MICHIGAN RETIREMENT AND DISABILITY RESEARCH CENTER
UNIVERSITY OF MICHIGAN

Promoting research on retirement, disability, and Social Security policy

Insurance Purchases of Older Americans

Karolos Arapakis, Eric French, John Bailey Jones, and Jeremy McCauley

MRDRC WP 2023-463

UM21-09

Insurance Purchases of Older Americans

Karolos Arapakis
Center for Retirement Research
at Boston College

Eric French
University of Cambridge
Institute for Fiscal Studies

John Bailey Jones
Federal Reserve Bank of Richmond

Jeremy McCauley
University of Bristol

March 2023

Michigan Retirement and Disability Research Center, University of Michigan, P.O. Box 1248. Ann Arbor, MI 48104, mrdrc.isr.umich.edu, (734) 615-0422

Acknowledgements

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium through the University of Michigan Retirement and Disability Research Center Award RDR18000002-03. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the federal government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States government or any agency thereof.

Insurance Purchases of Older Americans

Abstract

Using data from the Health and Retirement Study (HRS) matched to administrative Medicare and Medicaid records, and the Medical Expenditure Panel Survey (MEPS), we estimate the determinants of Medigap insurance purchases of older Americans. Our data and framework allow us to consider multiple reasons individuals do not purchase Medigap insurance: (i) adverse selection caused by those in poorer health purchasing Medigap; (ii) crowd out due to Medicaid insurance coverage or the option to default on future medical debt; (iii) behavioral factors such as risk tolerance and cognition. We find that those who purchase Medigap spend approximately $2,300 more than those who do not. However, we find no evidence that this higher spending is caused by adverse selection in this market and only modest evidence that crowd out and behavioral factors are important for understanding Medigap purchases. Our results are consistent with the view that this higher spending is caused by moral hazard driven by the lower out-of-pocket prices for additional care faced by those with Medigap.

Citation

Arapakis, Karolos, Eric French, John Bailey Jones, and Jeremy McCauley. 2023. “Insurance Purchases of Older Americans.” Ann Arbor, MI. University of Michigan Retirement and Disability Research Center (MRDRC) Working Paper; MRDRC WP 2023-463. https://mrdrc.isr.umich.edu/publications/papers/pdf/wp463.pdf

Authors’ acknowledgements

Preliminary, please do not quote. We thank Chris Firth, Pauline Sow, and Erin Henry for exellent help analyzing the data. Funding from the Social Security Administration through the Michigan Retirement Research Center (MRDRC grant UM21-09) and the Economic and Social Research Council for grants “Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies" (ES/T014334/1) and “Intergenerational transfers, insurance, and the transmission of inequality” (ES/V001248/1). Correspondence to eric.french.econ@gmail.com, jeremymccauley@gmail.com. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA, any agency of the Federal Government, the Center for Retirement Research, the Federal Reserve Bank of Richmond, or the Federal Reserve System. Any errors are the authors’ own.

1 Introduction

The great majority of Americans aged 65 and older receive health care benefits from Medicare, a mostly free single-payer program that covers most health care services. Traditional Medicare does not cover the complete cost of health services, however, as it imposes deductibles, co-insurance, and co-payments. As a result, there is an active market in Medicare Supplement Insurance, referred to as "Medigap", which is privately-provided health insurance designed to cover these "gaps" in traditional Medicare. Due to Medicare's relatively high cost-sharing requirements, most Medicare beneficiaries have Medigap or other coverage that supplements Medicare. Understanding who buys Medigap and how Medigap affects medical spending is important for understanding how individuals manage medical expense and other risks in old age (Koijen et al. 2016; Laitner et al. 2018). Furthermore, because Medigap makes health care almost free for consumers, it potentially induces an inefficiently high level of health care consumption. As a result, there have been proposals to limit Medigap either by mandating minimum deductibles or by enacting surcharges on Medigap premiums (Jacobson et al., 2014).

This paper uses data from the Health and Retirement Study (HRS), matched to administrative Medicare and Medicaid records, and also the Medical Expenditure Panel Survey (MEPS), to estimate the determinants of Medigap insurance purchases among older Americans. Our data and framework allows us to consider multiple reasons for why individuals do or do not purchase Medigap insurance, in particular: (i) adverse selection (i.e., those who are less healthy and more likely to need health care have an incentive to buy more insurance); (ii) crowd-out by publicly-provided Medicaid insurance or medical debt default (i.e., those who believe they can default on debt or think they may be eligible for Medicaid coverage in the future may see no need for additional insurance); and (iii) behavioral factors such as risk tolerance and cognition (i.e., individuals' attitudes to risk may impact whether they buy insurance).

We combine administrative Medicare and Medicaid records with out-of-pocket spending data found in the HRS and imputations for other payors based on MEPS, which results in a nationally representative long panel of medical spending by all payors for older individuals, the first of its kind. These data are then linked to the broad set of covariates available in the HRS. This dataset allows us to assess whether Medigap enrollees have higher or lower medical expenses than non-enrollees, and whether there are health, behavioral or financial factors that are also correlated with the Medigap purchase decision. Relative to the datasets used in previous analyses of the Medigap purchase decision, these data form a longer panel and cover a broader range of variables.

We find that annual total medical spending of those who purchase Medigap insurance is approximately $2,300 more than those who do not. However, we find no evidence that this higher spending is driven by adverse selection. In fact, we find that those who purchase Medigap are in better health, consistent with the hypothesis of "advantageous selection" in this market (Fang et al., 2008). Turning to alternative explanations, we find modest evidence that crowd out and behavioral factors such as risk tolerance and cognition are important for understanding Medigap purchases. However, these factors do not explain the higher medical spending of those purchasing Medigap. Instead, our results are consistent with moral hazard, where those with Medigap pay little out-of-pocket at the margin for additional health care, which may lead these individuals to demand more than those without Medigap.

2 Literature Review

Within the literature that tests for the determinants of insurance choice, our paper seeks to bridge two approaches.

The first approach is to examine the Medigap purchase decision through the lens of asymmetric information. Our paper is closely related to empirical investigations of the "positive correlation property," which posits that there should be a positive association between insurance coverage and ex-post risk.[1] Surprisingly, evidence for the positive correlation property is rather mixed. In the paper closest to ours, Fang et al. (2008) find evidence that those purchasing Medigap spend *less* for their total medical care than those not purchasing coverage.[2] Such a finding might reflect "advantageous selection", the case where healthier (and/or less expensive) individuals are more likely to purchase insurance. Fang et al. (2008) indeed find that Medigap purchasers are in better health. Advantageous selection in health insurance can arise when there are factors that simultaneously make individuals more healthy and more inclined to purchase insurance. Fang et al. (2008) identify several such factors, finding, for example, that that those who purchase insurance are more likely to have high cognitive ability. Another concern is moral hazard, as the subsidies embodied in Medigap coverage encourage increased spending. Keane and Stavrunova (2016) find that even though Medigap policyholders are in better health, they spend more than those without Medigap. They take this combination of outcomes as an indicator of moral hazard.

[1] See Cardon and Hendel (2001) for an early test of "positive correlation" and Chiappori and Salanie (2000) for a theoretical treatment. Cohen and Siegelman (2010) review the empirical literature.

[2] Papers testing for assymetric information in the Medigap insurance market include Wolfe and Goddeeris (1991), Ettner (1997), Hurd and McGarry (1997), Khandker and McCormack (1999), Dardanoni and Li Donni (2012) and Keane and Stavrunova (2016). See Fang et al. (2008) for a review.

The second approach is to investigate whether alternative forms of insurance are crowding out private coverage. A number of studies have shown that Medicaid has the potential to crowd out private insurance (Brown and Finkelstein 2008; Braun et al. 2019, Laitner et al. 2018). For those eligible, Medicaid pays bills not covered by Medicare and Medigap. Thus Medicaid payments displace payments made by Medigap. Thus Medicaid potentially negates any value of Medigap for individuals who either currently receive Medicaid or would likely receive Medicaid in the event of a negative health shock. Another possible source of crowding out exists in the implicit insurance provided by bankruptcy and other forms of medical debt forgiveness. Mahoney (2015) finds that individuals with less wealth to lose under bankruptcy are less likely to purchase health insurance. It bears noting that crowding out can induce advantageous selection. Richer individuals, who are less likely to rely on Medicaid or bankruptcy protection, are both healthier and more likely to purchase Medigap insurance.

Our paper extends the literature by considering both approaches simultaneously, using a dataset that combines accurate measures of medical spending with accurate measures of the health, behavioral and financial factors that potentially influence the Medigap purchase decisions. For example, Fang et al. (2008) consider the sources of advantageous selection by regressing the probability of purchasing Medigap against health, cognition, risk aversion, and planning horizon and a measure of medical spending. The HRS dataset they use for this exercise, however, did not contain a comprehensive measure of total medical spending, leading them to rely on imputations constructed using data from the Medicare Current Beneficiary Survey (MCBS). With administrative records now available, we are able to replace imputations with observed spending. Similarly, Mahoney (2015) works with the MEPS dataset, which has high quality spending measures, but lacks many of the correlates found in the HRS.

3 Institutional Background

3.1 Medicare

Medicare has four parts. Medicare originally consisted of Hospital Insurance (Part A) and Supplemental Medical Insurance (Part B). Since then it has expanded to include Medicare Advantage (Part C), which allows beneficiaries to instead enroll in a private HMO or PPO plan, and drug coverage (Part D).

Part A covers inpatient hospital services and is available for free to virtually all individuals aged 65 or older. Most Part A beneficiaries are also enrolled in Part B. Part B mainly covers doctor office

visits and other outpatient services. Part B beneficiaries are usually charged a premium, which can vary with their circumstances. Part D, which became effective in 2006, covers prescription drugs. Part D also imposes premia, but as with Part B, coverage is heavily subsidized and most eligible households participate (Kirchhoff, 2020). Although traditional Medicare covers the costs of many health care services, it imposes high cost-sharing requirements, including deductibles for Parts A and B services, 20% coinsurance for most Part B services, co-insurance for inpatient hospital and skilled nursing facility stays exceeding 20 days. There is no maximum on potential out-of-pocket costs each year.

As an alternative to Parts A and B, individuals can enroll in Medicare Advantage (Part C) plans, and have their coverage provided by HMOs or PPOs approved by Medicare. Part C plans cover the same services as Parts A and B, but may include additional benefits. The Medicare Advantage program is not separately financed, but is instead funded by transfers from the Part A and Part B trust fund. Many individuals buy bundled plans that include Part D coverage (U.S. Centers for Medicare and Medicaid Services, 2022). Over our sample period, the fraction of Medicare beneficiaries enrolled in a Medicare Advantage plan rose from 18% in 1999 to 27% by 2012. Since then, the fraction has continued to grow (Jacobson et al., 2016).

3.2 Medigap

Medicare Supplemental Insurance, or Medigap, is privately-provided insurance designed to cover some of the copayments, coinsurance and deductibles associated with Part A and B coverage, along with a few services not covered by Parts A and B.[3] Medigap users must be enrolled in Parts A and B and pay part B premia. Medigap does not cover the out-of-pocket charges associated with Part C, and insurers cannot sell Medigap plans to Part C users unless the users are switching back to traditional Medicare.[4]

A feature of the Medigap program that makes it extremely useful for the study of asymmetric information is that the coverage and pricing of Medigap policies are extremely restricted. Insurers are allowed to offer at most 10 standardized plans, and during the annual open enrollment period they cannot deny applicants coverage or charge them more for pre-existing conditions (Fang et al., 2008). This means that individuals with different health risks choose from the same menu of policies. By law, insurers can only price Medigap policies based on age, gender, state of residence

[3]New Medigap plans cannot cover prescription drugs, but customers can remain enrolled in plans that cover drug costs if the plans were purchased before the advent of Part D in 2006 (U.S. Centers for Medicare and Medicaid Services, 2023).

[4]See U.S. Centers for Medicare and Medicaid Services (2023) for more details.

and smoking status (Keane and Stavrunova, 2016). Thus, expenditure risk due to other factors, such as health status, can be viewed as "private" information for purposes of the analysis.

3.3 Medicaid

Many older individuals also receive Medicaid, public insurance available to the those with low income and wealth, or those impoverished by their medical expenses. Medicaid covers a number of services not covered by Medicare, most notably the cost of long-term care. Medicaid also affects what beneficiaries have to pay out of pocket for Medicare coverage. Many traditional Medicare beneficiaries are "dual-eligibles," who in addition to receiving Medicare coverage also receive Medicaid benefits that pay for services Medicare does not cover in addition to payments of their Medicare premia, co-pays, and deductibles. In addition, traditional beneficiaries who do not qualify for full Medicaid benefits may instead qualify for Medicare Savings Programs that cover premia and/or cost sharing. De Nardi et al. (2012) provide additional detail.

Like Medigap, Medicaid covers payments not covered by Medicare. For this reason, payments made by Medicaid merely displace payments by Medigap. As a result, individuals eligble for Medicaid may find little benefit from purchasing Medigap.

4 Data and Descriptive Statistics

Our main data source is the HRS survey data linked to restricted administrative Medicare and Medicaid records. The HRS data thus provide measures of Medicare, Medicaid, and out-of-pocket medical payments. Our second data source is the MEPS, which we use to impute medical payments from other sources; these include medical payments made by Medicare Part C, private insurers, and other smaller payors such as the Veterans Administration and state or local health departments. In the rest of this section we discuss each data source separately, describe our sample selection criteria, and examine our medical spending measure.

4.1 The HRS

The HRS is a nationally representative biennial survey of the over-50 U.S. population and their spouses. The survey started in 1992 and is based on a steady-state sampling design, with a new cohort of individuals aged 51-56 entering every six years. The HRS asks questions about a broad range of topics, and the richness of its data makes it highly appropriate for studying selection in Medigap insurance.

The HRS contains rich information about the respondents' demographics, health status (including cognitive ability), insurance choices, risk attitudes, longevity expectations and financial planning

horizons. For most of these variables we follow the approach in Fang et al. (2008), although for the risk tolerance measure we follow Kimball et al. (2008), who estimate the risk tolerance for each respondent in the HRS by maximum likelihood.[5] The HRS also contains detailed data on out-of-pocket medical spending, including payments for drugs, hospital stays, nursing home care, doctor visits, dental visits, and outpatient care. Notably, the expenditure data include expenses incurred during the last year of life. Appendix A describes many of the HRS variables used in our study.

4.2 Administrative Medicare and Medicaid Records

The Centers for Medicare and Medicaid Services (CMS) provide confidential administrative payment and enrollment records for Medicare and Medicaid which we link to consenting HRS respondents.

The Medicare records include Medicare part A, B and D payments, the respondent liability amount (i.e., the co-pays and deductibles from Medicare part A and B covered care) and enrollment (Medicare part C) information for every year between 1991 and 2016.[6] Specifically, the records include reimbursement amounts for inpatient, skilled nursing facility, home health, and hospice claims made under Medicare Part A, as well as outpatient, carrier (non-institutional providers such as individual or group practitioners, non-hospital labs, and ambulances), and durable-medical-equipment payments made under Medicare Part B. They also include drug-related spending made under Medicare Part D once it began in 2006. Although the records do not contain payments made through part C, they report for each year how many months the respondent was enrolled to Medicare Part C.

The Medicaid records include payment and enrollment information (distinguishing fee-for-service and HMO) for every year between 1999 and 2012. For those using Fee-For-Service Medicaid, we observe Medicaid payments to providers. For those using Medicaid HMOs, we observe the capitation payment Medicaid paid the HMO. Appendix B describes the Medicare and Medicaid data in more detail.

Appendix B describes the Medicare and Medicaid records in more detail.

[5] While Fang et al. (2008) used a pre-publication version of those estimates we use the published estimates; these incorporate more waves of information and as a result provide a smoother (arguably, more realistic) distribution of risk aversion for the relevant HRS respondents.

[6] The Medicare records contain information on the respondent's insurance liabilities (e.g. co-pays, deductibles and co-insurance) due to Medicare Parts A and B that must be paid either out of pocket or by a third party. In Section 4.5, we discuss how we use this information, along with estimates of payments made by third parties (e.g. Medigap, other smaller public and private payors), to infer the amount of medical care that is defaulted upon.

4.3 The MEPS

To impute medical payments not recorded in our main HRS dataset, we use data from the 1996-2017 waves of the MEPS. The MEPS is a nationally representative survey of non-institutionalized households. MEPS respondents are interviewed up to 5 times over a 2 year period, forming short panels. We aggregate the data to an annual frequency. MEPS respondents are asked about their (and their spouses') demographic information, health status, health insurance, and the health care expenditures paid out-of-pocket, by Medicaid, by Medicare, private insurance, and by other smaller public and private sources. The survey responses are matched to medical spending information provided by health care providers.

Although the MEPS does not capture certain types of medical expenditures, such as nursing home expenditures, comparing the MEPS data to the aggregate statistics shows that MEPS captures most types of spending very well (Pashchenko and Porapakkarm, 2016).

4.4 Sample Selection

For our analysis, we use the data covering 1999 to 2012, the years for which we have information about every medical spending payor. In order to focus on Medigap, we exclude individuals younger than 65. Also, we exclude individuals who do not consent to provide their Medicare and Medicaid information or are otherwise unable to be matched with administrative records.

We restrict our main sample to people who are covered by basic fee-for-service 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 try to limit the sample to people who would be charged more than a nominal premium for supplemental coverage. We also exclude those covered by employer-provided insurance and who are still working, as their employer-provided insurance will be the primary payor and not Medicare. This leaves us with 37,110 person-year observations for the full sample. Appendix C documents our sample selection criteria in greater detail.

In addition to our main sample, we construct an alternative sample consisting of people who currently receive Medicaid. As we discuss below, Medigap enrollment rates are much lower for Medicaid recipients, probably because Medicaid likely crowds out Medigap purchases. We also construct a sample of people who are enrolled in Medicare Part C as a robustness check. Again, Medigap enrolment is very low as Medigap is not intended to cover part C co-payments.

4.5 The Medical Spending Measure

Total medical spending is the sum of payments made by Medicare, Medicaid, private payors, other smaller payors (e.g., state or local health departments), out-of-pocket spending and care that was defaulted upon. To the best of our knowledge, this is the first paper that includes all types of coverage over a nationally-representative long panel.[7] Spending by all payors is measured on an annual basis, except for out-of-pocket expenses, which are biennial quantities that we annualize.

To calculate total Medicare (Part A, B and D) and Medicaid payments, we sum over all the categories of spending that are available in the administrative data. As payments made by Medicare Part C are missing from the HRS, we impute them using conditional mean matching, which is two-step distribution preserving procedure. First, for the subset of individuals receiving Part C, we use the MEPS data to regress these payments on a set of observable characteristics available in both datasets. These variables include household income, a fourth order age polynomial, employment, education, marital status, doctor and hospital visits, race indicators, health measures, out-of-pocket spending and interactions.[8] Second, we apply the MEPS regression coefficients to the HRS data and obtain predicted values for each HRS respondent. To this predicted level of medical spending we add spending residuals drawn from MEPS respondents with a similar levels of predicted spending. We describe our approach in more detail in Appendix D.

We use the same conditional mean matching method to impute payments made by private insurers and other smaller payors, which are also missing from the HRS.[9]

To calculate the amount of medical charges that are defaulted upon, we begin with the respondent's Medicare liability amount, which is the amount of co-pays for Medicare Part A and Part B. From this we subtract the categories of out-of-pocket spending that fall under Medicare Parts A and B,[10] and the imputed payments made by private payors (including Medigap) and other smaller payors. Any remaining liability is the total default amount.

In Appendix E we show that the distribution of medical spending in our data matches up well against the MCBS. French et al. (2017) find that out-of-pocket medical spending in the HRS,

[7] Because we are imputing insurance payments from MEPS, we are missing private payments for long-term care. We believe our measure captures all other payors and types of care.

[8] In both the HRS and the MEPS, we assume that private insurance payments exist only for those who report holding private insurance.

[9] We impute payments from private insurance and other payors separately and set private payments to 0 for those without private insurance. We use the same covariates as the ones we use to impute Part C.

[10] In the HRS, out of pocket medical spending is the sum of co-pays and deductibles for multiple types of care. We include payments to hospitals, outpatient surgery, doctor visits, home based health care, and other health care costs since these are the types of care covered by Medicare Parts A and B. We exclude out of pocket payments for nursing homes, dental care, and drug costs since these tend to be the types of care not covered by Medicare Parts A and B.

MCBS, and MEPS also match up well. We extend their exercise by comparing Medicare and Medicaid payments in the HRS restricted data to Medicare and Medicaid payments in the MCBS.

4.5.1 Descriptive statistics

The first three columns of Table 1 show means and standard deviations (for non-categorical variables) for our main sample. Column (1) shows that of roughly 37,100 person-year observations, 21,200, or 57% purchase Medigap.

Columns (2) and (3) show sample means for those not purchasing and purchasing Medigap, respectively. Those without Medigap are less educated, less likely to be married, and much more likely to be Black or Hispanic. Those without Medigap are slightly more likely to receive Medicaid in the next two years, 4% vs. 1%, and are 3 percentage points more likely to default on medical payments in the next two years. Consistent with this difference, the second and third columns show that Medigap purchasers are signficantly richer, holding an additional $320,000 of wealth and receiving an additional $6,600 of annual income.

Individuals purchasing Medigap spend roughly $12,200 per year on medical services, or $2,300 more than those not purchasing Medigap. Interestingly, there is very little difference in Medicare or out-of-pocket payments between those with and without Medigap. The key difference between the two groups is that those who purchase Medigap receive more in the way of private insurer payments. Given that many of the private insurance policies held by older individuals are in fact Medigap policies, it is hardly surprising that Medigap holders have higher private insurance payments.

There are at least three potential reasons why those with Medigap might have higher total medical spending than those without. First, there could be adverse selection in this market, with less healthy individuals being more likely to purchase Medigap. As we show below, however, Medigap purchasers, if anything, appear to be healthier. Second, there could be moral hazard. Because those who have purchased insurance pay less out of pocket for a given amount of medical care, they may be purchasing more overall care. Third, they may also be facing lower prices because they are more likely to default or qualify for Medicaid, allowing them to avoid payment. However, note that those covered by Medigap are *less* likely to default or qualify for Medicaid. Fourth, it could be that those who purchase Medigap consume more care because they better understand the value of greater healthcare due to higher cognitive capacity or because they are more forward looking.

The higher likelihood of future default and future Medicaid receipt among those those not receiving Medigap suggests that these forms of implicit insurance may be crowding out Medigap. To further explore these differences, columns (4) - (6) of Table 1 provide summary statistics for

Table 1: Summary Statistics by Medigap Status

	A. Main Sample			B. Medicaid Recipients		
	All	No Medigap	Medigap	All	No Medigap	Medigap
	(1)	(2)	(3)	(4)	(5)	(6)
Medigap	0.57	0.00	1.00	0.13	0.00	1.00
Female	0.61	0.58	0.62	0.68	0.68	0.72
Black	0.11	0.21	0.03	0.30	0.33	0.11
Hispanic	0.04	0.09	0.01	0.20	0.22	0.06
Ever married	0.74	0.70	0.76	0.50	0.50	0.49
Currently married	0.54	0.50	0.58	0.28	0.29	0.17
Widower	0.19	0.21	0.18	0.22	0.21	0.31
Less than High School	0.29	0.37	0.23	0.65	0.66	0.52
High School	0.37	0.32	0.40	0.23	0.22	0.34
Some college	0.19	0.16	0.20	0.06	0.06	0.09
College	0.15	0.14	0.17	0.03	0.02	0.04
Working	0.17	0.17	0.17	0.07	0.07	0.04
Future Medicaid eligibility	0.02	0.04	0.01	0.80	0.80	0.78
Future default	0.32	0.34	0.31	0.53	0.53	0.49
Age	76.09	75.90	76.23	76.73	76.12	80.85
	(7.66)	(7.96)	(7.42)	(8.32)	(8.20)	(7.96)
Household wealth ($1,000s)	591.7	407.1	727.9	40.5	37.1	63.2
	(1,423.1)	(1,088.0)	(1,613.2)	(97.7)	(82.8)	(162.7)
Household income	22,522	18,739	25,359	6,835	6,334	10,235
	(77,184)	(87,682)	(68,127)	(10,771)	(8,046)	(21,253)
Total medical spending	11,226	9,911	12,211	25,663	24,591	32,925
	(19,659)	(17,697)	(20,957)	(34,995)	(33,748)	(41,866)
Medicare spending	6,303	6,085	6,467	13,107	12,598	16,552
	(12,980)	(13,154)	(12,846)	(21,801)	(21,654)	(22,516)
Medicaid spending	1	2	0	7,770	7,936	6,645
	(26)	(33)	(2)	(16,104)	(15,795)	(18,054)
Out-of-pocket spending	3,020	3,034	3,009	2,942	2,628	5,072
	(9,398)	(8,528)	(10,000)	(10,998)	(10,208)	(15,171)
Private insurer spending	1,307	126	2,192	449	21	3,351
	(5,911)	(1,202)	(7,630)	(2,731)	(230)	(6,939)
Part C payments	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Medicare liability	1,163	1,061	1,241	1,988	1,867	2,809
	(1,852)	(1,864)	(1,840)	(3,264)	(3,102)	(4,113)
Total default amount	596	666	544	1,394	1,408	1,304
	(1,497)	(1,632)	(1,386)	(2,861)	(2,761)	(3,463)
Observations	37,110	15,901	21,209	2,014	1,755	259

Notes: Expenditures expressed in 2014 dollars. Standard errors for non-categorical variables are in parentheses; other quantities are averages. Medicare spending is the sume of expenditures for Parts A, B, and D. Private insurance excludes Medicare Part C and Medicaid. Total Medical Expenditure equals sum of Medicare, Medicaid, out-of-pocket, Private insurance, and Part C payments, and default amounts. Future Medicaid eligibility measures whether the individual, who is currently not enrolled in Medicaid, will be enrolled within the next two years. Future default is a measure for whether an individual will default on medical payments within the next two years.

Medicaid recipients. Given that 13% of Medicaid recipients purchase Medigap insurance, as opposed to 57% in the main sample, Medicaid-related crowd-out seems very likely. As in the main sample, Medigap purchasers are less likely to be Black or Hispanic and have higher wealth and income. Once again, Medigap users have higher total medical spending, but in contrast to the main sample, less than half of the gap, $3,300 out of $8,300, is attributable to private insurer payments; in the Medicaid sample, Medigap users have significantly higher Medicare and out-of-pocket payments. These differences suggest that Medigap selection dynamics differ markedly between those receiving and not receiving Medicaid, and therefore in what follows we focus only on our main sample and exclude those currently receiving Medicaid.

4.5.2 Health differences

Table 1 shows that those with Medigap have approximately $2,300 more in total medical expenditures per year. In Table 2, we investigate whether this difference in spending is driven by differences in health. Put differently, we look for evidence of whether there is adverse selection in the Medigap market. If there is adverse selection, then those who are less healthy are more likely to purchase insurance. An attractive aspect of the HRS is that it has many measures of observed health to use in such a test. Columns 1 and 2 of Table 2 present some selected health measures both for those with Medigap and for those without. The third column of the table presents the difference in means between the two groups, and the fourth displays the p-values associated with the difference. Table 2 shows that those purchasing Medigap are healthier along a number of dimensions. For example, they are less likely to have diabetes or high blood pressure, are less likely to have difficulties dressing or walking, and have higher cognition scores. Differences in health therefore appear unlikely to explain differences in insurance purchases, a result we explain in greater detail in the next section.

Table 2: Health Statistics, by Medigap Status

	No Medigap (1)	Medigap (2)	Difference (3)	p-value (4)
Self reported health: Excellent	0.075	0.091	-0.016	0.00
Smoker	0.112	0.069	0.043	0.00
Diabetes	0.232	0.170	0.062	0.00
High blood pressure	0.636	0.611	0.025	0.00
Stroke	0.104	0.079	0.025	0.00
Difficulty dressing	0.137	0.101	0.036	0.00
Difficulty walking	0.112	0.089	0.022	0.00
Cognition: TICS score	8.85	9.36	-0.50	0.00

Notes: This table shows mean values for select health variables for those with no Medigap insurance and those with Medigap, along with differences in the means.

5 Estimation Results

The comparisons in Table 2 show that across a variety of measures, those with Medigap insurance are in better health than those without Medigap. This suggests that health cannot explain the higher medical spending of those covered by Medigap. In fact, given that those who purchase Medigap are healthier, it suggests that, holding health status constant, the gap in medical spending between those with and without insurance would be larger than the gap reported in Table 1. To show this more formally, we turn to regression analysis. If those who purchase insurance are healthier (causing them to have lower medical spending, holding all else equal), as suggested by Table 2, then controlling for observed health should increase the gap in spending.

Column (1) of Table 3 presents estimates from a regression of total medical spending on a Medigap indicator, gender, an age polynomial, regional dummies, and a set of year dummies. Because in any given year, insurers can price Medigap policies based on age, gender, state of residence and smoking status, we control for these covariates. After conditioning on these variables, the spending gap between those with and without Medigap is $1,200. This gap is smaller than the gap reported in Table 1, showing that these variables do have predictive power for medical spending. Columns (2) and (3) present the coefficient estimates from estimating the Column (1)-regression separately for men and women, and show that the gap is $1,900 for men and $800 for women. Panel B shows results from the same regression as in Panel A, but includes numerous health controls. Comparing columns (1) and (4) reveals that adding the health controls doubles the gap in total spending, from $1,200 to $2,400. In other words, it is healthier people who are purchasing

Medigap, a tendency commonly referred to as "advantageous selection".

Table 3: OLS Results of Total Medical Expenditure on Medigap Status

	A. Without Health Controls			B. With Direct Health Controls		
	All	Female	Male	All	Female	Male
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Medigap	1,209***	786**	1,903***	2,425***	2,107***	3,347***
	(287)	(369)	(434)	(268)	(340)	(422)
Female	355	0	0	341	0	0
	(288)	(.)	(.)	(273)	(.)	(.)
Age	323**	112	598***	269**	151	424***
	(144)	(189)	(178)	(124)	(167)	(164)
Age-squared	-3.80	11.46	-22.25	-15.61	-4.37	-33.15**
	(12.86)	(17.25)	(13.63)	(11.20)	(15.32)	(13.25)
Age-cubed	0.21	-0.04	0.43	0.18	-0.10	0.64**
	(0.32)	(0.42)	(0.30)	(0.28)	(0.38)	(0.30)
Observations	37,931	23,253	14,678	36,852	22,685	14,167
Adjusted R^2	0.0210	0.0255	0.0150	0.207	0.220	0.200

Notes: This table presents estimates from a regression of total medical spending on a Medigap indicator. Panel A controls for gender, an age polynomial, census region, smoker status, and a set of year dummies. Panel B additionally controls for numerous health controls (i.e., self-reported health measures; change in self-reported health meaures; any diagnoses; any treatments; difficulties with ADLs/IADLs; and whether they get help for ADLs/IADLs). For the full list of health variables see Appendix Table A. Standard errors in parentheses are heteroskedasticity robust. * p<0.1, ** p<0.05, *** p<0.01.

Table 4 provides estimates of the relationship between Medigap and medical spending, controlling for additional determinants of Medigap coverage. It presents estimates from a regression of Medigap coverage on total medical expenditure and a number of covariates that have previously been suggested to be important in explaining Medigap purchases (e.g., Fang et al. 2008). Columns (1)-(3) present the estimated coefficients on total medical spending for different subsamples. Column (1) presents estimates from our full sample. Column (2) presents estimates for the sample of respondents with observed values of our risk tolerance measure, predicted variance of medical spending, education and income. Only a subsample of HRS sample members were asked about risk preferences, leading to a significant loss of observations when restricting the sample to respondents with these variables. Column (3) presents estimates for the sample that, in addition to the previous variables, also has measures of cognition, longevity expectations, planning horizon, future Medicaid eligibility, and future medical spending default.

Our goal is to examine which variables, if any, can explain the higher medical spending of those

Table 4: OLS Results of Medigap Status on Total Medical Expenditures

	Coeffient Estimates for Expenditures$_j$/10,000			Conditioning Variables										
	Full Sample (1)	First Subsample (2)	Second Subsample (3)	Risk Tolerance (4)	$\widehat{Var_j}$ (5)	Risk Tolerance $\times \widehat{Var_j}$ (6)	Education (7)	Income (8)	Cogn. (9)	Life Expec-tancy (10)	Planning Horizon (11)	Future Mdcd (12)	Future Default (13)	Obs (14)
A. First Imputation Method: Using Observations With No Medigap														
1	0.0007*** (0.0002)	0.00099*** (0.0002)	0.0008*** (0.0002)	N	N	N	N	N	N	N	N	N	N	37,913
2	...	0.00098*** (0.0002)	0.0008*** (0.0003)	Y	N	N	N	N	N	N	N	N	N	17,178
3	...	0.0027*** (0.0003)	0.0025*** (0.0003)	Y	Y	Y	N	N	N	N	N	N	N	17,178
4	...	0.0024*** (0.0002)	0.0022*** (0.0003)	Y	Y	Y	Y	N	N	N	N	N	N	17,178
5	...	0.0023*** (0.0002)	0.0021*** (0.0003)	Y	Y	Y	Y	Y	N	N	N	N	N	17,178
6	...	...	0.0020*** (0.0001)	Y	Y	Y	Y	Y	Y	N	N	N	N	14,526
7	...	...	0.0020*** (0.0002)	Y	Y	Y	Y	Y	Y	Y	N	N	N	14,526
8	...	...	0.0020*** (0.0002)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	14,526
9	...	...	0.0021*** (0.0002)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	14,526
10	...	...	0.0021*** (0.0002)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	14,526

Notes: This table presents estimates from a regression of Medigap coverage on total medical spending. All regressions control for gender, an age polynomial, census region, smoker stuatus, and a set of year dummies. Additional variables listed in each column are added as controls sequentially. Subsample 1 consists of observations where the variables listed in columns (4)-(8) are all observed. Subsample 2 consists of observations where the variables listed in columns (4)-(13) are all observed. Standard errors in parentheses are heteroskedasticity robust. * p<0.1, ** p<0.05, *** p<0.01.

who purchase Medigap. Column (1), row 1 of the table shows that for every $10,000 in medical spending, the probability of purchasing Medigap rises by 0.07 percentage points. Column (2), row 1 shows that the coefficient rises modestly to 0.10 percentage points when restricting the sample to respondents that have measures of risk tolerance, the predicted variance of medical spending, education and income. Column (3), row 1 shows that the estimated coefficient falls to 0.08 percentage points when using the sample is further restricted to respondents who have measures of cognition, longevity expectations, planning horizon, and future Medicaid eligibility and medical spending default. Restricting the sample therefore appears to not significantly impact the estimated effect of medical spending.

The remaining rows of the table add covariates to the regression. Row 2 shows that adding the risk tolerance measure has little impact on the estimated coefficient. Row 3 adds in the predicted variance of future medical spending and its interaction with risk tolerance. It shows that these variables do not help explain the higher medical spending of those who purchase Medigap. In fact, controlling for these variables increases the positive relationship between medical expenditures and Medigap: the coefficient roughly triples. Rows 4 and 5 show that adding education and income partly explains the higher medical spending of those purchasing Medigap, but also that the reduction in the coefficient on expenditures is small, equalling 0.03-0.04 percentage points. Rows 6-8 show that cognition, life expectancy, and planning horizon, all variables that were emphasized in Fang et al. (2008), have little explanatory power. While those with Medigap coverage do score higher on cognitive measures, adding these variables has little effect on the estimated coefficient. Finally, rows 9 and 10 show that future Medicaid receipt and and default have little explanatory power for the gap, suggesting that for people in our sample, the potential receipt of future care from these sources does not significantly crowd out Medigap coverage.

6 Conclusion

Using high quality administrative medical spending data linked to the HRS, we rule out several explanations for why some individuals purchase Medigap and others do not. In Tables 2 and 3 we show that those with Medigap coverage are healthier than those without coverage, rejecting the hypothesis of adverse selection in this market. Instead, the evidence is consistent with the hypothesis of advantageous selection. In Table 4 we investigate other non-health-related sources of selection into Medigap. We show that education explains a small amount of the higher medical spending of those with Medigap coverage. However, we find that many factors previously reported to be important, such as risk aversion and cognition, are not important. We also examine whether

crowd out provided by future Medicaid receipt and the potential to default on future medical debt plays a role, but again find that these factors do not appear to be very important.

This raises the question of why those with Medigap have higher medical spending. A key potential mechanism is moral hazard. Those with Medigap face lower out-of-pocket prices for additional medical care, which may increase demand. We cannot directly measure moral hazard. However, as we have exhausted most of the other candidate explanations, this suggests that prices are likely important in determining care choices.

References

Braun, R. A., Kopecky, K. A., and Koreshkova, T. (2019). Old, frail, and uninsured: Accounting for features of the us long-term care insurance market. *Econometrica*, 87(3):981–1019.

Brown, J. and Finkelstein, A. (2008). The interaction of public and private insurance: Medicaid and the long term care insurance market. *American Economic Review*, 98(5):837–880.

Cardon, J. H. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *RAND Journal of Economics*, pages 408–427.

Chiappori, P.-A. and Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1):56–78.

Cohen, A. and Siegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and insurance*, 77(1):39–84.

Dardanoni, V. and Li Donni, P. (2012). Incentive and selection effects of Medigap insurance on inpatient care. *Journal of Health Economics*, 31(3):457–470.

David, M., Little, R. J. A., Samuhel, M. E., and Triest, R. K. (1986). Alternative methods for CPS income imputation. *Journal of the American Statistical Association*, 81(393):29–41.

De Nardi, M., French, E., and Jones, J. B. (2016a). Medicaid insurance in old age. *American Economic Review*, 106(11):3480–3520.

De Nardi, M., French, E., Jones, J. B., and Gooptu, A. (2012). Medicaid and the elderly. 36(1):17–34. Economic Perspectives.

De Nardi, M., French, E., Jones, J. B., and McCauley, J. (2016b). Medical spending on the US elderly. *Fiscal Studies*, 37(3-4):717–747.

De Nardi, M., French, E., Jones, J. B., and McGee, R. (2021). Why do couples and singles save during retirement? Technical report, National Bureau of Economic Research.

Ettner, S. L. (1997). Adverse selection and the purchase of Medigap insurance by the elderly. *Journal of Health Economics*, 16(5):543–562.

Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of advantageous selection: Evidence from the Medigap insurance market. *Journal of Political Economy*, 116(2):303–350.

French, E. and Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732.

French, E., Jones, J. B., and McCauley, J. (2017). The accuracy of economic measurement in the Health and Retirement Study. *Forum for Health Economics and Policy*, 20(2).

Hurd, M. D. and McGarry, K. (1997). Medical insurance and the use of health care services by the elderly. *Journal of health economics*, 16(2):129–154.

Jacobson, G., Casillas, G., Damico, A., Neuman, T., and Gold, M. (2016). Medicare Advantage 2016 spotlight: Enrollment market update. *The Henry J. Kaiser Family Foundation Website*.

Jacobson, G., Huang, J., and Neuman, T. (2014). Medigap reform: Setting the context for understanding recent proposals. *The Henry J. Kaiser Family Foundation Website*.

Keane, M. and Stavrunova, O. (2016). Adverse selection, moral hazard and the demand for Medigap insurance. *Journal of Econometrics*, 190(1):62–78.

Khandker, R. K. and McCormack, L. A. (1999). Medicare spending by beneficiaries with various types of supplemental insurance. *Medical Care Research and Review*, 56(2):137–155.

Kimball, M. S., Sahm, C. R., and Shapiro, M. D. (2008). Imputing risk tolerance from survey responses. *Journal of the American Statistical Association*, 103(483):1028–1038.

Kirchhoff, S. M. (2020). *Medicare Part D Prescription Drug Benefit*. Congressional Research Service Washington (DC).

Koijen, R. S., Van Nieuwerburgh, S., and Yogo, M. (2016). Health and mortality delta: Assessing the welfare cost of household insurance choice. *The Journal of Finance*, 71(2):957–1010.

Laitner, J., Silverman, D., and Stolyarov, D. (2018). The role of annuitized wealth in post-retirement behavior. *American Economic Journal: Macroeconomics*, 10(3):71–117.

Mahoney, N. (2015). Bankruptcy as implicit health insurance. *American Economic Review*, 105(2):710–46.

Nicholas, L. H., Langa, K. M., Iwashyna, T. J., and Weir, D. R. (2011). Regional variation in the association between advance directives and end-of-life Medicare expenditures. *JAMA*, 306(13):1447–1453.

Pashchenko, S. and Porapakkarm, P. (2016). Medical spending in the US: facts from the Medical Expenditure Panel Survey data set. *Fiscal Studies*, 37(3-4):689–716.

U.S. Centers for Medicare and Medicaid Services (2022). Understanding medicare advantage plans.

U.S. Centers for Medicare and Medicaid Services (2023). Supplements & other insurance.

Wolfe, J. R. and Goddeeris, J. H. (1991). Adverse selection, moral hazard, and wealth effects in the medigap insurance market. *Journal of Health Economics*, 10(4):433–459.

Appendix A: Table Describing HRS Survey Variables

Table A.1: HRS variables

Variable	Description
A. Health expenditure:	
Total expenditure	Total annual health care expenditure for 12 months of the survey year. Sum of Medicare Parts A, B, D payments, Medicaid, survey responses for out-of-pocket expenditures, and imputed other payments
B. Insurance:	
Medigap	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
C. Demographics:	
Race	Indicators for self-reported black, other, and nonresponse
Hispanic	Indicators for self-reported Hispanic and nonresponse
Education	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	Indicators for married, widowed, divorced, separated, and nonresponse
Number of children	The number of children the respondent has ever had
Income	Indicators for self-reported total household income in \$5,000 intervals from \$5,000 to \$50,000 and \$50,000 plus
Work status	Indicators if currently working for pay and for nonresponse
D. Health:	
Height	Self-reported height, in inches, and height squared
Body mass index	Self-reported [weight (kg)]/[height $(m)^2$]
Ever a smoker	Indicator if respondent has "ever smoked" tobacco
Current smoker	Indicator if respondent now smokes tobacco and for nonresponse

Variable	Description
D. Health (continued):	
Diagnoses	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	Indicators for respondent ever having cataract surgery or a hearing aid
(Instrumental) ADLs	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	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
E. Cognition:	
Word recall	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	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	Number of times respondent can subtract the number 7 sequentially, starting from 100
Numeracy	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
F. Expectations:	
Longevity	Most recent answer to the question "What is the percent chance you will live to 75 or more?"
G. Planning horizon:	
Financial	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
H. Risk attitudes:	
Risk tolerance	Estimates of risk tolerance from Kimball et al. (2008), using 1992 and 1992 responses to hypothetical income gambles

Appendix B: Our Medicare and Medicaid Data

Medicare

We link restricted Medicare fee-for-service (Parts A and B), and Part D data for the years 1999-2012 (2006 was the first year of Medicare Part D and thus our Part D data begins then) to our HRS survey data for respondents who consent to allow their Medicare data to be linked to their survey responses (approximately 64.7% percent of persons in our study population). These records have enrollment information and data on reimbursement amounts for inpatient, skilled nursing facility, home health, and hospice claims (Medicare Part A), as well as outpatient, carrier (non-institutional medical care providers such as individual or group practitioners, non-hospital labs, and ambulances), and durable medical equipment claims for Medicare Part B.

We use the Beneficiary Annual Summary File (BASF), which summarizes information from the micro-level claims records. The BASF contains annual information for each individual on the number of months of enrollment in Medicare Part A, Part B, and non-fee-for-service plans. The BASF has information on Medicare fee-for-service (FFS) claims. Almost all claims for services used by non-FFS Medicare patients are not observed in these data, so all analyses exclude an individual in a given year if they were enrolled in a non-FFS Medicare plan for more than half the year.

Medicare Part D is the prescription drug benefit. We calculate the Medicare Part D payment using the Part D event files. For the Part D contribution we subtract from the gross drug cost the payments paid by the beneficiary, family, or friends the drug costs at point of sale over the whole year.

Medicaid

As with the Medicare data, we are able to link restricted Medicaid data (CMS Medicaid Analytic eXtract, or "MAX" files) for those in the HRS who gave permission, allowing us to measure Medicaid expenditures for the Medicaid beneficiaries in our dataset for the years 1999-2012. The MAX files contain personal summaries (which contain eligibility, enrollment, and demographic information) and claims data across four service categories (inpatient, long-term care, prescription drugs, and other services). Other services include a variety of services (e.g., physician services and lab work) that do not fit under the other three service categories. The inpatient, long-term care, prescription drugs, and other services files contain the primary variable of interest, "Medicaid Payment Amount," which is the total amount of money paid by Medicaid for a particular service. We sum over all the claims for all the different service categories for a particular individual in each year.

Appendix C: Sample Selection

Our initial sample comprises of 42,240 individuals. We drop those that are not merged to the administrative data or are under 65 years old. We drop those that are covered by employer-provided insurance, Medicare Part-C and government insurance. We drop the person-year observations that the Medigap coverage status is missing or are on Medicaid. Table C.1 below denotes the HRS sample size after every drop. Our final sample size is 9,231 unique individuals or 37,110 person-year observations.

Table C.1: Sample Selection

Selection criteria (reason for dropping)	Remaining sample size
Not merged to the administrative data	25,681
Under 65 years old	24,510
Covered by employer or government insurance	23,944
Enrolled in Medicare Part-C	23,437
Medigap coverage stats is missing	9,729
Enrolled in Medicaid	9,231

Notes: Remaining sample size refers to the number of unique individuals after the sample selection criteria.

Appendix D: Imputing Missing Medical Expenditures

Our goal is to measure all medical spending, including out-of-pocket spending, Medicare and Medicaid payments, and private (including Medigap) and other public (such as Veterans Administration benefits, and care provided by local and state health departments) insurance payments. While the HRS includes information on out-of-pocket spending and can be linked to Medicare and Medicaid payments, it does not include private, or other public insurance payments. In this appendix we describe how we use data from the Medical Expenditure Panel Survey (MEPS) to impute these payments in the HRS. Although the MEPS has extremely high quality information on all payors for all household members, it lacks the long-panel dimension of the HRS and does not have many of the variables available in the HRS. Our imputation procedures allows us to exploit the best of both datasets.

Our imputation procedure has two steps. First, we use the MEPS to infer private and other public insurance payments, conditional on variables that are observed in both datasets. Second, we impute private and other public insurance payments in the HRS data using a conditional mean matching procedure (which is a procedure very similar to hot-decking).

First Step of Imputation Procedure

We use the MEPS to infer payments of other payors, conditional on the observable variables that exist in both the MEPS and the HRS datasets.

Let i index individuals in the HRS and j index individuals in the MEPS. Define M_{it}^{obs} as out of pocket, Medicaid, and Medicare (Part A, B, and D, but not Part C) payments which are observed in both the HRS and MEPS datasets, M_{it}^{miss} as the components of medical spending that are missing in the HRS but observed in the MEPS, and $M_{it} = M_{it}^{miss} + M_{it}^{obs}$ as total medical spending. To impute M_{it}^{miss}, which is missing in the HRS, we follow David et al. (1986), French and Jones (2011), and De Nardi et al. (2021) and use a predictive mean-matching regression approach. There are two steps to our procedure. First, we use the MEPS data to regress M_{it}^{miss} on observable variables that exist in both datasets. This regression has an R^2 statistic of 0.15 for private insurance payments and 0.18 for other payors. Second, we impute M_{it}^{miss} in the HRS data using a conditional mean-matching procedure, a procedure very similar to hot-decking.

First, for every member of the the MEPS sample, we regress the variable of interest M^{miss} on the vector of observable variables z_{jt}, yielding $M_{jt}^{miss} = z_{jt}\beta + \varepsilon_{jt}$. Second, for each individual j in the MEPS we calculate the predicted value $\widehat{M^{miss}}_{jt} = z_{jt}\hat{\beta}$, and for each member of the sample we calculate the residual $\widehat{\varepsilon}_{jt} = M_{jt}^{miss} - \widehat{M^{miss}}_{jt}$. Third, we sort the predicted value $\widehat{M^{miss}}_{jt}$ into

deciles and keep track of all values of $\hat{\varepsilon}_{jt}$ within each decile. We use this procedure separately to impute private payments, and other payments.

In practice we include in z_{jt} a fourth-order age polynomial, marital status, gender, self-reported health (=1 if self reported health is good, very good, or excellent), race, visiting a medical practitioner (doctor, hospital or dentist), out-of-pocket medical spending, education of head (high school, some college, college), death of an individual, and total household income. We estimate this regression two times: once for the privately insured, and once for other payors.

Because the measure of medical spending in the HRS is medical spending over two years, we divide HRS out-of-pocket medical spending by 2 and assume that medical spending is equal across the two years.

Second Step of Imputation Procedure

For every observation in the HRS sample with a positive Medicaid indicator, we impute $\widehat{Med}_{it} = z_{it}\widehat{\beta}$, using the values of $\widehat{\beta}$ estimated from the MEPS. Then we impute ε_{it} for each observation of this subsample by finding a random observation in the MEPS with a value of $\widehat{Med}_{jt}$ in the same decile as $\widehat{Med}_{it}$, and setting $\hat{\varepsilon}_{it} = \hat{\varepsilon}_{jt}$. The imputed value of Med_{it} is $\widehat{Med}_{it} + \hat{\varepsilon}_{it}$.

As David et al. (1986) point out, our imputation approach is equivalent to hot-decking when the "z" variables are discretized and include a full set of interactions. The advantages of our approach over hot-decking are two-fold. First, many of the "z" variables are continuous. Second, to improve fit we use a large number of "z" variables. We find that adding extra variables is very important for improving fit when imputing payments. Because hot-decking uses a full set of interactions, this would result in a large number of hot-decking cells relative to our sample size. Thus, in this context, hot-decking is too data intensive.

Appendix E: Validating the Administrative Medical Spending Data

Here, we examine in greater detail the accuracy of the administrative medical spending data, as well as the out-of-pocket spending found in the AHEAD cohort of the HRS, comparing them to data from the Medicare Current Beneficiary Survey (MCBS) and Medical Expenditure Panel Survey (MEPS). See De Nardi et al. (2016a) and De Nardi et al. (2016b) for more details of the MCBS data and for example Nicholas et al. (2011) for details of the HRS linked data.

The MCBS is a nationally representative survey of Medicare beneficiaries, consisting of Disability Insurance recipients and Medicare recipients aged 65 and older. The survey contains an over-sample of beneficiaries older than 80 and disabled individuals younger than 65. Respondents are asked about health status, health insurance, and health care spending (from all sources). The MCBS data are matched to Medicare records, and medical spending data are created through a reconciliation process that combines information from survey respondents with Medicare administrative files. As a result, the survey is thought to give extremely accurate data on Medicare payments and fairly accurate data on out-of-pocket and Medicaid payments. As in the HRS survey, the MCBS survey includes information on those who enter a nursing home or die. Respondents are interviewed up to 12 times over a 4 year period. We aggregate the data to an annual level. In both samples, we applied only modest sample selection restrictions. The key sample selection issue shown in Table C.1 is that in the HRS we drop households with missing or erroneous Medicare or Medicaid records.

Table E.1 compares individual level distributions of total, out-of-pocket, Medicare, and Medicaid payments between the MCBS and the HRS data. Medical spending is higher in our HRS sample than in the MCBS sample. Furthermore, this higher level of spending is driven by higher out-of-pocket spending, Medicare, and Medicaid spending. These differences potentially are an advantage of the HRS data since, as noted in De Nardi et al. (2016b), the MCBS clearly understates aggregate Medicare and especially Medicaid spending, potentially due to the issue that the MCBS does not have administrative data on Medicaid spending, and thus relies heavily on imputation.

The next set of benchmarking exercises that we perform is for out-of-pocket medical spending, Medicaid recipiency and income between the AHEAD cohort of the HRS and MCBS. For both the HRS and MCBS, we restrict the sample to singles (over the sample period) who meet the HRS/AHEAD age criteria (at least 70 in 1994, 72 in 1996, ...) and who are not working over the sample period. Because the MCBS sample lacks spousal information, for this analysis we focus only on singles. We use the De Nardi et al. (2016a) measure of permanent income and construct a measure of permanent income that is the percentile rank of total income over the period we observe

Table E.1: Individual Medical Spending Percentiles: HRS versus MCBS

	Total Spending				OOP			
	HRS		MCBS		HRS		MCBS	
Total Spending Percentiles	Average Expenditure	Percentage of Total	Average Expenditure	Percentage of Total	Average Expenditure	Percentage of Total	Average Expenditure	Percentage of Total
All	17,091	100	14,120	100	3,825	100	2,740	100
95-100%	114,238	33.4	97,880	34.6	45,643	59.6	26,930	49.1
90-95%	59,000	17.3	48,890	17.3	8,619	11.3	6,700	12.2
70-90%	26,870	31.4	20,540	29.1	3,480	18.2	2,920	21.3
50-70%	9,025	10.6	7,750	11	1,394	7.3	1,360	9.9
0-50%	2,502	7.3	2,250	8	178	3.6	420	7.6

	Medicare				Medicaid			
	HRS		MCBS		HRS		MCBS	
Total Spending Percentiles	Average Expenditure	Percentage of Total	Average Expenditure	Percentage of Total	Average Expenditure	Percentage of Total	Average Expenditure	Percentage of Total
All	11,343	100	7,720	100	1,896	100	1,320	100
95-100%	85,268	37.6	67,560	43.7	33,773	89.1	24,980	94.7
90-95%	41,731	18.4	28,370	18.4	4,092	10.8	1,360	5.2
70-90%	17,251	30.4	10,280	26.6	230	0.2	10	0.1
50-70%	5,031	8.9	2,980	7.7	0	0	0	0
0-50%	1,076	4.7	550	3.5	0	0	0	0

Table E.2: Income, Out-of-pocket Spending, and Medicaid Recipiency Rates: HRS versus MCBS

	HRS/AHEAD Data				MCBS Data		
Income Quintile	Total Income	Annuity Income	Out-of-pocket Expenses	Medicaid Recipiency	Total Income	Out-of-pocket Expenses	Medicaid Recipiency
Top	33,580	26,300	7,000	3.0	44,150	8,020	5.4
4th	19,290	14,390	6,360	5.6	19,710	7,300	8.0
3rd	15,500	10,900	5,050	11.0	13,740	6,470	15.5
2nd	10,290	8,270	4,270	28.1	10,020	5,340	41.8
Bottom	7,740	4,820	2,550	60.9	6,750	4,050	69.9

Source: Table A.2 of De Nardi et al. (2016a).
Notes: Calendar years 1996-2010, for those age 72 and older in 1996.

these individuals (the MCBS asks only about total income). The first four columns of Table E.2 show sample statistics from the full HRS/AHEAD sample while the final three columns of the table shows sample statistics from the MCBS sample. The first statistics we compare are income. Total income in the HRS/AHEAD data (including asset and other non-annuitized income) lines up well with total income in the MCBS data, although income in the top quintile of the MCBS is higher than in the HRS/AHEAD. Next, we compare out-of-pocket medical spending in the MCBS and HRS/AHEAD. Out-of-pocket medical spending (including insurance payments) averages $2,550 in the bottom PI quintile and $7,000 in the top quintile in the HRS/AHEAD. In comparison, the same numbers in the MCBS data are $4,050 and $8,020. Overall, out-of-pocket medical spending in the MCBS and HRS/AHEAD are similar, which may be surprising given that the two surveys each have their own advantages in terms of survey methodology.[11] The share of the population receiving Medicaid transfers is also very similar in the HRS/AHEAD and MCBS. Sixty-one percent and 70 percent of those in the bottom PI quintile are on Medicaid in the the HRS/AHEAD and MCBS, respectively. In the top quintile, 3% of people are on Medicaid in the HRS/AHEAD whereas 5% are in the MCBS.

[11]There are more detailed questions underlying the out-of-pocket medical expense questions in the HRS, including the use of "unfolding brackets". Respondents can give ranges for medical expense amounts, instead of a point estimate or "don't know" as in the MCBS. The MCBS has the advantage that forgotten medical out-of-pocket medical expenses will be imputed if Medicare had to pay a share of the health event.