health activities proxy for other unmeasured preference-related characteristics that themselves have a causal effect on nursing home utilization. We find, for example, that individuals who engage in more preventive health activity tend to systematically overestimate their risk probability relative to actual experience. This suggests the existence of an unobserved “pessimism” factor; Kostas Koufopolous (2003) conjectures that pessimism simultaneously increases demand for insurance and demand for risk reducing activities which lower risk type.²⁵

One concern with our findings, however, is that preventive health measures may themselves be determined by insurance coverage or health status. We suspect this is unlikely as everyone in our sample is covered by Medicare (which covers the expenses from these preventive health measures at least to some extent), virtually our entire sample of elderly individuals has seen a doctor in the last two years, and our analysis includes detailed controls for health status. However, as a further check, we turn to a different type of precautionary behavior that is less likely to be affected by health status or health insurance coverage: seat belt use. The results in panel C of Table 5 indicate that individuals who report that they always wear their seat belt in a car (which 77 percent do) are also less likely to go into a nursing home and more likely to own insurance.²⁶

Interestingly, other than preventive activity and wealth, the characteristics of the individual that we can measure and that insurance companies do not use in pricing appear to have the *same*

correlation with insurance coverage and risk occurrence. For example, individuals with more children are both less likely to have insurance and less likely to use nursing home care. The same is true for non-whites relative to whites, Hispanics relative to non-Hispanics, and less educated individuals relative to more educated.

IV. Conclusion

In this paper we document the existence of multiple forms of private information in an insurance market, and demonstrate how these factors can have offsetting effects on the correlation between insurance coverage and risk occurrence, thus invalidating the standard test of asymmetric information. Our empirical work focuses on the private long-term care insurance market but the ideas we develop have broader applicability.

We begin by using data on self-assessments of nursing home risk, along with elaborate controls for the risk classification used by the insurance company, to demonstrate directly that private information about risk type exists, and is positively correlated with insurance coverage. We then show that despite this private information, insurance coverage and risk occurrence are not positively correlated, and thus the standard test of asymmetric information fails to reject the null of symmetric information. We reconcile these findings by presenting evidence of a second type of heterogeneity – heterogeneity in preferences for insurance – that offsets the selection based on private information about risk. For example, we find that both wealthier individuals and individuals who behave more cautiously in terms of preventive health activities or seat belt use are both more likely to own insurance and less likely to use long-term care; these factors are not used by insurance companies in pricing insurance.

Our findings highlight that when individuals have private information about risk preferences as well as risk type, an asymmetric information equilibrium in an insurance market can look very

different from the standard theoretical case with heterogeneity in risk type alone. As in the standard uni-dimensional case, the equilibrium with multiple forms of private information it is unlikely to be efficient relative to the first best *even if* the market does not appear “adversely selected” in aggregate. However, an unanswered question, and an important avenue for further work, is under what conditions this added dimension of heterogeneity reduces or increases the efficiency cost of asymmetric information about risk type.

It is worth noting that although preference-based and risk-based selection act in offsetting directions in the long-term care insurance market, they may reinforce each other in other insurance markets. Indeed, in a recent paper, Alma Cohen and Liran Einav (2005) provide evidence that risk type and risk aversion are positively correlated in the automobile insurance market. Differences across insurance markets in the relationship between preference-based selection and actual risk may therefore provide a potential unifying explanation for the apparent differences across insurance markets in whether the insured are above-average in their risk type.

Finally, our analysis suggests a more robust approach to testing for asymmetric information in insurance markets than the widely used “positive correlation” test. Moreover, it is possible to implement this test without the rich data on subjective beliefs available in the HRS. Specifically, the econometrician can reject the null hypothesis of symmetric information if, conditional on the information used by the insurer in setting prices, she observes some other characteristic of the individual that is correlated with both insurance coverage and ex-post risk occurrence that is unknown (or unused) by the insurer. There are many examples of information not priced by insurance companies that the econometrician may observe in survey data, such as wealth which is not priced for annuities, occupation which is often not priced for auto insurance, and preventive health measures which are often not priced for health insurance. These types of

disparities between the information collected and used by the insurance company and that available to the econometrician suggests that this test may find widespread applicability.

Appendix A: The AHEAD Sample and Variable Definitions.

Sample definition: Our sample is drawn from the original Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS). The AHEAD is a representative sample of individuals born in 1923 or earlier, and their spouses. The AHEAD respondents were interviewed in 1993, 1995, 1998 and 2000. We restrict our analysis to data from 1995 to 2000, as 1995 is the first year in which we have a reliable measure of long-term care insurance (see below). We exclude the 3 percent of original respondents who were in a nursing home in 1995. We also exclude from our analysis the approximately 13 percent of the 1995 respondents for whom the interview was completed by a proxy respondent; these individuals are not asked the question about self-reported beliefs of nursing home use that is central to our analysis. Our analyses that do not use this beliefs variable are robust to including the proxy interviews (see Finkelstein and McGarry 2003 for these results). Non-death (i.e. “real”) attrition is just over 4 percent from 1995 and 2000. All of our estimates from the AHEAD data are weighted using the 1995 household weights.

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

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

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

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

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

The question thus does not specifically target long-term care insurance coverage. It yields an estimated coverage rate of just over 2 percent, substantially below what other analyses have indicated for this time period. By contrast, the reported coverage rate using the 1995 measure (10 percent) matches other existing estimates (see e.g. Cohen, 2003 and citations therein). Our concern about the accuracy of the 1993 long-term care insurance measure was corroborated in email correspondence with David Weir, Assistant Director of HRS (April 2002).

Construction of some of the health measures collected by insurance companies

Cognition: We follow Kahla M. Mehta et al. (2002) who work specifically with AHEAD and define an individual as cognitively impaired if he has a score of 8 or less (out of 35) on the Telephone Interview for Cognitive Status (TICS).

Depression: We again follow Meta et al. (2002) and define depression as a score of 3 or greater (out of 8) on the CES-D8.

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

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

Appendix B: The positive correlation test applied to proprietary company data

An alternative way of implementing the positive correlation test that more closely mirrors the existing literature is to use administrative data from a specific insurance company on the contracts chosen by individuals and their ex-post risk experience. Such data provide more detailed information on the coverage choices made than is available in the AHEAD. These data also have the added advantage that all individuals face the same set of product choices, at least within the company studied, and the risk classification and premium assigned to each individual can be observed directly. We therefore obtained such a data set and verified that, as in the AHEAD data, there is no evidence of a positive correlation between insurance coverage and risk occurrence in the insurance company data.

Our data consist of all 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. We also observe all claims incurred through December 31, 2001. The company is among the top-five companies in this market.

Although these data come from a single company, they appear comparable to the broader market on a variety of dimensions. These include the average age at purchase, the gender-mix of the purchasers, the average daily benefit, and the average length of the benefit period.²⁷ In addition, this particular company has experienced similar growth rates in policy sales to the industry as a whole over our sample period (LIMRA 2001). Finally, the company follows the standard risk classification practices of the industry (see e.g. ACLI 2001, Weiss 2002), and varies the premium based on the date of purchase, the individual's age at purchase, and three health-based risk classifications: preferred, standard, or substandard. However, there are a few dimensions along which the company differs from the industry average. Almost all of the

policies sold cover both home care and nursing home care, whereas in the industry as a whole only about three-quarters of recent policies do (HIAA 2000a). In addition, the policies tend to have larger deductibles than industry averages.

To test the positive correlation prediction, we examine the relationship between the quantity of insurance purchased and subsequent nursing home utilization. Long-term care policies typically have a deductible equal to a specified number of days. The relevant risk for the insurance company is the risk that the individual stays in a nursing home beyond the length of the deductible; we also only observe nursing home use in these data if it exceeds the length of the deductible and results in a claim. The modal deductible in the sample is 100 days. We therefore define a “failure” in our hazard model as at least 100 continuous days of nursing home care and 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). 87 percent of our sample has a deductible equal to exactly 100 days. The average failure rate in our sample, 0.3 percent, is quite low, but is consistent with market-wide and population statistics on rates of nursing home utilization of 100 days or more (SOA 1992, 2002).²⁸

We estimate a proportional hazard model of the relationship between policy characteristics and a measure of nursing home utilization (the failure rate)

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

The proportional hazard model assumes that the hazard at time t , $\lambda(t, x_i, \beta, \lambda_0)$, can be decomposed into a baseline hazard $\lambda_0(t)$ and a “shift factor” $\exp(x_i' \beta)$ which represents the proportional shift in the hazard caused by the vector of explanatory variables x_i with unknown coefficients β .

Specification of the baseline hazard $\lambda_0(t)$ is a key issue in estimating models like (4).

Following Cox (1972, 1975), we estimate a continuous-time, semi-parametric, partial likelihood proportional hazard model. This method allows us to estimate the β coefficients without imposing any parametric assumptions about the form of the baseline hazard function $\lambda_0(t)$, which is partialled out of the estimation equation. The model therefore produces a globally concave and well behaved likelihood function. Right-censoring is easily handled in this framework. To control for the insurance company's risk classification, we include indicator variables for issue year, rating category, and issue age.²⁹

Finkelstein and Poterba (2004) show that selection can occur along many dimensions of an insurance contract. We therefore include measures of each of the four main policy characteristics that determine the quantity of insurance in the contract. These are: (1) the deductible, (2) the total number of days for which benefits may be received over the lifetime of the policy ("benefit period"), (3) the maximum amount of incurred nursing home care expenditures that the policy will reimburse per day in care ("maximum daily benefit"), and (4) the degree of escalation of the nominal maximum daily benefit ("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, and analogously, decreasing in the size of the deductible. We also control for the remaining policy features (see notes to Appendix Table B).

The first two columns of Appendix Table B report results for the entire sample of policies. Some of these policies have been in effect for an extremely short period of time. In principle, this should not present an issue for our estimation since the assumption of the proportional hazard model is that the covariates affect the baseline hazard multiplicatively. Nonetheless, to

investigate whether our results are affected by the length of the panel, the next pair of columns shows that the results are quite similar if we restrict the sample to the approximately one-third of policies issued in 1997 or 1998, all of which have between three and five years of exposure.

The top portion of the table shows the estimated coefficients for issue age and rating category, which reflect the insurance company's risk-categorization. These variables are all statistically significant. As expected, the hazard rate increases monotonically with both issue age and risk categorization, and pairs of adjacent issue age categories or rating categories are statistically significantly different from each other.³⁰

The remainder of the table reports the coefficients on the covariates used to test the positive correlation property; the predicted sign for each covariate is summarized in the right-most column. There is little evidence to support the positive correlation prediction. The coefficients on the benefit escalation and benefit period variables tend to have the *opposite* sign from what is predicted. The coefficients on the deductible and daily benefit variables typically have the expected sign, but the estimated effects are almost uniformly insignificantly different than zero and the magnitudes are 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 substantially smaller than the change in hazard associated with any 5-year increase in issue age

REFERENCES

American Council of Life Insurers. *Life Insurers Fact Book, 2001*. Washington, DC, 2001.

Arnott, Richard J. and Joseph E. Stiglitz. "The basic analytics of moral hazard." *Scandinavian Journal of Economics*, 1988, 90 (3), pp. 383-413.

Brown, Jeffrey R. and Amy Finkelstein. "Supply or Demand: Why is the Market for Long-Term Care Insurance So Small?" NBER Working Paper 10782, 2004a.

Brown, Jeffrey R. and Amy Finkelstein. "The Interaction of Public and Private Insurance: Medicaid and the Long Term Care Insurance Market." NBER Working Paper 10989, 2004b.

Cawley, John and Tomas Philipson. "An empirical examination of information barriers to trade in insurance." *American Economic Review*, 1999, 89 (4), pp.827-846.

Cardon, James H. and Igal Hendel. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics*, 2001, 32 (3), pp. 408-427.

Chassagnon, Arnold and Bertrand Villeneuve. 2005. "Efficient risk-sharing under adverse selection and subjective risk perception." *Canadian Journal of Economics*, 38(3): 955-978.

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. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 2000, 108 (1), pp. 56-78.

Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie. "Asymmetric Information in Insurance: General Testable Implications." *Rand Journal of Economics*, forthcoming.

Cohen Alma and Liran Einav. "Estimating risk preferences from deductible choice." Unpublished mimeo. Stanford University, 2005.

Cohen, Marc. "Private Long-Term Care Insurance: A look ahead." 2003. *Journal of Aging and Health* (15): 74-98.

Congressional Budget Office. *Financing Long-Term Care for the Elderly*. Washington DC: US Government Printing Office, April, 2004.

Cox, David. "Regression Models and Life-Tables (with discussion)." *Journal of the Royal Statistical Society*, 1972, 187-220.

Cox, David. "Partial Likelihood" *Biometrika*, 1975, 62(2): 269-276.

Crocker, Keith J. and Arthur Snow. "The efficiency of competitive equilibria in insurance markets with asymmetric information." *Journal of Public Economics*, 1985, 26 (2) pp. 207 – 219.

Cutler, David and Richard Zeckhauser. "The Anatomy of Health Insurance," in A.J. Culyer, and J. Newhouse eds., *The Handbook of Health Economics*, 2000, pp. 563-643

de Meza, David and Webb, David C. "Advantageous selection in insurance markets." *Rand Journal of Economics*, Summer 2001 32 (2), pp. 249-262

Dionne, Georges and Louis Eeckhoudt. "Self-Insurance, Self-Protection, and Increased Risk Aversion." *Economic Letters* 1985, 17, pp. 39-42.

Dionne, Georges, Christian Gourieroux and Charles Vanasse. "Testing for evidence of adverse selection in the automobile insurance market: a comment." *Journal of Political Economy* 2001, 109(2) pp. 444-451.

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

Finkelstein, Amy, Kathleen McGarry and Amir Sufi. "Dynamic Inefficiencies in Insurance Markets: Evidence from long-term care insurance." *American Economic Review Papers and Proceedings*. May 2005, (95): 224-228.

Finkelstein, Amy and James Poterba. "Selection Effects in the Market for Individual Annuities: New Evidence from the United Kingdom." *Economic Journal*, 2002, 112 (476), pp. 28-50.

Finkelstein, Amy and James Poterba. "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market." *Journal of Political Economy*, 2004, 112(1), pp. 193-208.

Gan, Li, Michael Hurd and Daniel McFadden. "Individual subjective survival curves" 2005, in David Wise (ed) *Analyses in Economics of Aging*, University of Chicago Press: 377-411.

Grabowski, David and Jonathan Gruber. "Moral hazard in nursing home use." NBER Working Paper 11723, 2005

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

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

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

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

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

Health Insurance Association of America (HIAA). *Who Buys Long-Term Care Insurance in the Workplace?* 2001.

Jeleva, Meglena and Bertrand Villeneuve. "Insurance contracts with imprecise probabilities and adverse selection." *Economic Theory* 2004, 23 (4) pp 777-794.

Jullien Bruno, Bernard Salanie and Francois Salanie. "Should More Risk-Averse Agents Exert More Effort?" *The Geneva Papers on Risk and Insurance Theory*, 1999, 24, pp. 19-28.

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

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

Lewis, Stephanie, John Wilkin and Mark Merlis. *Regulation of Private Long-term Care Insurance: Implementation Experience and Key Issues*. Prepared for The Kaiser Family Foundation, March 2003.

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

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

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

Mellor, Jennifer M. "Long-term care and nursing home coverage: are adult children substitutes for insurance policies?" *Journal of Health Economics* 2001, 20 (4) pp.527-547.

Murtaugh, Christopher, Peter Kemper and Brenda C. Spillman. "Risky Business: Long-term care insurance underwriting." *Inquiry* 1995, 32 (3) pp. 371-284.

Murtaugh, Christopher, M., Peter Kemper, Brenda C. Spillman, and Barbara L. Carlson. "Amount, distribution, and timing of lifetime nursing home use." *Medical Care* 1997, 35 (3) pp. 204-218.

National Association of Insurance Counselors (NAIC). 2002a. *Executive Summary of Amendments to the Long-Term Care Insurance Model Regulation*, January 2002a.

National Association of Insurance Counselors (NAIC). *State Survey of Long-Term Care Insurance Rating Practices and Consumer Disclosure Amendments*. September 2002b.

National Center for Health Statistics. *Health, United States, 2002 With Chartbook on Trends in the Health of Americans*. Hyattsville, Maryland, 2002.

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

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

Rothschild, Michael and Joseph Stiglitz. "An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 1976, 90 (4) pp. 629-649.

Sloan, Frank A. and Edward C. Norton. "Adverse selection, bequests, crowding out, and private demand for insurance: evidence from the long-term care insurance market." *Journal of Risk and Uncertainty*, 1997, 15 (3) : 201-219.

Smart, Michael. "Competitive insurance markets with two unobservables." *International Economic Review*, 2000, 41(1) pp. 153-169.

Smith, V. Kerry, Donald Taylor and Frank Sloan. "Longevity expectations and death: can people predict their own demise?" *American Economic Review* September 2001, 91 (4) pp. 1126-1134.

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

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

Wambach, Achim. "Introducing Heterogeneity in the Rothschild-Stiglitz Model." *Journal of Risk and Insurance* 2000, 67 (4) pp.579-92.

Weiss Ratings Inc. *Weiss Ratings' Consumer Guide to LTC Insurance*. Palm Beach Gardens, FL, 2002.

Table I: Relationship between individual beliefs and subsequent nursing home use

	(1)	(2)	(3)	(4)
Individual Prediction	0.091*** (0.021)		0.043** (0.020)	0.037* (0.019)
Insurance Company Prediction		0.400*** (0.020)	0.395*** (0.021)	
Controls for application information	NO	NO	NO	YES
pseudo-R ²	0.005	0.097	0.099	0.183
N	5,072	5,072	5,072	4,780

Notes: Reported coefficients are marginal effects from probit estimation of equation (1).

Dependent variable is an indicator for any nursing home use from 1995 through 2000 (mean is 0.16). Both individual and insurance company predictions are measured in 1995.

Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively. Column 4 – which includes controls for “application information” – includes controls for age (in single year dummies), sex, marital status, age of spouse, over 35 health indicators, and a complete set of 2-way and 3-way interactions for all of the variables used in the insurance company prediction (age dummies, sex, limitations to activities of daily living, limitations to instrumental activities of daily living, and cognitive impairment); see text for more details.

Table II: Relationship between individual beliefs and insurance coverage

	(1)	(2)	(3)	(4)
Individual Prediction	0.086*** (0.017)		0.099*** (0.017)	0.083*** (0.016)
Insurance Company Prediction		-0.125*** (0.023)	-0.140*** (0.023)	
Controls for application information	NO	NO	NO	YES
pseudo-R ²	0.007	0.010	0.019	0.079
N	5,072	5,072	5,072	4,780

Notes: Reported coefficients are marginal effects from probit estimation of equation (2).

Dependent variable is an indicator for whether individual has long-term care insurance coverage in 1995 (mean is 0.11). Both individual and insurance company predictions are measured in 1995. Heteroskedasticity-adjusted robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively. Column 4 – which includes controls for “application information” – includes controls for age (in single year dummies), sex, marital status, age of spouse, over 35 health indicators, and a complete set of 2-way and 3-way interactions for all of the variables used in the insurance company prediction (age dummies, sex, limitations to activities of daily living, limitations to instrumental activities of daily living, and cognitive impairment); see text for more details.

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

	No Controls	Controls for Insurance Company Prediction	Controls for Application Information
	(1)	(2)	(3)
Correlation coefficient from bivariate probit of LTCINS and CARE	-0.105*** (p=0.006)	-0.047 (p=0.25)	-0.028 (p=0.51)
Coefficient from probit of CARE on LTCINS	-0.046*** (0.015)	-0.021 (0.016)	-0.014 (0.016)
N	5,072	5,072	4,780

Notes: Top row reports the correlation of the residual from estimation of a bivariate probit of any nursing home use (1995-2000) and long-term care insurance coverage (1995); p values are given in parentheses. Bottom row reports marginal effect on indicator variable for long-term care insurance in 1995 from probit estimation of equation (3). The dependent variable is an indicator variable for any nursing home use from 1995 through 2000; heteroskedasticity-adjusted robust standard errors are in parentheses. For all rows, control variables are described in column headings; see text for more information. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively. Means of CARE and LTCINS are 0.16 and 0.11 respectively.

Table IV: Relationship between LTCINS and CARE; sample restricted to individuals with same choice set

	No Controls	Controls for Insurance Company Prediction	Controls for Application Information
	(1)	(2)	(3)
Correlation coefficient from bivariate probit of LTCINS and CARE	-0.123* (p=0.08)	-0.122* (p=0.10)	-0.191** (p=0.017)
Coefficient from regression of CARE on LTCINS	-0.032* (0.018)	-0.028* (0.015)	-0.033** (0.012)
N	1,504	1,504	1,438

Notes: Sample is limited to individuals in the top quartile of the wealth and income distribution and who have none of the health characteristics that might make them ineligible for private insurance. Top row reports the correlation of the residual from estimation of a bivariate probit of any nursing home use (1995-2000) and long-term care insurance coverage (1995); p values are given in parentheses. Bottom row reports marginal effect on indicator variable for long-term care insurance in 1995 from probit estimation in equation (3). The dependent variable is an

indicator variable for any nursing home use from 1995 through 2000; heteroskedasticity-adjusted robust standard errors are in parentheses. For all rows, control variables are described in column headings; see text for more information. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively. Means of CARE and LTCINS are 0.09 and 0.17 respectively.

Table V: Preference-based Selection

	No Controls		Control for Insurance Company Prediction		Control for Application Information	
	NH Entry	LTC Insurance	NH Entry	LTC Insurance	NH Entry	LTC Insurance
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Wealth						
Top Wealth Quartile	-0.095*** (0.013)	0.150*** (0.020)	-0.038** (0.014)	0.131*** (0.020)	-0.018 (0.015)	0.139*** (0.022)
Wealth Quartile 2	-0.073*** (0.013)	0.104*** (0.020)	-0.025* (0.014)	0.089*** (0.020)	-0.013 (0.014)	0.092*** (0.020)
Wealth Quartile 3	-0.030** (0.015)	0.062*** (0.020)	0.0004 (0.016)	0.052*** (0.019)	0.006 (0.015)	0.057*** (0.020)
Bottom Wealth Quartile (Omitted)	-----	-----	-----	-----	-----	-----

Individual	0.086***	0.089***	0.042**	0.098***	0.035*	0.086***
Prediction	(0.021)	(0.017)	(0.020)	(0.017)	(0.019)	(0.017)

Panel B: Preventive Health Activity

Preventive Activity	-0.106***	0.066***	-0.054***	0.052***	-0.016	0.016
	(0.0118)	(0.017)	(0.018)	(0.017)	(0.019)	(0.017)

Individual	0.095***	0.082***	0.047**	0.095***	0.037*	0.082***
Prediction	(0.021)	(0.017)	(0.020)	(0.017)	(0.020)	(0.017)

Panel C: Seat Belt Use

Always Wear	-0.059***	0.053***	-0.031**	0.048***	-0.018	0.029***
Seatbelt	(0.014)	(0.010)	(0.013)	(0.010)	(0.012)	(0.010)

Individual	0.092***	0.084***	0.044**	0.097***	0.038*	0.082***
Prediction	(0.021)	(0.017)	(0.020)	(0.017)	(0.019)	(0.016)

Note: Table reports marginal effects from probit estimation of equations (1) and (2). Additional controls are given in column headings; see text for more information. In Panel A, omitted wealth category is quartile 4. For panel A, income controls are omitted from the “application information” controls since they are highly multi-collinear with assets. In Panel B, “preventive activity” measures the proportion of gender-appropriate preventive health behaviors undertaken; all estimates in Panel B include an additional control for gender. Heteroskedasticity-adjusted

robust standard errors are in parentheses. ***, **, * denotes statistical significance at the 1 percent, 5 percent, and 10 percent level respectively.

Appendix Table B: Hazard of Receiving Nursing Home Care for 100th Consecutive Day

Covariates in Regression	Policies Issued 1997 - 2001		Policies Issued 1997 or 1998		Prediction of “positive correlation” property
	Coefficient	Standard error	Coefficient	Standard error	
<i>Issue Age Category:</i>					
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)	
<i>Rating Category:</i>					
High risk (omitted)	--	--	--	--	
Rated standard risk	-0.535***	(0.175)	-0.562***	(0.200)	
Rated low risk	-1.100***	(0.259)	-0.964***	(0.322)	
<i>Deductible:</i>					
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 and larger than coefficient on 60-day deductible
<i>Daily Benefit:</i>					
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 and larger than coefficient on \$100 daily benefit
<i>Benefit Period:</i>					
1-4 years, no 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 and larger than coefficient on “5+ years, no extension”
Unlimited	0.168	(0.153)	0.075	(0.175)	Positive and larger than coefficient on

					“5+ years, possible extension”
<i>Escalation of Benefits:</i>					
“Index option” (omitted)	--	--	--	--	
5 percent compound escalation	-0.102	(0.236)	-0.254	0.288	Negative
5 percent “simple” escalation	0.111	(0.131)	0.016	(0.154)	Negative and less than coefficient on 5 percent compound escalation
No escalation	0.335	(0.333)	--		Negative and less than other coefficients
Failure Rate	0.3%		0.6%		
N	144,798		49,888		

Note: Estimates from a Cox proportional hazard model. Failure defined as at least 100 days of continuous nursing home use. Also included are: indicators of issue year, whether the policy is tax qualified and frequency of policy premium payments. All covariates shown in table are indicator variables. Daily benefit categories were chosen to divide the sample into approximately equal sized groups. Indicator variables to measure the benefit period distinguish among benefit

periods of 1-4 years, 5+ years (but finite), and unlimited; policies with finite benefit periods are further they further distinguished by whether benefit can be extended under certainty circumstances. Indicates for the four possible benefit escalation options are (in order of increasing generosity): constant nominal benefits, increases of 5 percent per year of the original benefit (“simple” escalation), 5 percent per year compounded (“compound” escalation), and the greater of 5 percent compounded annual over 3 years or CPI-growth over the last 3 years at the option of the policy holder (“indexed”).

* Finkelstein: MIT Department of Economics E52-262F, 50 Memorial Drive, Cambridge MA 02142 USA (email: afink@mit.edu), and National Bureau of Economic Research; McGarry: UCLA Department of Economics, 8283 Bunche Hall, Box 951477 Los Angeles, CA 90095-1477 USA (email: mcgarry@ucla.edu). We are grateful to Daron Acemoglu, Doug Bernheim, Jeff Brown, Pierre-Andre Chiappori, Raj Chetty, Janet Currie, David Cutler, David de Meza, Seema Jayachandran, Jerry Hausman, Ginger Jin, Larry Katz, Ben Olken, Sarah Reber, Casey Rothschild, Michael Rothschild, Bernard Salanie, Jesse Shapiro, Jonathan Skinner, Jonathan Wright, three anonymous referees, and numerous seminar participants for helpful comments and discussions. We are especially grateful to Jim Robinson for generously providing us with the actuarial model of long-term care utilization. Finkelstein thanks the NIA (R01 AG021900), and McGarry thanks the NIA (R01 AG016593) for financial support.

¹ See e.g. John Cawley and Tomas Philipson, 1999; Pierre-Andre Chiappori and Bernard Salanie 2000; and James H. Cardon and Igal Hendel, 2001.

² Of course, this prediction – and the appropriate empirical test of it – applies only to 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). Although we will not always state this qualification explicitly in our discussion, it is implicitly always present. In the empirical work below, we take great care to condition on the risk classification and option set of the individual.

³ Chiappori et al. (forthcoming) show that for there to be asymmetric information about risk type and no positive correlation between insurance coverage and risk occurrence requires – if more than one type of contract is purchased – that there be not only an additional dimension of heterogeneity but also a markup of price above expected claims.

⁴ For an example of an equilibrium with private information about risk type and risk preferences in which there is scope for Pareto improvement through government intervention, see de Meza and Webb (2001).

⁵ Although there is no direct regulation of prices or policies, the government is a major presence in long-term care insurance provision through Medicaid, the public insurance program for the indigent. In the empirical work below we discuss how we account for the potential effects of Medicaid.

⁶ Another interpretation of these focal responses is that the responses may convey ordinal information about their beliefs of their risk (e.g. low, medium, or high) but not cardinal information about the scale. Almost 50 percent of respondents report a five-year nursing home entry probability of zero, and 14 percent report a 50 percent probability.

The working paper version of this paper therefore reports a parallel set of results to those shown here in which we use a set of indicator variables for low risk (reports of 0 probability), medium risk (1-49 percent) and high risk (50-100 percent) to measure beliefs; the main findings of the paper are robust to this alternative parameterization of individual's beliefs (Finkelstein and McGarry 2003).

⁷ Although we report results from estimating equations (1) and (2) independently, the findings are robust if we instead estimate the equations simultaneously.

⁸ Chiappori and Salanie (2000) discuss several alternative parametric and non-parametric approaches to implementing the test for residual correlation. In the interest of space we do not discuss and report them here, but we have verified that our findings are robust to these alternative approaches.

⁹ The model we use predicts care utilization for typical individuals in the population and makes no adjustment for potential moral hazard effects of the insurance

¹⁰ We also tried introducing the variables used in this prediction directly into the regression in a fully flexible and interacted manner. And we tried including indicator variables for the decile of the insurance companies' prediction. Both produced results that were virtually indistinguishable from those obtained using the prediction itself. We report the specification with the actuarial prediction directly on the right hand side since this single prediction of risk can be compared directly to the individual's own prediction.

¹¹ According to our conversations with industry actuaries, insurance companies collect more detailed information than they currently use in order to build a detailed claims database for future improvements in actuarial modeling.

¹² These indicator variables are: limitations with respect to each of five activities of daily living (bathing, eating, dressing, toileting, and walking across a room) and two instrumental activities of daily living (grocery shopping and managing medication); use of each of five devices (a wheelchair, walker, crutches, cane and oxygen); low body mass index; high body mass index; whether the respondent is incontinent, uses prescription drugs regularly, smokes, is depressed, has a drinking problem, suffers from cognitive impairment; whether the respondent has or had: diabetes, diabetes treated with insulin, kidney failure associated with diabetes, a stroke, a heart condition, medication for a heart condition, a heart attack, congestive heart failure, high blood pressure, hip fracture, lung disease, cancer, psychiatric problems, arthritis, prior nursing home use, prior home health care, or has been injured in a fall. Finkelstein and McGarry (2003) provide summary statistics for these covariates. Appendix A provides a description of the construction of some of the variables.

¹³ This company asked only whether the individual had less than \$30,000 in financial assets, presumably to screen for likely Medicaid eligibility.

¹⁴ To deal with the issue that insurance companies deny coverage to some observably poor health individuals, we collected information from applications and underwriting guides on criteria used to deny coverage; all of these criteria are included in the “application information” specification. In the sensitivity analysis below we also show that the results are robust to excluding individuals who might have been denied coverage based on these criteria.

¹⁵ See Jeleva and Villeneuve (2004) for analysis of a market equilibrium when insurance companies are better at predicting risk than consumers but consumers have residual private information about risk.

¹⁶ The small R-squared of the regression analysis, even with the extremely detailed controls in column (4), suggests that there is a large amount of genuine uncertainty about subsequent nursing home use, and thus a potentially large value of insurance that covers this risk.

¹⁷ Our findings also raise the question of why insurance companies do not collect additional information to reduce the informational advantage of the consumer. Our analysis suggests the answer may be that the collection of additional available information is unlikely to reduce the consumer’s residual private information. We added additional control variables in equations (1) and (2) for all the information we can observe in the AHEAD that the insurance companies do not collect – including the number, sex, and proximity of the individual’s children, the individual’s race, religion and education, information on a spouse’s health, and individual investments in a variety of potentially risk reducing behaviors (described in more detail in Section III.C). Although jointly statistically significant, these additional control variables did not attenuate either the magnitude or statistical significance of the coefficient on the individual’s prediction.

¹⁸ For other forms of health insurance, whether the current or former employer of the individual or his spouse offers insurance can be an essential element of the option set. However, although the group long-term care insurance market began to grow early to mid 1990s and is currently about 20 percent of the total market, it was virtually non-existent at the time when the elderly individuals in our data would have been in the workforce (HIAA 2000b, HIAA 2001).

¹⁹ An examination of applications from the major long-term care insurance companies and several of their underwriting guides indicates that coverage is often denied to individuals who have limitations with respect to any ADLs (bathing, eating, dressing, toileting, walking, and maintaining continence), use mechanical devices

(wheelchair, walker, crutches, quad cane, oxygen) or suffer from cognitive impairment. All of these factors are controlled for in our “application information” specification. The practice of denying observably high risk individuals coverage is common in other insurance markets as well. It may reflect issues of reputation or brand name, greater concerns about asymmetric information for individuals of higher observable risk, or lower variability in care utilization for these individuals.

²⁰ This approach is likely to yield an over-estimate of the number of ineligible individuals because individuals classified as ineligible in 1995 may have been previously eligible for insurance. We use this over-inclusive definition to be sure that the remaining sample would have been eligible for insurance coverage.

²¹ We verified that the point estimates in Tables 1 and 2 were not sensitive to the same sample restriction, although in a few specifications standard errors increased to the point where statistical significance no longer obtained.

²² The downside to these data is that they do not contain the rich information on individuals’ beliefs about their risk type or about other characteristics not used by the insurance company that is critical for our analysis.

²³ Finkelstein et al. (2005) provide a more detailed discussion of this dropping behavior as well as an empirical exploration of some of the factors that may be behind it.

²⁴ We verified that these results are robust to instead measuring preventive health activity subsequent to the insurance contracting in 1995 (i.e. over the period 1998 – 2000).

²⁵ See also Arnold Chassagnon and Villeneuve (2005) for the characteristics of equilibrium with adverse selection as well as heterogeneity in pessimism.

²⁶ To verify that the variation in seat belt use does not merely reflect differences in state laws (or the enforcement of these laws) requiring seat belt use, we verified that the results are robust to including state fixed effects.

²⁷ Finkelstein and McGarry (2003) provide summary statistics for these data in Table 1. HIAA (2000a) provides comparable industry-wide statistics.

²⁸ 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).

²⁹ For ease of presentation, we include indicator variables for age in five roughly equal size bins corresponding to less than 60, 60-64, 65-69, 70-74, and 75+. Including separate indicator variables for each age rather than five-year intervals does not affect the coefficients of interest. We do not directly control for the premium because we have controlled for all of the characteristics of the individual and the policy that determine it.

³⁰ The precision of our estimates, and specifically our ability to detect these statistically significant differences in the hazard rate across issue age and rating categories, helps alleviate any concerns that the failure rate may be too low to provide sufficient power to identify significant differences in the hazard rate across the covariates of interest, should they exist.