Sources of Advantageous Selection: Evidence from the Medigap Insurance Market

Hanming Fang

Duke University

Michael P. Keane

University of Technology, Sydney and Arizona State University

Dan Silverman

University of Michigan

We provide evidence of advantageous selection in the Medigap insurance market and analyze its sources. Conditional on controls for Medigap prices, those with Medigap spend, on average, \$4,000 less on medical care than those without. But if we condition on health, those with Medigap spend \$2,000 more. The sources of this advantageous selection include income, education, longevity expectations, and financial planning horizons, as well as cognitive ability. Conditional on all these factors, those with higher expected medical expenditures are more likely to purchase Medigap. Risk preferences do

We thank Pierre-André Chiappori, Amy Finkelstein, Alessandro Lizzeri, Jim Poterba, Bernard Salanié, Aloysius Siow, Robert Town, seminar and conference participants at Brown, Concordia, Duke/University of North Carolina, Federal Reserve Bank of New York, Johns Hopkins, Minnesota, Queens, Texas–Austin, Toronto, University of California at Los Angeles, Virginia, Wharton, Wisconsin–Madison, Yale, the National Bureau of Economic Research, and the Stanford Institute for Theoretical Economics for helpful suggestions. We especially thank two anonymous referees and the editor, Canice Prendergast, for their thoughtful comments that significantly improved the paper. Mike Chernew and Tami Swenson assisted us with the access to and clarification of the MCBS data, and John Robst gave us access to Medigap plan pricing data from Weiss Ratings. All remaining errors are our own. Fang and Silverman gratefully acknowledge financial support from the Economic Research Initiative on the Uninsured. Keane's research was supported by Australian Research Council grant FF0561843.

[*Journal of Political Economy,* 2008, vol. 116, no. 2]

not appear as a source of advantageous selection; cognitive ability is particularly important.

I. Introduction

Asymmetric information is central to modern economic models of insurance pioneered by Arrow (1963) and Pauly (1974). The classic equilibrium models developed by Rothschild and Stiglitz (1976) and Wilson (1977) assume that potential insurance buyers have one-dimensional private information regarding their risk type. They choose from a menu of contracts, each specifying a premium and amount of coverage, the one best suited to their type. These models predict a positive correlation between insurance coverage and ex post realizations of loss. The reason is ex ante adverse selection: the "bad risks" (i.e., those relatively likely to suffer a loss) have an incentive to buy more insurance. Allowing for ex post moral hazard only strengthens the positive correlation between coverage and ex post loss. This "positive correlation property" of classic asymmetric information models forms the basis for empirical tests of asymmetric information in several recent papers (see Chiappori and Salanié 2000).

These empirical tests, reviewed further in Section II below, generate mixed results. In some markets, such as automobile insurance (e.g., Chiappori and Salanié 2000) and long-term-care insurance (Finkelstein and McGarry 2006), there was no statistically significant evidence of the positive correlation property. Findings like these have fueled an emerging literature on the possibility that multidimensional private information may lead to what has been called "advantageous selection." On the theoretical side, de Meza and Webb (2001) postulate that individuals have private information about both their risk type and their risk aversion.[1] They argue that selection based on risk aversion is advantageous if those who are more risk averse both buy more insurance coverage and have lower risks. Failure to condition on risk aversion may then mask the positive correlation between insurance coverage and ex post risk predicted by one-dimensional models. Following de Meza and Webb, the existing literature points to risk preferences as the primary suspect behind advantageous selection. In general, however, any private information could function as a source of advantageous selection if it is positively correlated with insurance coverage and at the same time negatively correlated with risk.

In this paper, we examine the evidence for and sources of advanta-

[1] Hemenway (1990), which used the term "propitious selection," seems to contain the first description of this phenomenon in the economics literature.

geous selection in the "Medigap" insurance market. The Medicare program provides limited health insurance for U.S. senior citizens. A Medigap policy is health insurance sold by a private insurer to fill "gaps" in coverage of the basic Medicare plan (e.g. co-pays, prescription drugs).[2] The Medigap market is ideal for studying multidimensional private information and advantageous selection because of two key features.

First, since 1992, the coverage and pricing of Medigap policies have been highly regulated by the U.S. government. Specifically, in all but three states, insurance companies can sell only 10 standardized Medigap policies; moreover, within the 6-month Medigap open-enrollment period—which starts when an individual both is older than 65 and is enrolled in Medicare Part B—an insurer cannot deny Medigap coverage, place conditions on a policy, or charge more for pre-existing health conditions. As shown in the theoretical analysis of Chiappori et al. (2006), in order for multidimensional private information to manifest itself in the form of a violation of the positive correlation property, the supply side of the insurance market has to be noncompetitive in the sense that the insurance companies are not free to offer any insurance contract they choose. Thus, the standardization of Medigap policies and the restrictions on medical underwriting make this market especially well suited to studying the evidence for multidimensional private information.

Second, the Medigap market is closely linked to the Medicare program. As a result, one can obtain detailed administrative data on diagnoses, treatments, and expenditures of consumers in the Medigap market. Specifically, our analysis relies in part on the Medicare Current Beneficiary Survey (MCBS), which combines survey data and Medicare administrative records. The Medicare administrative data on medical expenditures provide perhaps the most accurate measure of health expenditure risk for a large sample of the entire Medicare population. The MCBS also contains extensive health measures that allow us to obtain accurate measures of ex post expenditure conditional on age and health. Though the MCBS itself does not contain detailed information about risk aversion and other potential sources of advantageous selection, the Health and Retirement Study (HRS), a longitudinal data set covering a large sample of the Medicare-eligible population, has information about such variables. Our empirical strategy uses the MCBS and HRS jointly to examine the sources of advantageous selection.

We find strong evidence of multidimensional private information and advantageous selection in the Medigap market. Conditional on controls for the price of Medigap, we find that medical expenditures for senior

[2] See Sec. IV for more details about the Medicare program and the Medigap insurance market.

citizens with Medigap coverage are, on average, about $4,000 less than for those without. This strong negative correlation between ex post risk and coverage cannot be consistent with "no private information" or "one-dimensional private information" models of the insurance market; thus it directly indicates the presence of multidimensional private information, as well as advantageous selection. Indeed, once we condition on health (which cannot, by law, be used in Medigap pricing), expenditures for seniors with Medigap are about $2,000 more than for those without. These findings indicate that those who purchase Medigap tend to be healthier; that is, there is advantageous selection.

Equally important, we investigate several potential sources of this advantageous selection. This analysis points to variation in cognitive ability as a prominent source of advantageous selection. We find that elderly citizens with higher cognitive ability both are more likely to purchase Medigap and are healthier. We also investigate the potential pathways through which cognitive ability may act as a source of advantageous selection.

Interestingly, we find that variation in risk preferences, which was much discussed in the previous literature, does not appear to be a primary source of advantageous selection in the Medigap insurance market. Specifically, we find that even though direct measures of risk tolerance are significant predictors of Medigap insurance purchase, those who are more risk averse are not particularly healthy; as a result, risk preferences do not much contribute to advantageous selection.[3]

Our paper is most closely related to Finkelstein and McGarry's (2006) study of the long-term-care (LTC) insurance market. Using panel data from a sample of Americans born before 1923 (the Asset and Health Dynamics among the Oldest Old [AHEAD] study), they find no statistically significant correlation between LTC coverage in 1995 and use of nursing home care in the period between 1995 and 2000, even after controlling for insurers' assessment of a person's risk type. This evidence, alone, is consistent both with "no asymmetric information" and with "multidimensional private information." To distinguish between these stories, they first eliminated the no asymmetric information interpretation. Specifically, they found that a subjective probability assessment contained in the 1995 AHEAD questionnaire—"What do you think are the chances that you will move to a nursing home in the next five years?"—is positively correlated with both LTC coverage and nursing home use in 1995–2000, even after controlling for insurers' risk assessment. Since this variable is presumably unobserved by the insurer, these

[3] In this regard, our result is similar to that in Cohen and Einav (2007). The authors inferred both accident risk and risk aversion using an estimated structural model of automobile insurance deductible choice and found that they are positively correlated, contrary to what is required for risk aversion to be a source of advantageous selection.

positive correlations suggest private information and adverse selection by the insured.[4] Second, they developed a proxy for risk aversion, using information on whether respondents undertake various types of preventive health care. They found that people who are more risk averse by this measure are both more likely to own LTC insurance and less likely to enter a nursing home—consistent with multidimensional private information and advantageous selection based on risk aversion. In fact, their findings suggest that, on net, adverse selection based on risk and advantageous selection based on risk aversion roughly cancel out in the LTC insurance market.

Our paper examines the Medigap market, which, as we argued above, is especially well suited for a study of advantageous selection. In doing so, our paper makes three new contributions to the literature.

First, our method of inference for the presence of multidimensional private information differs from Finkelstein and McGarry's. We find a statistically significant and quantitatively large negative correlation between ex post medical expenditure and Medigap coverage, controlling for individual characteristics that are used in pricing. The large negative correlation between Medigap coverage and ex post medical expenditure is inconsistent with either no asymmetric information or single-dimensional private information, thus leading us directly to an interpretation of the results as evidence of multidimensional private information and at the same time as evidence of advantageous selection.

Second, our paper is, to our knowledge, the first to examine directly multiple potential sources of advantageous selection. Specifically, instead of using behavioral proxies for risk aversion as Finkelstein and McGarry did, we exploit the direct measures of risk aversion elicited from the respondents in the HRS data. More important, we examine not just risk preferences as the source of advantageous selection, but also several other potential sources.

Third, the empirical evidence in our paper suggests that for the Medigap insurance market, risk preferences, which were much discussed in the previous literature, do not appear to be a main source of advantageous selection; instead, our results suggest that cognitive ability plays a prominent role. We also explore various channels through which cognitive ability may lead to advantageous selection.

The remainder of the paper is structured as follows. Section II reviews additional related literature. Section III presents a simple conceptual framework to illustrate the idea of advantageous selection. Section IV provides some detailed background about Medicare and the Medigap

[4] Finkelstein and Poterba (2006) proposed such use of characteristics of insurance buyers that are observable to the econometrician but not used by insurers in setting prices as a general strategy to test for asymmetric information in insurance markets.

insurance market. Section V describes the MCBS and HRS data sets used in our analysis. Section VI provides direct and indirect evidence of advantageous selection using the MCBS data. Section VII examines the sources of advantageous selection by combining the MCBS and HRS and presents our main results. Finally, Section VIII presents conclusions.

II. Related Literature

Within the large literature that tests for the presence of asymmetric information, our paper is most closely related to empirical investigations of the "positive correlation property," the robust prediction of standard equilibrium models that there should be a positive association between insurance coverage and ex post risk.[5] The positive correlation property has been tested in several recent studies of a variety of markets. The empirical results are mixed and differ by market. For example, in a life insurance market, Cawley and Philipson (1999) found that the mortality rate of U.S. males who purchase life insurance is below that of the uninsured, even when controlling for many factors such as income that may be correlated with life expectancy. In an auto insurance market, Chiappori and Salanié (2000) found that accident rates for young French drivers who choose comprehensive automobile insurance are not statistically different from the rates of those opting for the legal minimum coverage, after controlling for observable characteristics known to automobile insurers. In contrast, Cohen (2005), using data from Israel, finds that new auto insurance customers choosing a low deductible tend to have more accidents, leading to higher total losses for the insurer. Others have examined the evidence of asymmetric information in the choice of insurance contracts such as deductibles, copayments, and so forth. For example, Puelz and Snow (1994) studied automobile collision insurance and argued that, in an adverse selection equilibrium, individuals with lower risk will choose a contract with a higher deductible, and contracts with higher deductibles should be associated with lower average prices for coverage. They found evidence in support of each of these predictions using data from an automobile insurer in Georgia.[6] In an annuity insurance market, Finkelstein and Poterba (2004) found systematic relationships between ex post mortality and annuity characteristics, such as the timing of payments and the possibility of payments to the annuitant's estate, but they do not find evidence of substantive mortality differences by annuity size. Yet a third

[5] See Chiappori and Salanié (2000). Chiappori et al. (2006) generalize this empirical prediction to a larger class of models. Chiappori (2001) provides an extensive survey of the theoretical and empirical literature.

[6] However, see Chiappori and Salanié (2000) and Dionne, Gouriéroux, and Vanasse (2001) for critiques of the Puelz and Snow study.

approach is to estimate structural models of health insurance and health care choice. For example, Cardon and Hendel (2001) estimated such a model using data from the National Medical Expenditure Survey. They find that estimated price and income elasticities, as well as demographic differences, can explain the expenditure gap between the insured and the uninsured. Thus they judge the role of adverse selection to be economically insignificant.

Our paper is also related to a literature that looks specifically for evidence of adverse selection in the Medigap market. The results from that literature are also mixed. For example, Wolfe and Goddeeris (1991), using data from the Retirement History Survey, found that respondents with better self-reported health were more likely to purchase Medigap and that those with Medigap did not spend more on hospital stays, physician care, and prescription drugs.[7] In contrast, Lillard and Rogowski (1995), Ettner (1997), and Hurd and McGarry (1997) found little evidence of variation in the probability of purchasing private insurance by health status. Also related to our paper, Khandker and McCormack (1999), using MCBS 1991 and 1993, found that those with Medigap tended to incur higher levels of Medicare-reimbursed spending, particularly Part B services. However, because Medigap plans cover Medicare co-pays, they reduce the out-of-pocket price of Medicare-covered services. Thus, it is possible that those with basic Medicare alone may incur more total expenditures, despite smaller Medicare-reimbursed expenditures, because they spend more out of pocket. (Indeed, our table 1 in Sec. V.*E* is consistent with this view.) Thus, we argue that, to study selection in the Medigap market, a better measure of health expenditure risk is total medical expenditure, not just expenditure reimbursed by Medicare or Medigap.[8]

III. Multidimensional Private Information and Advantageous Selection

It is now well understood that, given multidimensional private information, the correlation between ex post risk realizations and coverage may be negative, zero, or positive (see Hemenway 1990; de Meza and Webb 2001; Jullien, Salanié, and Salanié 2007). These papers all focus on private information about risk aversion as the source of advantageous

[7] Khandker and McCormack (1996) estimated multinomial logit models of insurance choice and found that individuals reporting better health were significantly more likely to enroll in Medigap plans.

[8] Unfortunately, the waves of the MCBS data used in Khandker and McCormack's (1999) analysis did not contain information about out-of-pocket costs as well as expenses paid by supplementary insurers. We are grateful to Tami Swenson for the clarification on this data issue.

selection. In this section, we present a simple conceptual framework that shows how multidimensional private information may lead to advantageous selection.

Consider an individual age 65 or older (and thus covered by Medicare) who, with probability $p \in [0, 1]$, will experience a health expenditure shock of size $L > 0$, over and above what is covered by Medicare.[9] For simplicity, assume that the individual can choose to purchase Medigap insurance at a premium m that will reduce the out-of-pocket expenditure to zero. Let $\boldsymbol{\gamma}$ denote a vector of characteristics that may also affect the individual's probability of purchasing Medigap. The variables in $\boldsymbol{\gamma}$ may include risk aversion, but also other characteristics such as cognitive ability, planning horizons, and so forth. Suppose that in the population $(p, \boldsymbol{\gamma})$ is distributed according to a joint cumulative distribution function F. For illustrative purposes, suppose that $Q(p, \boldsymbol{\gamma}; L, m)$ is the probability that an individual with risk type p and characteristics $\boldsymbol{\gamma}$ will purchase Medigap when the expenditure shock is L and the premium is m. For example, if $\boldsymbol{\gamma}$ measures risk aversion, we can show that $Q(\cdot, \cdot; L, m)$ is increasing in risk type p and risk aversion $\boldsymbol{\gamma}$; that is, more risky and more risk-averse individuals are more likely to purchase Medigap (see Fang, Keane, and Silverman 2006). In standard insurance models in which risk type p is the only dimension of heterogeneity, $\boldsymbol{\gamma}$ is implicitly assumed to be constant in the population. For simplicity, we suppress the dependence of Q on L and m below.

We can describe the "positive correlation property" test of Chiappori and Salanié (2000) within this simple conceptual framework. The test compares the average health shock occurrence for those with and without Medigap insurance. In our framework, the average health shocks among those with and without Medigap insurance are respectively given by

$$A_{\text{Medigap}} = \frac{\iint pQ(p, \boldsymbol{\gamma})dF(p, \boldsymbol{\gamma})}{\iint Q(p, \boldsymbol{\gamma})dF(p, \boldsymbol{\gamma})} \tag{1}$$

and

$$A_{\text{No\_Medigap}} = \frac{\iint p[1 - Q(p, \boldsymbol{\gamma})]dF(p, \boldsymbol{\gamma})}{\iint [1 - Q(p, \boldsymbol{\gamma})]dF(p, \boldsymbol{\gamma})}, \tag{2}$$

where the denominator in (1) (and respectively in [2]) is the measure of individuals who purchase (respectively, do not purchase) Medigap, and the numerator is the expected number of health shocks that occur to those who purchase (respectively, do not purchase) Medigap. The

[9] We assume away the price effect, also called "moral hazard," by assuming that the expenditure level L does not depend on health insurance status.

test of the positive correlation property asks whether $A_{\text{Medigap}} > A_{\text{No\_Medigap}}$. This condition obviously holds if $\boldsymbol{\gamma}$ is constant in the population and $Q(\cdot, \boldsymbol{\gamma})$ increases in p. However, if $\boldsymbol{\gamma}$ is heterogeneous in the population and if at least one element $\gamma \in \boldsymbol{\gamma}$ satisfies the following two conditions,[10] then the sign of $A_{\text{Medigap}} - A_{\text{No\_Medigap}}$ is ambiguous.

PROPERTY 1. γ is positively correlated with insurance coverage; that is, $Q(p, \boldsymbol{\gamma})$ is increasing in γ.

PROPERTY 2. γ is negatively correlated with risk p.

Moreover, the average probability of insurance purchase for a given risk type p (after integrating out $\boldsymbol{\gamma}$), namely, $\tilde{Q}(p) = \int Q(p, \boldsymbol{\gamma}) dF_{\boldsymbol{\gamma}|p}(\boldsymbol{\gamma}|p)$, may not be monotonic in p if at least one element of $\boldsymbol{\gamma}$ satisfies the above properties. We will say that private information item γ is a source of advantageous selection if it satisfies these two properties.

The above discussion is merely illustrative; we only analyze individuals' insurance purchase decisions assuming a particular equilibrium (i.e., a particular menu of insurance options) and do not consider a more complete model in which insurance companies may compete by offering different insurance contracts. There does not yet exist an equilibrium model of an insurance market with multidimensional private information.

In our empirical analysis, we first provide, in Section VI, evidence that is akin to $A_{\text{Medigap}} < A_{\text{No\_Medigap}}$; that is, we find that the health expenditure risk for those with Medigap insurance is lower than the risk of those without Medigap insurance. This immediately suggests the presence of multidimensional private information generating advantageous selection. Then, in Section VII, we examine the sources of advantageous selection; that is, we look for elements of $\boldsymbol{\gamma}$ that may account for the earlier finding that $A_{\text{Medigap}} < A_{\text{No\_Medigap}}$.

IV. Institutional Background

A. *Medicare*

Medicare is the primary health insurance program for most seniors in the United States. All Americans age 65 and older who have, or whose spouses have, paid Medicare taxes for more than 40 quarters are eligible. During the period we study, the original Medicare plan consisted of Parts A and B. Part A, the hospital insurance program, covers inpatient hospital, skilled nursing facility, and some home health care. Almost all

[10] Of course, any $\boldsymbol{\gamma}$ that is negatively correlated with insurance coverage and positively correlated with risk p will also work. Moreover, the negative correlation between $\boldsymbol{\gamma}$ and risk p may arise either exogenously or endogenously (in the sense that the would-be insured with a higher $\boldsymbol{\gamma}$ may take an action to reduce p). For our purpose this distinction is unimportant.

retirees are automatically enrolled in Part A when they turn 65, and there are no premiums for this coverage. Medicare Part B, also called Medicare Insurance, covers Medicare-eligible physician services, outpatient hospital services, certain home health services, and durable medical equipment. Part B enrollees have to pay a monthly premium, which was $67 in 2004. Almost all people choose to enroll in Part B when they turn 65. Under Part B, individuals were responsible for a $110 deductible in 2005 and face a 20 percent co-pay for all Medicare-approved services after exceeding the deductible. The basic Medicare plan is available everywhere in the country.[11] Some areas also offer what are now called Medicare Advantage Plans, which are managed care plans (either health maintenance organizations [HMOs] or preferred provider organizations). In 2001, approximately 15 percent of Medicare beneficiaries were enrolled in such Medicare HMOs. We discuss Medicare HMOs further in Sections V.*C* and VI.*C*. During the period we analyze, Medicare did not cover prescription drugs.

B. Medigap

While Medicare is the primary health insurance for most older Americans, the program leaves seniors at significant risk of health expenditures. On average, basic Medicare benefits cover about 45 percent of the personal health care expenditures of aged beneficiaries in the United States (Kaiser Family Foundation 2005). To insure Medicare beneficiaries against some of that risk, private insurance companies sell "Medigap" policies that cover some of the co-pays, deductibles, and uncovered expenses, that is, the "gaps" in the basic Medicare plan.

One reason the Medigap insurance market is ideal for studying multidimensional private information is that, as a result of the Omnibus Budget Reconciliation Act of 1990 (OBRA 1990) effective from 1992, Medigap policies are standardized into 10 plans, A–J, in all states except Massachusetts, Minnesota, and Wisconsin.[12] The basic plan, A, covers all co-pays for hospital stays longer than 60 days and all co-pays for physician visits and outpatient care (but not the deductible). All other plans offer these basic benefits, and more. Details of the different constellation of benefits from plans A–J can be found in CMS (2005, 33–35). The restrictions on the Medigap plans have important implications for whether multidimensional private information will manifest itself through a violation of the positive correlation property. Without restrictions on the insurance contracts that insurance companies can offer,

[11] For further details about the coverage offered by Medicare Parts A and B, see CMS (2005, 55–64).

[12] These states received waivers that allow them to offer somewhat different standardized plans.

Chiappori et al. (2006) showed that a suitably modified version of the positive correlation property still holds even with multidimensional private information.

OBRA 1990 also regulated Medigap pricing and enrollment in ways that tend to amplify the asymmetries of information favoring the insured. Most important, Medigap policies are required by law to have an open-enrollment period. This 6-month period begins after the first day of the first month an individual both is age 65 or older and is enrolled in Medicare Part B. During this period, insurers cannot deny or delay Medigap coverage.[13] Moreover, during open enrollment, the federal legislation prohibits insurers from medical underwriting in pricing Medigap policies; they can effectively price only on age, gender, smoking status, and state of residence.[14]

C. Medigap Pricing

While federal regulations prohibit insurers from medical underwriting in Medigap pricing during the open-enrollment period, in a market served by multiple insurers, different firms may price the same policies differently. Indeed, using Medigap premium data from Weiss Ratings, Maestas et al. (2006) documented substantial price variation for the same Medigap policies. They argued that search costs play an important role in explaining the price dispersion for these seemingly homogeneous products. Robst (2006), using the same data, found that the average premium an individual faces for a given Medigap plan depends almost exclusively on his or her age, gender, and state of residence.

In our analysis below, we use age, gender, and state of residence to control for Medigap pricing.[15] The dispersion in Medigap prices documented by Maestas et al. (2006) does not invalidate these controls for Medigap pricing because, given the federal regulations, during the open-enrollment period all individuals are offered the same set of Medigap plans, and hence they face the same distribution of prices condi-

[13] States also have the option of going beyond the federal regulations regarding the 6-month open-enrollment period. Connecticut and New York, e.g., have indefinite open enrollment (i.e., no medical underwriting is ever permitted), whereas California, Maine, and Massachusetts have an annual open-enrollment period of 1 month around a person's birthday (Lewin Group 2001).

[14] In principle, insurers are free to vary prices by zip code or county. However, the most comprehensive Medigap premium data from Weiss Ratings show that, for any given insurer, most premium variation occurs across rather than within states. The reason could be that the number of policyholders per zip code or county is too small for risk pooling (see Maestas, Schroeder, and Goldman 2006).

[15] Robst (2006, table 3) showed that less than 10 percent of all available Medigap plans offered a small average \$100 discount for nonsmokers (from Weiss Ratings data). We do not control for smoker status in our basic analysis because such smoker discounts are available in only a small number of states.

tional on these variables. It may be that, as Maestas et al. argued, differences in search costs lead some individuals to face different effective prices than others in the same market. If so, it would be important to understand the sources of this variation in search costs. Most relevant for this paper, we would like to know whether cognitive ability and other factors function as sources of advantageous selection through their effects on costs of price search. In Section VII.*D*, we consider in greater depth this and other pathways through which cognitive ability and other factors induce advantageous selection.

The government also requires that, after the open enrollment, as long as the would-be insured has had Medigap coverage for the past 6 months, enrollment in a different plan offered by the same company be guaranteed by law. However, if an individual did not enroll in Medigap in the open-enrollment period or if his or her Medigap coverage lapsed, then insurance companies may subsequently impose both coverage and pricing rules different from those that apply during the open-enrollment period. In these circumstances, our Medigap price controls (age polynomial, gender, and state of residence) may not reflect the prices such individuals face. Any potential problems caused by such situations, however, are still consistent with our interpretation of advantageous selection (see Sec. VI.*D* for the details).

V. Data and Descriptive Statistics

Our analysis relies on two large data sets, the MCBS and HRS. Here we provide a basic description of these data and descriptive statistics of the selected samples in each.

A. *Medicare Current Beneficiary Survey*

The MCBS began in September 1991 and is a continuous panel survey of a nationally representative sample of the Medicare population.[16] Beneficiaries sampled from Medicare enrollment files (or appropriate proxies) are interviewed in person, three times a year, using computer-assisted personal interviewing. All the MCBS survey data are linked to Medicare claims and other administrative data. The final file consists of survey, administrative, and claims data and thus provides a comprehensive view of respondents' heath care costs and use.

The central goal of the MCBS is to determine expenditures and sources of payment for all services used by Medicare beneficiaries, in-

[16] See http://www.cms.hhs.gov/mcbs/ for more details. A supplemental sample is added annually in the September–December round to replenish sample cells depleted by refusals and deaths.

cluding co-payments, deductibles, and uncovered services. This is important, since our focus is on the total health expenditure, that is, the combined expenditures that were covered by Medicare, other public insurance, or private insurance or paid out of pocket. In addition, the MCBS also contains extensive information on the health and demographics of respondents and whether respondents have supplemental insurance. The Data Appendix (table A1) provides a detailed description of how the variables used in our analysis are constructed.

B. *Health and Retirement Study*

The HRS began in 1992 as a panel survey of a nationally representative sample of people born between 1931 and 1941 and their spouses. This original cohort has been reinterviewed every other year since. In 1998, the objective of the HRS was expanded to include learning about the entire U.S. population over age 50. To achieve this goal, the original HRS survey was merged with an existing and related survey of individuals born in 1923 or before, the AHEAD. In addition, two more samples were added: the "Children of the Depression" (CODA) cohort born between 1924 and 1930 and the "War Baby" cohort born between 1942 and 1947. Our interest in those over age 65 leads us to limit our analysis to health and insurance data collected in the 2000 and 2002 waves of the HRS, the latest years for which a final version of the data is available. Whenever possible we include data from all HRS cohorts.[17]

The HRS is particularly well suited for studying advantageous selection in Medigap insurance. It contains detailed information about current and past health status of respondents as well as rich data on their insurance choices. The health information includes both self-reported health and a very large set of objective measures, such as diseases diagnosed, and a list of activities the respondent has difficulty performing. The insurance data include information on where the insurance was acquired, its premiums, and its coverage. The HRS also contains high-quality information about economic and demographic variables. However, the HRS does not have comprehensive information about total medical expenditure.[18] In Section VII we describe procedures to impute

[17] In the end, 37 percent of our combined HRS sample comes from the original HRS, 44 percent from AHEAD, and 19 percent from CODA cohort respondents. Because we use only the 2000 and 2002 waves of the HRS, we use only the 2000 and 2001 waves of the MCBS in our main analysis, though we also used lagged health measures from the 1999 MCBS in several specification checks.

[18] Even with the Medicare claims data that were recently linked to the HRS, we still lack comprehensive information about medical expenditure that is not paid by Medicare. As table 1 below shows for the MCBS sample, Medicare-reimbursed expenditure differs substantially from the total medical expenditure.

medical expenditure for the HRS sample using the information in the MCBS.

Equally important, the HRS is unusual in its attention to variables central to economic theory, including measures of risk attitudes, longevity expectations, and financial planning horizons; it also contains several measures of cognitive ability.[19] The measures of risk attitudes we use are based on HRS respondents' choices over a series of hypothetical gambles. The responses to these gambles place respondents into four ordered risk categories. Assuming that an individual's responses to these hypothetical income gambles are error-prone reflections of his or her fixed, constant relative risk aversion preferences, Kimball, Sahm, and Shapiro (forthcoming) estimate the risk tolerance for each respondent in the HRS by maximum likelihood. In the bulk of our analysis, we use their estimates to form our measures of risk aversion (see their table 6).[20]

Longevity expectations may also play a role in determining health insurance choices, though the net effect of higher longevity expectations on investment in health is theoretically ambiguous. Those who expect to live longer may want to spend more now on their health since such investment will pay dividends over a longer horizon (see Khwaja 2005). Thus, they would demand more insurance. However, the marginal value of additional current health investment may be lower when a long life already seems likely. The HRS asked of all respondents age 65 and younger, and repeated in every HRS wave, "What is [percent chance] you will live to 75 or more?" In our analysis, we use the most recent available response to this question as our measure of longevity expectations.[21]

Similarly, the length of financial planning horizons (which presumably reflects both uncertainty and the subjective rate of time discount) may influence insurance choices. Here the effect seems unambiguous, since those with longer horizons would be more willing to pay immediate costs (premiums) to avoid expected future costs. The HRS collects information on financial planning horizons by asking respondents how

[19] Details of these measures can be found in Fang et al. (2006, sec. 5.2).

[20] While risk attitudes solicited from choices over hypothetical gambles are undoubtedly measured with error, a number of studies have found the answers to these questions to be significant predictors of risk-taking behavior. See, e.g., Barsky et al. (1997) on smoking, drinking, and insurance purchase; Lusardi (1998) on wealth accumulation; Charles and Hurst (2003) and Kimball et al. (forthcoming) on portfolio choice; Kan (2003) on residential and job mobility; and Schmidt (forthcoming) on the timing of fertility and marriage. In Sec. VII.*C*, we will also find that it is a strong predictor of Medigap purchase.

[21] This question presumably measures longevity expectations with error and may reflect both beliefs about longevity and the degree of certainty about those beliefs. Evidence consistent with both error and uncertainty about beliefs is found in the heaping of responses around focal responses such as zero, 50, and, to a lesser extent, 100. See Kezdi and Willis (2005) for a thorough discussion of these measures.

far ahead they plan the family's saving and spending, ranging from "next few months" to "longer than 10 years" in five categories.

Finally, our measure of a person's cognition combines his or her performance on four different tests/questions: word recall, a Telephone Interview for Cognitive Status (TICS) score, subtraction, and numeracy. These scores may proxy for an individual's degree of economic "rationality," that is, his or her ability to think through the costs and benefits of Medigap insurance. Indeed, there is a large body of literature showing that many of the elderly have difficulty understanding the basic Medicare entitlement and/or the features of supplemental insurance.[22]

C. *Sample Selection and Medigap Insurance Status*

Both the MCBS and HRS contain detailed information about respondents' health insurance choices. Each reports whether the respondent is covered by Medicare, Parts A and B, and whether that coverage is provided by a Medicare HMO. We also know whether the respondent had supplemental coverage and, if so, its premium and its source. Given that our goal is to study the decision to buy supplemental insurance, our sample should include only people who are covered by basic Medicare (Parts A and B) and do not have access to free (or heavily subsidized) supplemental coverage provided by a former employer, Medicaid, or some other government agency (e.g., the Veterans Administration). That is, we want to limit the sample to people who would have to pay more than a nominal premium to obtain supplemental coverage. Owing to these considerations, in our analysis we include only respondents covered by basic Medicare (including those in a Medicare HMO) and exclude anyone covered by employer-provided health insurance, Medicaid, or other government insurance. In this selected sample, we define the Medigap status to be equal to one if the respondent has purchased additional private insurance that is secondary to Medicare.[23]

In our main analysis, we treat Medicare HMO enrollees as having basic Medicare.[24] This choice is motivated by three considerations. First,

[22] See, e.g., Harris and Keane (1998) for empirical evidence, Keane (2004) for a survey of the literature, and Sec. VII.*D* for additional references.

[23] In our working paper (Fang et al. 2006), we also report results from an alternative sample selection in which we retain respondents who have employer-provided supplemental insurance, provided that they must pay at least $500 per year in premiums for that insurance. We then define Medigap status as equal to one if the respondent either has purchased private insurance that is secondary to Medicare or has employer-provided insurance for which he or she pays at least $500 in premiums. None of our results, either qualitatively or quantitatively, depends on which definition of Medigap status we use.

[24] The general view is that Medicare HMOs exchange restrictions on the choices for medical treatment for additional coverage similar to that provided by typical Medigap policies. Enrollees of Medicare HMOs are discouraged, though not precluded, from purchasing additional Medigap insurance policies (see CMS 2005, 5).

60 percent of Medicare HMO enrollees do not pay an additional premium; thus their decision to enroll in a Medicare HMO is often really a trade-off between restrictions on provider choice and additional coverage for the gaps in Medicare; it is not a decision to pay for additional coverage. Second, Medicare HMOs are not available in all markets; they are typically not offered in rural areas. Third, several recent studies have shown that those who enrolled in Medicare HMOs are, on average, healthier than those on basic Medicare (see Cox and Hogan 1997; Banthin and Taylor 2001). Moreover, Desmond, Rice, and Fox (2006) suggest that greater Medicare HMO enrollment might cause adverse selection into Medigap. Thus, we think that it is conservative to classify Medicare HMO enrollees as choosing only basic Medicare, in the sense that it makes it less likely we will find advantageous selection into Medigap. However, panel B of table 5 below reports results in which we either code Medicare HMO enrollees as having Medigap or instead drop them from the analysis. The qualitative results do not change; indeed, as expected, the evidence for advantageous selection becomes even stronger when they are dropped.

D. *Measures of Health and Medical Expenditure Risk*

Both the MCBS and HRS have detailed and comparable measures of observable health indicators including self-reported health, activities of daily living (ADL) limitations, instrumental ADL limitations, various diagnoses/treatments, and others (see the category Health in the Data Appendix for more details).

For our measure of ex post health expenditure risk, we use "total medical expenditure," which corresponds to the variable *pamttot* in the MCBS. It is constructed by CMS from a variety of sources, including Medicare administrative records and survey responses.[25] In calculating *pamttot*, CMS includes, for any health care event identified either from the survey respondent or from the respondent's Medicare file, payments from 11 potential sources: Medicare fee-for-service, Medicare HMOs, Medicaid, employer-based private health insurance, individually purchased private health insurance, private insurance managed care, private insurance from unknown sources, the VA and other public insurance, out-of-pocket payments, and uncollected liability. Thus, the variable *pamttot* comes as close as possible to measuring total health expenditure from all sources.

An important question is whether total medical expenditure is indeed the most relevant measure of expenditure risk when a would-be insured

[25] See MCBS public use documentation on "Cost and Use," secs. 3 and 5, for more details. This documentation is available online at http://www.cms.hhs.gov/apps/mcbs/.

contemplates whether to purchase Medigap. One argument in favor of treating total medical expenditure as the relevant risk is that Medigap policies by law cover broad ranges of expenditure not covered by Medicare (i.e., Medicare deductibles, co-pays, and prescription drugs). A person in worse health would typically tend to have greater expenditure risk in all these areas. Thus, to a good approximation, Medigap plans can reasonably be thought of a simply providing "more" coverage in a unidimensional health risk framework, because by law Medigap plans cannot be structured to only cover particular health conditions.

E. *Descriptive Statistics*

Table 1 provides some descriptive statistics for the Medigap and no Medigap samples, separately for the MCBS and HRS. In both the MCBS and the HRS samples, there are no significant differences between the Medigap and no Medigap subsamples in gender and age, but they do differ significantly in their educational attainment and marital status. Interestingly, in the MCBS, the mean total medical expenditure is more than $12,000 for those with no Medigap, whereas it is only about $8,400 for those with Medigap. However, the medical expenditure reimbursed by Medicare is slightly higher for those with Medigap than for those without, consistent with the findings in the literature (see, e.g., Khandker and McCormack 1999).

Table 1 also shows that the MCBS and HRS samples are quite similar in the means of the common set of demographic variables. This suggests that using the MCBS to impute means and variances of medical expenditure for the HRS is reasonable.

VI. Evidence of Advantageous Selection

In this section we present a set of simple regressions that together provide strong evidence of advantageous selection in the Medigap market: those who purchase Medigap appear to be healthier and to have lower ex post medical expenditure. We also present direct evidence that healthier people are more likely to purchase Medigap insurance, conditional on observables that determine price.

A. *Basic Regression Results: Indirect Evidence of Advantageous Selection*

Table 2 reports two panels of results from regressing total medical expenditure on Medigap status along with controls for the determinants of price (gender, a third-order polynomial of age, and controls for state and year), with or without controlling for health status of the individuals.

TABLE 1
Descriptive Statistics of Selected Variables

	MCBS			HRS		
		By Medigap Status			By Medigap Status	
Variable	All	Medigap	No Medigap	All	Medigap	No Medigap
Medigap	.46	1.00	.00	.49	1.00	.00
	(.50)	(.00)	(.00)	(.50)	(.00)	(.00)
Female	.60	.60	.60	.57	.58	.56
	(.49)	(.49)	(.49)	(.49)	(.49)	(.50)
Age	75.8	75.6	76.0	75.3	75.7	75.0
	(7.7)	(7.1)	(8.2)	(7.1)	(7.1)	(7.1)
Black	.09	.03	.14	.08	.03	.12
	(.29)	(.18)	(.35)	(.26)	(.17)	(.33)
Hispanic	.08	.03	.11	.04	.01	.07
	(.27)	(.18)	(.32)	(.20)	(.11)	(.25)
Married	.49	.56	.42	.53	.55	.52
	(.50)	(.50)	(.49)	(.50)	(.50)	(.50)
Widowed	.38	.35	.40	.35	.35	.35
	(.48)	(.48)	(.49)	(.48)	(.48)	(.48)
Divorced	.08	.05	.10	.08	.07	.09
	(.27)	(.23)	(.30)	(.27)	(.25)	(.29)
Less than high school	.34	.25	.42	.31	.26	.35
	(.38)	(.33)	(.41)	(.36)	(.34)	(.38)
High school	.28	.31	.25	.36	.39	.34
	(.45)	(.46)	(.43)	(.48)	(.49)	(.47)
Some college	.21	.25	.18	.18	.19	.17
	(.41)	(.43)	(.38)	(.39)	(.39)	(.37)
College	.08	.10	.06	.08	.08	.08
	(.27)	(.30)	(.24)	(.27)	(.27)	(.27)
Working	.12	.15	.11	.14	.14	.14
	(.33)	(.35)	(.30)	(.34)	(.35)	(.34)
Total medical expenditure	10,679	8,362	12,646	. . .	. . .	. . .
	(18,962)	(14,535)	(21,843)			
Medicare reimbursement	4,638	4,761	4,553	. . .	. . .	. . .
	(12,534)	(11,462)	(13,393)			

Note.—Statistics are calculated using cross-section sample weights. Standard deviations are in parentheses. Number of observations varies by variable and sample.

Each panel reports results separately for the full sample and for male and female subsamples.

In panel A, where no health controls are included, we find a large and statistically significant relationship between total medical expenditure and Medigap status. Specifically, in the whole sample, those with Medigap have expenditures that are, on average, about $4,000 less than those without Medigap; the negative relationship between Medigap coverage and total medical expenditure is stronger for women (about $6,000) than for men (about $2,000).

The regressions in panel B are analogous to those in panel A, but with the addition of extensive controls for health, which we describe in detail in the Data Appendix. Conditional on observable (but not priced)

TABLE 2
Ordinary Least Squares Regression Results of Total Medical Expenditure on "Medigap" Coverage in the MCBS

	A. Without Health Controls			B. With Direct Health Controls		
Variable	All (1)	Female (2)	Male (3)	All (4)	Female (5)	Male (6)
Medigap	−4,392.7***	−6,037.4***	−1,863.4***	1,937.0***	1,677.3***	2,420.9***
	(346.5)	(455.5)	(538.8)	(257.2)	(348.0)	(395.8)
Female	270.0	. . .	. . .	−751.6***	. . .	. . .
	(356.2)			(283.3)		
Age − 65	387.5***	460.6***	292.9	394.5***	417.5***	355.4*
	(138.0)	(175.5)	(228.5)	(117.2)	(144.6)	(196.8)
$(\text{Age} - 65)^2$	1.9	−1.8	5.6	−27.5***	−32.0***	−22.8
	(10.6)	(13.2)	(18.8)	(9.2)	(11.4)	(16.2)
$(\text{Age} - 65)^3$	.12	.17	.07	.47**	.55**	.47
	(.22)	(.27)	(.43)	(.21)	(.25)	(.38)
State dummy	Yes	Yes	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,945	9,725	6,220	14,129	8,371	5,758
Adjusted R^2	.073	.092	.060	.211	.196	.252

Note.—The dependent variable is total medical expenditure. All regressions are weighted by the cross-section sample weights. Health controls included in panel B are described in detail in the Data Appendix under the category Health. A total of 71 health indicators are included. Robust standard errors clustered at the individual level are in parentheses.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

health indicators, in the full sample, those with Medigap have total health care spending of about $1,900 more, on average, than those without Medigap. The positive association between Medigap and total medical expenditure seems to be stronger for men (about $2,400) than for women (about $1,700).

As we argued in the introduction, the results in panel A alone indicate the presence of multidimensional private information. The results of panels A and B together imply, indirectly, that there is advantageous selection in the Medigap market; that is, those with better health are more likely to purchase Medigap. That is the only way to rationalize simultaneously the large negative correlation between Medigap and ex post health expenditure in panel A without health controls and the large positive correlation with health controls in panel B. The results in panel B indicate that once we condition on health status, those with Medigap have higher total health expenditures. This is what we would expect from the effects of moral hazard; for individuals with the same health, those with Medigap insurance face a lower price of health care.

B. More Direct Evidence of Advantageous Selection

To provide somewhat more direct evidence of advantageous selection, we now try to find a small number of health factors to summarize the extensive list of health variables used in the regressions reported in panel B of table 2 and then directly examine the partial correlations of the health factors and Medigap status. We analyzed the extensive list of health variables and found that there are four significant health factors that can capture the bulk of variance and covariance in the list of observable health variables. To be conservative, we include five factors for our subsequent analysis. Moreover, by examining the factor loadings (not reported), we can offer interpretations of these factors.[26] Factor 1 can be interpreted as a "nonresponse" factor, which loads heavily on variables that are indicators of nonresponse (i.e., there is a nonresponder type). Factor 2 loads negatively on self-reported health and positively on difficulties in instrumental ADLs and thus is an unhealthy factor. Factor 3 loads positively on self-reported health and negatively on measured medical conditions in the past 2 years and thus is a healthy factor. Factor 4 loads positively on self-reported health and self-reported health changes in the last year. It represents a part of self-reported health not captured by factors 2 and 3. Factor 5 does not appear to have a clear interpretation.

Table 3 reports results from regressions analogous to those in panel B of table 2, except that our extensive list of health variables is summarized by five health factors extracted using factor analysis. The coefficient estimates for Medigap in table 3 are qualitatively unchanged from those in panel B of table 2. As expected, factor 2 is positively related to health care expenditure, factors 3 and 4 are negatively related, and factors 1 and 5 are not related.[27] The standard deviations of the factors are presented in brackets in the Variable column. For example, a one-standard-deviation increase in the most important unhealthy factor, factor 2, will increase medical expenditure by about $4,800 for both men and women; a one-standard-deviation increase in the most important healthy factor, factor 3, will reduce medical expenditure by about $2,300 for women and about $3,200 for men. The other factors, factors 1, 4, and 5, are of much smaller magnitude and are often statistically insignificant.

These health factors allow us to examine more directly whether healthy individuals are more likely to purchase Medigap. Table 4 reports the partial correlations between Medigap coverage and the health factors, conditional on gender and age. The columns labeled EXP simply report the regression coefficients for the health factors from specification 1 in table 3. These coefficient estimates inform us whether the factor is "healthy" or "unhealthy." The columns labeled PCORR report the partial correlations and their level of significance.

[26] The factor loadings are available from the authors on request.

[27] The health factors for the three samples are separately estimated, and, as a result, the factors for the three samples are actually different factors.

TABLE 3
Ordinary Least Squares Regression Results of Total Medical Expenditure on Medigap Coverage in the MCBS with Controls for Health Factors

Variable	All (1)	Female (2)	Male (3)
Medigap	2,083.3***	1,601.1***	2,775.5***
	(280.6)	(353.8)	(426.3)
Female	−1,311.2***	. . .	. . .
	(271.5)		
Age − 65	421.8***	447.6***	436.0**
	(118.7)	(151.5)	(194.2)
$(\text{Age} - 65)^2$	−28.7***	−32.6***	−27.3*
	(9.2)	(11.9)	(16.0)
$(\text{Age} - 65)^3$	.48**	.53**	.54
	(.21)	(.26)	(.37)
Factor 1 [SD = 1.00]	321.0	−410.7***	1,097.1*
	(498.0)	(89.0)	(658.7)
Factor 2 [SD = .98]	4,917.5***	4,902.2***	4,880.6***
	(268.3)	(343.2)	(368.4)
Factor 3 [SD = .92]	−2,979.4***	−2,500.8***	−4,048.4***
	(306.2)	(300.7)	(623.9)
Factor 4 [SD = .86]	−652.1**	−684.3*	−1,073.5
	(311.1)	(384.3)	(911.2)
Factor 5 [SD = .88]	75.8	436.4*	−2,278.9*
	(305.0)	(253.8)	(1,340.1)
State dummy	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes
Other demographic controls	No	No	No
Observations	14,129	8,371	5,758
Adjusted R^2	.14	.13	.18

Note.—The dependent variable is total medical expenditure. All regressions are weighted by the cross-section sample weights. Robust standard errors clustered at the individual level are in parentheses. The standard deviations of the factors are in brackets in the Variable column. Factor 2 is the major unhealthy factor, factors 3 and 4 are the healthy factors, factor 1 captures nonresponse, and factor 5 does not have a clear interpretation.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

For the most part, table 4 reveals a consistent pattern: important unhealthy (healthy) factors tend to have negative (positive) and significant partial correlations with Medigap coverage. For example, column 2 shows that factor 2, the most important unhealthy factor, has a sizable negative correlation with Medigap (−.1166) and a p-value of almost zero; in contrast, factor 3, the most important healthy factor, has a positive correlation with Medigap (.0319) and, again, a p-value of almost zero. The factors that have Medigap correlations of the “wrong” sign are typically of two kinds: either the factor itself is not very important (with small and insignificant coefficient estimates) or the partial correlation is statistically insignificant. The partial correlations between factors 2 and 3 and Medigap status in the male and female subsamples have the same signs and magnitudes similar to those for the whole sample.

TABLE 4
Partial Correlation between Medigap Coverage and Health Factors in the MCBS, Conditional on Gender and Age

Factor	All EXP (1)	All PCORR (2)	Female EXP (3)	Female PCORR (4)	Male EXP (5)	Male PCORR (6)
Factor 1	321.0	.03	−410.7***	.03	1,097.1*	.03
		(.00)		(.01)		(.01)
Factor 2	4,917.5***	−.12	4,902.2***	−.13	4,880.6***	−.10
		(.00)		(.00)		(.00)
Factor 3	−2,979.4***	.03	−2,500.8***	.04	−4,048.4***	.02
		(.00)		(.00)		(.01)
Factor 4	−652.1**	−.02	−684.3*	−.02	−1,073.5	.02
		(.04)		(.12)		(.08)
Factor 5	75.8	.02	436.4*	.01	−2,278.9*	−.02
		(.01)		(.19)		(.12)
Observations	14,129		8,371		5,758	

Note.—The columns labled EXP are the regression coefficients from table 3. They are included in this table for the interpretation of the factors. The columns labeled PCORR list the partial correlations of Medigap with the corresponding factors. The numbers in parentheses are the significance levels of the partial correlations.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

C. Robustness Checks

We now show that our qualitative results about advantageous selection in the Medigap market are robust to a number of alternative data-coding choices.

Alternative unidimensional measures of health.—Our qualitative findings about advantageous selection remain unchanged if we replace total medical expenditure as a unidimensional measure of health with the level of medical expenditure that is not reimbursed by Medicare. We call this alternative measure “potential Medigap expenditure” because it is arguably a good approximation of the expenditure risk faced by a person who has only basic Medicare coverage.

Column 1 in panel A of table 5 simply reproduces the Medigap coefficients from table 3 (for the All sample), and column 2 presents the results using potential Medigap expenditure as the dependent variable. The first row in each column is the baseline specification that includes only gender, an age polynomial, and state of residence as controls for price, and the second row adds direct controls for health. As can be seen from panel A, if we use potential Medigap expenditure as an alternative measure of health expenditure risk, the basic message of advantageous selection persists: when we control only for variables used to price Medigap, the Medigap coefficients are in the range of −$4,000. But when we control for health, they turn positive. The positive values are not as large when we use potential Medigap expenditure as the measure of health. But the key point is that inclusion of the observable

TABLE 5
ROBUSTNESS OF THE EVIDENCE OF ADVANTAGEOUS SELECTION: MEDIGAP COEFFICIENTS

A. Alternative Measures of Health Expenditure Risk

Health Controls	Total Medical Expenditure (1)	Potential Medigap Expenditure (2)
No	−4,392.7***	−4,454.0***
	(347.0)	(202.2)
Yes	1,937.0***	80.4
	(257.6)	(132.1)

B. Alternative Treatment of Medicare HMO

Health Controls	Treated as Medigap (1)	Dropped from Sample (2)
No	−4,418.5***	−3,996.8***
	(364.8)	(298.7)
Yes	1,899.6***	2,011.3***
	(276.6)	(276.6)

C. Trimming Top 5 Percent of the Observations

Health Controls	Total Medical Expenditure (1)	Potential Medigap Expenditure (2)
No	−1,400.1***	−1,103.3***
	(183.1)	(94.4)
Yes	1,673.2***	247.7***
	(147.8)	(73.8)

D. Including Additional Demographic Controls

Health Controls	All (1)	Female (2)	Male (3)
No	−3,783.3***	−5,687.4***	−1,448.2***
	(375.4)	(485.7)	(569.8)
Yes	1,732.8***	1,426.2***	2,210.1***
	(272.4)	(358.4)	(418.9)

NOTE.—All regressions are weighted by the cross-section sample weights. The descriptions of the direct health controls can be found in the Data Appendix. The additional demographic controls used in panel D include race, education, marital status, income, working, and number of children. Robust standard errors in parentheses are clustered at the individual level.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

health variables significantly increases the magnitude of the Medigap coefficient in either case, thus indicating, once again, that those with Medigap are healthier.

Alternative treatments of Medicare HMOs.—In our data, approximately 15 percent of all Medicare beneficiaries chose to participate in a Medicare HMO. So far, we have treated them the same as those with basic fee-for-service Medicare. We now show the results with two alternative treatments of those with Medicare HMOs: we either code them as having Medigap or code them as missing (thus dropping them from our analysis).

Panel B of table 5 reports the Medigap coefficient estimates under the two alternative treatments of Medicare HMOs. Column 1 shows that the recoding of Medicare HMO participants as Medigap actually somewhat strengthens the findings of advantageous selection reported earlier. For example, the negative Medigap coefficient becomes somewhat more negative. This is not surprising given the consensus in the literature (noted earlier) that there is advantageous selection into Medicare HMOs. Similarly, dropping those observations from our analysis does not change the results, either qualitatively or quantitatively.

Trimming the outliers.—It is well known that the distribution of medical expenditure is right skewed. For example, in our selected sample, total medical expenditure has a skewness of 4.2, with a mean of $10,679 and a median of $3,467; potential Medigap expenditure has a skewness of 3.5, a mean of $6,040, and a median of $2,292. In our view, outlier medical expenditures are likely an important concern when individuals consider whether to purchase Medigap. Nevertheless, we show in panel C of table 5 that the findings of advantageous selection are not solely driven by the highest expenditure levels. In column 1 we report the Medigap coefficient estimates after we drop the observations whose total medical expenditure is above the ninety-fifth percentile. The coefficient estimates, not surprisingly, get smaller in absolute magnitude, but the qualitative conclusion regarding the presence of advantageous selection is not affected. Column 2 reports the results for potential Medigap expenditure after we trim the top 5 percent. Again, the coefficients are smaller in magnitude, but the qualitative results are the same.

D. Discussion

In this subsection we consider a few issues of potential concern regarding the interpretation of our findings.

Insurers' selection.—A natural question is whether the advantageous selection we documented for the Medigap insurance market is, as we interpret it, driven by consumers or is instead induced by insurers. Insurers have incentives to "cream-skim," that is, target their offerings at relatively good risks, because medical underwriting is prohibited. Three observations cast some doubt on the importance of cream-skimming in explaining the advantageous selection we observe. First, the best evidence for cream-skimming in related markets comes from selection into Medicare HMOs, not Medigap, and in our main analysis we classify those in Medicare HMOs as in basic Medicare. And as we saw in Section VI.*C*, dropping Medicare HMO enrollees from the analysis has little effect on our estimates of advantageous selection into Medigap. (Indeed, if anything, the evidence for advantageous selection becomes stronger, consistent with cream-skimming in the Medicare HMOs.)

Second, as Maestas et al. (2006) noted from the data from Weiss Ratings, about two-thirds of the Medigap policies sold are agent-solicited and only about one-third are sold by the insurance companies directly. It seems likely that the incentives of agents and insurance companies are not perfectly aligned: agents want to sell policies and care less about the risk to the insurance company of a given contract. Thus, without compensation schemes that reflect ex post risk, agents will choose to locate their offices and market their products where they can sell insurance and pay little attention to ex post risks to the insurance companies.

Third, we can ask whether the extent of advantageous selection is mitigated if we condition on observable demographics of the consumers, including race, education, marital status, income, working status, and number of children, which insurance companies could use to cream-skim to the extent that they may be correlated with the health expenditure risks of the insured.[28] Panel D of table 5 reports the Medigap coefficients with and without health controls, but including the above list of additional demographic controls. Controlling for the additional demographics does lower the magnitude of the Medigap coefficient, but only modestly. For example, without health controls, expenditures for those with Medigap are about $3,800 less than for those without Medigap for the whole sample, compared to $4,400 without the demographic controls. Similarly, with direct health controls, the Medigap coefficient estimate falls only slightly to $1,700 with the additional demographic controls for the whole sample, compared to $1,967 without. The effects of these additional demographic controls are similar in the female and male samples. This evidence suggests that our finding of advantageous selection is not mainly driven by insurers' selection.

Panel D of table 5 can also be seen to show that even though these additional demographic variables can explain part of the observed advantageous selection, the extent is rather limited. As a result, in the next section we will extend our search for the sources of advantageous selection.

Endogenous health measures.—Another possible concern is that the health measures we used in our analysis could be affected by Medigap

[28] It is important to emphasize that, in order to establish the existence of advantageous selection, we should not include these additional demographic variables because insurance companies are not allowed to price on them.

insurance status itself. It is possible that if those with Medigap are more likely to seek care, they may be more likely to have certain conditions diagnosed. This would make the Medigap population seem less healthy than they actually are (relative to the basic Medicare population). This would cause them to have lower than expected expenditure (conditional on health), thus biasing the Medigap coefficient in panel B of table 2 in a negative direction (i.e., we would understate the degree of advantageous selection).[29]

Selection in Medigap renewal.—Yet another potential concern is that our Medigap price controls (age polynomial, gender, and state of residence) do not reflect the prices faced by those who let their coverage lapse because the prohibition on medical underwriting in Medigap pricing applies in most states only in the open-enrollment period. In this case, our finding that expenditures for those with Medigap are about $4,000 less than for those without may, to some extent, reflect the following possibility: those without Medigap are less healthy because their coverage previously lapsed and, moreover, they are currently priced out of Medigap because of their poor health. Because the MCBS panel is relatively short, we do not have Medigap coverage information for a large number of individuals since their open enrollment. However, we argue here that if the mechanism described above is influencing our estimates, its effect is nonetheless consistent with our interpretation of advantageous selection.

Consider the group of individuals without Medigap who are unhealthy and are priced out of Medigap because of a lapse in their coverage. One possibility is that they never purchased a Medigap policy in the first place. This would be consistent with our interpretation of advantageous selection: less healthy individuals are less likely to purchase Medigap during open enrollment (when everyone is approximately the same age, and thus state and gender controls alone would be sufficient to control for Medigap pricing). The second possibility is that those currently without Medigap did purchase a policy during open enrollment but subsequently failed to renew. The standard adverse selection model with one-dimensional private information about risk would have suggested that less healthy people should be more likely to renew, just as they should be more likely to enroll in the first place. If, in contrast, less healthy people are less likely to renew, this is again a form of advantageous selection, albeit in the renewal stage.

[29] In any case, our results are not changed if we perform our analysis using lagged, instead of current, health indicators.

VII. Sources of Advantageous Selection

In this section we investigate the sources of advantageous selection. That is, we seek to identify dimensions of individuals' private information that satisfy the two properties we mentioned in Section III, that is, unpriced variables that both (i) make individuals more likely to purchase Medigap and (ii) are negatively correlated with their health expenditure risk.

A. *Empirical Strategy*

The ideal data set for our analysis would be the HRS augmented by links to Medicare administrative data containing information on total medical expenditure. Unfortunately, the HRS is not yet properly linked to the Medicare administrative records and has imperfect information on out-of-pocket spending or spending reimbursed by other sources relative to MCBS. On the other hand, the MCBS does not contain information about many suspected sources of advantageous selection. We now describe our empirical strategy, which combines the MCBS and the HRS to examine the sources of advantageous selection.[30]

The data in the MCBS can be written as

$$\{E_i,\ M_i,\ \mathbf{H}_i,\ \mathbf{D}_i\}_{i\in\mathcal{I}_{\mathrm{MCBS}}} \tag{3}$$

and the data in the HRS as

$$\{M_j,\ \mathbf{H}_j,\ \mathbf{D}_j,\ \mathbf{X}_j\}_{j\in\mathcal{I}_{\mathrm{HRS}}}, \tag{4}$$

where $\mathcal{I}_{\mathrm{MCBS}}$ and $\mathcal{I}_{\mathrm{HRS}}$ denote the MCBS and HRS samples, respectively. Note that the variables $\{M, \mathbf{H}, \mathbf{D}\}$, which denote Medigap coverage, health measures, and demographics, are common to both data sets. But E, total medical expenditure, appears only in the MCBS, and $\mathbf{X}$, the list of variables that we think are potential sources of advantageous selection, appears only in the HRS.

Our strategy is simple and consists of two steps. In the first step, we use the MCBS data to estimate prediction equations for total medical expenditure risk as well as its variance. (We describe our imputation strategies in the next subsection.) These equations will use only covariates that are also available in the HRS, so we can use the estimated prediction equations from the MCBS data to impute the mean ($\hat{E}_j$) and variance ($\widehat{\mathrm{Var}}_j$) of health expenditures for each person in our HRS

[30] There is a sizable literature on empirical methods to deal with the incomplete data by combining multiple data sets. For an excellent survey, see Ridder and Moffitt (2007).

sample. With the imputed $\hat{E}_j$ and $\widehat{\mathrm{Var}}_j$, our augmented HRS data can now be represented as

$$\{M_j, \mathbf{H}_j, \mathbf{D}_j, \mathbf{X}_j, \hat{E}_j, \widehat{\mathrm{Var}}_j\}_{j \in \mathcal{I}_{\mathrm{HRS}}}. \tag{5}$$

In the second step, we first regress Medigap coverage on expected expenditure and pricing variables:

$$M_j = \delta_0 + \delta_1 \hat{E}_j + \boldsymbol{\delta}_2 \mathbf{D}_j + \varepsilon_j, \tag{6}$$

where, as before, the variables in $\mathbf{D}_j$ include a third-order polynomial in age, gender, and state of residence to capture the pricing of Medigap insurance. As we report below, and consistent with our finding in Section VI, we obtain a negative and significant estimate for δ_1, the coefficient on expected expenditure, implying advantageous selection in the purchase of Medigap in the HRS. We then gradually add potential sources of advantageous selection from the list of variables contained in $\{\mathbf{X}_j, \widehat{\mathrm{Var}}_j\}$. We will show below that when we estimate the partial correlation between Medigap coverage and health expenditure risk, controlling not only for the determinants of price, $\mathbf{D}_j$, but also for $\{\mathbf{X}_j, \widehat{\mathrm{Var}}_j\}$, the partial correlation will turn positive. More precisely, when we estimate

$$\begin{aligned} M_j = {} & \theta_0 + \theta_1 \hat{E}_j + \theta_2 \mathrm{risktol}_j + \theta_3 \widehat{\mathrm{Var}}_j \times \mathrm{risktol}_j + \theta_4 \widehat{\mathrm{Var}}_j \\ & + \boldsymbol{\theta}_5 \mathbf{X}_j + \boldsymbol{\theta}_6 \mathbf{D}_j + \varepsilon_j, \end{aligned} \tag{7}$$

we find that $\hat{\theta}_1$ is positive and significant—consistent with the positive correlation property predicted by standard insurance models with unidimensional private information. This is the sense in which we say we have successfully identified several key sources of advantageous selection.

B. Imputation Strategies

Our empirical strategy requires a determination about which sample of the MCBS to use in estimating the prediction equations. Conceptually, we want a measure of expenditure risk for a person who has basic Medicare and is considering whether to buy Medigap. To obtain such a measure, it is not clear whether we should estimate prediction equations using only data for those without Medigap or whether we should instead include the whole MCBS sample. We follow a practical strategy and estimate the prediction equations both ways. We explain below that biases induced by each method may be slight and are likely to understate the extent of advantageous selection. Later we show that our results are robust to which method we use.

Imputation using the MCBS subsample with no Medigap coverage.—With the first method, we use only the subsample in the MCBS with no Medigap coverage to estimate the mean and variance of medical expenditures. Suppose that the mean and variance prediction equations obtained from the MCBS are

$$\hat{E}_{i1} = \hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}_1 \mathbf{H}_i + \hat{\boldsymbol{\alpha}}_2 \mathbf{D}_i \tag{8}$$

and

$$\widehat{\mathrm{Var}}_{i1} = (E_i - \hat{E}_{i1})^2 = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{H}_i + \hat{\boldsymbol{\beta}}_2 \mathbf{D}_i. \tag{9}$$

We can then impute the mean and variance of medical expenditures for the HRS sample as follows: for each $j \in \mathcal{I}_{\mathrm{HRS}}$, the imputed mean medical expenditure is

$$\hat{E}_{j1} = \hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}_1 \mathbf{H}_j + \hat{\boldsymbol{\alpha}}_2 \mathbf{D}_j, \tag{10}$$

and the imputed variance of medical expenditure is

$$\widehat{\mathrm{Var}}_{j1} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{H}_j + \hat{\boldsymbol{\beta}}_2 \mathbf{D}_j. \tag{11}$$

Imputation using the whole MCBS.—With the second method, we use the whole MCBS sample. In this case, we include in the regressions a Medigap status indicator M_i. That is,

$$\hat{E}_{i2} = \hat{\eta}_0 + \hat{\eta}_1 M_i + \hat{\boldsymbol{\eta}}_2 \mathbf{H}_i + \hat{\boldsymbol{\eta}}_3 \mathbf{D}_i \tag{12}$$

and

$$\widehat{\mathrm{Var}}_{i2} = (E_i - \hat{E}_{i2})^2 = \hat{\xi}_0 + \hat{\xi}_1 M_i + \hat{\boldsymbol{\xi}}_2 \mathbf{H}_i + \hat{\boldsymbol{\xi}}_3 \mathbf{D}_i. \tag{13}$$

We then impute the mean and variance for each member $j \in \mathcal{I}_{\mathrm{HRS}}$ of the HRS sample as follows:

$$\hat{E}_{j2} = \hat{\eta}_0 + \hat{\boldsymbol{\eta}}_2 \mathbf{H}_j + \hat{\boldsymbol{\eta}}_3 \mathbf{D}_j \tag{14}$$

and

$$\widehat{\mathrm{Var}}_{j2} = \hat{\xi}_0 + \hat{\boldsymbol{\xi}}_2 \mathbf{H}_j + \hat{\boldsymbol{\xi}}_3 \mathbf{D}_j. \tag{15}$$

Note that in the imputation equations (14) and (15), we set M_j equal to zero for the HRS sample. Thus the predictions above pertain to the mean and variance of medical expenditures for a person without Medigap coverage.

Discussion of the imputation methods.—If, conditional on $\mathbf{H}_j$ and $\mathbf{D}_j$, selection into Medigap were random, then either of the two approaches outlined above would be conceptually correct. The list of observable health variables that we include in our imputation is extremely detailed, so it may be that selection based on unobserved health is not quantitatively important.[31]

If, however, there is nonrandom selection into Medigap conditional on $\mathbf{H}_j$ and $\mathbf{D}_j$, each imputation method will have limitations. Consider the first method. If those with Medigap have systematically better unobserved health (just as they have better observed health), then $\hat{E}_{j1}$ will tend to overestimate the expected medical expenditure for those in HRS who actually have Medigap. This bias will cause us to understate the degree of advantageous selection in the HRS. Using the second imputation method, we need to include the Medigap status indicator M_i in prediction equations (12) and (13). Otherwise we will exaggerate the pre-Medigap purchase expenditure risk by including the "moral hazard" or price effect of Medigap coverage. However, given the selection into Medigap, the Medigap coefficient will be biased. For example, if those with Medigap have systematically better unobserved health (just as they have better observed health), the Medigap coefficient will be downward biased (i.e., we understate the price effect). This would cause $\hat{E}_{j2}$ to, again, overstate the pre-Medigap expenditure risk for those who actually have Medigap.

C. *Sources of Advantageous Selection: Main Findings*

This subsection reports our main results regarding the sources of advantageous selection. Panels A and B in table 6 report, respectively, results based on the two methods of imputing health expenditure risk described above. In each panel, columns 1–3 report estimates of the coefficient on predicted health expenditure $\hat{E}_j$ (with the corresponding p-values of the estimates in parentheses) in equations (6) and (7) for samples of varying sizes.[32] (The sample sizes change because of the additional variables that we explain below.) We enter $\hat{E}_j$ into the regressions in tens of thousands of dollars ($10,000) to make the results easier to interpret. Columns 4–11 indicate the list of control variables included in $\mathbf{X}$, with Y (N) indicating that a particular variable is (is not) included. Column 12 gives the sample size for the regression.

Here we explain the decline in the number of observations due to

[31] The (unadjusted) R^2 is around .3 in the regressions for E and around .2 in the regressions for Var. The relatively low R^2's are not surprising because medical expenditures are intrinsically highly random. Thus, low R^2's do not mean that the model does a poor job of predicting expected medical expenditures for a certain type of person. It should also be noted that this does not reveal whether the random selection assumption is right or wrong.

[32] We adjusted the standard errors for the coefficient estimates of $\hat{E}_j$ to account for the fact that $\hat{E}_j$ was estimated in the first step. Specifically, we calculated the variance of the first-step prediction errors, which is then multiplied by $\hat{\theta}_1^2$ and added to the estimated variance of the second-step regression residuals.

missing values as we add more variables to the regressions.[33] Two variables lead to substantial loss of observations: First, adding risk tolerance to the regression eliminates about two-thirds of the sample because the HRS asked these questions of only about one-third of the subsample. Second, when we include the cognition variables (in particular, cognition questions related to numeracy), we lose another half of the sample. In order to avoid the possibility that our results are driven by the changing samples, we rerun all our regressions on three samples. Column 1 reports results using the full sample, column 2 for the medium sample (for which the risk tolerance measure is available), and column 3 for the smallest sample (for which the cognition variables are also available).

Now we explain the key findings from table 6.

• Rows 1 and 9 show that, regardless of the imputation method, if we do not control for any of the $\mathbf{X}_j$ variables, individuals with higher health expenditure risk (measured by $\hat{E}_j$) are less likely to purchase Medigap. For the sample with the most complete data, the point estimates imply that a $10,000 increase in health expenditure risk reduces the probability of buying Medigap by 4.6–5.7 percentage points. This, of course, is simply a confirmation (from a different angle and using different data) of our finding of advantageous selection reported in Section VI.

• Rows 2 and 10 add risk tolerance alone to the regression (6). This specification is important because the previous literature we have cited, both theoretical and empirical, has focused almost exclusively on risk aversion as the source of advantageous selection. When rows 2 and 10, respectively, are compared with rows 1 and 9, the inclusion of risk tolerance only slightly reduces, if at all, the magnitude of the negative coefficient on predicted health expenditures.[34] However, risk tolerance, as predicted by standard economic models, is a significant predictor of Medigap purchase: when it is entered as a set of dummy variables, an F-test rejects the null hypothesis of joint insignificance with a p-value of less than .001; indeed more risk-tolerant individuals are less likely to purchase Medigap. Importantly, more risk-tolerant individuals are not especially unhealthy. Thus, in the language we used in Section III, what we find here is that risk aversion (the opposite of risk tolerance) satisfies property 1 but does not satisfy property 2, which is also necessary for it to be a source of advantageous selection in the Medigap insurance market.

• Results change, however, when we also control for the variance of

[33] Most of the missing values result from the relevant survey questions not being asked instead of nonresponse.

[34] This result does not depend on the linear specification for the effect of risk tolerance. If, instead, risk tolerance enters as a third-order polynomial or as a complete set of dummy variables, the point estimate of the coefficient on $\hat{E}_j$ is essentially unchanged from that without the inclusion of risk tolerance.

TABLE 6

SOURCES OF ADVANTAGEOUS SELECTION

	Coefficient Estimates of $\hat{E}_j/10{,}000$			Conditioning Variables								Observations
	(1)	(2)	(3)	Risk Tolerance (4)	$\widehat{\text{Var}}_j$ (5)	Risk Tolerance $\times \widehat{\text{Var}}_j$ (6)	Education (7)	Income (8)	Cognition (9)	Longevity Expectation (10)	Planning Horizon (11)	(12)
	A. First Imputation Method: Using Only Observations in the MCBS with No Medigap											
1	−.0391 (.001)	−.0558 (.001)	−.0574 (.116)	N	N	N	N	N	N	N	N	9,973
2	. . .	−.0558 (.001)	−.0570 (.118)	Y	N	N	N	N	N	N	N	3,467
3	. . .	−.0301 (.162)	.0393 (.121)	Y	Y	Y	N	N	N	N	N	3,467
4	. . .	−.0234 (.281)	.0508 (.063)	Y	Y	Y	Y	N	N	N	N	3,467
5	. . .	−.0039 (.843)	.0636 (.060)	Y	Y	Y	Y	Y	N	N	N	3,467
6	. . .	. . .	.0758 (.049)	Y	Y	Y	Y	Y	Y	N	N	1,696
7	. . .	. . .	.0781 (.055)	Y	Y	Y	Y	Y	Y	Y	N	1,695
8	. . .	. . .	.0783 (.061)	Y	Y	Y	Y	Y	Y	Y	Y	1,659

	B. Second Imputation Method: Using All Observations in the MCBS											
9	−.0275	−.0519	−.0465	N	N	N	N	N	N	N	N	9,973
	(.004)	(.002)	(.167)									
10	. . .	−.0519	−.0461	Y	N	N	N	N	N	N	N	3,467
		(.002)	(.169)									
11	. . .	−.0296	.0263	Y	Y	Y	N	N	N	N	N	3,467
		(.173)	(.307)									
12	. . .	−.0222	.0380	Y	Y	Y	Y	N	N	N	N	3,467
		(.313)	(.133)									
13	. . .	−.0034	.0490	Y	Y	Y	Y	Y	N	N	N	3,467
		(.862)	(.091)									
14	. . .	. . .	.0562	Y	Y	Y	Y	Y	Y	N	N	1,696
			(.068)									
15	. . .	. . .	.0580	Y	Y	Y	Y	Y	Y	Y	N	1,695
			(.074)									
16	. . .	. . .	.0587	Y	Y	Y	Y	Y	Y	Y	Y	1,659
			(.080)									

NOTE.—All regressions include controls for female, a third-order polynomial in age 65, and state of residence. *p*-values are in parentheses.

health expenditures. In rows 3 and 11 we report results from a regression that includes risk tolerance, the predicted variance, and the interaction between the two. The inclusion of these measures affects the coefficient estimate on $\hat{E}_j$.[35] In the medium samples, the estimated coefficient on $\hat{E}_j$ remains negative but is now statistically indistinguishable from zero. In the small samples, the coefficient on $\hat{E}_j$ reverses its sign from negative to positive, though it is statistically indistinguishable from zero.

Notably, the estimated coefficient on the level of risk tolerance and the interaction between variance $\widehat{\mathrm{Var}}_j$ and risk tolerance are both negative. The latter is as we would expect. The probability of buying insurance increases with greater risk aversion, and more so for a person who faces a higher variance of expenditures. According to our point estimates, at the mean level of $\widehat{\mathrm{Var}}_j$ in our sample, a one-standard-deviation increase in the risk tolerance measure decreases the probability of purchasing Medigap by 6.7 percentage points. What is puzzling, at least from the perspective of standard economic theory, is that the coefficient on the variance term $\widehat{\mathrm{Var}}_j$ is negative. As the interaction between variance and risk tolerance is also negative, this implies that individuals with higher uncertainty in their medical expenditures are less likely to purchase Medigap (with this negative effect being larger for people who are more risk tolerant).

Remark 1. Strictly in terms of mechanics, the negative coefficient we estimate for the variance term $\widehat{\mathrm{Var}}$ is not a surprise. To see this, note that because the inclusion of $\widehat{\mathrm{Var}}$ in the regression reduces the magnitude of the negative coefficient on $\hat{E}$, $\widehat{\mathrm{Var}}$ is in some sense acting as a "source" of advantageous selection. Thus it must satisfy the two properties listed in Section III. Because $\widehat{\mathrm{Var}}$ increases with $\hat{E}$ (i.e., those with higher mean health expenditure also have high variance of health expenditure), it has to be the case that $\widehat{\mathrm{Var}}$ reduces the probability of purchasing Medigap in order for it to "act" as a source of advantageous selection.

Remark 2. A possible explanation within the standard expected utility theory for the negative coefficient estimate on $\widehat{\mathrm{Var}}$, with mean expenditure $\hat{E}$ held fixed, goes as follows: Consider a simple static model in which consumers face a loss of L with probability p. Then a mean-preserving spread implies reducing p while increasing L. Why might someone with higher L (and lower p) demand less insurance? Under the current U.S. institutional structure, Medicaid may cover catastrophic losses, and emergency rooms are required to (and often do) treat the uninsured. Thus the incentives for purchasing Medigap may be lower

[35] The bulk of the change in the coefficient on $\hat{E}_j$ is attributable to the inclusion of a control for the variance in health expenditure. For example, in results not shown here, adding just a control for variance in the smallest sample increases the coefficient on $\hat{E}_j$ to .0370, with a p-value of .143.

for those with higher L (and lower p) because large losses are more likely to be eligible for public-funded health care.

REMARK 3. Several conventional departures from the expected utility theory postulated in the behavioral decision theory literature can also explain the negative coefficient on variance. In such models, a decrease in p, with pL held fixed, could in fact reduce demand for insurance. For example, in the rank-dependent utility model (see, e.g., Quiggin 1982; Starmer 2000), the probability weights in expected utility theory are replaced by probability weighting functions. Say these functions down-weight low-probability events. Then, in a comparison of two people with equal pL, but where one has a large probability of small losses and the other has a very small probability of catastrophic losses, the latter person may in fact demand less insurance (because he or she chooses to engage in "wishful thinking" and largely ignore the small chance of the catastrophic event).

REMARK 4. Yet another possible behavioral explanation is that individuals may underestimate the variance of their health expenditures. It is possible that, conditional on mean expenditures, those with higher variances will have a larger downward bias in their prediction of variance, reducing their demand for insurance.[36]

REMARK 5. The above discussions have assumed that variance is a fixed and exogenous trait of an individual. A final possibility is that it is instead endogenous. For example, the technology of health investment may be such that the same behaviors that lead to poor health cause both a higher mean and a higher variance in expenditures, as well as leading people to purchase less insurance.

- As we include more variables from $\mathbf{X}_j$, reported respectively in rows 4–8 and rows 12–16 for the two imputation methods, the coefficient on $\hat{E}_j$ eventually becomes positive and statistically significant at the 5–10 percent level. A key result is in row 6, where the inclusion of cognitive ability as a control variable causes the coefficient on predicted expenditure to turn to positive .0758 with a p-value of .049. The most complete specification 8 implies that a \$10,000 increase in predicted health expenditure increases the probability of buying Medigap by 7.8 percentage points. All the new $\mathbf{X}_j$ variables we include shift the partial correlation between health expenditure risk and Medigap coverage in a positive direction.

Among all the variables in $\mathbf{X}_j$, cognition and income are distinctively important in two senses: first, if only cognition or income is included in $\mathbf{X}_j$, it substantially changes the estimated coefficient on $\hat{E}_j$; and second, when all variables in $\mathbf{X}_j$ are included in the regression, both cognition

[36] Koufopoulos (2005) explores how overconfidence in risk perception might lead to advantageous selection.

and income stand out as significant predictors of Medigap purchase. For example, adding the cognition variables alone changes the coefficient on $\hat{E}_j$ from $-.057$ (p-value .12) to $-.012$ (p-value .68); and in the full regression that includes all elements of $\mathbf{X}_j$, a standard deviation increase in just the TICS score is associated with a 5.4-percentage-point increase in the probability of purchasing a Medigap policy.[37]

Similarly, adding income variables alone changes the coefficient on $\hat{E}_j$ from $-.057$ (p-value .12) to $-.022$ (p-value .45); and someone with an annual income between $45,000 and $50,000 is 12 percentage points more likely to purchase a Medigap policy than someone with an income between $15,000 and $20,000. Thus, both cognition and income provide some important part of the explanation for the otherwise negative correlation between mean expenditure risk and insurance purchase.

In summary, these results indicate that sources of the advantageous selection we documented in Section VI include a number of factors—income, education, longevity expectations, and financial planning horizon—that would typically enter a rich economic model of insurance purchase. In addition, however, we find evidence that factors typically omitted from economic models, such as levels of cognitive ability and financial numeracy, are also important sources of advantageous selection. (We will explore the pathways through which cognitive ability may act as a source of advantageous selection below.) We find no evidence, however, that risk preferences, which the theoretical literature has focused on and which are significant predictors of Medigap purchase, can explain advantageous selection in this market. These findings are robust to the methods we use to impute expected medical expenditure and its variance.

Our study is, to the best of our knowledge, the first to provide direct evidence on the sources of advantageous selection in health insurance markets; it is also the first study to identify a set of variables that are sufficient to explain away the negative correlation between ex post health expenditure and Medigap coverage documented in Section VI and obtain a positive partial correlation between health expenditure risk and the level of insurance coverage.

Robustness.—In our working paper (Fang et al. 2006), we showed that our findings regarding the sources of advantageous selection are robust to different choices regarding whether those with employer-provided health insurance are included, as well as alternative measures of health expenditure risk (i.e., potential Medigap expenditure as described in Sec. VI.*C*) and alternative measures of the degree of uncertainty in

[37] For reasons of space, these results are not presented here but are available from the authors on request.

health expenditures (i.e., using the ratio of ninetieth and tenth percentiles in place of the variance).

D. *Pathways for Cognitive Ability as a Source of Advantageous Selection*

Our analysis points to cognitive ability as an important source of advantageous selection in the Medigap market. Because this factor is typically omitted from economic models, we consider here potential pathways through which cognitive ability may induce advantageous selection. This subsection is exploratory. Further research is needed to quantify the importance of the postulated pathways.

• A first potential pathway through which cognitive ability may act as a source of advantageous selection is via its effect on individuals' ability to evaluate the costs and benefits of purchasing Medigap. This channel is consistent, for example, with earlier literature showing that many senior citizens have difficulty understanding Medicare and Medigap rules; in particular, many fail to understand Medicare cost-sharing requirements (see, e.g., Cafferata 1984; McCall, Rice, and Sangl 1986; Davidson, Sofaer, and Gertler 1992; Harris and Keane 1998).[38] The conceptual framework in Section III clarifies how this channel might work. Let (p, c) denote health risk and cognitive ability, respectively. Suppose that there is a negative correlation between c and p; that is, individuals with higher cognitive ability have lower health expenditure risk. Then one particularly simple model of this channel would have a threshold cognitive ability level c^* below which individuals are simply unaware of Medigap and therefore do not buy. Alternatively, c^* might represent the cognitive ability level below which the costly effort required to determine the optimal Medigap decision is too great. In this latter case, optimal rules of thumb or other psychological forces may lead consumers more often to choose the status quo of no Medigap.[39] In either case, those with $c > c^*$ will purchase Medigap if it is worthwhile, and presumably only those with high health expenditure risk, that is, $p > p^*$, will choose to purchase Medigap. As a result, the set of individuals who purchase Medigap insurance is given by $\Omega_{\text{Medigap}} = \{(p, c) : c \geq c^*, p \geq p^*\}$, and the set of individuals who do not purchase Medigap is given by $\Omega_{\text{No\_Medigap}} = \{(p, c) : c < c^*;\ \text{or}\ (c > c^*, p < p^*)\}$. Because of the negative correlation in the population between c and p, such that p tends to be

[38] There is also a literature showing that many consumers have difficulty understanding health insurance plans more generally. See, e.g., Gibbs, Sangl, and Burrus (1996), Isaacs (1996), Tumlinson et al. (1997), and Cunningham, Denk, and Sinclair (2001).

[39] See, e.g., Iyengar and Lepper (2000), Iyengar and Jiang (2003), and Choi, Laibson, and Madrian (forthcoming) for evidence of the effect of complexity on the likelihood of choosing defaults and making "no choice."

TABLE 7
Do High-Cognition Individuals Pay Lower Medigap Premiums?

	(1)	(2)	(3)	(4)	(5)	(6)
Word recall	−2.5				−6.9	
	(12.9)				(24.0)	
Numeracy		26.2			71.9	
		(95.3)			(110.6)	
TICS score			−63.2*		−82.9	
			(37.3)		(68.3)	
Subtraction				−13.5	−14.4	
				(22.1)	(64.6)	
Cognition factor						−40.6
						(94.4)
Income/1,000	.42	−1.42	.45	.18	−1.04	−.97
	(1.5)	(2.14)	(1.44)	(1.44)	(2.34)	(2.34)
Observations	1,096	571	1,113	1,126	549	549
R^2	.11	.07	.11	.10	.08	.08

Note.—All regressions include controls for gender, state of residence, a third-order polynomial in age, and Medigap plan letters. Robust standard errors are in parentheses.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

larger for those with $c < c^*$, it is possible that we observe advantageous selection if we do not condition on cognition c.

• A second potential channel through which cognitive ability may act as a source of advantageous selection is via its effect on search costs. Given the price dispersion in Medigap plans offered by different insurers, those with higher cognitive ability may be able to obtain lower effective prices and thus be more likely to purchase Medigap. We examined two testable implications of this pathway. First, if this pathway is important, Medigap premiums paid by individuals with higher cognitive ability should tend to be lower than those paid by individuals with lower cognitive ability. Second, the observed extent of advantageous selection should be less pronounced in states with less Medigap price dispersion. To test the first implication, we use a subset of individuals in the HRS who both have nonmissing information about cognitive ability and also report their Medigap premium. Table 7 provides the results from simple regressions of the paid Medigap premium on the four measures of cognitive ability we used in our earlier analysis, word recall, numeracy, TICS score, and subtracting 7, first separately in columns 1–4 and jointly in column 5. In column 6 we extract a cognitive ability factor from the four measures. We also included income in the regressions because earlier we also found income to be a source of advantageous selection. Thus it is also interesting to see whether higher-income individuals purchase Medigap with a higher probability because of lower effective prices (due to lower search costs, say). The evidence is rather weak: column 3 does show that individuals with high TICS

TABLE 8
Comparisons of the Extent of Advantageous Selection between States with Different Coefficients of Variation for Medigap Plan C Prices

	States with Coefficient of Variation below the Median		States with Coefficient of Variation above the Median	
	(1)	(2)	(3)	(4)
Medigap coefficients	−4,879.7***	1,748.0***	−3,774.1***	2,333.1***
	(459.0)	(331.6)	(528.5)	(415.0)
Health controls	No	Yes	No	Yes
Observations	8,889	7,894	7,056	6,280

Note.—All regressions are of the same specifications as the corresponding ones in table 2. They are weighted by the cross-section sample weights. The descriptions of the direct health controls can be found in the Data Appendix. Robust standard errors in parentheses are clustered at the individual level.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

scores pay slightly lower premiums, but this is significant only at the 10 percent level, and all the other coefficients are statistically insignificant.[40]

To test the second implication, we divide the states into two groups according to whether the coefficient of variation of Medigap Plan C (the most popular Medigap plan) in that state is below or above the median coefficient of variation (which is equal to .17).[41] We use the MCBS data and rerun the regressions reported in table 2 separately for the two groups of states. Table 8 shows the estimated Medigap coefficients with and without health controls for the two groups of states. The magnitude of advantageous selection is little affected by the degree of price dispersion in the Medigap market (and, indeed, it is slightly greater in the states with a low coefficient of variation).

Taken together, the results in tables 7 and 8 suggest that the search cost channel is not central to the relationship between cognitive ability and advantageous selection. However, we view this evidence as preliminary, and additional research on the relationship between cognition and search costs is needed.

• A third potential channel through which cognitive ability may act as a source of advantageous selection is via its effect on individuals' information about health risks. High–cognitive ability individuals may be healthier, but they may be more knowledgeable about potential health risks. Specifically, consider two individuals with different cognitive ability. The high–cognitive ability individual may have better health status now, but he or she is aware of all the potential risks to health;

[40] A one-standard-deviation increase in TICS score (1.3) leads to about a $78 lower Medigap premium, which is less than 10 percent of the average premium for even the cheapest Medigap Plan A.

[41] The coefficients of variation for different Medigap plans by state come from app. table 1 of Maestas et al. (2006), which used data from Weiss Ratings.

the one with low cognitive ability thinks that there will be no more health shocks beyond what he or she already experienced. The first individual may be more likely to purchase Medigap than the second, thus leading to advantageous selection.[42]

Understanding the pathways for cognitive ability and other variables to act as sources of advantageous selection has important policy implications. If the first channel is important, it would suggest a role for educational interventions to facilitate choice, or simplification of Medigap rules to make the cost-benefit calculations simpler (see, e.g., Harris 2002); if the second channel is important, then pamphlets with detailed price quotes (products that Weiss Ratings currently provides at a cost) directly sent to Medicare recipients may increase Medigap enrollment; if the third channel is important, it will call for yet a different kind of information campaign, which is not about Medicare or Medigap, but about various health risks the elderly may be facing.

V. Conclusion

In this paper we use data from the Medicare Current Beneficiary Survey to provide strong evidence of advantageous selection in Medigap insurance market. The first type of evidence comes from two sets of regressions. In one set, we regress total medical expenditure on Medigap status and control only for the determinants of price (gender, age, and state of residence). We find that those with Medigap incur, on average, about $4,000 less in total medical expenditure than those without Medigap. In the second set of regressions, we regress total medical expenditure on the determinants of price along with a rich set of controls for health status. Conditional on price and health, we find that those with Medigap incur about $2,000 more in medical expenditure, on average, than those without Medigap. These two sets of results can be reconciled only if those with better health are more likely to purchase supplemental coverage, that is, if there is "advantageous selection." These results are robust to different definitions of Medigap status and to separating the sample into male and female subsamples. We find that the magnitude of advantageous selection is larger for females than for males.

We then propose a simple empirical strategy to combine MCBS and HRS data and examine the sources of advantageous selection. Our findings indicate that these sources include factors, such as income, education, and planning horizons, that a rich economic model of insurance

[42] We thank an anonymous referee for suggesting this potential pathway for cognitive ability to act as a source of advantageous selection. Coelho and de Meza (2007) also argued that individuals with different cognitive abilities may have different abilities to predict the likelihood of illness, which will affect their insurance purchase.

purchase would typically accommodate. Interestingly, we find no evidence that variation in risk preferences, which is the primary focus of the theoretical literature on advantageous selection, explains the otherwise negative relationship between coverage and expenditure risk. Those who are less risk tolerant buy more insurance, but they are not particularly healthy.

In addition, we find that measures of cognitive ability and financial numeracy, which standard economic models do not accommodate, are important sources of advantageous selection; and we provide a preliminary exploration of the pathways through which cognitive ability may act as a source of advantageous selection. Our findings also suggest that heterogeneity in risk distributions, not merely expected levels of expenditure risk, should be included in models of asymmetric information. Specifically, we found that differences in the variance of health expenditure, which standard models of insurance purchase ignore, are related both to average expenditure risk and to insurance purchase. Our findings are robust to changes in the sample and the method of imputing expected medical expenditure.

Finally, while advantageous selection in an insurance market can cancel out the positive correlation between ex post risk and insurance coverage that arises in classic adverse selection models such as Rothschild and Stiglitz (1976), it is important to emphasize that this does not mean there is no inefficiency in such a market. The policy implications of multidimensional selection models are an important topic for future research.

Data Appendix

TABLE A1
DATA DESCRIPTION

Variable	Data	Description
Health expenditure:		
Total expenditure	MCBS	Total annual health care expenditure for 12 months of the survey year. Expenditure includes data from Medicare administrative files and survey responses for out-of-pocket and otherwise insured expenditures
Insurance:		
Medicare	MCBS/HRS	Indicators for whether the respondent is covered by Medicare Parts A and B
Medigap	MCBS	Indicator for whether repsondent with Medicare coverage also has self-purchased private health insurance. Those covered by employer-provided health insurance, Medicaid, or VA Champus (Tri-Care) are treated as missing
	HRS	Indicator for whether respondent with Medicare coverage also has private health insurance that is secondary to Medicare and is not purchased from a (spouse's) employer or union. Those covered by employer-provided health insurance, Medicaid, or VA Champus (Tri-Care) are treated as missing
Demographics:		
Race	MCBS/HRS	Indicators for self-reported black, other, and nonresponse
Hispanic	MCBS/HRS	Indicators for self-reported Hispanic and nonresponse
Education	MCBS/HRS	Indicators for highest grade completed less than eighth grade, some high school, high school graduate, some college, college graduate, at least some graduate school, and nonresponse
Marital status	MCBS/HRS	Indicators for married, widowed, divorced, separated, and nonresponse
Number of children	MCBS/HRS	The number of children the respondent has ever had
Income	MCBS	Indicators for self-reported total household income in $5,000 intervals from $5,000 to $50,000 and $50,000 plus

	HRS	Same indicators as above, except in this case we use reported as well as imputed values for total household income. Imputations are those generated by Rand. See http://hrsonline.isr.umich.edu/meta/rand/randhrse/randhrse.pdf for details of the imputation
Work status	MCBS/HRS	Indicators if currently working for pay and for nonresponse
Health:		
Self-reported	MCBS/HRS	Indicators for self-reported health excellent, very good, good, and fair
Height	MCBS/HRS	Self-reported height, in inches, and height squared
Body mass index	MCBS/HRS	Self-reported [weight (kg)]/[height $(m)^2$]
Ever a smoker	MCBS/HRS	Indicator if respondent has "ever smoked" tobacco
Current smoker	MCBS/HRS	Indicator if respondent now smokes tobacco and for nonresponse
Diagnoses	MCBS/HRS	Indicators for if a doctor has ever told the respondent he/she has arthritis, high blood pressure, diabetes, (nonskin) cancer, lung disease, heart attack, chronic heart disease, stroke, psychiatric illness, Alzheimer's disease, broken hip, and for each diagnosis, nonresponse
Treatments	MCBS/HRS	Indicators for respondent ever having cataract surgery or a hearing aid
(Instrumental) ADLs	MCBS/HRS	Indicators for if a respondent has at least some difficulty walking 2–3 blocks, stooping, reaching overhead, lifting 10 lbs., dressing, walking at all, bathing, eating, getting out of a chair, using the toilet, daily living, preparing meals, shopping, using the telephone, managing money and bills, and for nonresponse
Help with instrumental ADLs	MCBS/HRS	Indicators for if a respondent receives help dressing, walking at all, bathing, eating, getting out of a chair, using the toilet, preparing a meal, shopping, using the telephone, or managing money and bills, and for nonresponse
Risk attitudes:		
Risk tolerance	HRS	Estimate of risk tolerance from Kimball et al. (forthcoming), using responses to hypothetical income gambles from 1992 and 1994
Cognition:		
Word recall	HRS	Variables recording the number of words recalled from a list of 10, both immediately after the list was read and several minutes later
TICS score	HRS	Telephone Interview for Cognitive Status: number of correct answers on a test of knowledge, language, and orientation. Questions include naming objects, vocabulary questions, and basic knowledge such as the U.S. president's name
Subtraction	HRS	Number of times respondent can subtract the number 7 sequentially, starting from 100

TABLE A1 (*Continued*)

Variable	Data	Description
Numeracy	HRS	Number of correct answers to “word problems” of division and multiplication on topics of probability, compound interest, and division of assets; asked only in 2002
Expectations:		
Longevity	HRS	Most recent answer to the question “What is the percent chance you will live to 75 or more?”
Planning horizon:		
Financial	HRS	Indicators for whether the respondent’s most important period for planning saving and spending is the next few months, the next year, the next few years, the next 5–10 years, or more than 10 years

References

Arrow, Kenneth. 1963. "Uncertainty and the Welfare Economics of Medicare Care." *A.E.R.* 53 (December): 941–73.

Banthin, Jessica S., and Amy K. Taylor. 2001. "HMO Enrollment in the United States: Estimates Based on Household Reports, 1996." MEPS Research Findings no. 15, Agency for Healthcare Research and Quality, Rockville, MD.

Barsky, Robert B., F. Thomas Juster, Miles Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Q.J.E.* 112 (May): 537–79.

Cafferata, Gail L. 1984. "Knowledge of Their Health Insurance Coverage by the Elderly." *Medical Care* 22 (September): 835–47.

Cardon, James H., and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *Rand J. Econ.* 32 (Autumn): 408–27.

Cawley, John, and Tomas Philipson. 1999. "An Empirical Examination of Information Barriers to Trade in Insurance." *A.E.R.* 89 (September): 827–46.

Charles, Kerwin, and Erik Hurst. 2003. "The Correlation of Wealth across Generations." *J.P.E.* 111 (December): 1155–82.

Chiappori, Pierre-André. 2001. "Econometric Models of Insurance under Asymmetric Information." In *Handbook of Insurance*, edited by Georges Dionne. Boston: Kluwer.

Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié, and Francois Salanié. 2006. "Asymmetric Information in Insurance: General Testable Implications." *Rand J. Econ.* 37 (Winter): 783–98.

Chiappori, Pierre-André, and Bernard Salanié. 2000. "Testing for Asymmetric Information in Insurance Markets." *J.P.E.* 108 (February): 56–78.

Choi, James J., David Laibson, and Brigitte C. Madrian. Forthcoming. "Reducing the Complexity Costs of 401(k) Participation through Quick Enrollment™." In *Developments in the Economics of Aging*, edited by David A. Wise. Chicago: Univ. Chicago Press.

CMS (Centers for Medicare and Medicaid Services). 2005. *Choosing a Medigap Policy: A Guide to Health Insurance for People with Medicare.* Baltimore: U.S. Dept. Health and Human Services.

Coelho, Marta, and David de Meza. 2007. "Self Deception, Self Selection, Self Destruction: Experimental Evidence against Adverse Selection." Manuscript, London School Econ.

Cohen, Alma. 2005. "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market." *Rev. Econ. Statis.* 87 (June): 197–207.

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *A.E.R.* 97 (June): 745–88.

Cox, Donald F., and Christopher Hogan. 1997. "Biased Selection and Medicare HMOs: Analysis of the 1989–1994 Experience." *Medical Care Res. and Rev.* 54 (September): 259–85.

Cunningham, Peter J., Charles Denk, and Michael Sinclair. 2001. "Do Consumers Know How Their Health Plan Works?" *Health Affairs* 20 (March/April): 159–66.

Davidson, Bruce, Shoshanna Sofaer, and Paul Gertler. 1992. "Consumer Demand and Biased Selection in the Demand for Coverage Supplementing Medicare." *Soc. Sci. and Medicine* 34 (9): 1023–34

de Meza, David, and David C. Webb. 2001. "Advantageous Selection in Insurance Markets." *Rand J. Econ.* 32 (Summer): 249–62.

Desmond, Katherine A., Thomas Rice, and Peter D. Fox. 2006. "Does Greater

Medicare HMO Enrollment Cause Adverse Selection into Medigap?" *Health Econ., Policy and Law* 1 (1): 3–21.
Dionne, Georges, Christian Gouriéroux, and Charles Vanasse. 2001. "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment." *J.P.E.* 109 (April): 444–53.
Ettner, Susan L. 1997. "Adverse Selection and the Purchase of Medigap Insurance by the Elderly." *J. Health Econ.* 16 (October): 543–62.
Fang, Hanming, Michael P. Keane, and Dan Silverman. 2006. "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." Working Paper no. 12289 (June), NBER, Cambridge, MA.
Finkelstein, Amy, and Kathleen McGarry. 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *A.E.R.* 96 (September): 938–58.
Finkelstein, Amy, and James Poterba. 2004. "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market." *J.P.E.* 112 (February): 183–208.
———. 2006. "Testing for Adverse Selection with 'Unused Observables.'" Manuscript, Dept. Econ., Massachusetts Inst. Tech.
Gibbs, Deborah A., Judith A. Sangl, and Barri Burrus. 1996. "Consumer Perspectives on Information Needs for Health Plan Choice." *Health Care Financing Rev.* 18 (Fall): 55–73.
Harris, Katherine M. 2002. "Can High Quality Overcome Consumer Resistance to Restricted Provider Access? Evidence from a Health Plan Choice Experiment." *Health Services Res.* 37 (June): 551–71.
Harris, Katherine M., and Michael P. Keane. 1998. "A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Revealed Preference and Attitudinal Data." *J. Econometrics* 89 (November): 131–57.
Hemenway, David. 1990. "Propitious Selection." *Q.J.E.* 105 (November): 1063–69.
Hurd, Michael D., and Kathleen McGarry. 1997. "Medical Insurance and the Use of Health Care Services by the Elderly." *J. Health Econ.* 16 (April): 129–54.
Isaacs, Stephen L. 1996. "Consumers' Information Needs: Results of a National Survey." *Health Affairs* 15 (Winter): 31–41.
Iyengar, Sheena S., and Wei Jiang. 2003. "Choosing Not to Choose: The Effect of More Choices on Retirement Savings Decisions." Manuscript, Columbia Univ.
Iyengar, Sheena S., and Mark Lepper. 2000. "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *J. Personality and Soc. Psychology* 79 (December): 995–1006.
Jullien, Bruno, Benard Salanié, and Francois Salanié. 2007. "Screening Risk Averse Agents under Moral Hazard." *Econ. Theory* 30 (January): 151–69.
Kaiser Family Foundation. 2005. *Medicare Chart Book.* Washington, DC: Kaiser Family Found.
Kan, Kamhon. 2003. "Residential Mobility and Job Changes under Uncertainty." *J. Urban Econ.* 54 (November): 566–86.
Keane, Michael P. 2004. "Modeling Health Insurance Choice Using the Heterogeneous Logit Model." Manuscript, Yale Univ.
Kezdi, G., and R. J. Willis. 2005. "Who Becomes a Stockholder? Expectations, Subjective Uncertainty, and Asset Allocation." Manuscript, Univ. Michigan.
Khandker, Rezaul K., and Lauren A. McCormack. 1996. *Enrollment and Utilization*

across Medicare Supplemental Plans. Report submitted to the Health Care Financing Admin. Washington, DC: Center Health Econ. Res.

———. 1999. “Medicare Spending by Beneficiaries with Various Types of Supplemental Insurance.” *Medical Care Res. and Rev.* 56 (June): 137–55.

Khwaja, Ahmed. 2005. “The Dynamic Trade-off between Medical Expenditure Insurance, Moral Hazard and Mortality Risk: A Life Cycle Analysis of the Effects of Medicare on Individual Health Incentives and Health Outcomes.” Manuscript, Fuqua School Bus., Duke Univ.

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. Forthcoming. “Using Survey-Based Risk Tolerance.” *J. American Statis. Assoc.*

Koufopoulos, Kostas. 2005. “Asymmetric Information, Heterogeneity in Risk Perceptions and Insurance: An Explanation of a Puzzle.” Manuscript, London School Econ.

Lewin Group. 2001. “Restricting Underwriting and Premium Rating Practices in the Medigap Market: The Experience of Three States.” Report no. 2001-01. Washington, DC: AARP, Public Policy Inst.

Lillard, Lee A., and Jeannette Rogowski. 1995. “Does Supplemental Private Insurance Increase Medicare Costs?” Working Paper no. 95-16, Rand Corp., Santa Monica, CA.

Lusardi, Annamaria. 1998. “On the Importance of the Precautionary Saving Motive.” *A.E.R. Papers and Proc.* 88 (May): 449–54.

Maestas, Nicole, Mathis Schroeder, and Dana Goldman. 2006. “Price Variation in Markets with Homogeneous Goods: The Case of Medigap.” Working Paper no. 504, Rand Corp., Santa Monica, CA.

McCall, Nelda, Thomas Rice, and Judith Sangl. 1986. “Consumer Knowledge of Medicare and Supplemental Health Insurance Benefits.” *Health Services Res.* 20 (February): 633–57.

Pauly, Mark V. 1974. “Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection.” *Q.J.E.* 88 (February): 44–62.

Puelz, Robert, and Arthur Snow. 1994. “Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market.” *J.P.E.* 102 (April): 236–57.

Quiggin, John. 1982. “A Theory of Anticipated Utility.” *J. Econ. Behavior and Organization* 3 (December): 323–43.

Ridder, Geert, and Robert Moffitt. 2007. “The Econometrics of Data Combination.” In *Handbook of Econometrics*, vol. 6A, edited by James J. Heckman and Edward E. Leamer. Amsterdam: Elsevier.

Robst, John. 2006. “Estimation of a Hedonic Pricing Model for Medigap Insurance.” *Health Res. and Educational Trust* 41 (December): 2097–2113.

Rothschild, Michael, and Joseph E. Stiglitz. 1976. “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information.” *Q.J.E.* 90 (November): 629–49.

Schmidt, Lucie. Forthcoming. “Risk Preferences and the Timing of Marriage and Childbearing.” *Demography.*

Starmer, Chris. 2000. “Developments in Non–Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk.” *J. Econ. Literature* 38 (June): 332–82.

Tumlinson, Anne, Hannah Bottigheimer, Peter Mahoney, Elliot Stone, and Ann Hendricks. 1997. “Choosing a Health Plan: What Information Will Consumers Use?” *Health Affairs* 16 (May/June): 229–38.

Wilson, Charles. 1977. “A Model of Insurance Markets with Incomplete Information.” *J. Econ. Theory* 16 (December): 167–207.

Wolfe, John R., and John H. Goddeeris. 1991. "Adverse Selection, Moral Hazard and Wealth Effects in the Medigap Insurance Market." *J. Health Econ.* 10 (4): 433–59.