The Life Insurance Market: Asymmetric Information Revisited

Daifeng He[1]

June 12, 2008

College of William and Mary

Abstract

This paper finds evidence for the presence of asymmetric information in the life insurance market, a conclusion contrasting with the existing literature. In particular, we find a significant and positive correlation between the decision to purchase life insurance and subsequent mortality, conditional on risk classification. Individuals who died within a 12-year time window after a base year were 19 percent more likely to have taken up life insurance in that base year than were those who survived the time window. Moreover, as might be expected when individuals have residual private information, we find that the earlier an individual died, the more likely she was to have initially bought insurance. The primary factor driving the difference between our and the prior literature's findings is that we focus on a sample of potential new buyers, rather than on the entire cross section, to address the sample selection problem induced by potential mortality differences between those with and those without coverage.

Keywords: Asymmetric information, private information, life insurance, sample selection, differential mortality, new buyers, health status, pricing factors, risk classification.

JEL Classification number: D82

---

[1] Economics Department, Washington University in St. Louis, Campus Box 1208, One Brookings Drive, St. Louis, MO 63130; dhe@wustl.edu. I am grateful to the editor, two anonymous referees, Sebastian Galiani, Steve Buccola, Lee Benham, Edward Greenberg, Bart Hamilton, Stephanie Lau, Bruce Petersen, Tanika Chakraborty, Charles Courtemanche, Jeremy Meiners, and Srikanth Ramamurthy for helpful comments and suggestions. Partial financial support from the Olin Business School and CRES is gratefully acknowledged.

I. Introduction

Empirical testing of contract theory comprises a burgeoning area of economic research (see Chiappori and Salanie 2003 for a review). Especially important in this literature have been inquiries into whether asymmetric information prevails in particular insurance markets. Much of the literature has adopted the "conditional correlation" approach illustrated in Chiappori et al. (2006), in which the presence of information asymmetry implies that, conditional on risk classification, the risk outcome is positively correlated with insurance coverage. Evidence has been mixed.[2]

The life insurance market is of particular interest for asymmetric information tests. It is an important market on account of size alone. In 2004, 77% of American households held life insurance. The industry had overall assets of $4.5 trillion and invested $4 trillion in the economy, making it one of the most important sources of investment capital in the United States (ACLI 2007a, 2007b). Life insurance contracts also are relatively explicit and simple, and the risk outcomes – policyholders' deaths – are in principle easy to verify and measure.

In an important contribution, Cawley and Philipson (1999) use the Health and Retirement Study (HRS) data to examine cross-sections of individual term life insurance contracts and find a negative or neutral correlation between mortality risk and coverage.[3] This negative-or-neutral-correlation result, together with their evidence for bulk discounts, has been widely cited as evidence that life insurance markets are free of asymmetric information.[4]

We find, in contrast, evidence of asymmetric information in these very markets.[5] With the same HRS dataset, we recover a significant positive correlation between the mortality outcome and the decision to purchase individual term life insurance, conditional on risk classification. In particular, individuals with higher risk (those who died within a 12-year time window after a base year) were 19% more likely to have

[2] For example, see Chiappori and Salanie (2000) and Cohen (2005) for the auto insurance market; Finkelstein and Poterba (2004) for the annuity market; Cardon and Hendel (2001) for the employer-provided health insurance market; and Fang et al. (2006) for the Medigap market.
[3] Using aggregate mortality data from the U.S., U.K., and Japan, McCarthy and Mitchell (forthcoming) also find that life insurance buyers' mortality rates are the same as, or lower than, those in the general population.
[4] For example, see Chiappori and Salanie (2000), de Meza and Webb (2001), Hendel and Lizzerri (2003), Fang et al. (2006), Chiappori and Salanie (2008), and Cutler et al. (2008).
[5] Following Cawley and Philipson, our analysis also focuses on the individual term life insurance market.

purchased individual term insurance in that base year than were individuals with lower risk (those who survived beyond the window). Indeed, decomposing the mortality outcome into time-until-death categories, we find that the earlier an individual died, the more likely she was to have initially taken up insurance. Such monotonicity further suggests the prevalence of asymmetric information.[6]

The primary factor driving the difference between our and Cawley and Philipsons' earlier findings is that we focus on a sample of potential new life insurance buyers rather than on the entire cross sectional sample. Potential new buyers are the subset of the total sample who did not own life insurance at the beginning of the sample period. They are not subject to the sample selection problem inherent in cross-sectional samples for asymmetric information tests in life insurance markets. This sample selection problem is as follows. Suppose individuals do have residual private information about their mortality risk. Those for whom the information is unfavorable, and who thus decide to buy life insurance, then are more likely to die early and thus less likely to be found in a cross-sectional sample than are those for whom the information is favorable. High-risk individuals with coverage therefore are under-represented in cross-sectional samples. Sample selection induced by potential mortality differences between the covered and uncovered may bias estimates of the conditional correlation between insurance coverage and mortality risk.[7]

To illustrate, consider the following thought experiment. Four individuals are alive, with the same appearance of good health, at time t - 5. Individuals 1 and 2 choose not to obtain coverage because they know they are in good health. Individuals 3 and 4 do choose coverage because, despite their healthy appearance, they know they are in poor health. At year t - *1*, individual 4 dies. The remaining three are randomly drawn into a sample and survive the entire sample period from year t to $t+5$. A researcher examining this sample will conclude that asymmetric information is absent: observed

[6] Our test is a joint test for the presence of asymmetric information, which may take the form of either adverse selection or moral hazard. Moral hazard can largely be ignored in the life insurance industry because insurance is unlikely to be an incentive for an individual to die sooner than she otherwise would. We therefore believe our results suggest the presence of adverse selection. This claim is, however, based on intuitive insight rather than on rigorous evidence.

[7] In survival analysis, "left truncation" is used to describe the situation in which the existence of an individual is unknown to the researcher if she dies before the beginning of the observation period. In our case, left truncation cannot be ignored because the mortality risk of those observed in the sample may not be representative of the population of interest. See Kalbfleisch and Prentice (2002), pp. 13 - 14.

mortality in the t through $t + 5$ window does not differ between the two individuals without insurance and the one with insurance, inasmuch as all three have survived the five-year sample period. The real story, however, is that half of the covered, and neither of the uncovered, have died within the full ten-year (t - 5 to t + 5) horizon.

The remainder of the paper is organized as follows. Section II describes the dataset. Section III discusses the empirical strategy, in which we focus on the sample of potential new buyers together with proper risk classification controls and a 12-year-window ex-post mortality risk measure. Section IV presents the results. Section V concludes.

II. Data

We use the Health and Retirement Study (HRS) dataset. The HRS is a nationally representative longitudinal survey of the elderly and near-elderly in the United States. It contains rich information on health status, insurance coverage, financial measures, demographics, and family structure. Our analysis uses the HRS cohort, which consists of individuals born between 1931 and 1941. This cohort has been interviewed biennially since 1992. Our sample ends in 2004.

We obtain life insurance coverage data from two early waves, 1992 and 1994, in order to simplify comparison with the previous literature.[8] The 1992 and 1994 waves also are the only ones in which the HRS explicitly asked whether a respondent held individual term life insurance. Moreover, following up on sample individuals from early waves allows us to observe actual mortality outcome in a sufficiently long time window.

Tracker 2004 provides the mortality data. HRS divides a respondent's vital status in each wave into one of five categories: alive in current wave, presumed alive in current wave, death reported in the current wave, death reported in a prior wave, and vital status unknown.[9] We code a respondent as alive in 2004 if she falls into category 1 or 2 – and dead in 2004 if she falls into category 3 or 4 – in wave 2004. We treat those

[8] Cawley and Philipson (1999) obtain life insurance information for the HRS cohort from the 1992 wave.
[9] For the precise coding criteria, see HRS Tracker 2004, Version 2, January 2007.

in category 5 as missing observations.[10] In this way, we observe the actual mortality outcome during a 12-year time window.

Table 1 provides summary statistics of the relevant variables in our sample, based mainly on information from the 1992 wave. Twenty-four percent of the HRS cohort owned individual term life insurance in 1992 and 27% owned it in 1994. Nineteen percent of potential new buyers obtained individual term life insurance between 1992 and 1994.[11] By 2004, about 15% of the cohort had died. The sample was largely balanced in gender, and nearly three-quarters of the respondents were married. High blood pressure, arthritis, and back pain were the most commonly reported medical conditions. About a tenth of the sample had a hospital stay in the past year, and roughly the same portion had been diagnosed with heart disease. Less than 10% of the sample was diagnosed with diabetes, cancer, lung disease, stroke, or asthma. Nearly a third of the sample had healthy weight, 44% were overweight, and 22% were obese.

III. Empirical Strategy

An ideal sample would satisfy the following requirements for testing for the presence of asymmetric information in life insurance markets. First, it should constitute a random sample of the population below a certain age threshold such that no individual in the population younger than that age would consider purchasing life insurance. For example, age 20 could be such a threshold if, given the absence of dependents, no individual younger than 20 would demand life insurance. Second, the sample should follow every individual until the last one dies. At the end of the sample period, a researcher could then observe the coverage status, mortality outcome, and complete set of risk classification factors of every sample individual who is a potential customer in the life insurance market. In such a sample, differential mortality would not create a selection problem. A positive correlation, conditional on the risk classification factors,

[10] As a robustness check, we also code a mortality upper-bound and a mortality lower-bound variable, as in Cawley and Philipson (1999), treating those in category 5 as dead and alive, respectively. All the Section IV results remain qualitatively the same.

[11] Considerable measurement error may be associated with the self-reported life insurance ownership data because, assuming a moderate per-wave lapse rate of 4% (based on waves 1996 and 1998 HRS data) and ignoring expired polices, we would otherwise obtain a 37% [(0.76*0.19+0.24)*0.96], rather than 24%, coverage rate in 1994. Measurement error in a discrete binary dependent variable may produce inconsistent estimates (Hausman et al., 1998) and this is a potential concern with both our and Cawley and Philipson's analyses.

between the decision to purchase insurance and a proper measure of mortality risk would provide evidence of asymmetric information.

A. *Potential New Buyers*

Such an ideal sample does not, of course, exist. The HRS sample likely suffers from the selection bias arising from potential mortality differences between those with coverage and those without. The HRS respondents were between 51 and 61 years old at the time of first interview, an age by which many of them probably had owned life insurance for many years.[12] This cohort may consist disproportionately of individuals with relatively low mortality risk because higher-risk individuals with coverage are more likely to have died before the survey started and thus are less likely to be found in the sample. The conditional correlation between the mortality risk and coverage therefore would be biased downward. Such selection bias may be responsible for the negative or neutral conditional correlations found in the earlier literature.

Our main contribution in the present paper is to address this selection bias induced through differential mortality. We define *potential new buyers* as those who did not own individual term life insurance in the 1992 wave. By the time of the 1994 wave, some individuals ("new buyers") in this group had purchased coverage, while the rest ("non-owners") remained uncovered. Distinguishing between potential new buyers and the entire cross-sectional sample is key to our approach. With potential new buyers – those who potentially were customers at the beginning of the sample period – a researcher does not face the differential-mortality pitfall, since she can completely track mortalities within the permitted time window.

B. *Estimation Model*

We therefore estimate the following logit model:[13]

[12] Other ways in which the HRS sample might fall short are: (a) most respondents were still alive by 2004, the end of our sample period; and (b) we may not observe every risk classification factor a typical life insurer might consider. These two limitations are, however, not very serious. Regarding (a), we define mortality risk in terms of whether an individual had died during the observed sample period. In Section IV, we show that the period during which buyers are most likely to take advantage of their private information is four to six years before death. Regarding (b), the HRS already is one of the most comprehensive datasets available and we believe, therefore, we have controlled for the majority of the risk classification factors that underwriters consider.

[13] All Section IV estimates are nearly identical, in both magnitudes and standard errors, when we use a linear

$$newbuyer_i = \alpha_0 + \alpha_1 mortality_i + \mathbf{X}_i\mathbf{B} + e_i \qquad (1)$$

where binary dependent variable *newbuyer* is unity if an individual reported holding individual term life insurance in wave 1994 but not in 1992 (a new buyer), and 0 if she reported having individual term life insurance in neither 1994 nor 1992 (a non-owner).[14,15]

mortality indicates whether an individual had died by wave 2004. The long data window allows us to use this ex-post mortality risk measure rather than the self-perceived or estimated actual mortality risk as in Cawley and Philipson. The self-perceived probability to live to age 75, reported by the HRS respondents, is a controversial measure of individuals' private information about mortality risk because many people have difficulty understanding and answering probabilistic questions (Hurd and McGarry 1995, Gan et al. 2005).[16]

Vector **X** represents the set of risk classification variables. Any asymmetric information test should be conditioned on insurers' risk classifications because the prediction of a positive risk-coverage correlation under asymmetric information applies only *within* risk classes, not *across* risk classes. Because we do not observe the actual risk classification each sample individual faces, we control as exhaustively as possible for the factors – what we term "pricing factors" – influencing insurers' price offers and willingness-to-insure (see the Appendix for the life insurance industry's underwriting practices).[17] This exhaustive set of pricing factors **X** – which we call *full pricing controls* – includes: (a) the respondent's age, gender, and smoking status (whether an individual

---

probability version of model (1).

[14] For those individuals who died between waves 1992 and 1994, questions about life insurance ownership were answered by proxy interviewees, who supposedly were able to provide accurate information on the deceased respondents. The proxies usually were a surviving spouse, children, or other informants. We obtain qualitatively similar results when we exclude the deceased individuals from our sample of potential new buyers.

[15] Note that *newbuyer* is defined solely in terms of whether a respondent reported owning *individual* term life insurance in waves 1992 and 1994. Individuals with *newbuyer*=0 may have held *group* term insurance prior to 1992 and carried it into the 1992 - 1994 period, purchased it new during the 1992 – 1994 period, or never held it at all. If group and individual insurance are substitutable for one another, our way of coding should induce a bias against finding evidence for asymmetric information. Potential new buyers who bought group but not individual insurance between 1992 and 1994, and who thus are coded as non-owners, may be high-risk and thus would have purchased individual insurance if deprived of group insurance. Treating such individuals as "non-owners" rather than "new buyers" should dampen the mortality differences between the two groups.

[16] For example, the histogram of self-reported mortality probabilities in the HRS sample shows that such probabilities tend to anchor at appealing numbers known as "focal points." Nearly half the respondents reported either 50% or 100% as the chance they would live to age 75, an unlikely representation of true subjective mortality probabilities.

[17] Whenever possible, we obtain those variables from the RAND version of HRS.

has ever smoked, and whether she smokes now); (b) health status and medical history (whether she drinks alcohol now; whether she has been diagnosed with diabetes, high blood pressure, cancer, heart disease, arthritis, lung disease, stroke, asthma, kidney disease, ulcer, high cholesterol, or back pain; whether she has had a hospital stay in the previous 12 months; and whether her BMI indicates she has healthy weight, is overweight, or obese);[18] and (c) family history (whether her father or mother had died before age 60).[19,20]

A potential concern with the full pricing control set is that we may be "over-controlling" by including excessively detailed health-related controls.[21] Whether or not we over-control is unclear. On the one hand, we do over-control to the extent that our highly detailed health indicator variables provide more risk categories than life insurers typically employ. On the other hand we do not over-control to the extent that our controls likely are not specified in a correct or flexible enough functional form to perfectly mimic the life insurance industry's actual risk classifications. To investigate robustness, we also adopt an alternative set of pricing controls, which we call *limited pricing controls*. The limited pricing control set contains only part (a) of our preferred full pricing control set **X**, namely respondent's age, gender, and smoking status. This set likely represents an under-control of the risk classification because, for any given combination of age, gender, and smoking status, life insurers typically group individuals into distinct risk categories based on their health status and medical and family history.

Coefficient α_1 is our parameter of interest, measuring the correlation between mortality risk and coverage conditional on risk classification. A positive estimate of α_1 suggests the presence of asymmetric information.[22]

---

[18] In the BMI, healthyweight lies between 18.5 and 24.5 and overweight between 24.5 and 30. A BMI above 30 is obese.

[19] The HRS records whether the respondent's mother and father are alive and if so what their age is; and if they are not, the age when they died. We code a parent who died before 60 as an indicator for unfavorable family history. We thus have two indicator variables for family history, one being whether the father had died before 60 and the other whether the mother had died before 60. These are crude measures. The HRS data do not, however, provide information about parents' cause of death.

[20] Indeed, to allow for possible nonlinearity among pricing factors, we also, as a robustness check, estimate model (1) with a complete set of two-way interaction terms among these full pricing factors. Estimates of the risk-coverage conditional correlation are similar to those in models not containing the interaction terms.

[21] The author thanks the editor and an anonymous referee for this insightful discussion.

[22] One ideally would like to examine the conditional correlation between the amount of individual term insurance purchased and subsequent mortality. The HRS did not, however, ask a respondent for the amount of *individual* term insurance she held, even though it did ask her to report the amount of *term* insurance. Because term insurance can be

Note that model (1) excludes from the analysis those individuals who owned life insurance at the beginning of our sample period. Any given life policy holder must, at an earlier point in her life, have been a "potential new buyer." That is, she would at such time have entered the market in the sense of considering whether to purchase life insurance. In other words, the entire group of insurance holders consists of many disjoint "slices" of new buyers, each being a sub-group defined relative to time *t* when that sub-group's members first purchased a policy. Correspondingly, the entire population consists of many slices – sub-populations or cohorts – of potential new buyers, also defined relative to time *t*. When we limit our analysis to the potential new buyers defined relative to 1992, we are choosing to examine one cohort of the entire population. Because purchase time may not be random, evidence for or against asymmetric information among this cohort may not be representative of what one may find in other cohorts.[23]

IV. Results

We report the results in four steps. For the sake of comparison, the first step replicates the Cawley and Philipson results. The second step modifies the controls to reflect risk classification in life insurance markets. The third step adopts, together with proper risk classification controls, the 12-year mortality indicator as our mortality risk measure. The fourth step repeats the third step but restricts the analysis to the sample of potential new buyers. That final step produces our paper's main result.

A. Replicating Cawley and Philipson's Result

Columns (1)-(2) in Table 2 reflect Step 1. Column (1) reports the logit coefficients obtained using Cawley and Philipson's (1999) primarily demand-side controls – variables affecting individuals' life insurance demand but not necessarily insurers' willingness to supply – plus estimated actual mortality risk, in conjunction with the entire cross-sectional 1992 sample.[24,25] In contrast to *newbuyer* in model (1), the

either individual or group, and group-market underwriting procedures are different from individual-market procedures, we cannot investigate the conditional correlation on the intensive margin.

[23] The author thanks the editor for insights into this discussion.

[24] Cawley and Philipson's (1999) logit regression controls for age, gender, marital status, smoking status, income,

dependent variable in the column (1) regression is $individual\_term$, a binary indicator for whether the respondent held individual term life insurance in the 1992 wave. The logit α_1 estimate is negative and statistically insignificant, consistent with the findings in Cawley and Philipson (see Table 5 in Cawley and Philipson 1999).[26] Column (2) reports the marginal effects of this logit regression.

B. Controlling for Risk Classification

In Step 2, we control for risk classification. We first drop the demand-side controls from the Step 1 specification, such as income, wealth, marital status, and bequest motives. Those variables are irrelevant for asymmetric information tests because life insurers do not price on such factors even if the information is available to them. In particular, column (3) is our under-control specification, controlling only for the limited pricing factors, which include age dummies, gender, and smoking status. The column (3) α_1 estimate remains negative and insignificant. Column (4) adds the health-related control variables omitted in the previous literature.[27] With the latter – preferred – set of controls, the column (4) α_1 estimate remains negative, consistent with the previous literature and contrary to what asymmetric information would suggest.

C. Using the 12-year Mortality Outcome to Measure Mortality Risk

---

wealth, and the following proxies for bequest motives: number of grandchildren, number of children, age of the youngest child, average age of children, number of siblings, and age of spouse.

[25] Also following Cawley and Philipson (1999), we obtain the estimated actual mortality risk as the predicted death probability between 1992 and 1994 from a logit regression of the 1992-1994 death indicator variable on the following repressors: age, age squared, height, weight, three cognitive measures, and indicators for female, black, white, married, diabetes, smoke now, smoke ever, drinks alcohol, cancer, lung disease, heart disease, stroke, psychiatric problems, arthritis, asthma, back problems, kidney disease, ulcers, high cholesterol, broken bones after age 45, high blood pressure, eyeglasses, and a hospital stay in the last year.

[26] Cawley and Philipson report in their Table 5 two logit coefficient estimates, -1.28 and -0.4, corresponding respectively to their "upper bound" and "lower bound" mortality variables as defined in Footnote 10. To save space, we obtain all our estimates by treating individuals with unknown vital status as missing observations. When we follow Cawley and Philipson's use of upper- and lower-bound mortality variables, we obtain estimates -0.705 and -0.509. These are in line with, though not exactly the same as, Cawley and Philipson's results.

[27] It is not clear, *a priori,* how this omitted variable problem biases the mortality-coverage conditional correlation estimate. If an observable health condition reduces the insurer's willingness to supply coverage, so that the premium rises or insurance is denied, then omitting observable health information biases the mortality-coverage correlation *downward* because the insurer's reduced supply will incorrectly be interpreted as a decline in the individual's privately harbored demand. If the observable health condition instead increases the individual's insurance demand, then omitting this observable information biases the correlation *upward* because the increase in the buyer's demand induced by this publicly observable information will be interpreted as an increase in her privately harbored demand.

In Step 3, we use the mortality outcome across the 12-year time window to measure the mortality risk. For comparison, column (5) of Table 2 continues to use Cawley and Philipson's mainly demand-side controls. Column (6) is our "under-control" specification, and column (7) includes our preferred set of full pricing controls. α_1 estimates in all three specifications are negative and insignificant and much smaller than those obtained in Step 2.

D. Asymmetric information in the Life Insurance Market: Restricting the Sample to Potential New Buyers

Step 4 repeats the Step 3 specifications but restricts the sample to potential new buyers. This step produces our main results. While column (8) continues to employ the Cawley and Philipson controls, column (9) is our under-control specification, namely controlling for age dummies, gender, and smoking status. Column (10) is our preferred specification, which adds health-related controls. The α_1 estimates now are all positive. Furthermore, they become significant in columns (9) and (10), where we properly control for risk classification. In our preferred specification, namely column (10) with the set of full pricing controls, the point estimate of the conditional correlation between mortality risk and life insurance coverage is 0.034, significant at the 5% level. This estimate indicates that the take-up rate among higher-risk individuals – those who died within the 12-year time window – is 3.4 percentage points higher than that among the lower-risk individuals – those who survived the 12-year window. Because lower-risk individuals' take-up rate in our sample is 0.18, the 0.034 says that higher-risk individuals have a 19 percent [(0.034/0.18)100] greater take-up rate than do lower-risk ones. That is, once risk classification is properly accounted for, those with higher mortality risk are substantially more likely to buy life insurance than are those with lower risk, implying that, even in the presence of stringent underwriting practices, individuals hold private information about their risk.

E. Purchase Timing

When potential buyers hold residual private information about their mortality risk and incorporate such information into their life insurance purchase decisions, we may expect that the earlier one died, the more likely it is that she would have purchased insurance. To investigate this prediction, we break our 12-year-window mortality risk measure into indicators of mortality outcomes between given pairs of consecutive waves. We estimate the following logit model:

$$newbuyer_i = \beta_0 + \sum_{t=1994}^{2004} \beta_t mort\_t_i + \mathbf{X}_i \mathbf{B} + e_i \qquad (2)$$

where *newbuyer* and control set **X** are defined as in model (1). Dummy variable *mort_t* indicates whether a respondent died between waves *t - 2* and *t*, with *t* taking values 1994, 1996, 1998, 2000, 2002, and 2004 corresponding to the respective interview waves. The β_t's are the parameters of interest, measuring how the mortality risk affects the take-up decision. We expect them to be monotonically decreasing as *t* grows.

For ease of reference, column (1) in Table 3 reproduces column (10) in Table 2. Columns (2) and (3) present the model (2) results. Column (2) is our under-control specification and column (3) the preferred specification. In both specifications, estimates of the parameters of interests are monotonically decreasing. The first three column (3) estimates – 0.139, 0.090, and 0.070 – suggest that those who died respectively two, four, and six years after the 1992 baseline exhibited take-up rates 77%, 50%, and 39% higher than those who survived more than 12 years after 1992 (the last with a 0.18 take-up rate). The first two estimates are significant at the 5% level and the last, with p-value 0.12, is marginally significant. Coefficients of *mort_2000, mort_2002,* and *mort_2004* are insignificant.

The above discrete measure of mortality risk provides a useful diagnosis of how those with residual private information time their purchase decisions. The estimates suggest that buyers are most likely to take up individual term life insurance four to six years before death.[28] If so, one alternatively may measure the mortality risk with an indicator signifying whether a potential new buyer had died by wave 1998.[29] Columns (4) and (5) in Table 3 present the estimates under such a risk measure. Again, column (4) includes the limited pricing controls and column (5) the full pricing controls. From these

[28] This may explain the fact that 5-year Level Term insurance accounts for the highest market share (34.8%) of all individual term life insurance (LIMRA, 1997).
[29] The 12-year time-window length in the main analysis is somewhat arbitrary in the sense that it is determined by the sample period length.

columns, we estimate, as expected, a much larger coverage-risk conditional correlation than we do using the 12-year-window mortality risk measure. In our preferred column (5) specification, mortality's 0.089 point estimate (significant at 1%) implies that individuals who died within six years after a baseline year had a 49% higher take-up rate in that base year than did those who survived beyond year six (the latter with a 0.18 mean take-up rate).[30]

V. Conclusion

We find, contrary to the earlier literature, evidence for asymmetric information in the life insurance market. After risk classification is carefully taken into account, individuals with higher mortality risk are 19% to 49% more likely to buy individual term life insurance than are those with lower risk, depending on the length of the time window within which the mortality risk is defined. Moreover, buyers appear to employ this informational advantage by seeking to take up insurance four to six years before death. Such results provide an alternative view of the informational content of life insurance markets, calling into question the widely held notion that life insurance is free of asymmetric information. Furthermore, the failure of the life insurance industry's comparatively stringent underwriting practices to eliminate strategic purchasing suggests that informational asymmetry might be even more prevalent in other insurance markets.

Our focus on potential new buyers is what drives the difference between our and the earlier literature's findings. When individuals do hold private information, analysis based on the entire cross-sectional data produces downward-biased estimates of the coverage-risk correlation. This bias arises because high-risk individuals with coverage are under-represented in cross-sectional samples, so that observed mortality differences between those with and without coverage are dampened in such samples. Our main contribution has been to show that restricting the sample to potential new buyers solves this sample selection problem, providing unbiased asymmetric information tests.

[30] The self-reported probability of living to age 75 is on average substantially higher among new buyers who survived beyond 1998 than it is among those who died by 1998 (64 vs. 54 on a scale from 0-100).

References

American Council of Life insurers (ACLI), 2007a. Protectors and Investors: America's Life Insurers. Available at http://www.acli.com/NR/rdonlyres/F40056AC-871D-4E1A-B0E0-EDA1DEA1AB92/10868/ProtectorsInvestors_Jan2007.pdf

American Council of Life insurers (ACLI), 2007b. Life Insurers Fact Book 2007. Available at http://www.acli.com/NR/rdonlyres/A85A882F-F871-431D-976E-3316884C63EB/15016/FB_07_AllChapters2_w_insert.pdf

Cawley, J., Philipson T., 1999. An empirical examination of information barriers to trade in insurance. American Economic Review 89, 827-846.

Cardon, J., Hendel, I., 2001. Asymmetric information in health insurance: evidence from the national medical expenditure survey. Rand Journal of Economics 32, 408-427.

Chiappori, P., Salanie, B., 2008. Modeling competition and market equilibrium under asymmetric information: empirical issues. 2008 AEA Annual Meeting, session titled "Empirical Work on Asymmetric Information in Insurance Markets."

Chiappori, P., Salanié, B., 2003. Testing contract theory: a survey of some recent work. In: Dewatripont, M., Hansen, L., Turnovsky, P. (Eds.). Advances in Economics and Econometrics - Theory and Applications, Eighth World Congress, Econometric Society Monographs. Cambridge: Cambridge University Press, 115-149.

Chiappori, P., Salanié B., 2000. Testing for asymmetric information in insurance markets. Journal of Political Economy 108, 56-78.

Chiappori, P., 2000. Econometric models of insurance under asymmetric information. In: Dionne, G., (Eds.). Handbook of Insurance, Boston: Kluwer, 365-94.

Chiappori, P., Jullien, B., Salanié, B., Salanié, F., 2006. Asymmetric information in insurance: general testable implications. Rand Journal of Economics 37,783-798.

Cohen, A., 2005. Asymmetric information and learning: evidence from the automobile insurance market. Review of Economics and Statistics 87, 197-207.

Cummins, J. D., Smith, B.D., Smith, R.N., Vanderhei, J.L., 1983. Risk classification in life insurance. Kluwer-Nijhoff Publishing.

Cutler, D., Finkelstein A., McGarry, K., 2008. Preference heterogeneity in insurance markets: explaining a puzzle. American Economic Review Papers and Proceedings 98, 157-162.

de Maza, D., Webb, D., 2001. Advantageous selection in insurance markets. Rand Journal of Economics 32, 249-262.

Fang, H., Keane, M., Silverman, D., 2006. Sources of advantageous selection: evidence from the Medigap insurance market. Working paper.

Finkelstein, A., McGarry, K., Sufi A., 2005. Dynamic inefficiencies in insurance markets: evidence from long-term care insurance. American Economic Review Papers and Proceedings 95, 224-228.

Finkelstein, A., Poterba, J., 2004. Adverse selection in insurance markets: policyholder evidence from the U.K. annuity market. Journal of Political Economy 112, 183-208.

Finkelstein, A., McGarry, K., 2006. Multiple dimensions of private information: evidence from the long-term care insurance market. American Economic Review 93, 938-958.

Gan, L., Hurd, M., McFadden, D., 2005. Individual subjective survival curves. In: Wise, D. (Eds.). Analyses in Economics of Aging. University of Chicago Press: 377-411.

Hausman, J., Abrevaya, J., Scott-Morton, F., 1998. Misclassification of a Dependent Variable in a Discrete Response Setting. Journal of Econometrics 87, 239-269.

Hendel, I., Lizzeri, A., 2003. The role of commitment in dynamic contracts: evidence from life insurance. Quarterly Journal of Economics 118, 299-327.

Hurd, M., McGarry, K., 1995. Evaluation of the subjective probabilities of survival in the Health and Retirement Study. Journal of Human Resources (Special Issue on the Health and Retirement Study: Data Quality and Early Results) 30: S268-S292.

Kalbfleisch, J., Prentice, R., 2002. The Statistical Analysis of Failure Time Data, 2nd Edition. John Wiley & Sons, Inc.

Kimball, M. S., Saham, C. R., Shapiro, M. D., 2007a. Imputing risk tolerance from survey responses. Journal of the American Statistic Association: Applications, forthcoming.

Kimball, M. S., Saham, C. R., Shapiro, M. D., 2007b. User's guide for risk preference parameters. Available at http://www-personal.umich.edu/~shapiro/data/risk_preference/ImputationUsersGuideJASA.pdf

LIMRA, 1997. Buyer's Study. Connecticut.

Lizzeri, A., Hendel, I., 2008. Who commits to long-term contracts? How much does it cost? The case of life insurance. 2008 AEA Annual Meeting, session titled "Empirical Work on Asymmetric Information in Insurance Markets."

McCarthy, D., Mitchell, O., forthcoming. International Adverse Selection in Life Insurance and Annuities. In: Tuljapurkar, S., Ogawa, N., and Gauthier, A. (Eds.). Riding the Age Waves: Responses to Aging in Advanced Industrial States. Elsevier. *Forthcoming*. (Also see NBER working paper #9975, http://www.nber.org/papers/w9975.)

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Table 1 Sample Summary Statistics

variable	definition	mean	std. dev
individual_term	whether reported owning individual term life insurance	0.24	0.43
individual_term (in year 1994)	whether reported owning individual term life insurance in 1994	0.27	0.45
term	whether reported owning term life insurance	0.51	0.50
life	whether reported owning life insurance	0.72	0.45
newbuyer	=1 if reported owning individual term life insurance in wave1994, but not in 1992; =0 if reported owning in neither waves.	0.19	0.39
mortality (entire 1992 sample, 2004)	whether dead by wave 2004	0.15	0.35
mortality (potential new buyers, 2004)	whether dead by wave 2004	0.15	0.35
mortality (entire 1992 sample, 1998)	whether dead by wave 1998	0.06	0.23
mortality (potential new buyers, 1998)	whether dead by wave 1998	0.06	0.23
newbuyer (higher risk, 2004)	whether being a new buyer (among those who died before 2004)	0.21	0.41
newbuyer (lower risk, 2004)	whether being a new buyer (among those who survived beyond 2004)	0.18	0.39
newbuyer (higher risk, 1998)	whether being a new buyer (among those who died before 1998)	0.26	0.39
newbuyer (lower risk, 1998)	whether being a new buyer (among those who survived beyond 1998)	0.18	0.44
age	age	55.56	3.25
gender	=1 if male, =0 if female	0.48	0.50
smoke_ever	whether smoke now	0.64	0.48
smoke_now	whether smoke ever	0.27	0.44
drink	whether drink now	0.63	0.49
diabetes	whether diagnosed with diabetes	0.08	0.28
HBP	whether diagnosed with HBP	0.33	0.47
cancer	whether diagnosed with cancer	0.05	0.21
heart	whether diagnosed with heart disease	0.10	0.31
arthritis	whether diagnosed with arthritis	0.34	0.47
lunge	whether diagnosed with lung disease	0.06	0.23
stroke	whether diagnosed with stroke	0.02	0.15
asthma	whether diagnosed with asthma	0.06	0.24
kidney	whether diagnosed with kidney disease	0.10	0.30
ulcer	whether diagnosed with ulcer	0.09	0.28
cholesterol	whether diagnosed with high cholesterol	0.25	0.43

back_pain	whether suffering from back pain	0.35	0.48
hospital_stay	whether had a hospital stay in the previous 12 months	0.11	0.31
BMI	Body mass index	26.98	5.00
underweight	whether BMI<=18.5	0.01	0.11
healthyweight	whether BMI<=24.5 and BMI>18.5	0.32	0.47
overweight	whether BMI<=30 and BMI>24.5	0.44	0.50
obese	whether BMI>30	0.22	0.41
history_father	whether father died before 60	0.20	0.40
history_mother	whether mother died before 60	0.12	0.33
income	household income	49,717	54,349
wealth	household wealth	235,626	485,259
married	whether married	0.74	0.44
num_grandkid	number of grandchildren	2.28	3.97
num_kid	number of children	3.18	2.05
min_age_kid	age of the youngest child	24.46	7.09
mean_age_kid	average age of children	28.58	6.61
num_sibling	number of siblings	2.85	2.40
age_spouse	spouse's age	55.12	6.92

Note: Summary statistics are, unless otherwise noted, based on HRS 1992 and are weighted by the 1992 individual sampling weights.

Table 2 Take-up Decisions: Adverse Selection in the Life Insurance Market

	entire cross-sectional sample							potential new buyers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	CP controls (coefficients)	CP controls	limited pricing controls	full pricing controls	CP controls	limited pricing controls	full pricing controls	CP controls	limited pricing controls	full pricing controls
mortality1992-1994	**-0.515**	**-0.093**	**-0.150**	**-0.351**						
	(1.135)	**(0.206)**	**(0.148)**	**(0.222)**						
mortality1992-2004					**-0.006**	**-0.004**	**-0.005**	**0.032**	**0.027***	**0.034****
					(0.019)	**(0.014)**	**(0.015)**	**(0.022)**	**(0.016)**	**(0.018)**
Observations	5989	5989	8737	8737	6191	8803	8794	4336	6116	6113
Pseudo R^2	0.012	0.012	0.006	0.008	0.012	0.006	0.008	0.007	0.005	0.010

Note: All columns, except for column (1), report marginal effects from logit regressions. The dependent variable in columns (1)-(7) is *individual_term*, indicating whether an individual reported owning individual term life insurance in the 1992 wave. The dependent variable in columns (8)-(10) is *newbuyer*, indicating whether an individual is a new buyer or a non-owner as defined in the text. $mortality^{1992-1994}$ is estimated actual mortality risk – the predicted death probability between 1992 and 1994 from a logit regression, as in Cawley and Philipson (1999) (see footnote 23 for the regressors in this logit regression). $mortality^{1992-2004}$ indicates whether an individual died by wave 2004. Columns (1)-(2), (5), and (8) control for age, gender, smoking status (whether smokes now and whether ever smoked), marital status, income, wealth, and proxies for bequest motives (number of grandchildren, number of children, age of youngest child, average age of children, and number of siblings). Columns (3), (6), and (9) control for age dummies, gender, and smoking status. Columns (4), (7), and (10) add the following health-related controls: whether drinks alcohol now; whether has been diagnosed with diabetes, high blood pressure, cancer, heart disease, arthritis, lung disease, stroke, asthma, kidney disease, ulcer, high cholesterol, or back pain; whether had a hospital stay in the previous 12 months; whether BMI indicates healthy weight, overweight, or obese; and whether father and mother had died before age 60. Please refer to the text and Table 1 for detailed variable definitions. All control coefficient estimates are suppressed. All regressions are weighted by HRS 1992 individual sampling weights. Robust standard errors are reported in parentheses. ***, **, and * indicates significance level at 1%, 5%, and 10%, respectively.

Table 3 Take-up Decisions: Alternative Mortality Risk Measures

VARIABLES	(1) full pricing controls (baseline)	(2) limited pricing controls	(3) full pricing controls	(4) limited pricing controls	(5) full pricing controls
mortality1992-2004	**0.034****				
	(0.018)				
mort_1994		**0.109***	**0.139****		
		(0.058)	**(0.062)**		
mort_1996		**0.074***	**0.090****		
		(0.043)	**(0.046)**		
mort_1998		**0.064**	**0.070**		
		(0.043)	**(0.044)**		
mort_2000		**-0.025**	**-0.016**		
		(0.030)	**(0.031)**		
mort_2002		**0.002**	**0.007**		
		(0.031)	**(0.032)**		
mort_2004		**0.015**	**0.019**		
		(0.036)	**(0.037)**		
mortality1992-1998				**0.077*****	**0.089*****
				(0.027)	**(0.029)**
Observations	6113	6076	6073	6302	6299
Pseudo R^2	0.009	0.007	0.011	0.007	0.011

Note: All columns report marginal effects from logit regressions. The dependent variable in all columns is *newbuyer*, indicating whether an individual is a new buyer or a non-owner as defined in the text. $mortality^{1992\text{-}1994}$ is the estimated actual mortality risk – the predicted death probability between 1992 and 1994 from a logit regression, as in Cawley and Philipson (1999) (see footnote 23 for the regressors). $mortality^{1992\text{-}2004}$ indicates whether an individual died by wave 2004; *mort_1994* indicates whether an individual died between waves 1992 and 1994; *mort_1996* indicates death between waves 1994 and 1996; *mort_1998, mort_2000, mort_2002,* and *mort_2004* are similarly defined; $mortality^{1992\text{-}1998}$ indicates whether an individual died by wave 1998. Columns (2) and (4) control for age dummies, gender, and smoking status (whether smokes now and whether ever smoked). Columns (1), (3), and (5) add the following health-related controls: whether drinks alcohol now; whether has been diagnosed with diabetes, high blood pressure, cancer, heart disease, arthritis, lung disease, stroke, asthma, kidney disease, ulcer, high cholesterol, or back pain; whether had a hospital stay in the previous 12 months; whether BMI indicates healthy weight, overweight, or obese; and whether father and mother had died before age 60. Please refer to the text and Table 1 for detailed variable definitions. All control coefficient estimates are suppressed. All regressions are weighted by HRS 1992 individual sampling weights. Robust standard errors are reported in parentheses. ***, **, and * indicates significance level at 1%, 5%, and 10% respectively.

Appendix: Underwriting in Individual Term Life Insurance Markets

The life industry's underwriting procedure for individual term policies is quite uniform across states. Basic pricing factors include age, gender, personal habits (e.g., tobacco, alcohol, or drug use), health status and medical history, family history, and some vocations and hobbies. Other factors may include driving records, aviation activities, residence, and frequency and destination of foreign travel.[31] Premiums are higher for the elderly, males, those with a history of smoking, drinking, or drug abuse, those with unfavorable health status and/or an unfavorable medical or family history, those in hazardous vocations, and those with high-risk hobbies.

A life insurance agent typically interviews the applicant after the application is received. Standard questions about health status and medical history are whether one has been diagnosed with high blood pressure, stroke, cancer, diabetes, high cholesterol, or a series of other conditions. The common question about family history is whether one or both parents died before 60 or 70 of cardiovascular disease or cancer.[32] The insurer typically also requires a medical examination of the applicant and her permission to release medical records. During the medical examination, a paramedic usually collects blood and urine samples, measures blood pressure, height, and weight, and records the applicant's medical history. The higher the face value of the insurance policy, the more detailed is the information required.

After gathering the applicant's information, the insurer adds to or deducts from a common base score points for favorable or unfavorable information. Based on the final score, the insurer classifies an applicant into a risk category such as "preferred plus," "preferred," "standard plus," or "standard."[33] Sub-categories often are available within these categories. The premium is largely similar for applicants in the same risk category given the same age, gender, and smoking status. The insurer usually would decline as uninsurable those individuals with five times the base score (McGill's Life Insurance 2000).

The tables below illustrate the underwriting guidelines for individual term life insurance provided by QuickQuote.com, an online quoting system. It shows how applicants with alternative pricing characteristics would be generally grouped into alternative risk categories. The first table includes most pricing factors except medical history; the second includes medical history.[34] In both tables, column (1) lists the requirements an applicant needs in order to qualify for the best risk category, "preferred plus." Columns (2), (3), and (4) refer respectively to "preferred," "standard plus," and "standard." For example, an applicant usually will not qualify for "preferred plus" if she has ever received high blood pressure treatments or her blood pressure readings have ever exceeded 140/85 (see column 1, "blood pressure" row, Appendix A, first table). This individual may, however, still qualify for "preferred" if her blood pressure is now under control and her readings have not exceeded 145/88 in the past two years (column 2, "blood pressure" row, same table).

[31] See McGill's Life Insurance (2000), Cummins et al., (1983) and records of the author's phone conversations with state insurance departments.
[32] The weight placed on family history has, except for cardiovascular-renal diseases, been declining in recent years on account of the difficulty of verifying the information. See McGill's Life Insurance (2000), p 520-521.
[33] Some companies have three categories: preferred, standard, and substandard. Category names can vary.
[34] Age, gender, and factors like height, weight, and BMI are not listed here, as they are self-explanatory.

	Prefered Plus	Preferred	Standard Plus	Standard
	(1)	(2)	(3)	(4)
Family History	No cardiovascular disease or cancer in either parent or siblings prior to age 60.	No death from cardiovascular disease or cancer in either parent or siblings prior to age 60.	Not more than one parent death from cardiovascular disease or cancer prior to age 60.	Not more than one parent death from cardiovascular disease or cancer prior to age 60.
Cholesterol / HDL Ratio	May not exceed 5.0	May not exceed 6.0	May not exceed 7.0	Levels above 7.0 may qualify
Cholesterol Level	May not exceed 220	May not exceed 240	May not exceed 260	Levels above 260 may qualify
Blood Pressure	No history of treatment. May not exceed 140/85.	Currently controlled. Current and historic readings over last two years may not exceed 145/88	Currently controlled. Current and historic readings over last two years may not exceed 150/92	Currently controlled. Current and historic readings over last two years may not exceed 150/92
Alcohol / Substance Abuse	No history.	No history in the past 10 years.	No history in the past 7 years.	No history in the past 7 years.
Driving History	No DUI, DWI or reckless driving in the past 5 years. No more than 1 moving violations in the last 3 years.	No DUI, DWI or reckless driving in the past 5 years. No more than 2 moving violations in the last 3 years.	No DUI, DWI or reckless driving in the past 3 years. No more than 3 moving violations in the last 3 years.	No DUI, DWI or reckless driving in the past 2 years. No more than 3 moving violations in the last 3 years.
Aviation	Commercial airline pilots may qualify. Not available for private pilots.	Commercial airline pilots may qualify. Private pilots may qaulify with an exclusion rider or extra premium.	Commercial airline pilots may qualify. Private pilots may qualify with an exclusion rider or extra premium.	Commercial airline pilots may qualify. Private pilots may qualify with an exclusion rider or extra premium.
Hazardous Avocation**	Not available.	May be available with extra premium.	May be available with extra premium.	May be available with extra premium.
Residence and / or Citizenship	Must be a U.S. resident for the past 3 years. Must be a US citizen or have permanent Visa.	Must be a U.S. resident for the past 3 years. Must be a US citizen or have permanent Visa.	Must be a U.S. resident for the past 3 years. Must be a US citizen or have permanent Visa.	Must be a U.S. resident for the past 3 years. Must be a US citizen or have permanent Visa.
Military	No active duty.	May be on active duty.	May be on active duty.	May be on active duty.
Foreign Travel	No travel to countries under State Department Advisory. Varies by company.	No travel to countries under State Department Advisory. Varies by company.	No travel to countries under State Department Advisory. Varies by company.	No travel to countries under State Department Advisory. Varies by company.

* The information here is from QuickQuote.com, a popular online life insurance quoting system and represents a collective sample of underwriting guidelines. The original table

is available at http://www.quickquote.com/uwGuideLines.html. (Last accessed on April 19th, 2008.)
** Examples include, but are not limited to, scuba diving, jet, snow, and water skiing, snowboarding, hang gliding, skydiving, paragliding, bungee jumping, mountain climbing, and amateur racing. Rules can vary by company.

Medical History				
Condition	**Preferred Plus**	**Preferred**	**Standard Plus**	**Standard**
	(1)	(2)	(3)	(4)
Alcohol / Drug Abuse Dependancy History	No	Yes	Yes	Yes
Anxiety	No	No	Yes	Yes
Arthritis (rheumatoid)	No	Yes	Yes	Yes
Asthma	No	Yes	Yes	Yes
Chronic Bronchitis	No	Yes	Yes	Yes
Cancer	No	No	No	Yes
Cardiovascular/ Heart Disease	No	No	No	Yes
Cholesterol Treatment	No	Yes	Yes	Yes
Chronic Obstructive Pulmonary Disease	No	Yes	Yes	Yes
Crohn's Disease	No	No	No	Yes
Depression	No	No	Yes	Yes
Diabetes Type I * **	No	No	No	No
Diabetes Type II	No	No	No	Yes
Emphysema	No	No	No	Yes
Epilepsy	No	No	Yes	Yes
Hypertension (High Blood Pressure)	No	Yes	Yes	Yes
Kidney / Liver Disease (chronic)	No	No	No	Yes
Melanoma	No	No	No	Yes
Multiple Sclerosis	No	No	No	Yes
Sleep Apnea	No	No	No	Yes
Stroke (including TIA)* **	No	No	No	No
Ulcerative Colitis	No	No	No	Yes
Vascular Disease	No	No	No	Yes

*** A substandard rating may be available for these medical conditions, depending on individual circumstances and insurance company guidelines