Policy and Methods in Health Services Research

Thesis adviser: Professor David Grabowski Portia Y. Cornell

Policy and Methods in Health Services Research

Abstract

This dissertation consists of two policy papers and one methods paper, all grounded in applied, empirical health-services research. The first two papers concern the influence of medical underwriting in the market for long-term care insurance in the U.S. The third presents a framework for evaluating provider preference as an instrumental variable in comparative effectiveness research with multiple treatments, using as an empirical demonstration a study comparing the safety of five atypical antipsychotics and their effects on chronic disease incidence.

In the first paper, I estimate an empirical model of the factors on which firms make decisions to underwrite individuals for long-term care insurance, using data on the health and coverage decisions for applicants at two U.S. firms. I apply the model parameters to a population-based sample to determine what proportion of households of prime ages to purchase long-term care insurance would be able to qualify for policies if they were to apply for coverage. Among the general population, I estimate that 40 percent of individuals would have their applications rejected if they were to apply for long-term care insurance—a rejection rate substantially higher than the rejection rate of 20–25 percent of applicants in the actual market.

The second paper examines policy ramifications of the bounds on coverage in the individual market for long-term care insurance that I establish in paper one. I study two types of policies designed

to encourage individuals to purchase long-term care insurance, tax incentives and state Partnership programs, estimating how the effects of differ with respect to individuals' underwriting probabilities. I exploit variation in the timing of states' implementation of these policies with difference-in-difference models, and estimate the demand elasticity of long-term care premiums using simulated statewide marginal tax prices as in instrumental variable for individual prices. I find that the response to these policy incentives is highly dependent on individuals' underwriting probabilities: conditional on wealth and income, tax and Partnership have no apparent effect on insurance purchase among low-approval households, and program effect appears to be concentrated among healthier individuals with high approval probabilities. In evaluating reforms for long-term care financing and their potential to increase private insurance rates, as well as reduce financial pressure on public safety-net programs, policy makers need to consider the role of underwriting in the market for long-term care insurance.

The third paper proposes a framework for assessing provider preference as instrumental variables in comparative effectiveness research and describes diagnostic tools to validate (or debunk) a candidate instrument. In an applied analysis, I show how to use these tools to compare multiple treatments. I compare the safety of commonly prescribed atypical antipsychotics using physician prescribing preference as an instrumental variable. Widespread adoption of a basic protocol and road map for validating potential instruments, particularly the use of sensitivity tests and compliers analysis, would improve the quality of comparative effectiveness research.

Contents

To the grandparents and elders.

Acknowledgments

First and foremost, I would like to thank my adviser and the members of my committee. David Grabowski, Mary Beth Landrum, and John Hsu gave me intellectual and professional guidance, support, and encouragement, all with good humor, acuity, and patience. I am grateful to the Agency for Healthcare Research and Quality, the National Institute on Aging, the Social Security Administration, and the National Institute of Mental Health for contributing to the funding of these projects and my graduate studies. I would like to thank Vicki Fung, David Stevenson, Marc Cohen, Maggie Price, Jie Huang, and Xiaomei Shi for helping me bring these projects to fruition. Bruce Margolis, Denise Liston, Malcolm Cheung, and Eileen Tell read and gave helpful comments on early drafts. I thank the editors at *Health Affairs* and anonymous reviewers for suggesting improvements to paper one. Participants in the Health Policy Research Seminar provided valuable feedback on all three papers as they progressed.

I was fortunate to matriculate to Harvard with an exceptionally cohesive class in the Health Policy Program. The potluck dinners, study sessions, and celebrations created a community that allowed us to thrive as graduate students. Rebecca Haffajee was my study partner in Evaluative Sciences and Statistics, my co-teaching fellow, and a role model as well as peer. Prachi Sanghavi is not only a dear friend and fellow "sewist" but someone who inspires me with her acumen and passion

for honest, rigorous research. I spent many amicable evenings, particularly in the first three years of graduate school, with Daria Pelech and Nick Leiby, working problem sets and workshopping fellowship proposals. Jacob Wallace, Aaron Schwartz, and Hannah Neprash also contributed their economists' perspectives to my research. Craig White was my study buddy for the general exam and workout buddy, who kept me going to the gym. David Kim's culinary skills and intellectual passion delighted and inspired. Kirstin Scott never failed to light up my day with her warm and effusive nature, and Alison Hwong's wise and quirky humor was always welcome. Paula Chu was there as a friend for an ice cream cone and a hug when I most needed it and Emily Largent was always ready with incisive wit. Loren Saulsberry and Anas El Turabi gave helpful feedback on presentations and grants. Debbie Whitney, the program director, and her staff did their utmost to foster our community and to make sure we had the financial support to see us through our graduate years. I would also like to express my gratitude to my many teammates in the Harvard University Cycling Association for their friendship and camaraderie that helped keep my life healthy and balanced. They have been verifast friends indeed.

My husband, Andrew Sparling, has applied his editor's pen to my drafts from the very beginning, when I crafted a research statement to apply to Harvard, to the final production of this dissertation. Arguments are stronger and clearer where he has provided help. Andrew has has helped me to be a better writer and a more rigorous thinker at every turn. I would not have earned this Ph.D. without his support and encouragement.

It has been seen that the good cars may be driven out of the market by the lemons. But in a more continuous case with different grades of goods, even worse pathologies can exist. For it is quite possible to have the bad driving out the not-so-bad driving out the medium driving out the not-so-good driving out the good in such a sequence of events that no market exists at all.

George Akerlof

1

Medical Underwriting In Long-Term Care Insurance: Market Conditions Limit Options For Higher-Risk Consumers

1.1 Abstract

A key feature of private long-term care insurance is that medical underwriters screen out would-be buyers who have health conditions that portend near-term physical or cognitive disability. We applied common underwriting criteria based on data from two long-term care insurers to a nationally representative sample of individuals in the target age range (50–71 years) for long-term care insurance. The screening criteria put upper bounds on the current proportion of Americans who could gain coverage in the individual market without changes to medical underwriting practice. Specifically, our simulations show that in the target age range, approximately 30 percent of individuals whose wealth meets minimum industry standards for suitability for long-term care insurance would have their application for such insurance rejected in the underwriting stage. Among the general population-without considering financial suitability-we estimated that 40 percent would have their application rejected. The predicted rejection rates are substantially higher than the rejection rates

of about one-fifth to one-quarter of applicants in the actual market. In evaluating reforms for long-term care financing and their potential to increase private insurance rates, as well as to reduce financial pressure on public safety-net programs, policy makers need to consider the role of underwriting in the market for long-term care insurance.

1.2 Introduction

Most Americans are unprepared for the financial risk posed by the potential need for care as they age. When faced with significant disability, older adults tend to rely on family caregivers, exhaust modest personal savings, and rely on the Medicaid program as the payer of last resort. As the older population grows and the cost of long-term care continues to increase, one potential remedy might be to expand private insurance coverage for long-term care.

Long-term care consists of a wide range of services, including those of paid homemakers and aides, adult day care, assisted living, and nursing home care. In 2015 the median national cost for a year of nursing home care in a semiprivate room was $80,300, the cost for a year of single occupancy in an assisted living facility was $43,200, and the annual cost for the services of a home health aide for four hours per day was $29,200 (Genworth Financial 2015). Such costs are growing faster than the rate of inflation (Stewart, Grabowski, and Lakdawalla 2009), and the number of Americans expected to need paid long-term care is projected to increase from twelve million people in 2012 to twenty-seven million in 2050, or from 3.7 percent to 6.8 percent of the US population (Congressional Budget Office 2013; Bureau of the Census 2014).

The majority of households have insufficient personal savings to maintain their living standards

during a healthy retirement, much less a cushion to pay for potential long-term care needs (Center for Retirement Research 2016). A typical long-term care insurance plan might provide coverage for up to three years of nursing home care and even greater amounts of home care (usually with a delay after the onset of disability) to beneficiaries who need assistance with at least two activities of daily living or who need supervision because of cognitive impairment. The average age of individual long-term care insurance buyers in 2010 was fifty-nine (LifePlans 2012). Most current sixty-five-year-olds will never reside in a nursing home, but 4–10 percent of them will live in a nursing home for five years or more (Brown and Finkelstein 2009)—which would exhaust the savings of most households.

Insurance provides value by protecting people from catastrophic financial risk without their needing to set aside funds to cover the maximum possible expense (instead, they pay regular premiums). Yet few people avail themselves of long-term care insurance. According to our analysis of data from the Health and Retirement Study (Health and Retirement Study 2016), only about 10 percent of Americans ages 60–65 had a long-term care insurance policy in 2010. A variety of market failures have restricted the reach of long-term care insurance to a narrow slice of relatively healthy and affluent buyers.

Proposals for subsidies, outreach, and tax preferences to make it easier to buy long-term care insurance need to be understood in the economic context that shapes the individual (nongroup) market, which in 2014 accounted for roughly two-thirds of the long-term care insurance policies in force and 86 percent of new policies (Fisherkeller 2015a). Adverse selection—the disproportionate enrollment of people with expensive needs in an insurance plan—can create instability in any market for insurance where consumers buy products voluntarily. Discrepancies between individuals' and actu-

aries' knowledge about need and preference for long-term care can dissuade relatively healthy and low-risk consumers from buying insurance for that care, which results in premium prices that reflect higher-than-average risk-in turn, making the insurance appeal to an even narrower market of consumers. One strategy for broadening the risk pool is to offer guaranteed- or simplified-issue plans to a defined group, such as through an employer. In guaranteed issue, coverage does not require applicants to undergo an exam or answer any questions about medical history. In simplified issue, no exam is required, but applicants typically must answer some questions about their medical history. However, few employers offer long-term care benefits, and the group market for such insurance has contracted in recent years (Fisherkeller 2015b).

Instead, in the context of the individual market, underwriters attempt to correct for information asymmetry by screening would-be buyers for health conditions that portend current or near-term (within 5–7 years) physical or cognitive disability. Underwriting accuracy confers a competitive advantage on an insurance company, and companies protect their underwriting protocols as confidential assets (for a description and visual representation of the underwriting process, see Appendix Section A.1 and Exhibit A.1).

Factors specific to long-term care exacerbate uncertainty for insurers. People purchase long-term care insurance as protection against financial risk that may be decades in the future. This long time horizon exposes firms not only to the risk of changing longevity and volatile interest rates, but also to fluctuations in health costs and disability trends (Tracer and Davison 2014). In contrast to life insurance, where mortality is more stable and predictable over the long term, with long-term care insurance underwriters have to consider how changes in population health and functional status,

health technology, and consumer preferences (for example, the shift away from nursing home care and toward home- and community-based care) will change expected claims. The added uncertainty makes long-term care insurance exceptionally challenging to underwrite and insure at a stable price, more so than other types of insurance offered in voluntary markets.

Moreover, in an extended period of low interest rates and rising long-term care costs, firms selling long-term care insurance have experienced steady losses when insurers' income from investments is insufficient to meet obligations to policyholders. Insurers have responded by increasing prices and tightening underwriting requirements or, in many cases, exiting the market altogether (Fisherkeller 2015a; Cutler 1996). In considering long-term care insurance reforms, policy makers would benefit from a deeper understanding of how medical underwriting can affect who has access to insurance. About one-fifth to one-quarter of applicants for long-term care insurance are disqualified from purchasing it (LifePlans 2010). Existing analyses simulate underwriting practices by applying heuristics from industry experts or from field underwriting guides for insurance agents (Temkin-Greener, Mukamel, and Meiners 2000–2001; Murtaugh, Kemper, and Spillman 1995; Hendren 2013). There is a need for empirical evidence that can inform policy. To meet that need, we used the actual coverage decisions of insurance firms to model insurance eligibility of Americans who are at prime ages for purchasing long-term care insurance. Our analysis was based on data collected from those firms, on which they based their underwriting decisions.

1.3 Study Data And Methods

Data We developed coverage-approval models from a data set composed of application decisions for 15,659 individuals who applied for long-term care insurance policies from one of two carriers in 2008–12. The underwriting variables used in our analysis were demographic and socioeconomic characteristics, cognitive and functional abilities, diagnosed health conditions, use of health care, health behaviors, and body mass index (BMI) (Exhibit 1). For additional details on our data collection, see Appendix Section 2.1.

Exhibit 1: Effects of characteristics used in underwriting on an applicant's probability of being approved for long-term care insurance in the United States, 2010–11

	Applicant pool		HRS sample
Characteristic	Prevalence of characteristic in sample (%)	Characteristic's effect on probability of approval	Prevalence of characteristic in sample (%)
Age (years)			
18–49	12.2	Ref	-a
50–59	34.1	-0.007	52.5
60–69	48.4	-0.037****	41.7
70–71	5.3	-0.081****	5.8
Female	44.9	0.016**	52.4
Socioeconomic characteristics			
College degree	47.6	0.020***	31.3
Employed	62.5	0.030****	58.5
Cognitive and functional ability			
Delayed word-recall score less than 7/10	30	-0.035****	83
Experienced memory loss	23.6	-0.020**	21.7
Difficulty taking medication	1.6	-0.094***	2.5
Difficulty with activities of daily living	0.4	-0.522****	12.2
Diagnosed health conditions			
High blood pressure	50.2	-0.078****	49.5
Back pain	40.9	-0.101****	38.9

Exhibit 1: (continued): Effects of characteristics used in underwriting on an applicant's probability of being approved for long-term care insurance in the United States, 2010–11

	Applicant pool		HRS sample
Characteristic	Prevalence of characteristic in sample (%)	Characteristic's effect on probability of approval	Prevalence of characteristic in sample (%)
Arthritis	24.5	-0.111****	46.9
Diabetes	20.1	-0.415****	18.3
Heart problems	19.9	-0.130****	16.1
Psychiatric illness	18.4	-0.123****	19.6
Lung problems	10.2	-0.086****	8
Cancer	5.7	-0.111****	10
Stroke	1.6	-0.528****	4.6
Health care use in previous 2 years			
Hospitalization	53.3	-0.065****	21.3
Long-term care	1.6	-0.083**	5.5
Health behaviors			
Drinks alcohol	89	0.024**	65.5
Ever been a smoker	37.9	-0.013*	57
Currently a smoker	8.5	-0.114****	18.7
Body mass index (BMI)			
Underweight (BMI less than 18)	0.7	-0.174****	0.9
Normal or overweight (BMI 18–30)	59.5	d	61.4
Obese (BMI 30–40)	37.4	-0.046****	32.2
Extremely obese (BMI 40 or more)	2.4	-0.268****	5.5

SOURCE: Authors' analysis of data from a pool of 15,659 applicants for long-term care insurance from two US firms in 2008–12 and of data from the 2010–11 wave of the Health and Retirement Study (HRS). NOTES The estimated probabilities of approval were modeled in a multivariate regression and represent the difference in probability compared to the reference group. Where the reference group is not indicated in the exhibit, it is the complement of the reported category (for example, females are 1.6 percentage points more likely to be approved than males). The constant for the multivariate model, representing the mean predicted probability when all characteristics are set to the reference category, is 92.0 percent. Estimated probabilities from the 13,770 people ages 50–71 in the HRS sample were weighted to correspond to the American Community Survey, a sample of noninstitutionalized US adults. The approval rate was 75.8 percent for the applicant pool. Standard errors are reported in Appendix Exhibit A.4 (see Note 11 in text). aNot applicable. *p < 0:10 **p < 0:05 ***p < 0:01 ****p < 0:001

Data on the US population were taken from the public use files of the Health and Retirement Study (HRS), a nationally representative survey of Americans over age 50 and older and their spouses (Health and Retirement Study 2012). We analyzed a cleaned and processed version of the data file furnished by the RAND Center for the Study of Aging (RAND 2015). The HRS includes information on the health, living arrangements, employment, income and assets, and insurance status of respondents. The detailed information in the HRS allowed us to align the underwriting variables described above with the HRS survey questions. To compose a snapshot of the current long-term care insurance market, in the analysis we used responses to the 2010–11 wave of the HRS, which roughly overlapped with the timing of the insurance applications in our study.

1.3.1 Analysis

We estimated an empirical model of underwriters' coverage decisions to identify factors that determined whether a firm offered coverage to an applicant (for the full model specification, see Appendix Exhibit A.5). We used the model to impute underwriting probabilities for each respondent in the HRS ages 50–71 (the HRS sample was representative of the US population age fifty and older, and the maximum age in the insurance data was seventy-one). We designated applicants whose predicted probability of approval was 50 percent or less as "likely disqualified" and those whose predicted probability was greater than 50 percent as "likely approved." We used the assets, income, ethnicity, and race information in the HRS to refine the average predicted rates of approval (for example, by differentiating among socioeconomic groups). Industry guidelines, which are included in most state regulations and are almost universally followed by insurance sellers, compel agents to con-

firm that applicants meet minimum financial benchmarks before proceeding to sell them a policy (National Association of Insurance Commissioners 2014; 2009). To reflect this practice, we created a subsample of individuals for whom long-term care insurance would be considered financially suitable. The subsample was composed of individuals whose income and assets exceeded $20,000 and $30,000, respectively (or $30,000 and $50,000, for a couple). We also report results for members of the subsample whose assets, excluding housing, exceeded $250,000-a benchmark that we chose because it represented the approximate cost of three years of care in a nursing home.

1.3.2 Limitations

Our study provides the first detailed analysis of how underwriting policies of companies that sell long-term care insurance may limit market size for that insurance. It is important, however, to note the study's limitations. First, when we extrapolated our underwriting results to the general population, we generated predictions for a sample representative of the national population from a model estimated on an applicant pool that differed from the general population in both observed aspects (better health, less use of health care use, and more education, compared to the general population) and unobserved ones (generally better off and more financially savvy but perhaps with greater demand for paid care, compared to the general population). In generalizing from the applicants to the HRS respondents, we described a hypothetical scenario that was based on the best available empirical evidence. Second, our applicant sample represented approximately 5 percent of the market for long-term care insurance during the study period. Thus, caution should be taken in extrapolating our findings to the industry as a whole. The disqualification rates that our model predicted were

similar to industrywide rates among applicants of ages similar to those in our sample. We discuss the generalizability of our model in Appendix Section 4.2 and Exhibit A.9.

Third, data on some health characteristics that were important components of the underwriting model were not available in the HRS. For instance, although both the underwriters and the HRS used multi-item instruments to evaluate cognitive function, the HRS cognitive function questions do not correspond exactly with the items on the underwriters' proprietary instrument. We matched the model on self-reported memory and ability to recall a ten-word list, items that were present in both instruments. In addition, data on several "knockout conditions" such as multiple sclerosis and other degenerative chronic diseases—which would usually disqualify an applicant at the field underwriting stage—were not available in the HRS. These discrepancies make the underwriters' criteria more sensitive than the HRS data in identifying the risk of needing long-term care. Without information about these knockout conditions, we might have overestimated the general population's probability of obtaining underwriting approval.

Fourth, some of our results were counterintuitive, which might reflect limitations in our analytic approach. For instance, a history of long-term care use would normally be an automatic disqualifier. We were unable to obtain a good estimate of this effect, however, because few applicants in our sample reported such a history. That may be because such individuals had already been disqualified in field underwriting. Furthermore, the positive estimate of the influence of alcohol intake on approval most likely does not signal that drinking improves the likelihood of approval. Instead, people who abstain may be more likely to have other disqualifying health conditions, compared to people who drink (Lang et al. 2007; Koppes et al. 2005).

Fifth, although we extrapolated our findings to compare predicted approval rates for whites, blacks, and members of other minority groups, these findings were based solely on the health and demographic profile of those populations in the aggregate. Our underwriting data did not contain information on applicants' race or ethnicity.

1.4 Study Results

1.4.1 Underwriting And Approval For Long-Term Care Insurance

The sample of 15,659 applicants for long-term care insurance consisted of 3,782 individuals (24 percent) who were not approved in the underwriting process and 11,877 (76 percent) who were qualified to purchase a policy. Older applicants had lower approval rates than younger ones: When we controlled for health and socioeconomic characteristics, each ten-year increase in age significantly decreased approval probability, and the average applicant ages seventy and older had a .081 lower probability of approval than an applicant younger than fifty in similar health (Exhibit 1). Applicants with a college degree were more likely to be approved than those with no college education. Those who were employed were more likely to be approved than those who were unemployed.

Having any one of a series of diagnosed chronic conditions significantly lowered the probability of approval, in comparison with not having the condition. Among the most influential conditions were diabetes and a history of stroke, with back pain, arthritis, heart problems, psychiatric illness, and cancer somewhat less influential.

Current smokers were less likely to be approved than nonsmokers. However, people who drank

any alcohol were more likely to be approved than nondrinkers.

Having needed long-term care services or having been hospitalized in the previous two years reduced an applicant's likelihood of approval (it should be noted that very few applicants had a history of long-term care use). Applicants who were dangerously underweight (BMI less than 18) or extremely obese (BMI 40 or more) were much more likely to be disqualified than those who were normal weight or overweight (BMI 18–30). For a person 5 feet 8 inches tall, BMI between 18–30 corresponds to weight between 118–197 pounds, and BMI of 40 to 263 lbs.

1.4.2 Underwriting Applied To The Representative US Population

In the 2010–11 wave of the HRS, 13,770 individuals fell within our targeted age range of 50–71. Forty-five percent of that sample met the minimum recommended standards for financial suitability for long-term care insurance, and 46 percent of the financially suitable group had nonhousing assets of at least $250,000. For people who met the financial-suitability guidelines, the mean probability of approval was 0.611 (95% confidence interval: 0.60, 0.62) (Exhibit 2). Seventy percent (95% CI: 69, 72) of the applicants in the financially suitable category were likely to be approved, as were 75 percent (95% CI: 73, 77) of the applicants whose households had nonhousing assets of at least $250,000. In the full HRS sample ages 50–71, only 60 percent (95% CI: 58, 61) were likely to be approved.

In both the applicant pool and HRS sample, approval rates declined steeply with age, with the largest decrease occurring at ages sixty and older in the financially suitable HRS sample (Exhibit 3). Approval probabilities increased steadily with household wealth, even above the fortieth percentile (net assets greater than $100,000) (Exhibit 4). Among applicants in their fifties, mean ap-

Exhibit 2: Effects of characteristics used in underwriting on an applicant's probability of being approved for long-term care insurance in the United States, 2010–11

	Mean approval probability	Likely to be approved (a)	Sample size	Population, millions (b)
Suitable	0.611	0.701	6166	37.98
Assets $30,000–250,000 (c)	0.58	0.656	3349	19.5
Assets $250,000 or more (c)	0.644	0.749	2817	18.48
Not suitable	0.441	0.474	7604	33.08
Full sample	0.532	0.596	13770	71.07

SOURCE Authors' analysis of data from the 2010–11 wave of the Health and Retirement Study (HRS). NOTES The estimated probabilities for the 13,770 people ages 50–71 in the HRS sample were predicted from a multivariate logistic regression model (Exhibit 1) estimated from approval information on the insurance applicants. The full model, with odds ratios, appears in the appendix. Financial suitability is defined as having a yearly household income of more than $20,000 for a single person (more than $30,000 for a couple) and nonhousing assets of more than $30,000 for a single person (more than $50,000 for a couple). HRS sample estimates were weighted to correspond to the American Community Survey, a sample of noninstitutionalized US adults. Estimates with standard errors are available in Appendix Exhibit A.6 (see Note 11 in text). aRespondents were designated as "likely to be approved" with an imputed approval probability of >0:5. Thus, the percentages represent the estimated proportion of the population predicted to be approved with that imputed probability or greater. (b) Population extrapolated to the national level. cAssets are the net total of all nonhousing assets (property, business assets, other real estate, and financial wealth, including retirement accounts) less nonmortgage debt.

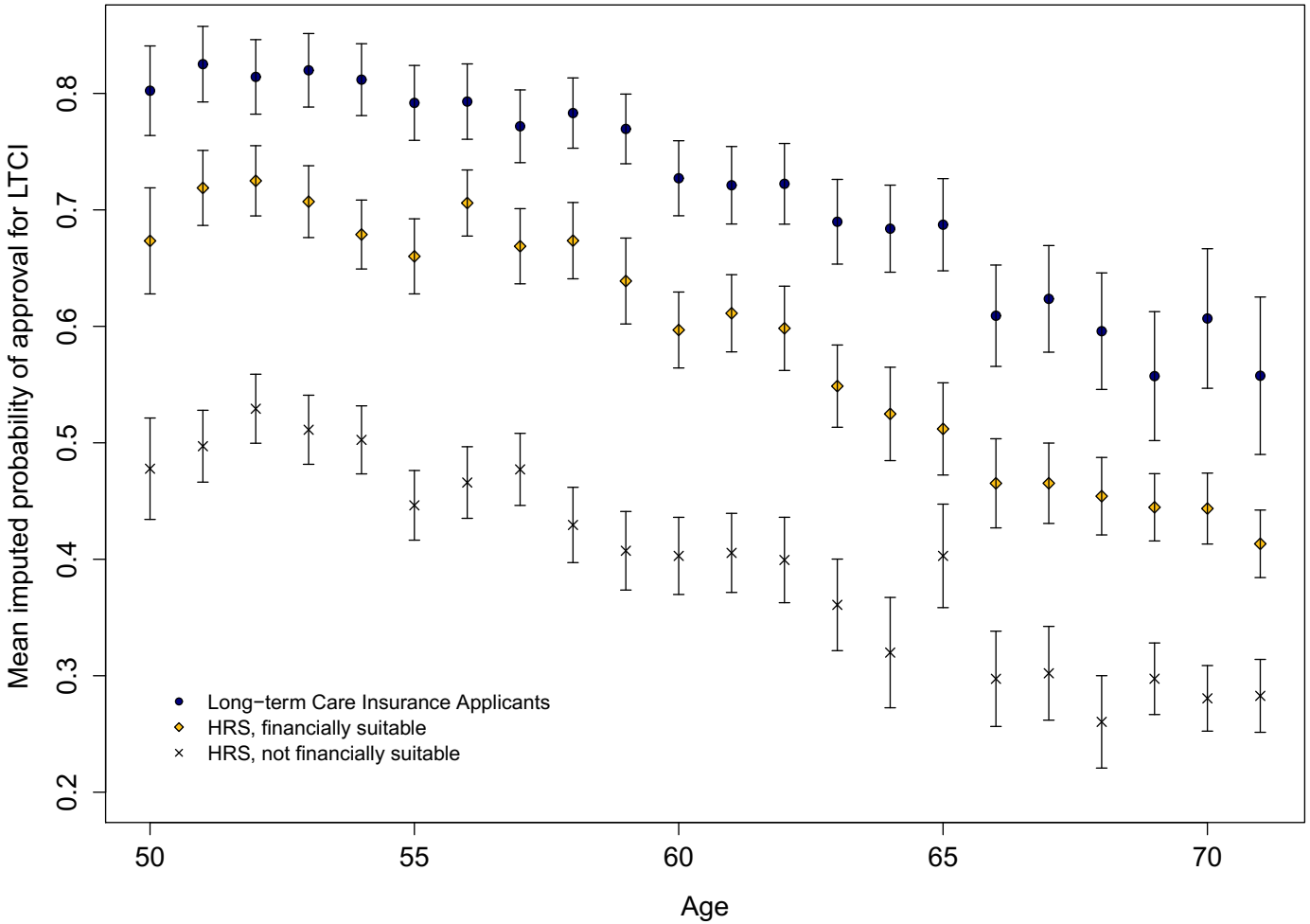


Exhibit 3: SOURCE Authors' analysis of data from the 2010–11 wave of the Health and Retirement Study (HRS). NOTES The estimated probabilities for the 13,770 people ages 50–71 in the HRS sample were predicted from a multivariate logistic regression model estimated from approval information on the insurance applicants. HRS statistics were weighted to correspond to the American Community Survey, a sample of noninstitutionalized US adults. "Suitable" means the applicant was financially suitable for approval, defined as having a yearly household income of more than $20,000 for an single person (more than $30,000 for a couple) and nonhousing assets of more than $30,000 for a single person (more than $50,000 for a couple). The applicant pool is described in the text. Error bars represent 95% confidence intervals.

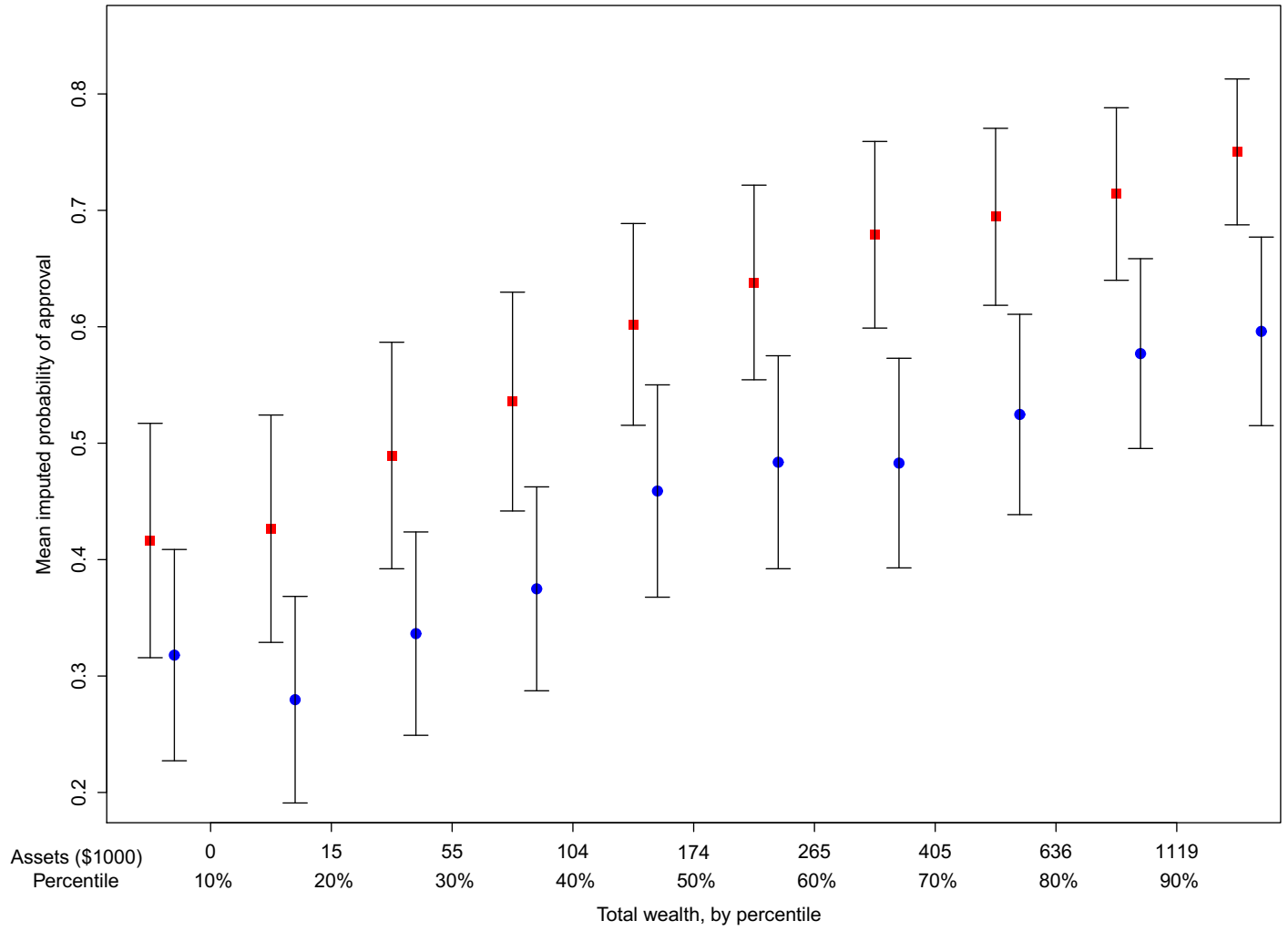


Exhibit 4: SOURCE Authors' analysis of data from the 2010-11 wave of the Health and Retirement Study (HRS). NOTES This exhibit shows estimated approval only for individuals in their 50s and 60s in order to highlight the effects of underwriting and wealth among those of prime purchasing ages. Elsewhere the analysis also reports estimated approval for people ages 70 and 71. The estimated probabilities for the HRS respondents depicted here were predicted from a multivariate logistic regression model estimated from approval information on the insurance applicants. HRS statistics were weighted to correspond to American Community Survey, a sample of noninstitutionalized US adults. Total wealth includes all housing, property, and financial wealth, less debt.

proval probabilities increased from approximately 0.4 in the lowest wealth decile to over 0.7 in the highest wealth decile, and among those in their sixties, from approximately 0.3 to 0.6.

When we applied underwriting parameters to different racial and ethnic populations in the United States, we found differential approval rates resulting from underlying differences in health status across groups. For instance, using information from the HRS, we found that 59.6 percent of whites were likely to be approved (95% CI: 57.7, 61.2), compared to only 45.0 percent (95% CI: 42.1, 47.9) of blacks and 52.0 percent (95% CI: 46.7, 57.3) of people from other minority groups (data not shown). As we expected, approved applicants had health profiles that were substantially different from those of applicants who were disqualified. The differences translated to an approximately threefold higher probability of disability within five years among those applicants whom we identified as likely to be disqualified, relative to those who were likely to be approved (for details, see Appendix Section 4.3 and Exhibit A.10).

1.5 Discussion

Current medical underwriting would exclude a large proportion of Americans from being able to buy long-term care insurance in the voluntary, private market. Because of self-selection and field underwriting, applicants for this insurance are considerably healthier and wealthier than the general population. We estimated that qualification rates in the general population would be lower than the 76 percent approval rate observed among applicants: only 60 percent of Americans ages 50–71 would most likely be approved if they applied for long-term care insurance and experienced underwriting standards comparable to those experienced by individuals in our sample. Underwrit-

ers would likely disqualify slightly more than one-third of people with assets between $30,000 and $250,000, even if subsidies, information campaigns, or other inducements managed to encourage them to apply for coverage. Individuals with assets in that range are presumably the ones who can afford premiums and are motivated to protect their assets, instead of spending down their savings if they need long-term care before they are eligible for Medicaid, but who are not necessarily wealthy enough to pay for long-term care out of pocket. Medical underwriting limits access to insurance regardless of the affordability of the policies. Thus, the ability of approaches that make premiums more affordable to spur purchase of long-term care insurance is limited by medical underwriting, although the limits will diminish somewhat as greater numbers of individuals apply for insurance. In fact, the modest impact to date of strategies such as tax credits for purchasing long-term care insurance or enhanced Medicaid asset protection for individuals who purchase policies should be interpreted in the context of an insurance market in which underwriting plays an important role (Goda 2011; Lin and Prince 2013). In particular, underwriting limits the access to long-term care insurance of people living with chronic conditions that are relatively common among middle-aged adults. Having diabetes or a history of stroke, being extreme obese, and having difficulty with at least one activity of daily living were the most important factors in the underwriting decision (Exhibit 1). Having diabetes and having difficulty with activities of daily living are widespread, affecting an estimated 18 percent and 12 percent, respectively, of people in this age group (Appendix Exhibit A.3). Because of underlying health differences across racial and ethnic groups, there are racial and ethnic differences in access to long-term care insurance. These differences are a function of disparities in health status for racial and ethnic minorities in the aggregate-disparities that we observed in

the HRS data and that have been well documented elsewhere (Smedley, Stith, and Nelson 2003). Nonwhite Americans are at higher lifetime risk than whites of entering a nursing home, and if they do, they more likely than whites to be in homes of relatively low quality (Mor et al. 2004). Blacks are less likely to purchase long-term care insurance than whites and are less sensitive to changes in premium prices (Goda 2011), a fact that is consistent with our finding that underwriting may make it more difficult for minorities than for whites to buy long-term care insurance at any price. Our analysis suggests that underwriting may explain part of that discrepancy in insurance purchase, though consumer preference may also play a role.

1.5.1 Policy Implications

Although adverse selection creates market conditions that necessitate medical underwriting for voluntary insurance, these conditions are not static. For instance, reforms could mitigate adverse selection by expanding the risk pool and encouraging the purchase of long-term care insurance at younger ages than is typically the case now, or by offering reinsurance to protect the insurer from large claims. This could be achieved by offering a universal long-term care benefit after a set waiting period, or by mandating the purchase of long-term care insurance. Underwriters might relax their criteria in response to an expanded applicant pool that contained younger people than is now the case, if they believed that adverse selection had lessened-and such a relaxation could lower rejection rates. Insurers might also respond by reducing premiums and loosening underwriting criteria. But the task of convincing households to forgo present consumption to protect against risks that could be decades in the future would still be tremendously difficult, even if the economic advantages were

substantial. Incentives targeted toward the group- and employer-based markets, where insurers can offer guaranteed-issue policies or relaxed underwriting standards, could encourage people to plan for long-term care needs at younger ages, when they are less likely to have disqualifying conditions. However, recent evidence indicates that the group market for long-term care insurance has stagnated and that many long-term care insurers have discontinued sales of the insurance and marketing efforts to new groups (Fisherkeller 2015a). The Community Living Assistance Services and Supports (CLASS) Act-a provision of the Affordable Care Act (ACA) that would have created a voluntary, public long-term care insurance program-was an effort to enroll younger populations in the insurance. Yet projections showed low enrollment and spiraling costs because of adverse selection. The administration of President Barack Obama declared the law actuarially unsustainable, and Congress quietly repealed it in 2013 (Senate Commission on Long-Term Care 2013). It is difficult to imagine a voluntary program that could surmount the problem of adverse selection without incurring the same limitations that currently affect the private market. One recent proposal put forward to address the need to finance long-term care includes a catastrophic social insurance program that would cover nursing care after a two-year waiting period and offer a lifetime benefit thereafter, covering 95 percent of the population while substantially offsetting Medicaid and out-of-pocket spending (Favreault, Gleckman, and Johnson 2015). If such a program existed, the role of voluntary long-term care insurance would shift: Such insurance would become a more limited product, covering the interim period after the onset of disability and before the catastrophic insurance took effect. Lower premium prices for this limited product could attract a healthier pool of consumers and alter underwriting considerations, which could result in lower underwriting rejection rates. A catastrophic

social insurance program would represent a large new program, however, and securing the necessary revenue would be challenging in the current fiscal and political environments. For individual health insurance, the ACA addressed adverse selection by mandating that individuals either buy private insurance or face penalties. Although a mandate for public or private insurance could address adverse selection in long-term care insurance, public support for such a mandate is weak and likely to remain so in the near term.

1.5.2 Conclusion

The public policy goals of bringing more private dollars into the long-term care system are to provide financial protection to older people and their families, reduce the growth in public spending for long-term care, and support the service infrastructure to meet growing demand. Making progress toward these goals will grow more urgent as the US population ages. Our findings do not preclude a role for the private market in long-term care reform. In fact, politically viable solutions are likely to consist of some combination of public safety-net programs together with incentives for increased personal savings and the purchase of long-term care insurance (Stevenson et al. 2010). Unless policy makers can find ways to broaden the risk pool significantly, however, current underwriting practices are likely to persist. Thus, simply subsidizing the voluntary nongroup market, without addressing the market conditions that necessitate underwriting, will provide protection only to people with the lowest risk for long-term care. It will not achieve the goal of expanding insurance protection to as many Americans as possible and ensuring that they have access to a secure financing system for long-term care.

2

The Role of Underwriting in Policy Incentives for Private Long-Term-Care Insurance

2.1 Introduction

Private insurance currently plays a small role in financing long-term services and supports in the United States. In 2012 long-term services and supports cost nearly $220 billion, or 9 percent of all health spending; of this, 61 percent were paid by Medicaid, the public safety-net program for the poor; 22 percent were out-of-pocket; and only 13 percent were covered by private insurance and other public sources (O'Shaughnessy 2014). In response, state policymakers have taken steps to promote the purchase of private long-term care insurance (LTCI). The aim of such efforts is to bring more private dollars into the long-term-care system in order to reduce the growth in public spending for long-term care, to provide beneficiaries with increased access and independence, and protect households from the financial risk posed by extensive long-term care stays—priorities that will grow more urgent as the U.S. population ages. With the majority of LTCI policies being sold on the individual market, the voluntary LTCI market has lacked the broader risk pool that the employer-

sponsored market has historically provided for health insurance. To combat losses from adverse selection, firms require would-be buyers to pass strict medical underwriting requirements. Those restrictions mean that LTCI is actually available only to a narrow slice of Americans who are not only wealthy enough to afford the premiums, but also healthy enough to be at low risk of eventually needing the insurance. Underwriting practices put an upper bound on the potential for subsidies to increase the proportion of Americans who are covered by LTCI, as well as subdue the offsets available from any increases that might occur: approximately 40 percent of the U.S. population ages 50–70 would likely be disqualified from private long-term care insurance for medical reasons (Cornell et al. 2016). Not surprisingly, research generally suggests most policy incentives to increase LTCI purchase only have a modest effect. However, these previous studies consider the impact across all individuals, regardless of their likelihood of gaining approval to purchase a policy via underwriting. In this paper, I estimate how two prominent LTCI subsidies—tax incentives and partnership programs—impact LTCI purchase in the context of the medical underwriting process.

2.2 Background

Reasons for the low penetration in the market for long-term-care insurance include both demand- and supply-side limitations. Long-term catastrophic risks, costs and possible need for care, and consumers do not have a good understanding of what public services are covered by public programs (LifePlans 2012). Many expect to rely on Medicaid, especially those whose assets are low enough that they may spend down to eligibility levels (Brown and Finkelstein 2008; Brown, Coe, and Finkelstein 2007). On the supply side, firms are beset by high transaction costs, imperfect competition, long-

term dynamic contracting problems, and asymmetric information (Brown and Finkelstein 2007). The actuarial challenges of offering LTCI policies are uniquely difficult: insurers in early years underpriced premiums, incorrectly estimating how fast LTC costs would rise, how many buyers would use benefits, how few would let their policies lapse, and the low interest rates that cut into profit margins. As with any insurance market, consumer information about their risk of needing long-term care that is unknown to actuaries and not accounted for in premium calculations can lead to adverse selection. Would-be consumers who expect higher costs will be more willing to buy policies, and will tend to buy more generous policies. Insurers respond to the problem by disqualifying some applicants through medical underwriting. Firms address the information asymmetry by collecting information on health risks. In practice, firms offer an age- and gender-rated standard premium rate (with some discounts, for example, to married couples or to individuals with blood pressure in the healthy range).

Both federal and a growing number of state tax codes offer tax incentives to subsidize LTCI premiums. In 1997, the Health Insurance Portability and Accountability Act (HIPAA) allowed deductions of LTCI premiums for taxpayers who itemized their health expenditures. Also in the 1990s, states began passing tax credits and deductions for LTCI premiums: in 1994, only one state provided tax subsidies, and by 2016, 30 states had them. Additionally, many states have addressed the potential for Medicaid to substitute for (or "crowd out") private insurance by allowing individuals who buy "partnership-qualified" policies to keep some assets in the event they should exhaust their insurance policies and turn to Medicaid. These rule changes are known as Long-Term-Care Partnership Programs. Persons who purchase and receive benefits under a "partnership qualified" policy may be

entitled to dollar-for-dollar asset protection under Medicaid, up to $100,000, if they use their benefits and subsequently apply for Medicaid. For example, if an individual uses $100,000 of insurance coverage, they would be able to retain that amount of assets (in addition to the modest assets already allowed under state rules) and apply for Medicaid. Although not a direct subsidy of premiums, Partnership programs encourage purchasers to buy "shorter and fatter" policies more aligned with their individual financial risk (i.e., the amount of net worth they wish to protect). As of 2016, 43 states had either adopted a Partnership program or begun filing applications, and more than 100,000 new partnership-qualified policies were in force.

Prior analysis of tax subsidies for long-term care have used a difference-in-differences approach to examine the effect of tax subsidies and Partnership programs on long-term care (Goda 2011; Lin and Prince 2013; Greenhalgh-Stanley 2012; Stevenson, Frank and Tau 2009). Overall, these studies suggest a modest impact of these policies on LTCI purchase, with stronger responses among those with higher education, income, wealth and health. Previous work has also estimated the supply-side restrictions that underwriting practices place on the long-term-care market (Murtaugh, Kemper and Spillman 1995; Hendren 2013). To my knowledge, no published studies explicitly examine underwriting as a key factor in determining the potential effect of higher rates of LTCI take-up on Medicaid expenditures. The impact on Medicaid expenditures depends crucially on those who respond to incentives (be they tax incentives, Partnership allowances, public-information campaigns or other enticements) counterfactual likelihood of eventually relying on Medicaid to pay for long-term care; and the willingness and ability of firms to extend LTCI to new and riskier markets.

The theoretical relationship between health status and insurance purchase has forces pushing the

correlation in two competing directions. On the demand side, individuals in poorer health have a higher expected return from $1 worth of insurance (if those health differences are not reflected in premium price). But the underwriting criteria that firms use to disqualify risky buyers will also disqualify many of those buyers whose risk is observable to insurers. The heterogeneity of how LTCI coverage rates respond to these policy incentives, therefore, is ambiguous.

This paper will examine how the demand response to tax subsidies and Partnership programs differs with individuals' probability of being offered coverage from underwriters. To accomplish that, this paper exploits variation among states over time in adoption of tax subsidies and Partnership agreements. I hypothesize that the effect of adopting these programs should intensify among potential buyers that are more likely to be approved by underwriters in the private, individual market. Although high-risk individuals may be able to qualify for insurance through their employers or family members in the group market, where underwriters' requirements are generally less strict, lack of access is likely to explain the low coverage rates more than price sensitivity. Although it is possible that firms might relax underwriting rules in response to subsidies in order to expand the market, the underlying dynamics of adverse selection and information asymmetries still make this market segment less attractive. It is more plausible that we would see the most action among low-risk prospective buyers, either extensively (higher probability of purchase) or intensively (higher-value policies). The purpose of this paper will be to examine how subsidies affect insurance purchase on the extensive margin in the context of medical underwriting.

2.3 Methods

2.3.1 Identification strategy and econometric model

We use a difference-in-differences approach to examine the effect of two state policies—tax subsidies and spend-down asset protection—on the long-term care insurance purchase decisions of individuals age 50–69. Exhibit 1 shows the variation over time in the proportion of the Health and Retirement Study sample living in states with tax subsidies and Partnership programs.

The general specification is as follows:

$$LTCI_{ist} = \gamma POLICY_{st} \times UNDERWRITING_{ist} + \lambda UNDERWRITING_{ist} + \beta X_{ist} + \omega t + \sigma s + \varepsilon_{ist} \tag{2.1}$$

The analysis is conducted at the individual survey wave level. Each wave corresponds to a two-year time period, and the data used in this study span from 1996–2012. The dependent variable, $LTCI_{ist}$, is an indicator that individual i residing in state s has long-term-care insurance coverage in year t. The *POLICY* variable is a binary indicator of whether state s had an active policy incentive—either tax subsidy or Partnership program.

$UNDERWRITING_{ist}$ is a continuous measure between 0 and 1 of an individual's probability of being approved in the medical underwriting process for long-term care insurance. In alternate specifications, it is a vector of four binary variables indicating the quartiles of the individual's predicted probability of being approved buy underwriters for long-term care insurance, π, from "very low" ($0 < \pi_1 < .25$) to "very high" ($.75 < \pi_4 < 1$). That measure is constructed from a multi-

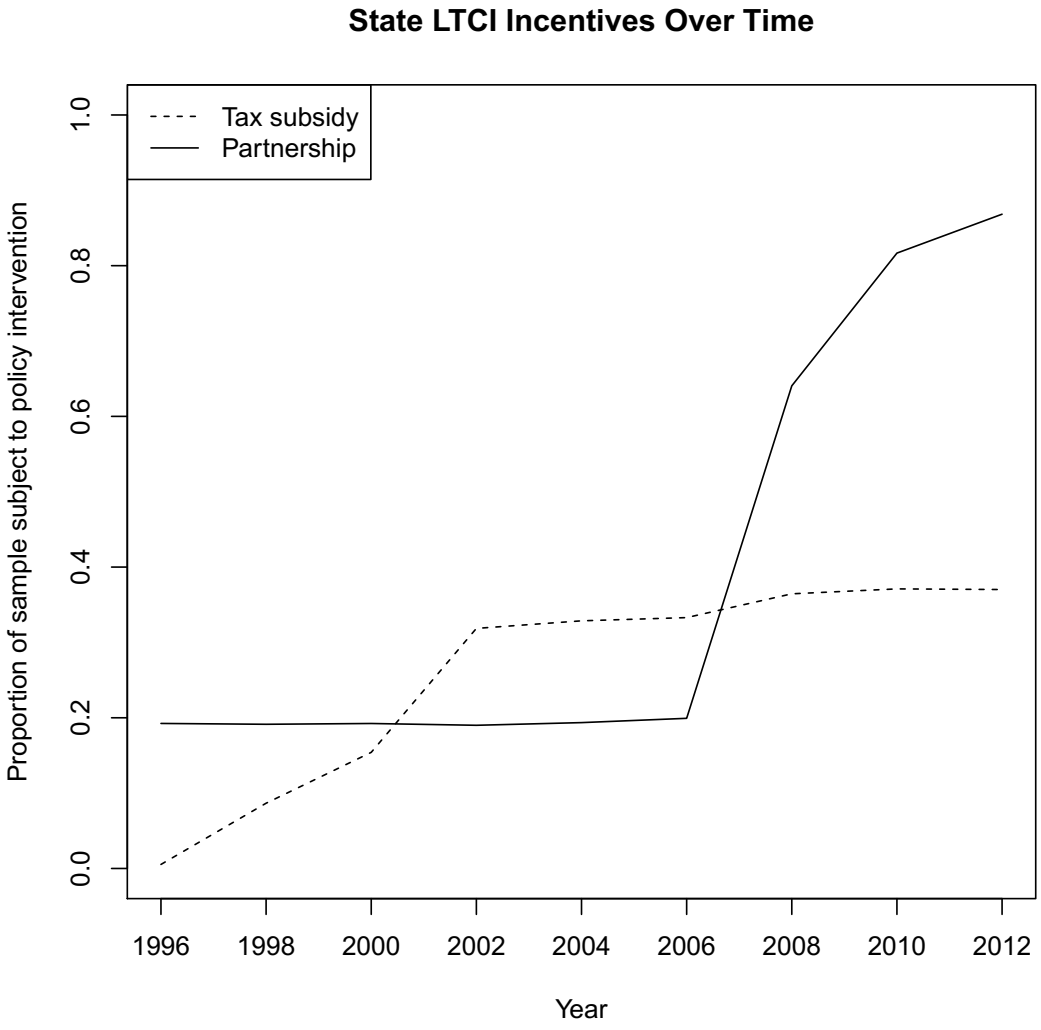


Exhibit 1: Change over time in the proportion of the analytic sample residing in a state with either a state tax subsidy (credit or deduction) for long-term care insurance or a Partnership program allowing holders of qualified long-term care policies additional asset allowances when qualifying for Medicaid long-term care services.

variate model of underwriting approval using a dataset of over 15,000 decisions by underwriters in two American long-term care insurance firms. The predictors in the model were characteristics commonly used in medical underwriting decisions that are also available in the Health and Retirement Study (HRS) data: demographic and socioeconomic characteristics, cognitive and functional abilities, diagnosed health conditions, previous use of health care, health behaviors, and body mass index (BMI). The coefficients estimated from the dataset of insurance applicants were then applied to the HRS responses to generate a predicted approval probability for each HRS respondent.

State and year fixed effects σ_s and ω_t control for state-specific levels of insurance and national trends in insurance rates. The preferred specification also includes μ_i, within-person fixed effects, controlling for all time-constant characteristics of individuals in the sample. Standard errors are clustered at the state level. Time-varying individual characteristics, X_{ist}, include controls for age, income and assets, education, race, marital status, number of children, and retirement status; as well as state-level factors nursing facility occupancy rates, nursing home beds per person over age 65, percent of the state population over age 65, and a Medicaid generosity composite measure of a state's asset and income retention rules and nursing home reimbursement rate. The heterogeneous effect of the policy depending on the individual's underwriting score in equation (2.1) is $\gamma + \lambda \times$ *UNDERWRITING*.

The identifying assumption in equation (2.1), where state fixed effects σ_s are included, is that there are no excluded, state-specific events that are correlated with the introduction of a tax incentive or Partnership program that would have caused changes in the rates of long-term care insurance. When individual fixed effects μ_i are included, the assumption is that there are no person-specific

inducements to change insurance behavior correlated with policy changes in the state where they live.

The second analysis estimates the effect of the after-tax price of $1 of long-term care insurance, where differences in the generosity of state subsidies cause variation in the realized price of insurance. I estimate the following equation:

$$LTCI_{ist} = \gamma PRICE_{ist} \times UNDERWRITING_{ist} + \lambda UNDERWRITING_{ist} + \beta X_{ist} + \omega t + \sigma s + \varepsilon_{ist} \tag{2.2}$$

In this model, $PRICE_{ist}$ varies by state and year with the implementation of the tax subsidy and its generosity (credit or deduction, percent of premium that is subsidized, and allowed maximum) as well as the individual's marginal state tax rate. Estimates of γ and λ will be biased in an ordinary-least-squares regression if individuals' demand for insurance is endogenous to their marginal tax rates in ways that are not captured. For instance, households with high financial literacy that also place a high value on insurance may take steps to reduce their marginal tax rate. To address this potential bias, I follow previous work (Goda 2011; Currie and Gruber 1996) by simulating the after-tax price of long-term care insurance in each state and year for a nationally representative sample of households in the target age range, and use the average of this simulated price as an instrumental variable (IV) to predict variation in individuals' realized price of $1 of long-term care insurance. Because the average price is calculated for the same group of individuals in each state, the only variation comes from changes in tax policy. This IV estimate isolates the changes in demand for long-term care insurance that are attributable to changes in generosity of the tax policy, independent of po-

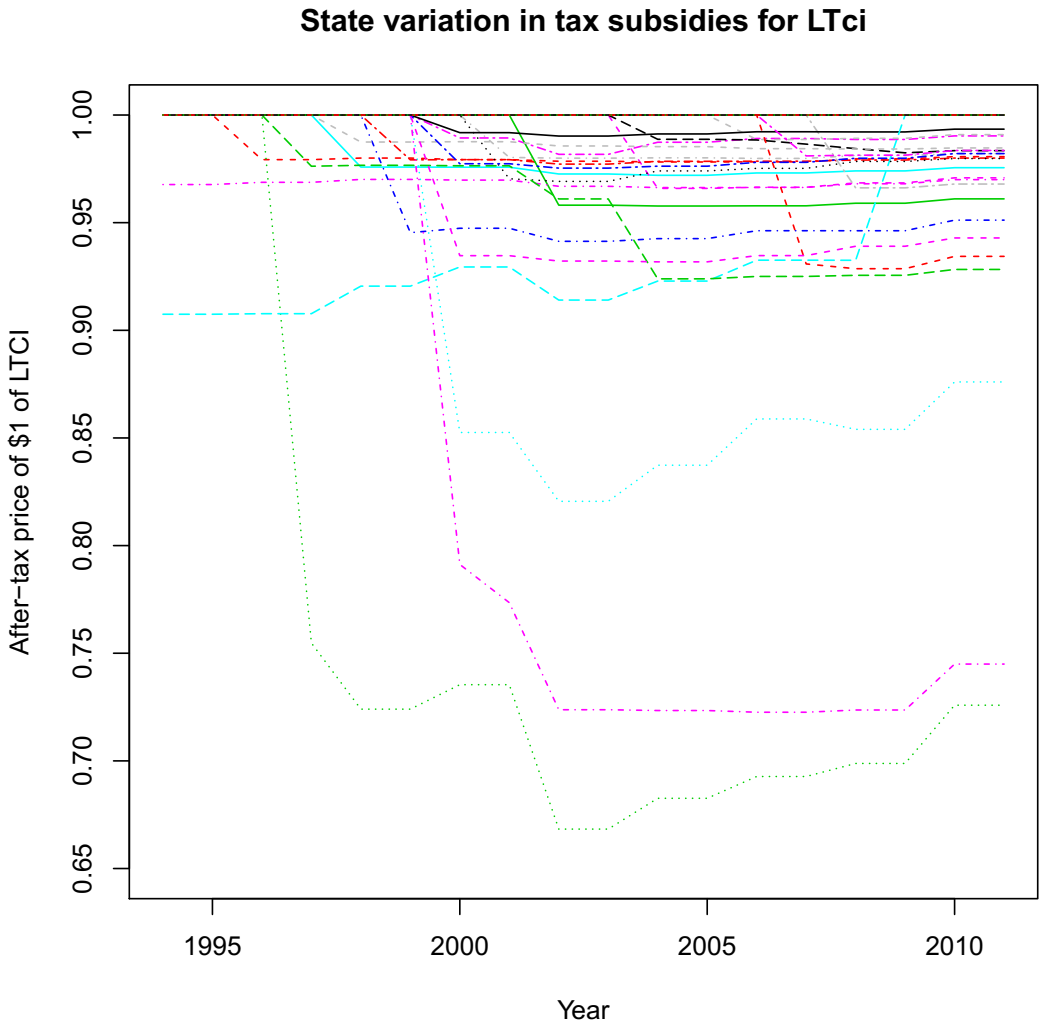


Exhibit 2: Change over time in the after-tax price of long-term care insurance. The vertical axis is the average after-tax price of $1 of long-term care insurance experienced by a nationally representative sample subject to tax rules. Each line represents a U.S. state. These values are used to construct an instrumental variable to predict variation in individuals' after-tax price, based on their state of residence, exogenously to their personal income and asset profiles.

tential confounders. In the appendix to this paper, I describe the two-stage estimation and show that the average simulated after-tax price is highly correlated with PRICE. I also show descriptive statistics that when respondents are sorted by simulated tax price the groups are well balanced with respect to covariates, suggesting the instrument is as good as randomly assigned.

To calculate tax price, each person in the HRS was assigned a premium amount based on their state, year and age. These premiums were calculated from mean annual premiums for policies sold in 2002, from Weiss Ratings, Inc. Premiums for ages not reported were interpolated, and assumed to grow by 3 percent per year (Johnson et al. 2007, Table V-4). Marginal tax rate simulations were done using the TAXSIM program from the National Bureau of Economic Research (Feenberg and Coutts 1993). The value of the state-year credit, deduction, or both were calculated according to each individual's marginal state tax rate, up to their state tax liability. The tax price of $1 of long-term care insurance was the proportion of the premium paid after subtracting the value of these credits and deductions. For individuals who did not file a tax return, the amount was set to 1.

Simulated average tax prices were simulated using the full, nationally-representative HRS sample subject to that state and year's tax rules. Because the HRS is a biennial survey in which the bulk of interviews are completed in even-numbered years, average marginal tax rates for odd-numbered years were calculating using the income and filing variables of HRS sample from the wave corresponding to the preceding year.

2.3.2 Data and sample

The data used in this analysis come from the Health and Retirement Study (HRS), sponsored by the National Institute on Aging. The HRS is a biannual panel survey of U.S. residents over age 50 and their spouses that began in 1992. The survey contains questions on demographics, health, wealth, family structure, housing, employment, disability, retirement, and insurance coverage. Only waves three (1996) and forward are used in these analyses because questions about long-term care insurance are worded ambiguously in the initial survey years. I used the publicly available version of these data available from Rand—a respondent-level database with consistent variable naming and imputations for wealth and income—merged with restricted identifiers for state of residence, which I obtained from the HRS.

Using those state identifiers, HRS data were merged with state-level policy information regarding the implementation of tax incentives and the Partnership program, by state. The binary Partnership indicator was defined as 1 if the respondent lived in a state with the program and the interview was completed after the official start of the program, and 0 otherwise. Similarly, a tax incentive was defined as 1 if a state deduction or credit existed for long-term care insurance in that year, and 0 if there was neither a deduction nor a credit.

The analysis samples were limited to respondents between the ages of 50 and 69. The HRS is a representative population sample only for the 50-and-over demographic (younger people in the sample are spouses and partners of age-eligible respondents; and few insurance firms market their products to customers over age 70, instead steering them away from long-term care insurance prod-

ucts before they even submit an application to the underwriting process.

Different samples were used for each of the two policies evaluated here. The tax-incentive analysis sample is limited to HRS waves 3–8 (survey years 1996–2006), because the bulk of policy changes occurred within this time span. The sample excluded self-employed persons, whose tax treatment of health insurance premiums differs. State level data on the population over age 65, nursing facility occupancy rates, and nursing home beds per 1,000 people age 65 and over were included . The final sample included 53,503 observations on 16,080 unique respondents. In the individual-fixed-effects model, respondents with only one observation were dropped, leaving 50,708 observations on 13,285 respondents.

Exhibit 3 shows summary characteristics of the sample calculated using sampling weights. Weights were structured to match the Current Population Survey, so respondents living in a nursing home or outside of the U.S. were assigned weights of equal to zero. Across waves, 10 percent of respondents in each year have long-term care insurance, on average. Twenty-three percent of responses occur in a state and year where there is either a tax credit or deduction for LTCI, and they experience a mean after-tax price of $0.98 per $1 of insurance. With the exception of four pilot states that implemented Partnership programs in 1994, state policy changes in the Partnership program occurred in 2006 and later. Therefore I restrict the sample for this analysis to waves 6–12 (sample years 2002 through 2012). Here I also restrict to ages 50–69, for a final sample size of 57403 observations on 19139 individuals. Forty-seven percent of the sample has experienced the implementation of the Partnership program.

Exhibit 3: Summary statistics for Health and Retirement Study samples

	A. Tax incentive analysis				B. Partnership analysis			
	mean	sd	min	max	mean	sd	min	max
Has LTC insurance	0.1	-0.3	0	1	0.11	-0.31	0	1
Underwriting approval probability	0.56	-0.31	0.00	0.97	0.54	-0.32	0.00	0.97
Partnership					0.51	-0.5	0	1
Tax subsidy	0.23	-0.42	0	1				
After-tax price $1 LTCI	0.98	-0.099	-0.059	1				
Assets (USD1000)	354.4	-909.8	-3637	90648	466	-1152.1	-2246	90648
Income (USD1000)	65.8	-106.7	0	7904	86.4	-253.7	0	25360
Female	0.55	-0.5	0	1	0.52	-0.5	0	1
Married	0.7	-0.46	0	1	0.68	-0.47	0	1
Age	59.4	-5.09	50	69	59.7	-5.01	50	69
College or above	0.23	-0.42	0	1				
Years of education					13.3	-2.91	0	17
Number of living children	3.08	-2.01	0	20	2.84	-1.89	0	19
Hispanic	0.077	-0.27	0	1	0.085	-0.28	0	1
African American	0.1	-0.31	0	1	0.11	-0.31	0	1
Retired	0.31	-0.46	0	1				
Self-reported health	2.71	-1.14	1	5	2.69	-1.11	1	5
Difficulty with 1+ ADL	0.12	-0.32	0	1				
Body mass index					28.7	-6.04	7	83
Years	1996–2006 (HRS waves 3–10)				2002–2012 (HRS waves 6–12)			
Observations	53503				57403			
Unique respondents	16080				19139			

Notes: Statistics are calculated with population weights furnished by the Health and Retirement Study, which are calibrated to correspond to the U.S. community-dwelling population. To facilitate fixed-effects analysis, each individual is assigned a constant weight forward from their first interview year that they appear in the sample. Self-reported health is an ordinal measure 1–5, with 1 being excellent and 5 being poor.

2.4 Results

2.4.1 Tax incentives

The results of the regression described in equation (1), with the presence of tax incentive as the policy treatment, are summarized in Exhibit 4 and confirm that the impact of tax incentives on long-term care insurance coverage is significantly altered by the underwriting scores of consumers. All models include year and state fixed effects.

Models (1)–(5) are variations on the models that Goda (2011) estimates for the effect of tax subsidies, where model (5) reflects her preferred specification and is similar in direction and order of magnitude to her main result: it suggests that the presence of a tax incentive (credit or deduction) increases participation in private long-term care insurance by 1.8 percentage points, or about 18 percent. In models (6) and (7), I show the interaction effect of tax subsidy and individuals. The state-fixed-effects estimate suggests that the effect of the tax subsidy is 0.1 percentage point greater with every 10 percentage-point increase in underwriting-approval probability (though the result is not statistically significant at the 0.05 level). In the fixed-effects model, the modifying effect is more striking: with a coefficient on the interaction term of 5.9 percentage points suggests that a 10 percentage-point increase in a person's underwriting approval probability increased their probability of holding long-term care insurance by about 6 percent (given the average LTCI prevalence of 10 percent), a result that is statistically significant at the 0.05 level.

Exhibit 4: Effect of presence of a tax incentive on purchase of long-term care insurance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Tax subsidy	0.0273**	0.0190**	0.0190**	0.0193**	0.0180*	0.0138	-0.0129		
	(0.0109)	(0.00754)	(0.00739)	(0.00910)	(0.00948)	(0.0108)	(0.00965)		
Subsidy X Approval probability						0.00919	0.0587**		
						(0.0194)	(0.0232)		
Subsidy X v. low approval								0.0128	-0.000942
								(0.0101)	(0.00924)
Subsidy X low approval								0.0227*	0.00586
								(0.0127)	(0.0107)
Subsidy X high approval								0.0173	0.0285**
								(0.0120)	(0.0123)
Subsidy X v. high approval								0.0211**	0.0362**
								(0.00967)	(0.0151)
Underwriting approval probability						0.00743	-0.00593		
						(0.00715)	(0.0122)		
Low approval								0.00711	0.0177***
								(0.00510)	(0.00625)
High approval								0.00697	0.0131*
								(0.00536)	(0.00757)
V. high approval								0.00344	0.00309
								(0.00556)	(0.00892)
State time trend	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes

Exhibit 4: (continued) Effect of presence of a tax incentive on purchase of long-term care insurance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual fixed effects	No	No	No	Yes	Yes	No	Yes	No	Yes
Assets & income	No	No	Yes	No	Yes	Yes	yes	yes	yes

* p<.1, ** p<.05, *** p<.01.

Notes: Standard errors in parentheses. All models include state and year fixed effects.

2.4.2 Tax price

Tables 5 and 6 give the instrumental-variable estimates of $\hat{\lambda}$, the effect of tax price on long-term care insurance coverage, in state-fixed-effects and person-fixed-effects models, respectively. First-stage estimates confirm that the simulated after-tax price Models (1) in both tables 5 and 6 are comparable to the main result found by Goda (2011). Table 3, Model (4), controlling for assets and income, suggest that that the level effect of tax price is more strongly negative with higher underwriting probability—an effect that is statistically significant at the 0.01 level in the individual-fixed-effects model (table . Table 3, Model (5) suggests, furthermore, that with each increasing quartile of underwriting approval probability, the price response intensifies. Among the highest underwriting-approval quartile, a $1 increase in the price of $1 of long-term care insurance decrease the rate of insurance purchase by 0.276 percentage points. In elasticity terms, with average insurance holding of about 12 percent in this group (see appendix), that suggests that a 1 percent increase in premium prices decreases insurance holding by 0.00276/.12 = 2.3 percent among the very-high-approval group.

2.4.3 Long-term care Partnership

The results for the effect of the Partnership program are summarized in Exhibit 7 and Figure 3(C).

Following Lin and Prince (2013), model (2), with individual fixed effects, estimates the effect of the Partnership separately by asset group; model (3) adds an additional two of data from the HRS, capturing additional program variation. Similar to Lin and Prince, the overall model finds that the

Exhibit 5: Effect of after-tax price on long-term care insurance: Panel A, state fixed effects

	(1)	(2)	(3)	(4)	(5)
After-tax price $1 LTCI	-0.143***	-0.132***	0.0122	0.0364	
	(0.0400)	(0.0399)	(0.0829)	(0.166)	
After-tax price X approval prob			-0.257**	-0.250*	
			(0.104)	(0.151)	
Tax price X v. low approval					0.0630
					(0.182)
Tax price X low approval					-0.0990
					(0.113)
Tax price X high approval					-0.125
					(0.0891)
Tax price X v. high approval					-0.177**
					(0.0734)
Underwriting approval probability			0.278***	0.250*	
			(0.103)	(0.149)	
V. low approval probability					-0.242*
					(0.138)
Low approval probability					-0.0745
					(0.0840)
High approval probability					-0.0489
					(0.0695)
Model	IV	IV	IV	IV	IV
Person_Fixed_Effects	No	No	No	No	No
Assets_Income	No	Yes	No	Yes	Yes

N=53501. Standard errors in parentheses.
* p<.1, ** p<.05, *** p<.01.

Exhibit 6: Effect of after-tax price on long-term care insurance: Panel B, individual fixed effects

	(1)	(2)	(3)	(4)	(5)
After-tax price $1 LTCI	-0.194***	-0.187***	0.0858	0.0596	
	(0.0498)	(0.0498)	(0.0896)	(0.102)	
After-tax price X approval prob			-0.457***	-0.421***	
			(0.128)	(0.132)	
Tax price X v. low approval					0.0210
					(0.101)
Tax price X low approval					-0.114
					(0.0870)
Tax price X high approval					-0.239***
					(0.0719)
Tax price X v. high approval					-0.276***
					(0.0755)
Underwriting approval probability			0.457***	0.420***	
			(0.127)	(0.131)	
V. low approval probability					-0.304***
					(0.108)
Low approval probability					-0.152
					(0.0990)
High approval probability					-0.0284
					(0.0760)
Person_Fixed_Effects	Yes	Yes	Yes	Yes	Yes
Assets_Income	No	Yes	No	Yes	Yes
Standard errors in parentheses * p<.1, ** p<.05, *** p<.01.					

Exhibit 7: Effect of Partnership on long-term care insurance purchase

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Partnership LTCI Program				0.003	-0.016**		
				(0.006)	(0.007)		
PartnershipXLow Assets	-0.019	-0.014	-0.015**			-0.024***	
	(0.012)	(0.008)	(0.007)			(0.008)	
PartnershipXMed Assets	-0.019*	-0.004	0.002			-0.010	
	(0.010)	(0.008)	(0.007)			(0.007)	
PartnershipXHigh Assets	-0.008	0.028***	0.039***			0.026***	
	(0.017)	(0.008)	(0.008)			(0.007)	
Partnership X Approval Probability (continuous 0-1)					0.035***	0.019*	
					(0.011)	(0.010)	
Partnership X V. low approval							-0.008
							(0.008)
Partnership X Low approval							-0.005
							(0.006)
Partnership X High approval							-0.001
							(0.007)
Partnership X V. high approval							0.018**
							(0.009)
Mid assets	0.032***	-0.002	-0.000	0.008	0.008	0.001	0.008
	(0.008)	(0.007)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
High assets	0.085***	-0.017*	-0.019**	0.007	0.007	-0.017*	0.007
	(0.011)	(0.009)	(0.009)	(0.007)	(0.007)	(0.009)	(0.007)
Underwriting approval probability				-0.000	-0.018	-0.011	
				(0.009)	(0.012)	(0.011)	
Individual fixed effects	N	Y	Y	Y	Y	Y	Y
Years included	2002-2010	2002-2010	2002-2012	2002 - 2012	2002 - 2012	2002 - 2012	2002 - 2012
Observations	47352	47352	57403	57403	57403	57403	57403

Standard errors in parentheses. * p<.1, ** p<.05, *** p<.01.

Partnership program appears to increase long-term care insurance purchase only among individuals with the highest assets, while among the medium- and low-asset groups the effect is not statistically different from zero and even negative.

I find that the overall effect of the Partnership program is 0.003 and not statistically different from zero, a finding that is consistent with previous research. When the Partnership variable is interacted with underwriting approval in Model 5, by contrast, there is a strong modifying effect of underwriting on the policy variable, such that the policy effect is actually negative (not statistically significant) among those least likely to be approved, and the effect of the policy is increasingly positive with higher likelihood of underwriting approval (illustrated in Figure 3(c)). Lin and Prince (2013), in their preferred model that disaggregates by wealth group, show that only the high-wealth households (those above the eightieth percentile of total wealth) respond to Partnership incentives. Because household wealth and health status are strongly positively correlated, in their model wealth is likely to also be a proxy for high underwriting approval. Among the highest-asset group (Model 3), the Partnership program leads to a 3.9 percentage point increase in LTCI purchase. In Model 5, I estimate that underwriting is also an important modifier of the program effect, with a coefficient of -0.016 on the Partnership main effect ($p<0.05$) and positive .035 on the interaction term ($p<0.01$).

2.5 Discussion

Adverse selection in the market for long-term care insurance leads insurers to impose strict medical underwriting requirements. Those restrictions mean that long-term care insurance is actually available only to a more narrow slice of Americans who are not only wealthy enough to afford the premi-

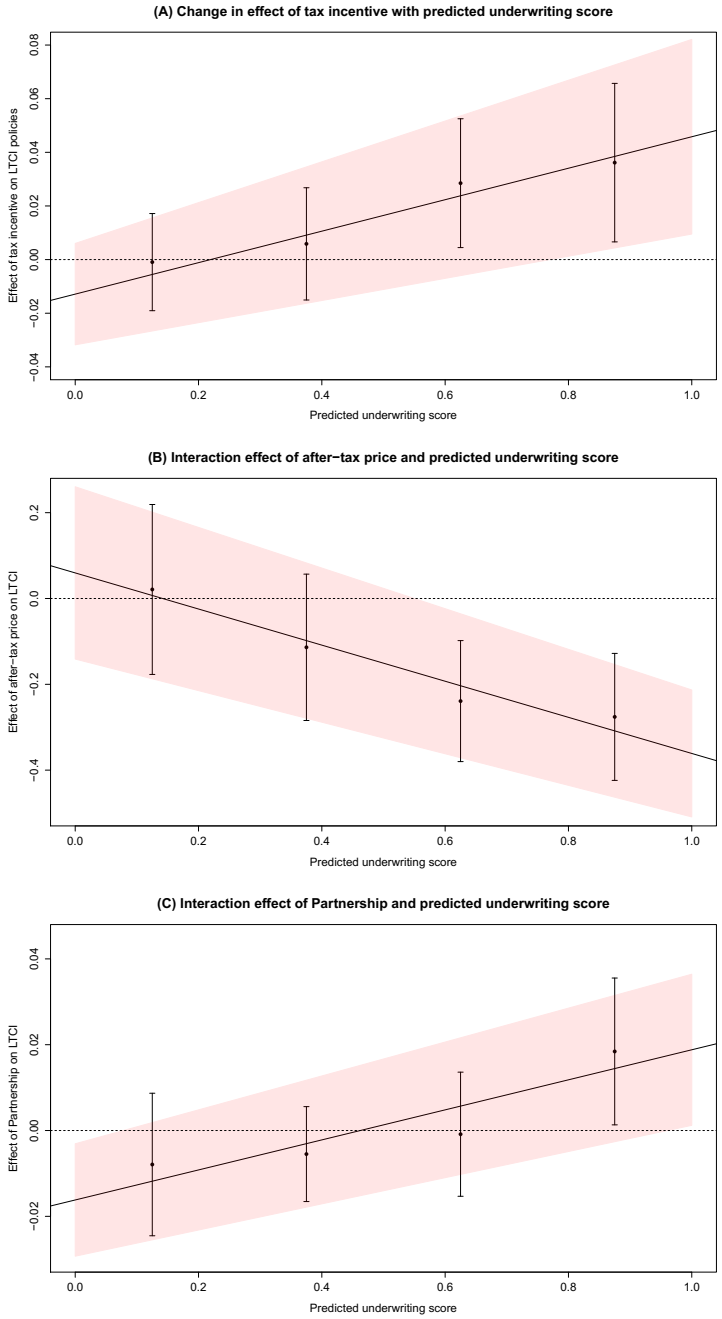


Exhibit 8: The horizontal axes represent underwriting score and the vertical axes represent the effect of the presence of the policy incentive in the preferred, individual-fixed-effects models. reported in Exhibit ??. Policy impact with a continuous interaction with underwriting are sloped lines with shaded 95% confidence interval and the differential policy effect by underwriting quartile. The shaded area represents the 95% confidence interval of the estimates.

ums, but also healthy enough to be at low risk of current or near future need for the services covered by the insurance. This study builds on previous research estimating the effect of policy incentives for long-term care insurance by assessing the role that medical underwriting plays in this market. It highlights that people's response to incentives is closely associated with their health status that predicts their need for long-term care. Those who may be most at risk of needing long-term care are the least likely to take advantage of state policy incentives because the supply-side forces shut them out of this market completely. I find that overall, the more likely an individual is to qualify for long-term care insurance, the greater their response to state incentives. Figure 3 gives a visual intuition for how the policies' effects intensify with higher underwriting probability. Predicted underwriting probability is on the horizontal axis, with the policy impact on insurance response on the vertical. For both the tax incentives and Partnership programs, less-healthy individuals more likely to be disqualified have no response statistically distinguishable from zero or even, for the least healthy, slightly negative response to the policy.

This analysis is the first to explicitly examine how medical underwriting may constrain the response of long-term care insurance buyers to incentives for insurance purchase. But limitations are important to note. First, though the inclusion of individual fixed effects controls for potentially endogenous factors that might be correlated with insurance purchase, the policy estimates could still be biased by economic or policy changes within states that correlate with the introduction of tax benefits or Partnership policies. The observed response in insurance purchase may not be to changes in price or Partnership per se, but instead to marketing and outreach from sellers and represent a one-time response. Furthermore, the HRS does not include any information on whether individuals

bought their policies through the private market or through employer-sponsored LTCI programs, where underwriting requirements are generally much less strict. Predicted underwriting scores represent estimates from underwriting decisions on a pool of applicants that self-selected and had already undergone field-underwriting review on the part of insurance agents. Imputed scores therefore represent a hypothetical probability that an individual will be able to purchase insurance, conditioned on their desire to seek it out and the affordability of the insurance product. As an aggregate measure of health status, it may represent consumer-side factors at play beyond the underwriting behavior of insurance companies. For instance, even though the models include controls for both income and assets, individuals in poorer health who have higher out-of-pocket medical expenses, making LTCI premiums unaffordable.

These results imply that responsiveness of demand for long-term care insurance to policy incentives needs to be interpreted in light of medical underwriting. Standard elasticities reported in previous research represent average effects across some individuals who would have no access to insurance at any price, even if they could afford it. Measuring heterogeneous effects gives a more accurate prediction of how different groups might respond. Where previous work has also looked at modifying variables, such as education and wealth, I argue that those characteristics were in fact proxies for the barriers to access that underwriting practices put in place. What seemed like a consumer-driven difference may, rather, be a supply-side phenomenon.

The effects of adverse selection and the fact of underwriting are important pieces of the puzzle to understanding the consumer behavior underlying trends in long-term care insurance purchase. Along with Medicaid crowd-out and low financial knowledge, underwriting practices also play

a role in limiting the overall prevalence of insurance. Policy makers continue to promote LTCI through tax incentives, Partnership programs, and encouragement campaigns. Among the market reforms suggested by the Senate Commission on Long-Term Care (2013) were to allow purchase of long-term care insurance with pre-tax dollars and to encourage financial products that combine annuities and long-term care insurance balancing opposing risks. But without addressing the underlying adverse selection issues that exclude a large portion of the population from being able to buy insurance at all, these types of reforms are unlikely to accomplish either the goal of protection Americans from potentially catastrophic long-term care costs, or substantially offset Medicaid expenditures.

Reforms that do address underwriting (such as guaranteed-issue insurance or reinsurance that would substantially decrease uncertainty for long-term care insurers) may have higher rates of take-up than would be previously thought. Previous studies have observed the impacts of these policies on insurance purchase to be modest. Those findings are consistent with the national trend, where the market for long-term care insurance is shrinking. For both tax incentives and the Partnership program, average effect of the presence of an incentive was substantially smaller than the differential effect with higher likelihood of passing underwriting standards, suggesting that if the difference is due to lower underwriting limits, then a combination of financial incentives could have a stronger impact than financial subsidies alone.

We demand rigidly defined areas of doubt and uncer-

tainty!

Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

3

Instrumental Variables for Multiple Treatments in Comparative Safety and Effectiveness Research

3.1 Abstract

Objectives. (1) To propose a checklist to evaluate a common class of instrumental variables (IVs), practice-pattern variation, which may be applied to a wide variety of comparative effectiveness research questions. (2) To demonstrate tests of instrumental validity in one application, namely using physicians' prescribing patterns as an instrument for antipsychotic medication choice. I extend a practical approach to IV analysis to research questions framed with multiple (>2) treatments.

Study design. A brief conceptual overview of common vulnerabilities of instrumental-variable analysis, together with examples of how to test for those vulnerabilities, accompanied by an empirical demonstration.

Empirical demonstration. An applied example estimates the comparative risks of five second-generation antipsychotics. Outcomes include new diagnoses of diabetes, hyperlipidemia, and being overweight. Physicians' prescribing preferences are used as an IV to predict treatment choices, and

IV estimates are compared to ordinary least squares.

Conclusions. Instrumental-variable analysis is a potentially powerful technique for use in comparative safety and effectiveness research, but uncritical use of the tool without case-by-case evaluation of the validity of the design can generate invalid findings. Widespread adoption of a basic protocol and road map for validating potential instruments, particularly the use of sensitivity tests and compliers analysis, would improve the quality of comparative effectiveness research.

3.2 Introduction

Comparative-safety and comparative-effectiveness research (CER) compares the benefits and risks of alternative treatments to inform clinical practice and health-care policies based on the best evidence. Randomized, controlled trials (RCTs) provide the best evidence of causal effects of one treatment or policy versus another, because randomization can assure that the treatment and comparator groups are statistically equivalent in terms of characteristics that affect the outcome. Nevertheless RCTs have limitations. Large-scale trials are expensive, and to attain the maximum statistical power per dollar spent, an RCT will often target a limited, homogeneous population, with the aim of securing approval from regulators and payers for the treatment's use among similar patients. Furthermore, Food and Drug Administration (FDA) approval of most new drugs requires demonstrating efficacy and safety against a placebo or against prevailing treatment practice. When multiple treatment options are available, therefore, head-to-head trials among all the treatment options may not exist. As a result, treatments can gain widespread use, even as gaps persist in knowledge about how similar treatments compare to one another on safety and effectiveness in real-world settings. Filling those

gaps in knowledge is important, because clinicians often face a decision not simply about whether to recommend a particular treatment on its own merits but about which treatment among a class of similar candidates will best suit the patient.

The increasing availability of electronic health information and administrative records, together with the tools to analyze them, present an opportunity to extract information about how drugs perform in real-world settings for patient who differ from the RCT sample in age, sex, main diagnosis, or complicating comorbidities. Administrative databases allow the researcher to study patients over several years and in a variety of provider settings. The classic challenge of observational research, however, remains. That is, where the researcher does not actively assign patients to treatment and control groups, the mechanisms that select subjects into the treatment groups, such as patient preference, differences among providers, and access to health care, create systematic differences between the treatment and control groups—differences that are not evident in the administrative data.

Instrumental variable (IV) analysis is one tool for addressing that weakness in observational studies. By exploiting a "natural" source of variation in who receives treatment, the method confers to observational studies some of the advantages of RCTs. One class of instrumental variables that have been used to address treatment-selection bias in CER is that of "preference-based" instruments. Variations among providers or groups of providers influence how or which medications or procedures are used. Those differences have formed (it is assumed) by some mechanism that is independent of an individual patient's health profile and potential outcomes. Some examples include differential distance to a specialty care provider (McClellan, McNeil, and Newhouse 1994); administrative instruments, such as insurance copayment amounts (Cole et al. 2006); and treatment patterns correlated

to geographic location, facility, or physician (Brooks et al. 2003, Wang et al. 2005, Brookhart et al. 2006, Rassen et al. 2009, Sanghavi et al. 2015). Preference-based instruments are potentially useful tools, but they should be approached with caution. Evaluating a candidate instrument is tricky, because several of the assumptions are not only unfamiliar to many researchers, but cannot be tested directly or definitively. Treatment effects can be difficult to interpret, because they estimate based on the slice of the treated population that responds to the instrument.

In answer to growing interest in the use of instrumental-variable methods to compare treatments in health studies, researchers have made several contributions that offer practical guidance on the use of instrumental variables (Landrum and Ayanian 2002; Brookhart and Schneeweiss 2007; Baiocchi, Cheng, and Small 2014; Pizer 2016). The purpose of the present paper is to build on that literature in two ways: first, to focus on the heuristics for applying knowledge and intuition to the task of validating a candidate instrument and interpreting results; and second, to outline a practical framework for CER researchers to follow, with attention to the challenges of administrative databases. I discuss descriptive analysis, which can clarify the characteristics of the population to which the study applies, and sensitivity analyses, which can set upper and lower bounds on a causal-effect estimate.

I demonstrate widely recommended assessment techniques using an example with multivalued treatments. To date this topic has not been described in pedagogically oriented papers on IV estimation, the vast majority of which focus on dichotomous or continuous treatments with monotonic effects. Specifying the clinical options in binary terms is more straightforward, both to estimate and to report, than more-complex conceptualizations. A dichotomous treatment/control model, however, may not reflect the real-world clinical decision process, when more than two treatment options

exist within a therapeutic class. Results may be difficult to interpret, in that the "control" reference group may represent several possible courses of treatment—especially when alternative treatments have heterogeneous potential outcomes for the patient (Brooks et al. 2003).

3.3 Applied example

The framework and methods explained in this paper are described in the context of a study of the adverse effects of the five atypical antipsychotic medications that are commonly prescribed in the U.S.: aripiprazole (Abilify), risperidone (Risperdal), olanzapine (Zyprexa), quetiapine (Seroquel), and ziprasidone (Geodon). The study compares the risk of diabetes, obesity, and dyslipidemia among new users of the five medications. The study illustrates the use of a preference-based instrument for a multi-treatment estimand, using administrative data.

Some background: These drugs are the treatments of choice for schizophrenia spectrum disorders, bipolar disorder, and psychotic depression (Pincus et al. 1998). Atypical antipsychotics are also prescribed "off-label," i.e., for purposes other than the federally approved ones, in populations that differ in many respects from the clinical-trial study populations, and often by physicians who are not trained psychiatrists (Alexander et al. 2011; Verdoux, Tournier, and Bégaud 2009). Clinical-trial evidence shows that among patients with schizophrenia, atypical antipsychotics, particularly olanzapine and risperidone, increase the incidence of metabolic syndrome X, a constellation of rapid weight gain, new-onset diabetes, and dyslipidemia, which increase the risk of cardiovascular disease (American Diabetes Association 2004; Casey 2005). Because patients with severe mental illness such as schizophrenia and bipolar disorder are already at higher baseline risk for heart disease and diabetes

and tend to have unhealthy behaviors like smoking, uncertainty exists about the relative contribution to the patients' comorbidities of treatment effects versus underlying disease risks. For that reason, it is important to assess what risks the drugs pose and how they compare with one another in real-world settings.

A naive comparison of side effects of atypical antipsychotics by individual treatment might not clarify the differences in risks among the drugs because of selection effects. For instance, physicians may be less-likely to prescribe olanzapine and quetiapine, which clinical trials suggest have drastic effects on weight gain (Gareri et al. 2014), to patients who are already overweight; or to prescribe risperidone, which has a high risk of ischemic stroke (Shin et al. 2013), to patients with risk factors for stroke. In both of those scenarios, the physician's prescribing behavior with respect to an individual patient's risks is likely to mask the average, causal effect of treatment choice on adverse outcomes. For that reason, methods often used in observational studies, such as multivariate-adjusted regression models or propensity weighting, which depend on the assumption that treatment is random, conditional on observed covariates, could produce biased results if not all of those selection factors were available in the data (which they rarely are).

Atypical antipsychotics are associated with risk factors, together known as metabolic syndrome, which drastically increase risk of cardiovascular disease and mortality: abdominal weight gain, glucose tolerance and diabetes, dyslipidemia (low blood levels of high-density lipoprotein and high triglycerides), and hypertension (Casey 2005). Exhibit 1 summarizes clinical-trial evidence of selected adverse effects that have been observed for the study drugs.

Olanzapine appears to have the largest effect on weight gain as well as diabetes and dyslipidemia,

Exhibit 1: Side effects associated with atypical antipsychotics

	Weight gain	Diabetes	Dyslipidemia
Aripiprazole	?/+	?	?
Olanzapine	+++	+	+
Quetiapine	++	D	D
Risperidone	++	D	D
Ziprasidone	?/+	?	?

Notes:
D indicates discrepant results;
? indicates no effect;
+ indicates mild effect;
++ indicates moderate effect;
+++ indicates strong effect

Table is adapted from American Diabetes Association 2004.

compared to the other atypical antipsychotics. Risperidone and quetiapine have moderate effects on weight, and the evidence as to their effects on diabetes and dyslipidemia is conflicting (American Diabetes Association 2004). Aripiprazole appears to have lower risk of inducing diabetes. The working hypothesis of the present study is that comparing the causal effects of these drugs to each other among a population-based sample of Medicare Advantage enrollees will mirror the findings from clinical trials.

This analysis defined patient outcomes in terms of having newly incident cases of obesity, diabetes, or dyslipidemia appearing in their Medicare claim record. Appendix C contains precise definitions of how the variables were constructed, with diagnostic codes. Subjects had a new case if there was no diagnosis observed in the 365 days prior to the prescription index date, and one was observed in the subsequent 365 days. In addition to claims diagnoses, patients were considered to have dia-

betes or dyslipidemia if they filled a prescription at least once for an antidiabetic or antilipemic drug, respectively.

3.3.1 Study population and data

This study was based on 2,049 new users of atypical antipsychotics diagnosed with schizophrenia, who were drawn from a population of enrollees in a Medicare Advantage-Part D plan. From the Medicare prescription drug event (PDE) database I retained individuals who had an atypical antipsychotic prescribed after a 180-day period with no prescription fills for any atypical antipsychotic. This six-month lead-in period was intended to make it more probable that changes in the outcomes would be attributable to initiating antipsychotic use, since effects on weight gain, blood cholesterol levels, and diabetes are most drastic in the first six to twelve months after initiating antipsychotic therapy. I defined treatment assignment as which of the study drugs the individual received (D=d1, d2, d3, d4, d5, for aripiprazole, olanzapine, quetiapine, risperidone, and ziprasidone, respectively), and each element dk of D is an indicator equal to 1 to indicate the drug that was prescribed and 0 for the others that were not. Provider identifier numbers in the PDE files were used to define prescribing physician's prescribing history among patients represented in the database.

3.4 Instrumental Variables Basics

To give unbiased estimates of a causal effect, an instrumental variable must satisfy at least four key assumptions (Angrist, Imbens, and Rubin 1996): (1) the treatment status of one patient does not affect the outcomes of other patients (no spillover); (2) the instrument affects whether the subject

receives treatment or not, and only in one direction (instrument relevance); (3) the instrument must be effectively random with respect to any patient characteristics that are related to their outcomes, that is, it does not share any causes with the outcome (ignorable assignment); and (4) the instrument is unrelated to outcomes except as it operates through the treatment assignment (the exclusion restriction). Technical discussions of the assumptions, including why they are necessary to achieve a valid IV estimate, have been supplied in the literature (Angrist, Imbens, and Rubin 1996; Imbens and Rubin 2015). The latter two assumptions, ignorable treatment assignment and the exclusion restriction, rely on the conceptual validity of the instrumental variable and can only be tested indirectly. All IV studies are vulnerable to violations of these assumptions, in that they will give invalid results if the assumptions are not satisfied. Yet as few as 6 percent of CER studies that use IV analysis consider variables that would measure violations of assumptions three and four (Garabedian et al. 2014). Furthermore, only some 20 percent use any falsification test whatsoever (Swanson and Hernán 2013).

In addition to carefully examining these vulnerable assumptions about internal validity, IV studies should also describe the external validity of the causal effect being estimated. A randomized trial with noncompliance offers some language to describe the types of subjects: "Compliers" are subjects whose treatment status depends on the randomized assignment, and "noncompliers" are those whose treatment is unaffected by assignment. The average treatment effect, also known as the intent-to-treat estimate, is the effect of the instrument on the outcome, averaged across all study subjects (both compliers and noncompliers). The local average treatment effect (LATE), which is the IV estimand, is the effect of treatment on the treated (compliers) only. In CER, the LATE is

usually reported, unless the causal effect of the instrument itself is of interest. It is not usually safeto assume that the treatment effect is homogeneous between the two types of subjects, since patients whose treatment is influenced by the instrument may differ in important ways from those who will. Therefore, it is generally important to characterize the subsample of compliers.

The purpose of this paper is to set up a framework for evaluating how well a candidate instrument satisfies the four criteria (instrument relevance, exclusion restriction, ignorable treatment assignment, and plausible external validity) in a CER study that uses provider preference as an instrument for evaluating causal effects in studies withadministrative data. For each assumption, I describe the relevant conceptual considerations given the choice of instrument, and I demonstrate what the empirical diagnostics are that can assess the validity of the assumptions being made.

3.5 Instrument Relevance

An IV is considered relevant if it is good at predicting among the choices of different treatments, and it is considered weak if its predictive power is low. Explicitly positing the theoretical basis for the instrument is an important pre-analytical step, because the heuristics for examining the theoretical basis for an instrument also generate hypotheses for validating the instrument against challenges to the assumptions.

3.5.1 Is the research question suited to a preference-based instrument?

When providers or groups of providers have different preferences or treatment algorithms dictating how medications or medical procedures are used, it may be possible to use observable variation

in treatment history at the physician, group, or facility level as a proxy for the underlying practice patterns. Several considerations in selecting a "preference-based" instrument apply. Area of clinical ambiguity: Preference-based treatment has better conceptual validity when there is no strong consensus in the medical literature about best practice among the treatments in question—which also of course makes the treatment choices attractive candidates for use in CER. If treatments require different levels of investment in equipment or skills, however, then preference patterns will depend on regional or individual access to capital and training, which will likely be correlated with other aspects of care quality and outcomes. Practice-pattern variation: Researchers should determine whether there is actually variation at the provider level, with patient clinical factors playing a minor role in how physicians select which treatment to recommend to their patient. Even if there are several treatment options available, if there is no difference or little difference among physicians in how often they choose a particular treatment then the instrument will be weak and will magnify even small sources of bias.

Idiosyncratic patient response: Economic theory suggests that physicians will apply therapeutic norms to their prescribing decisions when the benefits of customizing are lowest (Frank and Zeckhauser 2007). If there is heterogeneity among patients in drug efficacy and side effects, it may be difficult to predict for a given patient how he or she will react to treatment. In such a scenario, having a "ready-to-wear" first-line treatment may be a sensible physician response to complex decision-making. Physicians will then tend to begin patients on the treatment with which they are most familiar, before branching out to alternative treatment options.

Atypical antipsychotics are promising candidates for the physician-preference instrument. They

are therapeutically similar to each other, and what efficacy they will have and what adverse effects they will produce in particular patients are difficult to predict. Because the patients may respond well to one drug in the class but not others, payers tend to include all of the drugs in their formularies. There is evidence that physicians tend to gravitate in their prescribing behavior toward one drug over others; and furthermore that prescribing behavior of antipsychotics is highly variable among individual prescribers, but evens out and is less concentrated at county and regional levels (Taub et al. 2011). Therefore physician preference is likely to influence first-line treatment with antipsychotics strongly. Since the differences do not appear to arise regionally, it is less likely that the differences correlate to regional variation in health spending and quality than if physicians near each other tended to develop similar preferences.

3.5.2 How should the instrument be defined?

The goal in operationalizing an instrument is to get the best possible strength of identification, while nonetheless maintaining the integrity of the instrument as an effectively randomized conditional on observed covariates.

Different approaches to defining prescribing preferences are possible. One is to use the most-recent prior prescription that the physician has initiated as an instrument. Thus the variable would be equal to a vector of five indicators for patient i and physician j: $Z_{ij} = \{z_{ij1}, z_{ij2}, z_{ij3}, z_{ij4}, z_{ij5}\}$. For example, if a physician's most recent prescription for an atypical antipsychotic i were for quetiapine, then for the next patient we would have $Z_{ij} = \{0, 0, 1, 0, 0\}$. The advantage of such an approach is that the instrument updates over time, with every prescription that the physician issues. In a setting

where treatment practices are changing rapidly over the course of months, this type of definition allows the researcher to take advantage of trends as a source of treatment variation. That variation may be plausibly exogenous to patient characteristics when the population of patients receiving treatments within the therapeutic class is relatively stable. The disadvantage of the approach, however, is that relatively little information is contained in a single previous prescription event, so the measure is noisy and may not do a good job of predicting the next prescription choice.

Another approach is to define prescribing history over the entire study period, or within a defined calendar time. This approach allows the instrument to capture all available information in the database about provider preference, in order to get the strongest possible prediction. It does not, however, capture time trends; and furthermore it leads to the conceptually dubious choice of predicting treatment from events that postdate the index prescription. That may be acceptable if treatment patterns are stable, but the researcher should be clear about the assumptions being made.

3.5.3 Applied example

I combined approaches, by using all of a physician's observed prescription events prior to the index patient to define preference. One way to operationalize the history is with a vector of indicators ZijI, as in the previous example, where each Zijk indicates which drug was prescribed most often by physician j prior to patient i. Another is to designate continuous zik=p, the proportion of the time that drug k was prescribed among previous events. So if for an index patient we observed that the physician had prescribed risperidone 4 times and quetiapine once, then that would yield $Z'_{ij} = \{0.0, 0.0, 0.2, 0.8, 0.0\}$.

The first definition has the advantage of simplicity and is more amenable to commonly used diagnostics. The second definition contains more information about prescribing preference, in that it can distinguish a physician who almost exclusively prescribes one drug from a physician who tends to use one treatment marginally more often but also has a clearly observable second choice. For the best instrument strength, I chose the latter definition to estimate the causal effect of treatment. It is common when studying continuous treatments, however, to dichotomize into groups for the purpose of raw data description and instrument diagnostics. For such diagnostics, I used the multinomial "most-frequent" version, which allowed me to sort the sample into groups according to instrument value.

3.6 Ignorable treatment assignment and the exclusion restriction

A major potential vulnerability of instruments based on preference and area variation is that the variation in the instrument may arise from the care patterns that influence patient care in other ways, or be strongly correlated with persistent differences in the risk factors for the mix of patients that particular providers treat. Such potential confounders would violate the assumption (3), namely, that the instrument assignment is effectively random among patients who are similar in observable ways; and assumption (4), that the instrument only affects patients' outcomes by influencing their selection of individual treatment. Testing the ignorable-treatment-assignment and exclusion-restriction requirements demands thinking carefully about the theoretical basis for the choice of instrument. The researcher must ask, for instance, how physicians become concentrated in their prescribing behavior and how that behavior relates to their patient panels and other practice patterns. Exhibit 2

frames categories of potential confounders for the researcher to investigate with regard to his or her proposed instrument.

3.6.1 Ignorable treatment assignment

Differences in practice patterns might be correlated with the patient mix, which would threaten the ignorability assumption if some aspects of case-mix severity were unobservable. It is useful to apply researcher knowledge to measurable risk factors and see whether they differ between groups of providers with different preferences.

If the study is based on electronic health records (EHR), the database can provide a wealth of clinical information for estimating patients' baseline risk factors with more accuracy than can typically be extracted from claims data. In our study of the metabolic side effects, for example, information such as patient weight, blood pressure, and hemoglobin-A1c tests would offer nearly as much clinical information as would be available to the prescribing physician.

In an administrative database, baseline comorbidities need to be established based on billing codes associated with previous claims over some defined period, usually a year. Establishing patient history with claims is more difficult than with EHR, because coding is inexact and because being able to observe prior disease depends on the patient's use of the health-care system in the retrospective period. The decision about length of evaluation period involves a trade-off between statistical power and the sensitivity of the algorithm to identifying disease. Periods from six months to two years, for instance, could be chosen, depending on the typical accuracy of coding and on disease prevalence. In ideal scenarios, the researcher would be able to validate the diagnosis algorithm

Exhibit 2: Framework for assessing instrumental-variable assumptions

Assumption of ignorable/effectively-random assignment. Are there persistent differences among providers in the characteristics of their patient panels?

Patient characteristics that are potential confounders	*Strategies to measure confounders or proxies in administrative data*
Health status	
• Risk factors associated with outcome • Health comorbidities • Health behaviors • Clinical indication for treatment	• If electronic health record (EHR) data are available, identify patient diagnostics relevant to the outcome. • If using claims database, use ICD codes to identify baseline risk and pre-existing health conditions. • Identify recent prescriptions • Condition study sample on clinical indication
Access to health care	
• Insurance status • Quality of providers and facilities where the patient seeks care • Care coordination/primary care	• Insurance plan information from claims databases • Local provider quality metrics from Medicare scorecard data
Socioeconomic characteristics	
• Race • Education • Income • Urban/rural	• Race information sometimes available in Medicare membership files • Use Medicaid and low-income subsidies as proxy for income level • Census data to calculate income, education, and rural characteristics at the census-block level and calculate regional

Exclusion-restriction assumption. Are there mechanisms by which the instrument assignment could affect the outcome other by choice of one of the treatments being studied?

Provider experience	
• Experience with treatment options • Specializes in the conditions being treated	• Examine relationship between prescriber volume and preferences • Condition sample on provider specialty • Identify off-label prescribing • Provider specialty
Quality of care	
• Prescribes treatment when contraindicated • Appropriate screenings and follow-up care • Care coordination	• Recent and concurrent co-prescriptions with protective drugs and/or harmful drugs • Patients receive recommended screenings and monitoring for expected adverse effects

within the study population, by surveying a subsample of charts or electronic health records.

Socioeconomic status (SES) is closely linked at the population level with commonly studied outcomes like mortality, and in ways that are impossible to adjust for completely in observational studies. Yet only a fraction of CER studies that use instrumental variables attempt to control for those factors or to evaluate the choice of instrument with regard to such controls (Garabedian et al. 2014). One reason is that it is often difficult to tease out variables like income, education, and race from administrative data. If addresses or other fine-grained geographical codes for the study subjects are available, however, proxy measures for SES can be constructed from population-based data and linked by census block. Such geographic information can also be used to construct measures of urban density, as well as area-level measures of health-care quality. Medicare's Physician Compare and Hospital Care databases are amenable to such analysis.

Access to and quality of other health care that patients receive should also be examined by preference group. Billing databases often come from the payers themselves, and information on insurance status or type of insurance, as well as the details of particular plans, can be used to identify differences in provider networks or out-of-pocket costs. If the study sample has differences in pharmacy benefits, then administrative restrictions on treatments available, such as tiered formularies and requirements for prior authorization, may also be correlated with preferences. For example, a physician in a particular state with a large proportion of Medicaid patients in his or her case mix is likely to have more experience prescribing drugs that are favored in the formulary rules for that state, because formulary restrictions are commonly used tools for controlling costs. If care from other

providers is potentially important to patients' outcomes, it may be useful to determine other usual sources of care, besides the prescribing physician, for patients in his or her panel.

3.6.2 Exclusion restriction

The exclusion-restriction assumption rules out mechanisms by which provider preference can affect outcomes for an individual patient other than the choice of treatment. If providers' skills, experience, or technological resources differ along with preferences, then the instrument is invalidated, because the causal effect that the instrument seeks to estimate captures those other factors affecting patient health and falsely attributes whatever their effects may be to the treatment. The researcher can look at related provider behaviors to tease out concurrent effects.

Providers who have adopted new treatments ahead of their peers may also be quicker to use new technologies more generally, so if some treatments in the study are newer than others, it may be useful to look at providers' preferences for related treatments to see if they follow similar patterns. For instance, among the atypical antipsychotics analyzed in the applied example, ziprasidone and aripiprazole were more recently approved by the FDA (in 2001 and 2002 respectively) than the comparator treatments. Physicians who adopted those drugs might also be more likely to use weight-management protocols or prescribe newer diabetes-management drugs such as sitagliptin. Provider specialty and experience are further potential violators of the exclusion restriction. Concomitant and recent use of both protective and risk-increasing treatments can also vary with physician behavior and are important indicators of experience and quality. In the example, I examined baseline use of antidiabetic and cholesterol-lowering drugs. Concentrated prescribing behavior occurs most with

low-volume prescribers relative to high-volume prescribers. The latter tend to be more diverse in their treatment choices, presumably because they have more experience in prescribing the various drugs and assessing their effects on different patients.

Off-label prescribing can muddy the picture because patients who are receiving drugs for reasons other than the approved clinical indications will be different from those with conditions that are approved for treatment. Furthermore, physicians who are prescribing off-label may be not just be inexperienced but coprescribing harmful combinations or prescribing to contraindicated patients, for whom the drugs are dangerous. For those reasons, the researcher may want not just to inspect and control for indication but restrict the sample to patients with the same indication for treatment. Furthermore, physicians who prescribe off-label may often be low-volume prescribers, and thus, independently of other considerations, they may be expected to have less experience managing the adverse effects that the medications produce.Measured confounders can of course be controlled for in IV estimation. Apparent association between the instrument and observed confounders is nonetheless problematic, because the observed confounders might be merely proxies for an underlying characteristic that is only partially observable. If patients who are stratified by the instrument preference group are similar in terms of observed characteristics, that suggests that patients are also similar between instrument assignment on unobserved confounders. Conversely, however, imbalance on observed covariates should make the researcher suspicious of whether his or her instrument achieves ignorable treatment assignment.

Standardized difference (SD) compares the difference in means between the group of those patients who are encouraged by physician preference to take a treatment k (Zk=1) and the groups of

those not encouraged (Zk=0), in units of pooled standard deviation. Statistically significant differences may not necessarily disqualify an instrument, especially in large samples where clinically small differences can achieve statistical significance at common thresholds. Instead, simple standardized difference (SD) can be computed as follows:

$$SD_k = \sum \frac{(\bar{x}_k - \bar{x}_{j \neq k})}{\sqrt{\sigma_k^2 + \sigma_{j \neq k}^2}} \tag{3.1}$$

In the case where there are two treatments, *k* and *j*, just one SD per covariate can be reported. With more than two treatments, the research could choose to report the average SDs between treatment k and alternative treatments j for each covariate-treatment pair. While no strict cutoff exists for acceptability, a standardized difference of greater than 0.20 across instrument assignments should elicit further inspection.

3.6.3 Falsification tests

Falsification tests examine whether an instrument is correlated with an outcome that should be unaffected by the treatment, or with the outcome of interest in a closely related population that is not subject to treatment. Pizer (2016) provides a recent and thorough pedagogical discussion of how to conceptualize falsification tests for preference instruments. Good candidate tests might involve finding a similar population that should be unaffected by treatment choice, such as the population of patients who were treated by the prescribing physician but did not receive an antipsychotic; or limiting the sample to individuals who were prescribed an antipsychotic but discontinued treatment

after a thirty-day supply. Another type of falsification test involves estimating a "false" outcome that should not be affected by the treatment. Finding such a straw-man outcome in the case of atypical antipsychotics is difficult, however, since the constellation of factors associated with metabolic syndrome potentially are risk factors for many other patient outcomes, from heart disease to cancer. Like good instrumental variables, good falsification tests require a certain amount of cleverness and ingenuity on the part of the researcher—as well as cooperation from the data.

3.6.4 External validity checks: characterizing the compliers

Subjects whose exposure to treatment is affected by the instrument are likely to be different along several dimensions from those whose exposure is independent of the instrument. In the context of a preference-based instrument, the non-compliers might be patients who have clear contraindications to one of the treatments being studied, whose diagnosis has a stronger evidence base over other treatments, or who have a strong personal preference for one treatment over others, which leads them to override the norms that the physician or facility where they are receiving care would otherwise impose. It is not possible to identify from the data which subjects are compliers, since we can only observe the treatment under one possible assignment and cannot know the counterfactual under a different assignment. It is possible, however, to characterize the distribution of compliers, in terms of observed covariates. Extending Baiocchi, Cheng, and Small (2014) to the situation of multiple treatments, the mean of a covariate X_i among complier to treatment k can be written:

$$E[X|C = co] = \frac{E(D|Z = 1, X = x) - E(D|Z = 0, X = x)}{E(D|Z = 1) - E(D|Z = 0)} \tag{3.2}$$

Where D_k and Z_k are indicators of treatment k and the instrument assignment k, respectively. $E[X|C_k = co]$ is expectation of X, inverse-weighted by probabilities of compliance. With non-binary treatment, this results in characterizing compliers with each type of treatment. In practical terms, the steps can be operationalized as follows:

3.6.5 Sensitivity analysis

The purpose of sensitivity analysis is to put upper and lower bounds on how violations of either ignorable assignment or exclusion restriction will bias the IV estimate. The exclusion restriction states that, conditional on the actual atypical antipsychotic prescribed, if a patient would have developed diabetes had she or he been prescribed by the physician who preferred that drug, then he or she would also have developed diabetes if prescribed it by a physician who preferred a different drug. But suppose there is an unmeasured confounder, U, associated with physician preference, which exerts an additive effect on the probability that a patient will develop diabetes. Excluding that variable from either the OLS model or a two-staged–least-squares IV model would create bias in the estimators. But which would be further from the true causal effect? To start, supposing an unmeasured confounder U, exposure to a treatment D, and physician preference instrument Z. A simple model where the outcome is determined only by treatment D and a binary confounder, U, is the following:

$$Y = \alpha_0 + \alpha_1 D + \alpha_2 U + \varepsilon \tag{3.3}$$

If this is a true model of the way the world works, and ?1 is the true causal effect of D on Y, then

the bias from estimating an OLS model of that excludes U can be written in terms of the difference in prevalence of the confounder between levels of individual treatment:

$$BIAS(\hat{\alpha}_1^{ols}) = \alpha_2(E[U|D=1] - E[U|D=0]) \tag{3.4}$$

If the exclusion restriction is violated, then the confounder is related to the instrument in some way other than through exposure to treatment, so $E[U|Z,d] \neq E[U,d]$. The bias in the IV estimator is proportional to the difference in prevalence of U across levels of the instrument, over the probability of being a complier:

$$BIAS(\hat{\alpha}_1^{iv}) = \alpha_2 \frac{E[U|Z=1,d] - E[U|Z=0,d]}{E[D|Z=1] - E[D|Z=0]} \tag{3.5}$$

One way to evaluate the bias of the IV estimate is to compare it to OLS. The ratio of the bias of the IV estimate to the bias of the OLS can be written as follows (Baiocchi, Cheng, and Small 2014; Brookhart and Schneeweiss 2007):

$$Biasratio = \left| \frac{\frac{E[U|Z=1,d] - E[U|Z=0,d]}{E[D|Z=1] - E[D|Z=0]}}{E[U|D=1] - E[U|D=0]} \right| \tag{3.6}$$

For the IV estimator to have less asymptotic bias than the OLS estimator, the bias ratio has to be less than 1. The next step is to posit values for the missing confounder. One way to do that is to use measured variables as proxies for U and then calculate the bias in the IV that would result, if those measured covariates were left out. In the applied example, I estimated the probability of for each

atypical antipsychotic treatment, the association between physician preference and the covariate, and the effect of the covariate on incidence of diabetes, using the observed data.

3.7 Results

Summary statistics describing patients in the study sample are given in Exhibit 3. Fifty-four percent of the sample was female. The mean age was 59 and roughly half the sample was over 65. That distribution indicates that a sizeable fraction of the sample was qualified for Medicare through disability, which was expected, given that these individuals all had serious mental illness. In addition to schizophrenia, a large fraction of the sample also had other mental health diagnoses, including 63 percent with diagnoses of depression and 44 percent with bipolar disorder. Those additional diagnoses may have been true co-occurring disease or may have represented an evolution in diagnosis, since the diagnosis codes were assigned over two years.

Additional mental health diagnoses may by proxies for the severity of disease or for the complexity of the patient's case history. Therefore it can be useful to examine whether they vary by treatment. Despite the American Diabetic Association (ADA) recommendations that all patients who receive atypical antipsychotics should be screened for diabetes and dyslipidemia (ADA 2004), only a small portion of this sample received screenings in the year prior to receiving treatment: 7.6 percent and 9.9 percent respectively. Since such screenings represent recommended clinical-practice guidelines, differences in screening rates among treatment choices and among preference groups could be a red flag as to differences in how physicians manage their patients' side effects, pointing to a potential violation of the exclusion-restriction assumption.

Exhibit 3: Summary statistics sorted by by individual treatment

	All mean	Arip mean	Olan mean	Quet mean	Risp mean	Zipr mean
Female	0.54	0.59	0.47	0.51	0.56	0.54
Age	59.1	56.6	60.6	60.3	60.3	53.1
Age 18-49	0.28	0.34	0.26	0.27	0.24	0.36
Age 50-64	0.37	0.37	0.31	0.36	0.36	0.52
Age 65-79	0.27	0.24	0.34	0.26	0.29	0.10
Age 80+	0.085	0.045	0.10	0.11	0.099	0.012
Medicaid eligible	0.30	0.34	0.28	0.31	0.28	0.35
Part D low income	0.46	0.54	0.42	0.46	0.40	0.55
Medicaid add-on	0.36	0.41	0.34	0.37	0.33	0.43
Poverty Percent All Ages	15.0	15.0	15.0	15.0	15.0	15.0
Other paranoia	0.12	0.084	0.16	0.10	0.13	0.099
Other psychosis	0.42	0.32	0.47	0.45	0.44	0.33
Substance abuse	0.18	0.16	0.18	0.18	0.17	0.19
Dementia	0.15	0.075	0.16	0.20	0.16	0.062
Depression	0.63	0.67	0.62	0.65	0.59	0.57
Bipolar disorder	0.44	0.48	0.43	0.45	0.39	0.49
Diabetes screen	0.076	0.087	0.066	0.072	0.073	0.099
Dyslipidemia screen	0.099	0.10	0.097	0.10	0.090	0.11
Baseline obese	0.12	0.14	0.085	0.11	0.13	0.12
Baseline diabetes	0.52	0.56	0.47	0.50	0.52	0.54
Baseline dyslipidemia	0.53	0.50	0.50	0.54	0.53	0.56
Obesity	0.067	0.047	0.047	0.066	0.072	0.13
Diabetes	0.13	0.15	0.14	0.13	0.13	0.12
Dyslipidemia	0.13	0.15	0.16	0.12	0.13	0.14
Observations	2049	358	258	572	699	162

About half (52 percent) of the sample had diabetes at baseline, 12 percent were obese, and 53 percent had dyslipidemia. Patients who received aripiprazole had higher baseline prevalence of obesity and diabetes (14 and 56 percent), whereas patients who received olanzapine had lower prevalence (8.5 and 47 percent). That pattern would be consistent with physicians being more likely to prescribe aripiprazole, the lower-risk drug, to their diabetic and obese patients.

3.7.1 Covariate balance

Exhibit 5 shows the SD statistics plotted by individual treatment (white hollow shapes) and physician preference (red solid shapes). Each shape represents a different drug, and each line down the vertical axis is a covariate. An SD equal to zero represents balanced randomization of the instrument, so points further to the right represent worse covariate balance than points closer to the vertical axis. If the instrument works as it should to achieve better covariate balance, stratifying on physician preference groups (the solid red points) should bring the SD closer to the zero axis than the hollow white points. For instance, the instrument seems to do a good job of improving balance with regard to the patients' sex: for each shape, the red point is closer in than the white point. Aripiprazole and ziprasidone (circles and inverted triangles) are unbalanced on several characteristics, both on the treatment and on the instrument. Furthermore, it appears that aripiprazole stratified by preference (red circles) actually has worse balance than when stratified by instrument for sex, Medicaid eligibility, lipid screening, baseline diabetes, and baseline dyslipidemia.

3.7.2 Applied diagnostic: evaluating instrument strength

Table 4 shows linear regression models of individual treatment choices on the continuous measure of physician preference, i.e., the first-stage equation of a two-stage-least-squares IV estimating procedure. Each column is a regression model, where the dependent variable is a dichotomous indicator

Exhibit 4: Instrumental-variable first-stage estimates: effect of provider preference on individual treatment.

	Individual treatment				
	(1)	(2)	(3)	(4)	(5)
	Arip	Olan	Quet	Risp	Zipr
Arip % prescribed	0.233***	-0.00453	0.0361	-0.254***	-0.0103
	(4.50)	(-0.10)	(0.59)	(-3.94)	(-0.28)
Olan % prescribed	-0.0258	0.417***	0.0175	-0.401***	-0.00761
	(-0.53)	(10.03)	(0.31)	(-6.64)	(-0.22)
Quet % prescribed	-0.00931	0.00175	0.337***	-0.317***	-0.0124
	(-0.25)	(0.06)	(7.74)	(-6.86)	(-0.48)
Zipr % prescribed	-0.0743	0.0367	0.0711	-0.335***	0.302***
	(-1.29)	(0.74)	(1.05)	(-4.67)	(7.45)
Observations	2049	2049	2049	2049	2049
ftest	8.052	34.28	21.15	15.97	17.89

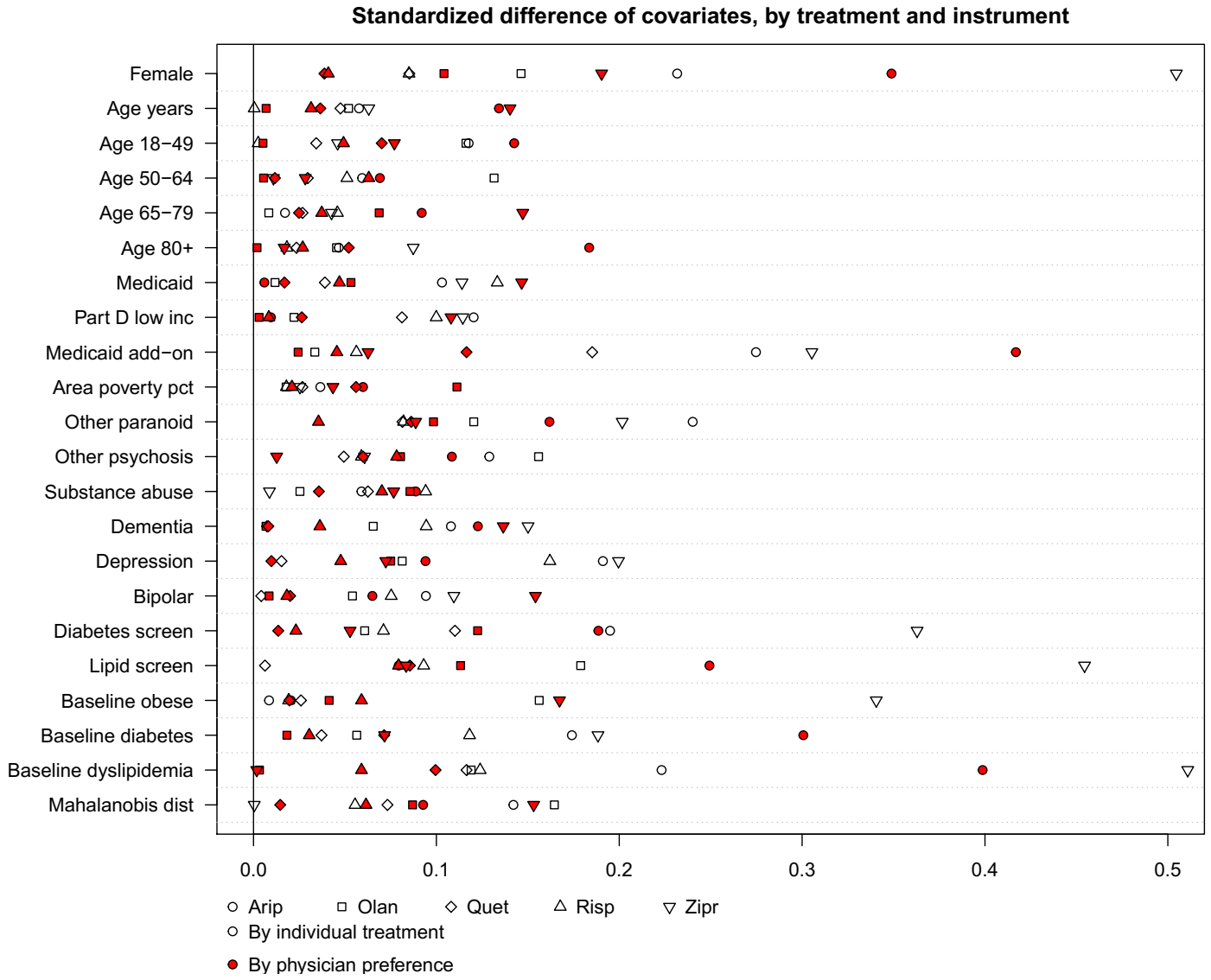
t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Each column contains a linear model of the effect of provider preference on whether the individual filled a prescription for the treatment indicated. Preference for risperidone is in the omitted reference category.

Prescriber preference for a treatment is defined by the proportion of previous prescriptions observed in the database within the therapeutic class. Preference is updated over time with each new prescription. The preference instrument was calculated from a sample of Medicare Part D enrollees (both MA-PD and Part-D-only plans) who were prescribed one of five atypical antipsychotics in 2008, 2009 and 2010. Covariates (not shown) included are age, means-tested health programs, comorbid mental-health diagnoses, and prior screenings for diabetes and dyslipidemia, and concomitant prescriptions for antidiabetic and antilipemic drugs (see Appendix Table 1).

Exhibit 5: Standardized difference (S.D.) is calculated for each covariate-treatment pair as the absolute difference between the standardized mean for individuals in the sample who received that treatment and individuals who did not. Smaller S.D. (closer to 0) indicates better covariate balance. S.D. by individual treatment and S.D. by provider preference were calculated and their ratio was taken. S.D.-ratio <1 implies improvement in covariate balance with the instrument and >1 implies worse covariate balance.



of whether the individual who received the treatment named in the column heading regressed on the value of the instrument. The models are linear estimates, so coefficients are level effects: e.g., in model 1, a one-percentage-point increase in the previous frequency of aripiprazole prescriptions is associated with a 0.23-percentage-point increase in the probability that the index patient will be prescribed aripiprazole. A rule-of-thumb test for adequate instrument strength is that the F-statistic be greater than 10. Here, all preference instruments except for the prediction for aripiprazole satisfy that guideline. Weaker instruments will magnify bias from confounders associated with physician preference, so the greater the potential for suspected instrument-outcome confounders, the greater the need for instrument strength.

3.7.3 Effect estimates

The OLS and IV estimates of the effect of atypical antipsychotic treatment on metabolic syndrome are presented in Exhibits 6 and 7. Both models adjust for all covariates listed in Exhibit 3. With risperidone as the reference group, coefficients in columns 1, 2, and 3 are interpreted as the level of increase in the probability of a new diagnosis of the outcome, over the average incidence for risperidone; and column 4 is the increase in the number of new cases with that treatment, relative to risperidone. In the OLS results in panel A, aripiprazole is associated with a significantly lower incidence of obesity and a higher incidence of diabetes than risperidone, while ziprasidone is associated with a significantly higher incidence of obesity. Panel B of Table 5 shows the causal instrumental-variable effect estimates. The IV analysis shows ziprasidone to have a much-greater effect on obesity and dyslipidemia than the other treatments: an increase of 0.257 and 0.268 respectively, over risperi-

Exhibit 6: Effect of atypical antipsychotic treatment on adverse outcomes: OLS estimates

	Obesity	Diabetes	Dyslipidemia
Arip	-0.0321*	0.0457*	0.0112
	(0.0162)	(0.0205)	(0.0204)
Olan	-0.0244	0.00111	0.0302
	(0.0180)	(0.0227)	(0.0226)
Quet	-0.0103	-0.00318	0.0000766
	(0.0139)	(0.0176)	(0.0175)
Zipr	0.0448*	0.00819	0.0224
	(0.0218)	(0.0275)	(0.0274)
Observations	2049	2049	2049
ftest	3.159	1.601	0.632
fprob	0.0134	0.172	0.640

Standard errors in parentheses
Risperidone is omitted reference group
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

done.

3.7.4 Bias ratio

In Exhibit 8, I show the relative bias from excluding the observed covariates from the IV model, in comparison with the OLS model. Values for the bias ratio less than one indicate that the IV estimate would be less biased than OLS from exclusion of that confounder. Because the IV bias is divided by the probability of compliance, the excluded variable bias is overwhelmingly larger than that of OLS.

Exhibit 7: Effect of atypical antipsychotic treatment on metabolic disease: instrumental-variable estimates

	Obesity	Diabetes	Dyslipidemia
Arip	0.0247	-0.362**	0.318**
	(0.0783)	(0.117)	(0.117)
Olan	-0.0297	0.0558	-0.0348
	(0.0727)	(0.109)	(0.108)
Quet	0.0379	0.130	-0.0558
	(0.0637)	(0.0954)	(0.0949)
Zipr	0.257**	-0.108	-0.368**
	(0.0884)	(0.132)	(0.132)
Constant	0.0359	0.163**	0.128*
	(0.0359)	(0.0537)	(0.0534)
Observations	2049	2049	2049

Standard errors in parentheses

Risperidone is omitted reference group.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Exhibit 8: Bias ratios

	Arip	Olan	Quet	Risp	Zipr
Female	5.20	2.65	1.64	8.23	2257.5
Age	14.0	0.14	6.96	3.57	0.022
Age 18-49	13.9	1.63	15.9	1.93	2.23
Age 50-64	19.0	1.35	6.23	22.8	3.02
Age 65-79	23.6	2.99	112.0	6.36	1.27
Age 80+	7.51	10.5	0.98	2.41	1.11
Medicaid eligible	5.44	0.79	38.9	1.81	8.21
Part D low income	3.90	4.64	5.26	2.20	2.22
Medicaid add-on	9.01	0.53	9.04	2.86	5.40
Poverty Percent All Ages	9.72	17.5	4.66	5.57	68.2
Other paranoia	6.64	2.51	10.1	9.49	1.33
Other psychosis	5.36	4.03	8.69	3.22	2.74
Substance abuse	13.3	32.1	17.2	8.65	10.6
Dementia	10.9	3.56	4.97	6.05	1.42
Depression	0.63	0.69	2.69	0.62	5.80
Bipolar disorder	0.45	22.7	3.57	2.63	7.90
Diabetes screen	33.9	0.21	18.4	10.9	1.10
Dyslipidemia screen	44.0	39.7	7.58	6.02	22.5
Baseline obese	8.70	0.22	3.28	9.27	15.8
Baseline diabetes	9.59	0.22	16.8	143.5	10.3
Baseline dyslipidemia	18.4	0.66	6.33	583.7	13.6

3.7.5 Complier characteristics

Exhibit 9 shows prevalence ratios of complier-weighted means to sample means. Each column represents a preference type. The probability of being a complier for that preference type was calculated for each individual and used to weight the sample mean. The interpretation of the statistic is the increase/decrease of the covariate for compliers to the instrument for whom the causal effect is being estimated, relative to the full sample. Values at or close to one indicate homogeneity between the compliers and the full sample. Most complier means are within ten percent of one, but a few exceptions stand out. Patients whose prescription choice is likely to change because their physician prefers aripiprazole are 25 percent more likely to suffer from dementia and 20 percent more likely to be eligible for Medicaid. Ziprasidone-compliant patients are 25 percent more likely to be over 65 and 15 percent more likely to suffer from dementia.

3.8 Discussion

This study illustrated techniques for validating a candidate for instrumental-variable estimation, by examining an applied example, where physicians' prescribing preference was used as an instrument to assess the causal effects of initiating treatment with atypical antipsychotic medications on the incidences of obesity, dyslipidemia, and diabetes. Clinical-trial evidence suggested the hypothesis that among five commonly prescribed drugs in that class, olanzapine would have the strongest effects on all three outcomes, followed by moderate or uncertain effects from quetiapine and risperidone and mild or no effect from ziprasidone and aripiprazole. Ordinary-least-squares and instrumental-

Exhibit 9: Prevalence ratios by provider preference: complier-weighted means to unweighted sample means

	Arip	Olan	Quet	Risp	Zipr
Female	0.97	1.06	1.02	0.98	0.97
Age	1.01	0.99	0.99	0.99	1.01
Age 18-49	0.93	1.04	1.03	1.08	1.00
Age 50-64	1.01	1.06	1.02	1.02	0.94
Age 65-79	1.06	0.94	0.99	0.94	1.25
Age 80+	1.20	0.92	0.89	0.91	1.32
Medicaid eligible	0.97	1.02	1.00	1.05	1.00
Part D low income	0.95	1.02	1.00	1.08	0.97
Medicaid add-on	0.96	1.02	1.00	1.06	0.98
Poverty Percent All Ages	1.00	1.00	1.00	1.00	1.00
Other paranoia	1.11	0.87	1.02	0.94	0.99
Other psychosis	1.09	0.97	0.97	0.96	1.06
Substance abuse	1.02	0.96	1.00	1.02	1.00
Dementia	1.25	0.99	0.87	0.95	1.15
Depression	0.98	1.00	0.98	1.04	1.03
Bipolar disorder	0.97	1.00	0.99	1.07	0.98
Diabetes screen	0.94	1.05	1.04	1.03	0.94
Dyslipidemia screen	0.98	1.02	0.98	1.06	0.92
Baseline obese	0.96	1.11	1.02	0.94	1.06
Baseline diabetes	0.98	1.03	1.00	1.00	0.97
Baseline dyslipidemia	1.02	1.01	0.99	1.00	0.97

Values further from 1 suggest greater difference
between compliers and the rest of the sample

variable estimates of the comparative safety yielded differing results, however, which conformed neither to each other nor to the hypothesis. In the OLS model, only aripiprazole appeared to have effects differing from the group, with slightly larger effects on diabetes and lower effects on obesity. In the IV model, aripiprazole appeared to protect strongly against diabetes (it was 36 percent lower than risperidone, the reference category) but increase the risk for dyslipidemia (it was 32 percent higher). Ziprasidone appeared to be a risk for obesity and protect against dyslipidemia: its use led to 25 percent higher incidence of obesity and 37 percent lower incidence of dyslipidemia, compared to risperidone. In both models, no statistically significant differences were detected among olanzapine, quetiapine, or risperidone.

Can estimates from either model be interpreted as causal effects? There are strong conceptual reasons for disbelieving that the OLS model conditioned the estimate on all aspects of treatment selection for an individual patient, which likely depended on clinical variables and chronic disease risk not observed in the claims data that were available. On that basis I proposed a commonly used type of instrumental variable, physicians' prescribing preference, and described diagnostic tests to determine whether the proposed instrument satisfied the instrumental-variable assumptions. Despite strong conceptual reasons, however, for thinking that physicians' prescribing preference was a promising source of exogenous variation in patients' treatment, it appears that physicians did develop concentrating behavior in ways that are associated with their patient panels and possibly also with follow-up care. When patients were stratified across levels of physician preference, the average covariate distance between groups increased for some variables and decreased for others, but overall it appears that physician preference was no better randomized among patients than the treatment

itself. I assessed the potential bias from unobserved confounders, by evaluating what would happen to the estimates if some of the observed variables were omitted. The bias ratio of expected bias from the IV to OLS estimate indicated that because the instrument tends to magnify any existing bias, IV estimates would indeed be more biased away from the true causal effect. The importance of the measured confounders suggests that remaining confounders also continue to lurk, unobserved. In particular, there are likely to be systematic differences between physicians who prescribe the newer drugs, ziprasidone and aripiprazole, which the FDA approved only in 2000 and 2001. One example of how such differences might arise would be variation in state Medicaid formularies, which increasingly put restrictions on some atypical antipsychotics as first-line treatments, in order to control spending. Different preferences might then arise for physicians whose patient mixes included mostly Medicaid patients, and those mixes would also be correlated with more chronic disease, worse health behaviors, and the other patient health problems that are associated with being poor enough to qualify for Medicaid.

The purpose of this paper has been to describe and demonstrate techniques for assessing the validity of instrumental variable analysis and interpreting the external applicability of its causal estimate. In particular, the paper has laid out several heuristics for assessing the assumptions that the instrument is assigned effectively randomly and that it operates on the outcome only through its effect on treatment choice. I have illustrated how some of those heuristics operate in practice, with reference to the applied analysis. Other tests, such as investigating differences across levels of physician preference in provider specialty and prescribing volume, would have provided additional insight into possible ways that physicians who prefer newer antipsychotics might differ in their treat-

ment patterns in other ways as well. Here I have limited the cohort to patients with schizophrenia for better instrument performance. Stratifying the analysis by provider specialty or other patient characteristics would be a tactic to get better identification and covariate balance, although previous work has shown that stratification does not necessarily increase partial-r-squared, yet reduces statistical power (Rassen et al 2009). To measure differences in patient SES, I used Medicaid and Part D low-income subsidy eligibility, as well as a county-level area poverty, as proxies for patient SES, but those are rather blunt tools. Additional area measures by census block would have added additional information about (or at least proxy measures of) race, class, and income. The atypical-antipsychotic analysis has illustrated many of the challenges of asking clinical questions of data created for administrative and billing purposes. Identifying chronic disease prevalence or incidence is notoriously difficult to do using only insurance claims, especially for individuals with multiple comorbid conditions (Hebert et al. 1999, Rector et al. 2004). Differences in coding accuracy, patients' use of health care for other purposes, and physician diagnoses create measurement error in trying to detect the underlying incidence of new disease. Ideally, algorithms to detect chronic disease should be validated against chart reviews within the dataset and population being studied.

3.8.1 Conclusions

When approached with appropriate skepticism and careful validation, instrumental-variable analysis can shed light on important clinical questions and generate hypotheses about the effectiveness and safety of treatments where randomized-trial evidence is lacking. Outside the context of a randomized trial or encouragement intervention, an IV is always vulnerable to suspicions that it may violate

the exclusion restriction or is not independent of unmeasured confounders. In such situations, however, other observational methods (such as propensity-score matching or weighting or multivariate regression) may also fail to satisfy the requirement that after controlling for observed covariates, the potential outcomes be independent of treatment status.

Rather than reject IV analysis altogether, CER researchers should treat it as a useful tool, to be employed with careful attention to the potential bias that can be introduced into estimates. The danger of IV analysis is that the weaker the instrument (that is, the lower the r-squared of the relationship between instrument and treatment), the more-magnified the bias and inconsistency in the estimates of causal effect. There is a tradeoff, therefore, between the balance of potential confounders that the instrument achieves on the one hand, and the amplification of residual unobserved confounders on the other. It is not sufficient to show that an instrument is relevant and improves covariate balance, because even when those statistics suggest that an instrument might work to satisfy the assumptions, the amplification effect can nonetheless produce estimates of causal effect that are less-reliable than propensity-score or multivariate regression estimates. Lamentably, perhaps due to the pressures of length restrictions on published articles, it has become customary in health-services research to publish IV studies without supplying sensitivity analyses or characterizing the compliers. As IV analysis becomes a regularly used tool in CER, it is crucial that researchers be familiar with the relevant diagnostics. Editors and reviewers must demand that sensitivity analysis be presented alongside IV models or in supplementary appendices. Establishing those norms will help ensure the scientific integrity of data science.

A

Supplemental Material to "Medical Underwriting In Long-Term Care Insurance"

This appendix provides supplemental and technical information about our methods for estimating underwriting decisions for Americans ages 50-71 in our paper, "Medical Underwriting In Long-Term Care Insurance: Market Conditions Limit Options For Higher-risk Consumers." We also include here sensitivity analyses and additional data summary.

A.1 Underwriting Process

Figure 1 shows steps that consumers go through before they become holders of long-term care insurance policies. While many people do not shop for policies because they do not think it is necessary or find it unaffordable, a large portion are excluded in the underwriting stages.

Because the application process is time-consuming and costly, agents typically do not market long-term care insurance to prospective buyers over the age of 70, where underwriting rejection rates can be high, and steer those who already exhibit some other easily determined disqualifying condi-

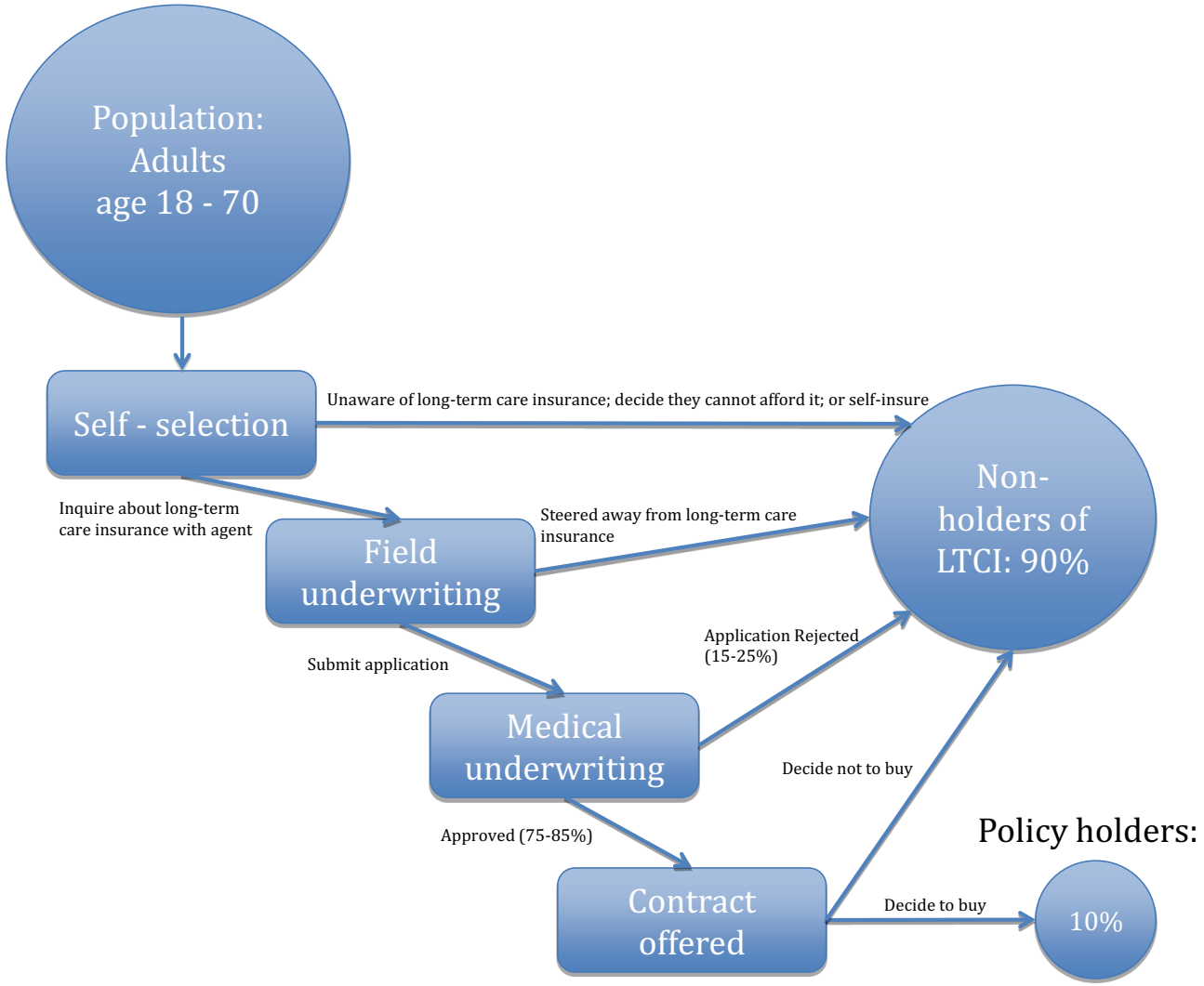
tion away from the process. For example, in a guide for insurance agents developed by one of the largest carriers of long-term care insurance, agents are instructed to discourage applications from individuals who are morbidly obese or who have been diagnosed with one of a list of conditions such as multiple sclerosis, Alzheimer's disease, cirrhosis of the liver, or Parkinson's disease (Genworth Financial 2013), conditions that would put the individual at high risk for immediate need of long-term care services. Most state regulations require agents to verify that their clients' income and assets meet minimum thresholds for the premiums to be financially suitable. As part of the initial meeting, carriers in these states (and most carriers even in states where it is not required by law) have their agents counsel clients—usually in the form of a personal worksheet—as to whether their assets and income are sufficient for long-term care insurance to be a suitable financial product for them. The rule of thumb proposed by the National Association of Insurance Commissioners (NAIC) is that that if the client is currently eligible for Medicaid or has less than $30,000 in assets, or if the premium amount would be more than 7% of their income, then long-term care insurance may not be appropriate (National Association of Insurance Commissioners 2014).

Underwriting accuracy confers a competitive advantage in the marketplace, and standards and protocols vary across companies and are protected as confidential company assets. These screens can include comprehensive screening of mobility, activities of daily living (ADLs) and instrumental activities of daily living (IADLs), cognitive screening, medical history, living environment and clinical observations. At the beginning of the long-term care insurance purchasing process, selling agents discourage applications from buyers who have easily determined disqualifying conditions or have insufficient income or assets for premiums to be financially suitable. For individuals who submit for-

mal applications, the underwriting assessment starts with a health history questionnaire. To verify applicants' information and collect more detailed information, insurers may request medical records or conduct telephone interviews or home visits. For qualifying applicants, firms offer a premium rate and coverage terms for consideration. Firms tend to offer age- and (more recently) gender-rated standard premium rates. Health is taken into account in deciding whether to offer coverage at all, and in some cases whether to provide a discount or added premium to a base premium rate.

After a reviewer with clinical training examines the applicant's file and makes a coverage recommendation, the applicant receives an offer of coverage with premium amounts and makes a decision to purchase. Firms tend to offer age- and (more recently) gender-rated standard premium rates, taking health into account only in deciding whether to accept or decline an applicant; they generally do not consider it in setting premiums for those they accept. When health is taken into consideration, some firms may offer discounts for being in a preferred risk class—for example, for no use of tobacco, having blood pressure and body weight in the healthy range, and being physically active—whereas the less healthy may be accepted into a substandard risk class at significantly higher premiums. We do not have information on whether the carriers in our study offered differentiated premiums.

Exhibit 1: Medical underwriting process



A.2 Methods

A.2.1 Data

Data Collection. The insurance data were collected by a full service third party administrator (TPA) that made underwriting decisions on behalf of these companies. Insurance firms collected information with written, self-administered questionnaires from applicants and sent them to the TPA where a clinician reviewed each applicant's file and offered a recommendation to "approve" or "not approve" the application. Although the applicants' answers were not independently verified, the insurer's right to rescind a policy based on fraud or material misrepresentation (generally only within the first two years after policy issue) gives applicants the incentive to report their health status in good faith. We use a sub-set of the characteristics that correspond to items in the Health and Retirement Study (HRS). Some reported health conditions are aggregated categories from several diagnoses or self-reported conditions (see Exhibit 2). Although employment and education may not be explicitly considered in underwriters' decisions, we nonetheless include these covariates because they can capture some aspects of health and functional status otherwise not measured in the HRS.

Where responses were missing one or more of the underwriting variables, we filled in these values using the Imputation by Chained Equations (ICE) method (Royston 2004).

Estimates from the HRS are weighted to correspond to the American Community Survey, a nationally representative sample of non-institutionalized adults. Therefore individuals in institutional settings (such as a nursing home) have a weight of 0, and all reported population proportions and prevalence have a denominator comprising individuals living in the community.

Exhibit 2: Underwriting and HRS variable alignment

Model variable	Underwriting question	HRS (RAND)	HRS coding
Age	Age in years at time of underwriting	ragey_e	Age in years at time of interview
Female	gender	ragender	0 (male), 1 (female)
College degree	>=16 years of education	radegrem	5 (BA)
Employment status	Employed	rwork	1 (currently working for pay)
Delayed word recall <7/10	delayed word recall score < 7/10	rdlrc	delayed word recall score <7/10
Take any medication for depression	Do you take any medication(s) for depression?	rcesd	cesd score >=6
Experiences memory loss	Do you ever experience Forgetfulness, Memory Loss or Confusion?	rmemry, ralzhe, rdemen	memory problems; Alzheimers problems; dementia problems
Difficulty taking medication	not "independent" for "taking medication"	rmedsa	some difficulty – taking medications
Difficulty with activities of daily living	not "independent" for any of the following: transferring, toileting, bathing, dressing, eating, mobiliy inside	radla	"some difficulty" with 1 or more of the following: bathikng, dressing, eating, getting out of bed, walking across a room
High blood pressure	High blood pressure	rhibpe	ever had high blood pressure
Back pain	Back or Spine Condition Pain andor Swelling in your Neck Back Spine Shoulders A	rback	had back problems
Arthritis	Degenerative Bone or Joint Disease/Arthritis	rarthre	ever had arthritis
Diabetes	Diabetes	rdiabe	ever had diabetes

Exhibit 2: Underwriting and HRS variable alignment

Model variable	Underwriting question	HRS (RAND)	HRS coding
Heart problems	Heart/Circulatory Problems AFib or Irregular heart beat Congestive Heart Failure Heart Attack Angina or heart related chest pain	rhearte	ever had heart problems
Psychiatric illness	Psychiatric Disorders Depression Anxiety	rpsyche	Ever had emotional / psychiatric problems
Lung problems	COPD / Emphysema / Asthma Shortness of Breath / Difficulty Breathing Sleep Apnea	rlunge	ever had lung disease
Cancer	Cancer Leukemia Lymphoma or Melanoma Hodgkins Disease	rcancre	ever had cancer
Stroke	Stroke / TIA / mini Stroke Peripheral Vascular Disease	rstroke	ever had stroke
Hospitalization, previous 2 years	Have you been hospitalized or received any medical care within the past 3 years?	rhosp	1 (hospital stay, prev. 2 years)
Long-term care, previous 2 years	In the Past 2 years have you been confined to a nursing home or received any adult day care, short term care or home care services?	rnrshom	Nursing home stay, prev 2 years
		rhomcar	Home health care, prev 2 years

A.2.2 Analysis

To estimate underwriting approval probabilities for the general population, we developed an empirical model of the coverage decision using underwriting data from the long-term care-insurance carriers and applied the model parameters to a nationally representative sample of older US residents. We report a linear probability model of underwriting approval, estimated using ordinary least squares, to facilitate an intuitive interpretation of the percentage-point effect on probability of approval of each characteristic and health condition. The model is as follows:

$Y_i = X_i\beta + \varepsilon_i$

Where Y is 1 for approved and 0 for disqualified applicants, and X is the vector of applicant characteristics (age categories, health conditions, etc.), and ε is a randomly distributed error term.

Those results, with standard errors, are displayed in Column 1 of Exhibit 4 with standard errors, and correspond to Exhibit 1 of the main article.

We estimated the probability of underwriting approval using the generalized linear model:

$Pr(Y_i = 1) = F(X_i\beta^*)$

Where F is the logistic function. Results with 95% confidence intervals are shown in Exhibit 5. Column 2 of Exhibit 4 shows the population-averaged marginal effects of X (using the Stata *margeff* command), with standard errors calculated using the delta method, where each estimate is the difference in approval rate for the entire sample between $X_{i,k} = 0$ and $X_{i,k} = 1$. These are similar to the OLS estimates.

To estimate individual probabilities we specified a model with indicators for each age-year value,

interaction terms of gender with age, and a variable for the number of health conditions (1, 2, and 3 or more), as reported in Exhibit 5, column 2. For each individual in the HRS sample, we calculate $\hat{p} = F(X_i^{HRS}\beta^*)$ for each respondent in the HRS sample to predict the probability they would be offered a policy, supposing they were to apply for insurance subject to similar underwriting conditions.

To summarize the results, for each sub-sample s we report both the mean of the predicted probabilities ($\bar{\hat{p}}^s$), and the percent of the sample that is likely approved ($\hat{\pi}^s_{appr}$), where:

$\bar{\hat{p}}^s = \frac{1}{n}\sum_{i=1}^{n}\hat{p}^s$

$\hat{\pi}^s_{appr} = \frac{1}{n}\sum_{i=1}^{n} I(\hat{p}^s >= 0.5)$

We generated the approval probability models with Stata version 13 and estimated survey statistics and generated figures with R version 3.1.3.

A.3 Results

A.3.1 Main results

Exhibit 3 gives the prevalence estimates of the underwriting variables for the insurance sample (disqualified, approved, and full sample) and for the HRS respondents. Exhibit 4 shows the differential, linear effect of a change from 0 to 1 for these variables and their standard errors. Column 1 shows the change in probability from a linear probability model estimated using ordinary least squares (reported in the main article, exhibit 1), and column 2 shows the average change in probability from the logistic regression model,with standard errors calculated using the delta method, using the MFX

command in stata. Exhibit 5 shows the odds ratio estimates from logistic models of the probability of approval. Column 2 contains the full specification of the model we used to impute probabilities, including interaction terms, fixed effects for each year of age, and indicators for 2 and 3-or-more of the chronic conditions in the model. Exhibit 6 summarizes the imputed estimates for the HRS sample corresponding to Exhibit 2 in the main article.

Exhibit 3: Summary statistics

	Disqualified	Approved	All Applicants	HRS
Age 18 - 49	0.122	0.231	0.205	0.000
Age 50 - 59	0.341	0.424	0.404	0.528
Age 60 - 69	0.484	0.320	0.360	0.415
Age 70 up	0.053	0.024	0.031	0.057
Female	0.449	0.486	0.477	0.523
Education 16+ years	0.476	0.550	0.532	0.313
Employed	0.625	0.756	0.725	0.587
Word recall score < 7	0.300	0.229	0.246	0.831
Self-reported memory loss	0.236	0.159	0.177	0.217
Difficulty taking medication	0.016	0.006	0.009	0.025
Difficulty with 1+ ADL	0.004	0.000	0.001	0.122
High blood pressure	0.502	0.279	0.333	0.494
Back pain	0.409	0.232	0.275	0.389
Arthritis	0.245	0.102	0.137	0.466
Diabetes	0.201	0.026	0.069	0.183
Heart problems	0.199	0.082	0.110	0.160
Psychiatric illness	0.184	0.092	0.115	0.195
Lung problems	0.102	0.050	0.062	0.080
Cancer	0.057	0.027	0.034	0.100
Stroke	0.016	0.001	0.004	0.046
Hospitalization, prev 2 years	0.533	0.353	0.396	0.213
Long-term care, prev 2 years	0.016	0.004	0.007	0.054
Drinks alcohol	0.890	0.903	0.900	0.655
Ever been a smoker	0.379	0.296	0.316	0.568
Current smoker	0.085	0.053	0.060	0.187
Underweight	0.007	0.005	0.005	0.010
Normal/Overweight	0.595	0.743	0.707	0.614
Obese	0.374	0.247	0.278	0.322
Extremely obese	0.024	0.005	0.010	0.054
Observations	3782	11877	15659	13770

Exhibit 4: Marginal effects

Marginal effects on probability of approval

	OLS	MFX
Age 50–59	-0.007	-0.014
	(0.008)	(0.009)
Age 60–69	-0.037***	-0.042***
	(0.009)	(0.010)
Age 70+	-0.081***	-0.073***
	(0.020)	(0.020)
Female	0.016**	0.015**
	(0.007)	(0.007)
Education 16+ years	0.020***	0.020***
	(0.006)	(0.006)
Employed	0.030***	0.028***
	(0.008)	(0.007)
Word recall score < 7	-0.035***	-0.033***
	(0.007)	(0.007)
Self-reported memory loss	-0.020**	-0.017**
	(0.008)	(0.008)
Difficulty taking medication	-0.094***	-0.071**
	(0.033)	(0.028)
Difficulty with 1+ ADL	-0.522***	-0.402***
	(0.092)	(0.097)
High blood pressure	-0.078***	-0.070***
	(0.007)	(0.006)
Back pain	-0.101***	-0.090***
	(0.007)	(0.006)
Arthritis	-0.111***	-0.086***
	(0.009)	(0.008)
Diabetes	-0.415***	-0.297***
	(0.012)	(0.010)
Heart problems	-0.130***	-0.104***
	(0.010)	(0.009)
Psychiatric illness	-0.123***	-0.105***
	(0.010)	(0.008)
Lung problems	-0.086***	-0.070***
	(0.013)	(0.011)
Cancer	-0.111***	-0.090***
	(0.017)	(0.015)
Stroke	-0.528***	-0.435***

Exhibit 4: Exhibit 4 (continued): Marginal effects

Marginal effects on probability of approval

	OLS	MFX
	(0.046)	(0.055)
Hospitalization, prev 2 years	-0.065***	-0.063***
	(0.007)	(0.006)
Long-term care, prev 2 years	-0.083**	-0.050
	(0.037)	(0.032)
Drinks alcohol	0.024**	0.023**
	(0.010)	(0.010)
Ever been a smoker	-0.013*	-0.014**
	(0.007)	(0.007)
Current smoker	-0.114***	-0.104***
	(0.013)	(0.012)
Underweight	-0.174***	-0.188***
	(0.042)	(0.050)
Obese	-0.046***	-0.044***
	(0.007)	(0.007)
Extremely Obese	-0.268***	-0.258***
	(0.031)	(0.038)
Constant	0.920***	
	(0.015)	
Observations	15659	15659
R-squared	0.22	
F-statistic	159	

Source: Authors' analysis data on applicants for long-term care insurance for two US firms in 2009 - 2011.

Notes: Table displays marginal effects of characteristics on probability of underwriting approval, with standard errors shown in parentheses. OLS is a linear probability model estimated with ordinary least squares (OLS). Model 2 is the marginal effects estimated from logistic regression shown in Table 5. The reference category for age is the 18 – 49 age group. Reference category for BMI categories is normal/overweight.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Exhibit 5: Models to estimate marginal effects and predict underwriting probabilities

	(1)	(2)
Age 50 - 59	0.901	
	[0.793,1.022]	
Age 60-69	0.743***	
	[0.648,0.852]	
Age 70+	0.613***	
	[0.479,0.783]	
Female	1.110**	0.931
	[1.013,1.216]	[0.751,1.154]
Education 16+ years	1.147***	1.156***
	[1.052,1.249]	[1.060,1.260]
Employed	1.213***	1.164***
	[1.097,1.341]	[1.050,1.291]
Word recall score < 7	0.795***	0.801***
	[0.723,0.873]	[0.728,0.880]
Self-reported memory loss	0.890**	0.880**
	[0.801,0.989]	[0.792,0.979]
Difficulty taking medication	0.609**	0.638**
	[0.413,0.899]	[0.432,0.941]
Difficulty with 1+ ADL	0.060***	0.054***
	[0.016,0.226]	[0.014,0.209]
High blood pressure	0.611***	0.745***
	[0.559,0.668]	[0.605,0.916]
Back pain	0.531***	0.632***
	[0.485,0.581]	[0.514,0.777]
Arthritis	0.548***	0.597***
	[0.491,0.613]	[0.484,0.736]
Diabetes	0.125***	0.143***
	[0.107,0.145]	[0.113,0.181]
Heart problems	0.483***	0.549***
	[0.429,0.544]	[0.443,0.679]
Psychiatric illness	0.478***	0.545***
	[0.425,0.538]	[0.439,0.675]
Lung problems	0.610***	0.664***
	[0.523,0.711]	[0.528,0.835]
Cancer	0.530***	0.618***
	[0.433,0.648]	[0.473,0.808]
Stroke	0.047***	0.053***
	[0.022,0.102]	[0.024,0.115]

Exhibit 5: Models to estimate marginal effects and predict underwriting probabilities

	(1)	(2)
Hospitalization, prev 2 years	0.643***	0.653***
	[0.590,0.701]	[0.599,0.712]
Long-term care, prev 2 years	0.705	0.675*
	[0.457,1.090]	[0.439,1.039]
Drinks alcohol	1.177**	1.164**
	[1.023,1.355]	[1.011,1.341]
Ever been a smoker	0.908**	0.930
	[0.826,0.997]	[0.846,1.023]
Current smoker	0.483***	0.481***
	[0.411,0.568]	[0.408,0.567]
Underweight	0.329***	0.312***
	[0.198,0.544]	[0.185,0.526]
Obese	0.740***	0.732***
	[0.675,0.811]	[0.668,0.803]
Extremely Obese	0.232***	0.228***
	[0.160,0.336]	[0.157,0.331]
1 health condition		0.539***
		[0.430,0.677]
2 health conditions		0.473***
		[0.318,0.704]
3 health conditions		0.575*
		[0.307,1.079]
Constant	10.563***	15.863***
	[8.522,13.093]	[12.370,20.341]
Observations	15659	15659
Pseudo R-squared	0.19	0.20
Akaike's Inf. Crit.	14113	14036
Bayesian Inf. Crit.	14328	14587
Log-likelihood	-7029	-6946

Source: Authors' analysis of the Health and Retirement Study.
Notes: We modeled probability of approval in a multivariate logistic regression. Exponentiated odds ratios are shown with 95% confidence intervals in brackets. Model 1 is the specification for marginal effects reported in Table 4. Model 2 is used to impute probabilities in the HRS sample, and includes fixed effects for each year of age, and age-female interactions (coefficients not shown). The reference category for age and age-female interactions is the 18 – 49 age group.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

A.4 Model checks

A.4.1 Sensitivity analyses

Exhibits 7 and 8 shows how the results vary with different probability thresholds for designating a respondent as "likely approved." We chose 0.5 as the cutoff because it has the strong advantage of appealing to common-sense intuition: above .5, an individual is more likely than not to be approved, and below, less likely. But some empirical context is also useful. The purpose of the analysis is to estimate an upper bound on how many individuals in the population would be able to pass underwriting requirements similar to these. Figure 8 give a picture of the sensitivity of the estimated approval rate to the threshold we assign to "approved." If the .5 threshold is applied to the insurance sample, then the predicted approval rate in that sample is approximately 88%. That approval rate is higher than both the actual acceptance rate for this sample of 76% and another recent estimate of the industry-wide average of 81%. As one would expect, the predicted approval rates for the general population (HRS) sample are especially sensitive to the choice of threshold probability, as these individuals are generally less healthy than the insurance sample population. Table 7 show how the estimates by financial category change with different threshold assumptions.

A.4.2 Generalizability of insurance sample to the industry

The companies in our insurance sample represent approximately 5% of market share for new policies issued over the study period. One firm may apply underwriting criteria differently from another, so our analysis assumes that the underwriting decisions that we model from this sample reflect, on ag-

Exhibit 6: Summary of imputed estimates of underwriting approval for long-term care insurance

Category of suitability of long-term care insurance for respondent[c]	Sample size	Population, millions (SE)[a]	Mean approval probability[a]	pct likely to be approved[a,b]
Suitable	6166	37.98 (2.43)	0.611 (0.006)	70.1 (0.83)
Assets $30,000 - $250,000	3361	19.50 (1.20)	0.580 (0.008)	65.6 (1.07)
Assets over $250,000	2805	18.48 (1.23)	0.644 (0.007)	74.9 (1.00)
Not suitable	7604	33.08 (1.64)	0.441 (0.005)	47.4 (0.78)
Full sample	13770	71.07 (4.07)	0.532 (0.005)	59.6 (0.71)

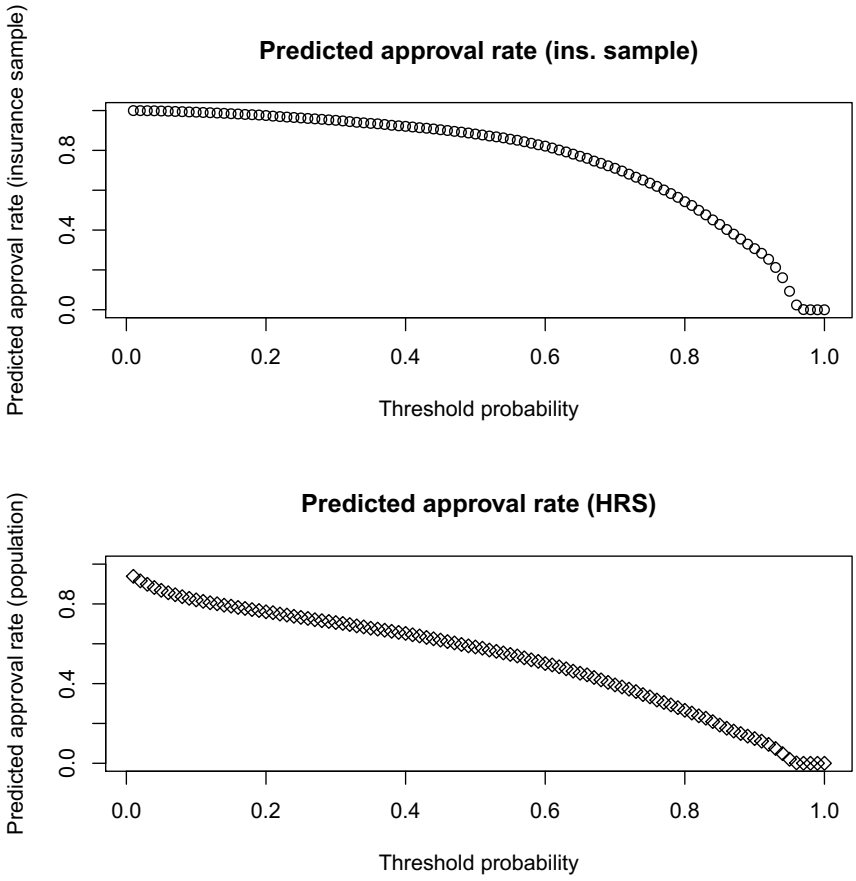
Source Authors' analysis of the Health and Retirement Study (HRS), 2010-2011, respondents ages 50 – 71.
Notes Probability estimates for HRS respondents were predicted from a multivariate logistic regression model, summarized in Exhibit 2, that the authors estimated from approval information on the insurance applicants. The full model with odds ratios is available in an online appendix (18). (a) Population estimates were weighted to correspond to the American Community Survey, a sample of non-institutionalized U.S. adults. Standard errors (in parentheses) account for the sampling design of the HRS and variance of the imputed probabilities and not modeling uncertainty from the prediction model. (b) Respondents were designated as "likely to be approved" with imputed approval probability > 0.5. Thus the percentage reported in this column represents the estimated population proportion with predicted approval > 0.5. (c) Financial suitability of long-term care insurance for the respondent is defined as household yearly income>$20,000 and non-housing assets>$30,000 for a single person, and respectively $30,000 and $50,000 for a couple.

Exhibit 7: Imputed estimates of underwriting approval for long-term care insurance: sensitivity analyses

Category of suitability of long-term care insurance for respondent[c]	pct likely to be approved[a,b]				
Threshold probability	0.4	0.45	0.5	0.55	0.6
Full sample	66.2% (0.72)	63.0% (0.75)	59.6% (0.71)	55.6% (0.72)	50.7% (0.70)
Suitable	76.2% (0.80)	73.5% (0.88)	70.1% (0.83)	66.2% (0.81)	61.0% (0.90)
Assets $30,000-250,000[d]	72.0% (1.12)	69.0% (1.18)	65.6% (1.07)	61.0% (1.09)	56.2% (1.16)
Assets $250,000 and over[d]	80.8% (0.82)	78.4% (0.89)	74.9% (1.00)	71.6% (1.06)	66.0% (1.17)
Not suitable	54.6% (0.86)	51.0% (0.83)	47.4% (0.78)	43.3% (0.73)	38.9% (0.72)

Source Authors' analysis of the Health and Retirement Study (HRS), 2011-2012, ages 50 – 71. Notes Probability estimates for HRS respondents were predicted from a multivariate logistic regression model, summarized in Appendix Table 5, that the authors estimated from approval information on the insurance applicants. (a) Population estimates were weighted to correspond to the American Community Survey, a sample of non-institutionalized U.S. adults. Standard errors (in parentheses) account for the sampling design of the HRS and variance of the imputed probabilities and not modeling uncertainty from the prediction model. (b) Respondents were designated as "likely to be approved" with imputed approval probability below the threshold, and as likely to be disqualified with imputed probability above the threshold. Thus the percentage reported in the first column represents the estimated population proportion with predicted approval >0.4, etc. (c) Financial suitability of long-term care insurance for the respondent is defined as household yearly income>20,000 and non-housing assets>$30,000 for a single person, and respectively $30,000 and $50,000 for a couple. (d) Assets include the net total of all non-housing assets: property, business assets, other real estate, and financial wealth (including retirement accounts), less non-mortgage debt.

Exhibit 8: Sensitivity of estimates to assumed approval threshold



Data source: Authors' analysis of data on applicants from two U.S. insurance firms in 2009-2012 (N=15659), the Health and Retirement Study, 2010-2011 (N=13770).
Notes: Figures show the proportion of the sample "likely approved" with change in assumed threshold for approval. Y-axis is the imputed probability, and X-axis is the designated threshold.

gregate, similar proportion of approved applicants to the industry as a whole. The rejection rate that we observe in this sample, 24%, is somewhat higher than the 19% rejection rate that another recent study found in a survey of companies that represented about 70% of the market (LifePlans 2010). Exhibit 9 shows the declination rates, by age group, of that industry sample Differences could originate from stricter underwriting standards at the firms that supplied our sample or more variation in health status among the applicant pool. Our model, however, predicts a rejection rate in the insurance sample of only 12%. That suggests that our model produces a conservative lower bound of the proportion of the population that would be disqualified if industry-wide criteria were applied to the population at large.

Exhibit 9: Comparison of analytic sample with industry declination rates by age group

	Industry Insurance Sample	Analytic Insurance Sample		
Age	Declined	Declined	Predicted probability	Predicted Declined
0-44	0.07	0.13	0.13	0.02
45-49	0.09	0.16	0.15	0.03
50-59	0.12	0.20	0.20	0.07
60-64	0.16	0.29	0.29	0.16
65-69	0.22	0.38	0.38	0.28
70-74	0.27	0.41	0.41	0.32
Mean	0.14†	0.24	0.24	0.12

Source: Industry sample is composed of survey responses from 15 long-term care insurance firms, constituting 70% of the total market‡.

Notes:

† The overall industry mean reported in the study was 0.19. The mean we give here is weighted by age category to match the age distribution of applicants in our insurance sample.

‡LifePLans, Inc. *Appendix J: A profile of declined long-term care insurance applicants.* In: *A Report on the Actuarial, Marketing, and Legal Analysis of the CLASS Program.* Washington, DC: Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Office of Disability, Aging, and Long-Term Care Policy. 2010 December 27 [cited 2015 May 1]. Available from: https://aspe.hhs.gov/sites/default/files/pdf/76321/appJa.pdf.

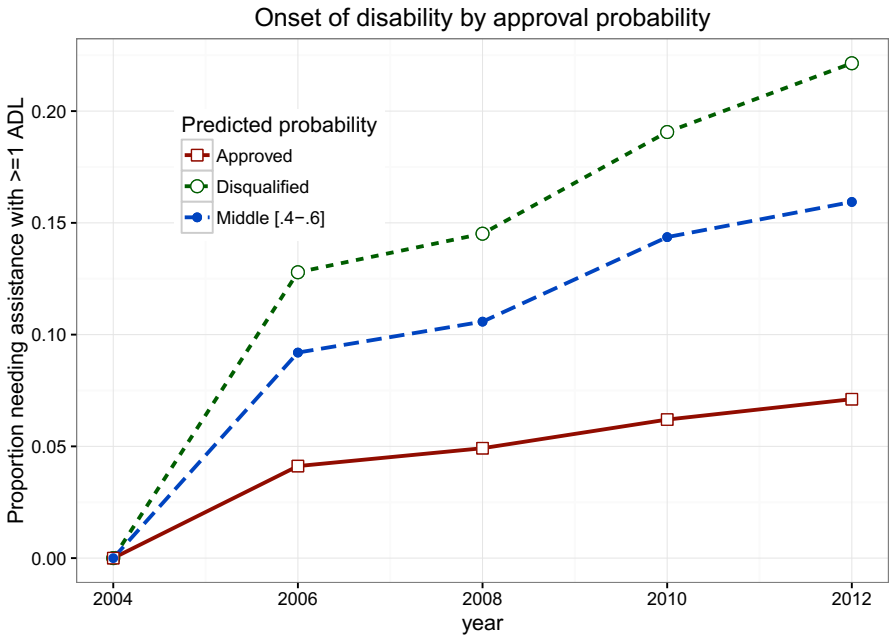
A.4.3 Disability and underwriting models

We plotted the incidence of disability in Figure 10, starting with a cohort reporting 0 ADLs (figure in appendix). Approved are individuals with predicted probability $>$ 0.5 (N=7545) and disqualified $<$ 0.5 (N=3778). Starting with a cohort of individuals with no ADL needs, the figure shows the cumulative incidence of disability (here defined as at least 1 ADL need) over 8 years. The figure gives some sense of how predictive the underwriting models are of near-term disability that might invoke long-term care benefits. The "marginal" group, with probabilities between 0.4 and 0.6, may be potentially insurable with lower risk than the "Disqualified" group, but is also higher risk. One interpretation of this figure is that firms are fairly accurate with underwriting determinations, and yet there may be some potential for policies that reduce adverse selection to allow that middle group to buy insurance.

A.4.4 Insurance purchase and underwriting models

As a check on our model (Table 11), we examine how purchase of new LTCI policies varies with predicted approval probability. When we look at the relationship between insurance take-up from the 2012 to 2014 period among individual who do not hold insurance in 2012, we find that higher approval probability is weakly associated with higher insurance take-up. Using individuals who report not owning an LTCI policy waves 3 – 11 of the HRS (spanning 1996 – 2014), we find that having approval probability over 0.5 is associated with 1%-point higher chance of buying insurance, approximately 33% higher than those with approval probability under 0.5 ($p < 0.01$).

Exhibit 10: Underwriting models and incidence of disability



Data source: Authors' analysis of the Health and Retirement Study, 2004 – 2012 (waves 7 – 11). Approved are individuals with predicted probability >0.5 (N=7545) and disqualified <0.5 (N=3778). Notes: Figure shows the cumulative incidence of disability, defined as needing assistance with at least 1 activity of daily living (ADL). Sample comprises individuals in the Health and Retirement Study. Estimates are weighted to match the Current Population Survey.

While the positive sign is encouraging, it is problematic to infer anything about the accuracy of the model from the magnitude of the relationship because demand is endogenous to health. With additional years of age or new health shocks, individuals probably become more aware of their potential need for long-term care insurance and more likely to seek it out, and also less likely to qualify if they do apply. To properly test the effect of underwriting on uptake, we would need some way to observe demand conditional on health and age, or a source of exogenous variation in demand (such as price shocks). Additionally, in the HRS we cannot distinguish between policies bought on the group and non-group markets, which have different underwriting standards.

Exhibit 11: Association of take up of new long-term care Insurance with high approval probability

	Dependent variable:
	New LTCI policy
Likely approved	0.010***
	(0.001)
Constant	0.031***
	(0.001)
Observations	141,012
R^2	0.001
F Statistic	98.791*** (df = 1; 141010)
Note:	*p<0.1; **p<0.05; ***p<0.01

B

Supplemental Material to "The Role of Underwriting in Policy Incentives"

B.1 Policy variables

Table 1 shows the policies examined in this paper and their implementation dates, including state Partnership programs, tax deductions, and tax credits.

Exhibit 1: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
AL	3/1/2009	X			1995	Deduction Individuals are allowed an itemized deduction for qualified long term care insurance contract to the extent that the amount does not exceed specified limitations. These amounts are indexed. Businesses, whether incorporated or not, may deduct LTC insurance as reasonable compensation expenses.
AK	Not Filed					No tax benefits presently.
AZ	7/1/2008					No tax benefits presently
AS	7/1/2008			X	1997	Deduction A deduction is allowed to the limits provided in the federal Internal Revenue Code (see above for details)
CA	Original Partnership			X	2003	Deduction A deduction is allowed to the limits provided in the federal Internal Revenue Code (see above for details)

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
CO	1/1/2008		X	X	2000	Credit A Credit is allowed for 25 percent of the premiums paid for long term care insurance during tax year for the individual and spouse. The Colorado credit is only applicable to thoise with federal taxable income of less than $50,000; to two individuals filing a joint return with a federal taxable income of less than $50,000 if claiming the credit for one policy; or less than $100,000 if claiming the credit for two policies.
CT	Original Partnership					No tax benefits presently
DE	11/1/2011					No tax benefits presently
DC	Not Filed	X			2005	Deduction A deduction for long term care insurance premiums paid annually is allowed from gross income provided that the deduction does not exceed $500 per year, per individual. It does not matter whether the individual files joiuntly and the LTC poilicy must meet District of Columbia's definitions.
FL	1/1/2007					No tax benefits presently
GA	1/1/2007					No tax benefits presently

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
HI	Pending			X	1999	Deduction Same as federal tax law, except subject to 7.5% of HI adjusted gross income, instead of federal adjusted gross income.
ID	11/1/2006	X			2003	Deduction A deduction is allowed for premium paid by a taxpayer for LTCi which is for the benefit of the taxpayer, a dependent of the taxpayer or an employee of a taxpayer and the amount can be deducted from taxable income to the extent the premium is not otherwise deducted by taxpayer.
IL	Pending					No tax benefits presently
IN	Original Partnership	X			2000	Deduction Deduction up to full cost of premium paid for qualified LTCi for taxpayer and taxpayer's spouse paid in the taxable year.
IA	1/1/2010			X	1997	Deduction A deduction is allowed to the limits provided in the federal Internal Revenue Code (see above for details)
KS	4/1/2007	X			2003	Deduction For tax years beginning in 2005,a subtraction from federal adjusted gross income for $500 in the tax year 2005, increasing each year by $100 until 2010. After 2010, it is a $1000 subtraction from the federal adjusted gross income for premium costs for qualified LTCi.

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
KY	6/16/2008	X			1998	Deduction Deduction from adjusted gross income allowed for any amount paid during the tax year for LTC premiums.
LA	10/1/2009		X		2002	Credit A credit against the individual income tax is allowed for amounts paid as premiums for eligible long term care insurance. The amount of the credit equals 10 percent of the total amount of premiums paid each year by each individual claiming the tax credit and the policy must meet the specific qualification requirements.
ME	7/1/2009	X			1990	Deduction The Superintendent of the State must certify the policy you purchase as a qualifying long term care policy. Then you are pemitted a deduction as long as the amount subtracted is reduced by the amount claimed as a deduction for federal income tax purposes. Sounds more complicated than it really is. Employers providing long term care benefits to employees may also qualify for a tax credit which follows a formula equal to the lowest of $5,000, 20 percent of the costs or $100 for each employee covered.

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
MD	1/1/2009		X		2000	Credit Taxpayer is allowed a one-time credit against the state income tax in an amount equal to 100% of eligible LTCi premium paid. The credit may not exceed $500 for each insured, may not be claimed by more than one taxpayer with respect to the same individual and may not be claimed if the insured was covered by LTCi before July 1 2000. No carryover is allowed. For employers, a credit up to an amount equal to 5% of the costs incurred by the employer during the taxable year for providing LTCi as part of the benefit package. The credit may not exceed $5000 or $100 for each employee covered by LTCi under the benefit package.
MA	Proposed					No tax benefits presently
MI	Work stopped					No tax benefits presently
MN	7/1/2006		X		2000	Credit A credit is allowed for LTCi premiums equal to the lesser of: (1) 25% of premiums paid to the extent not deducted in determining federal taxable income; or (2) $100 (which equals $200 for married couples who file joint tax returns.)

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
MS	Not Filed		X		2007	Credit A credit equal to 25% of premium costs paid during the taxable year for a qualified policy for self, spouse, parent, parent-in-law, or dependent. The credit cannot exceed $500.
MI	8/1/2008	X			2007	Deduction Taxpayers may deduct 100% of all non-reimbursed amounts paid for qualified LTCi premiums to the extent such amounts are not included in itemized deductions.
MT	7/1/2009	X	X		1997	Deduction - Credit Montana offers both a deduction for entire amount of qualified LTCi premiums covering taxpayer, taxpayer's parents, grandparents & dependents. A tax credit is now allowed for for premiums paid for long term care insurance coverage for a qualifying family member. The amount of the credit shall be based on the taxpayer's adjusted gross income and can not exceed $5,000 per qualifying family member in a taxable year. Or, $10,000 for two or more family members.

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
NE	7/1/2006	X			2006	Deduction Nevada now permits a tax deduction for Long Term Care Savings Plan contributions of up to $2,000 per married filing jointly return or $1,000 for any otrher return to the extent that it is not deducted for federal income tax purposes.
NV	1/1/2007					No tax benefits presently
NH	2/16/2010					No tax benefits presently
NJ	7/1/2008	X			1997	Deduction Deduction of LTC insurance premiums may be taken if they exceed 2% of adjusted gross income and cannot be reimbursed.

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
NM	Not Filed	X			2000	Credit / Deduction. New Mexico permits taxpayers who are age 65 and older and who are not a dependent of another taxpayer to claim a credit of $2,800 for medical care expenses which includes long term care insurance premiums paid for the filing taxpayer, spouse or dependents if expenses equal $28,000 or more within the particular taxable yeare (and so long as the expenses are nopt reimbursed). A deduction allows taxpayers an additional exemption of $3,000 for medical expenses if expenses (including the cost for LTC insurance) equal $28,000 or more within the taxable year and if expenses are not reimbursed or otherwise covered.

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
NY	Original Partnership	X	X		1996	Credit Credit for 20% of premium paid for qualifying LTCi premiums. Taxpayer is permitted to carry over to future tax years any credit amount in excess of taxpayer's tax liability for the year. Employers are eligible for a credit equal to 20% of the premiums paid during the tax year for the purchase of, or for continuing coverage under, a LTCi policy. The credit is not refundable and the credit may not reduce the tax to less than the minimum tax due. NY provided a tax deduction from LTCI premiums from 1996 through 2001, and a credit from 2002 onwards. It was doubled from 10% in 2002 to 20% in 2004.
NC	3/7/2011		X		1999	Credit A credit is allowed for premiums paid on LTC insurance for taxpayer, taxpayer's spouse or dependent in an amount equal to 15% of the premium costs, up to $350 for each policy on which the credit is claimed as long as adj. gross income meets the following limitations: Married Filing Separately <$50,000; Single <$60,000; Head of Household <$80,000; Married Filing Jointly or Qualifying Widower <$100,000.

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
ND	1/1/2007		X		1994 - 2009	Credit A credit is allowed for premiums paid on LTC insurance for taxpayer and or spouse up to $250 within any taxable year. "Utah and North Dakota have recently eliminated their tax incentives for long-term care insurance, Utah for tax year 2008 and North Dakota for tax year 2009."
OH	9/10/2007	X			1999	Deduction Deduction of federally qualified LTCi premiums for taxpayer, taxpayer's spouse and dependents to the extent deduction is not allowed in computing federal adj.gross income.
OK	7/1/2008					No tax benefits presently
OR	1/1/2008		X		2000	Credit Credit equal to the lesser of 15% of premiums paid during the tax year or $500 for LTC insurance coverage for individual, dependent or parents. For employers, a credit of $500 is allowed for each employee covered by an employer-sponsored policy.
PA	9/15/2007					No tax benefits presently
RI	7/1/2008					No tax benefits presently
SC	1/1/2009					No tax benefits presently
SD	7/1/2007					No tax benefits presently
TN	10/1/2008					No tax benefits presently
TX	3/1/2008					No tax benefits presently

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
UT	Not Filed				2000 - 2008	No tax benefits presently. "Utah and North Dakota have recently eliminated their tax incentives for long-term care insurance, Utah for tax year 2008 and North Dakota for tax year 2009."
VT	Not Filed					No tax benefits presently
VA	9/1/2007		X		2000	Deduction Virginia statutes permit a deduction from federal adjusted gross income for the amount paid in long term care insurance premiums provided the individual has not claimed a deduc tion for federal tax puposes or a credit under Virginia tax code 58.1-339.11. This code permits a credit against the individual's income taxes that shall not exceed 15 percent of the amount of long term care insurance premium paid during the taxable year. And, the credit can not be claimed to the extent that the individual has claimed a deduction for federal tax purposes. This one is worth having your CPA decide as a tax credit can be worth far more than a tax deduction.
WA	1/1/2012					No tax benefits presently

Exhibit 1 Continued: State policy implementation of long-term care insurance incentives

State	Partnership effective date	Unique State Deduction	Unique State Credit	Federal itemized deductions	Tax benefit year started	Tax details
WV	17/01/2010				2000	Deduction Deduction for LTCi premiums covering taxpayer, taxpayer's spouse, parents and dependents to the extent the amount paid for LTCi is not deducted in determining federal income tax.
WI	1/1/2009				1998	Deduction Deduction allowed for taxpayer & taxpayer's spouse for 100% of the amount paid for a LTCi policy to the extent the same deduction is not taken for federal income tax purposes.
WY	6/29/2009					No tax benefits presently

B.2 Additional summary statistics

Table 2 shows the summary statistics for approval probability and holding long-term care insurance policies, by quartile, for the samples used in the tax incentive and Partnership analyses. (Note that quartiles were calculated for the entire available HRS sample, 1996 - 2012, and are consistent between the two sets of analyses, which is why the quartiles do not contain equal numbers of observations.) The mean prevalence of insurance increases with imputed-approval quartile, from 6.5% to 12% for the tax sample, and from 7.5% to 13% for the Partnership sample.

B.3 Analysis of the effect of tax price

B.3.1 Calculation of insurance price after tax subsidies

One of the policy changes examined in this paper is the implementation of state tax deductions and credits for long-term care insurance premiums. The effect that these changes to the tax code have on the observed price for an individual or family, however, depend on their tax liability and marginal tax rates, so I examine both the binary effect of having any tax benefit, and a measure the price of $1 of long-term care insurance after tax deductions and credits have been accounted for (the after-tax price). I calculate marginal tax rates are calculated for Health and Retirement Study (HRS) respondents by running the respondents' income and demographic information through the National Bureau of Economic Research (NBER) TaxSim calculator (Feenberg and Coutts 1993). The variable inputs are shown in table 3.

Exhibit 2: Prevalence of long-term care insurance by approval quartile

	Tax incentive sample		Partnership sample	
Underwriting approval quartile	Approval probability	Has long-term care insurance	Approval probability	Has long-term care insurance
ltc_quart_1				
mean	0.048	0.065	0.048	0.075
sd	0.044	0.25	0.045	0.26
min	0.0000075	0	0.0000075	0
max	0.15	1	0.15	1
count	10746	10746	8752	8752
ltc_quart_2				
mean	0.37	0.097	0.37	0.10
sd	0.11	0.30	0.11	0.31
min	0.15	0	0.15	0
max	0.54	1	0.54	1
count	11884	11884	9615	9615
ltc_quart_3				
mean	0.67	0.11	0.67	0.12
sd	0.064	0.31	0.064	0.33
min	0.54	0	0.54	0
max	0.77	1	0.77	1
count	14481	14481	10499	10499
ltc_quart_4				
mean	0.87	0.12	0.87	0.13
sd	0.057	0.32	0.057	0.33
min	0.77	0	0.77	0
max	0.97	1	0.97	1
count	16392	16392	11367	11367
Years	1996-2006		2002 - 2012	

Exhibit 3: Variable inputs for TAXSIM

TAXSIM input	Description	HRS database	variable(s) used
pwages	Income of primary taxpayers	RAND	r#iearn
swages	Income of secondary taxpayer	RAND	s#iearn
dividends	Dividend income	RAND income and wealth files	hidivin
otherprop	Interest and other property income	RAND	h#iother
pensions	Taxable pension income	RAND	r#ipena + s#ipena
gssi	Gross social security benefits	RAND	r#isret + s#isret
transfers	Non-taxable transfer income	RAND	r#igxfr
rentpaid	Rent paid	RAND fatfiles	h079-h083
proptax	Property tax paid	RAND fatfiles	h075-h077, h186, h187
otheritem	Other itemized deductions	RAND fatfiles	q449-451 (medical), q454-456 (charity)
childcare	Child care expenses	imputed as 0	
ui	Unemployment compensation benefits	RAND	r#iunwc
mortgage	Mortgage interest paid	RAND	h#amort * 0.06
stcg	Short-term capital gain/loss	RAND	h#icap - hidivin
ltcg	Long-term capital gain/loss	imputed as 0	

B.3.2 Instrumental variable analysis

Ordinary least squares (OLS) model to estimate the effect of tax benefits on LTCI is likely to be biased and inconsistent. Income, and therefore the marginal tax rate, is correlated with demand for long-term care insurance in ways we cannot observe in the HRS data or control for in the OLS equation. Further muddying the waters, income is also related to health status and therefore underwriting probabilities. Following the strategy used by Goda (2011), I instrument for the after-tax price experienced by the respondent with a simulated average price for a nationally representative cohort (the full HRS sample in that year) subject to that state's tax laws.

B.3.3 Two-stage least squares estimate

We are interested in the causal relationship between prices and insurance update, as mediated by underwriting score. The structural equation to describe that relationship is as follows:

$$LTCI_{ist} = \alpha X_i + \gamma PRICE_{ist} \times UNDERWRITING_{ist} + UNDERWRITING + \omega_t + \sigma_s + \varepsilon_{ist} \quad \text{(B.1)}$$

where γ is the causal effect of price changes on long-term care insurance purchase. But since, in this case, tax price is endogenous to income, we start from the reduced-form and first-stage equations:

$$PRICE_ist = \beta_{01} X_i + \pi_1 SIMPRICE + \xi_{1ist} \quad \text{(B.2)}$$

$$LTCI_{ist} = \beta_{02} X_i + \pi_2 SIMPRICE + \xi_{2ist} \tag{B.3}$$

In (B.3), the parameter π_1 captures the first-stage effect of simulated tax price on the individual's observed after-tax price, after controlling for X_i. The parameter π_2 captures the reduced-form effect of simulated tax price on insurance purchase, also sometimes called the "intent to treat" effect. The covariate-adjusted IV estimator is the sample analog of the ratio π_2/π_1. Substituting the first-stage equation into the causal (structural) equation gives the two-stage estimate of γ:

$$LTCI_{ist} = \alpha X_i + \gamma[\beta_{01} X_i + \pi_1 SIMPRICE] + \xi_{2ist} \tag{B.4}$$

In a random sample, the first-stage values are fitted by:

$$\widehat{PRICE}_{ist} = \hat{\beta}_{01} X_i + \hat{\pi}_1 SIMPRICE \tag{B.5}$$

where β_{01} and π_1 are OLS estimates from equation (B.2).

The two-stage least squares (2SLS) estimate of γ can be constructed by estimating the coefficient on $\hat{s}_i$ in the regression of $LTCI_i$ on X_i and $\widehat{PRICE}_{ist}$.

$$LTCI_{ist} = \alpha X_i + \gamma \hat{s}_i + [\eta_i + \gamma(s_i - \hat{s}_i)] \tag{B.6}$$

B.3.4 Verifying IV assumptions

Several assumptions that form the basis for the IV framework should be verified or indirectly tested in order to make a candidate instrument a plausibly valid one. They include:

Instrument relevance: The instrument, simulated tax price, should explain variation in the premium price that households experience.

This is testable in the data by looking at the strength and precision of estimate for the coefficient $\hat{\pi}_1$ on *SIMPRICE* in equation [1]. When the instrument is weak, even small biases that result from any violation of the assumptions that follow are magnified. Table 4, shows the regression of observed after-tax price on simulated tax price. The F statistic for the coefficient on the instrument for both the state- and individual-fixed-effects models, exceeds the suggested critical value of 10 for a single instrument (Staiger and Stock 1997).

Independence: We assume that the instrument is as good as randomly assigned, conditional on the values of the observed X covariates. In this analysis, it is the assumption that controlling for observed covariates and year- and individual-fixed-effects, the introduction of a tax subsidy is independent of other factors that influence a household's decision to purchase long-term care insurance. This assumption might be violated if, for example, changes to the tax benefits for long-term care insurance are in response to shocks to the robustness of regional insurance markets that affect prices (since in this analysis, the year-to-year changes in premium price are presumed to be consistent across regional markets).

Exclusion restriction: Simulated tax price affects insurance take-up only through the premium

Exhibit 4: First-stage instrument strength

	(1) State FE	(2) Person FE
Simulated tax price	1.464***	1.126***
	(49.85)	(59.77)
N	53503	53503
N_g	51	16080
r2_a	0.315	0.275

F statistic for the instrument is shown in parentheses. Models regress the respondent's observed after-tax pricefor long-term care insurance premiums on the instrument, which is the state-average after-tax price for a nationally representative cohort in that year. State fixed effects and state-year time trend included in both models; model (2) includes individual fixed effects. Controls include gender, age, marital status, education, number of children, race (black, white, other), Hispanic ethnicity, retirement status, self-reported health status, state nursing-home occupancy, nursing home beds per person over age 66, and proportion of the state population over age 65.

discounts that it creates for households, and not through any other channel. An example of a violation of this assumption would occur if policy changes instigated a marketing campaign by insurance companies in that state. While still a downstream effect of the policy in a general sense, extra efforts on the part of insurance companies would not be strictly interpretable as a price elasticity.

Table 5 shows the standardized differences in means of the covariates, splitting the sample at the fiftieth percentile. In Column 1 the sample is split by the observed tax price, and in Column 2 the sample is split by the instrument, state-averaged tax price. Column 3 is the ratio of the S.D.'s of the IV to observed tax price. Mahalanobis Distance, in the final row, is a summary measure of covariate distance. Covariate balance is improved on all covariates except race. That suggests there is a strong correlation between race and state implementation of tax subsidies. While race is included as a control variable, that imbalance suggests the possibility that other important characteristics that are unobserved may also be correlated with state policies and with long-term care insurance. However, since the preferred model specification includes fixed effects for state and, in the preferred model, individual fixed effects, they are unlikely to create bias in this analysis.

Exhibit 5: Covariate Balance by Tax Price and Instrument.

	Tax Price	IV	Ratio
Female	0.03	-0.01	0.33
Married	-0.09	0.04	0.42
Age	1.94	-0.55	0.28
Some_HS	0.13	0.05	0.36
GED	0.01	-0.00	0.62
HS_grad	0.01	-0.02	1.46
Some_college	-0.03	0.00	0.07
College_grad	-0.12	-0.03	0.26
Children	0.34	0.14	0.41
Hispanic	0.06	0.05	0.78
White	-0.00	0.05	11.31
Black	-0.00	-0.05	20.76
Other_race	0.01	-0.00	0.13
Race_missing	0.00	0.00	0.65
Retired	0.17	-0.04	0.22
Health	0.23	0.02	0.07
Income	-36.90	-5.52	0.15
Assets	-167.20	-28.76	0.17
MahalDis	0.40	0.08	0.21

Columns 1 and 2 show standardized differences (S.D.) in means of covariates when the sample is split at the 50th percentile. The final row reports Mahalanobis distance, an overall measure of covariate distance.

C

Supplemental Material to "Instrumental Variables for Multiple Treatments"

C.1 Variable Definitions

Exhibit 1: Variable Definitions

Item	Definition	Data source
Atypical antipsychotic treatment	Study subjects received an atypical antipsychotic (AHFS therapeutic class code 28-16-08-04). Five study drugs were retained: aripiprazole, olanzapine, quetiapine, risperidone, and ziprasidone. The variable is a vector $D = \{d_1, d_2, d_3, d_4, d_5\}$ where each dk is an indicator that the individual treatment received treatment k, one of the study drugs.	PDE
Physician prescribing preference, multinomial (definition 1)	Vector of indicators $Z^1 = \{z_1, z_2, z_3, z_4, z_5\}$ where z_k is an indicator that the individuals physician prescribed treatment k most often for previously treated patients.	PDE
Physician prescribing preference, continuous (definition 2)	Vector of continuous $[0, 1]$ variables $Z^2 = \{z_1, z_2, z_3, z_4, z_5\}$ where z_k is the proportion of previous prescriptions.	PDE
Female	=1 if female	MMR
Age	Age in years at time of index prescription date	MMR
Medicaid eligible	Study subject is Medicare/Medicaid dual enrollee	MMR
Part D Low Income	Receives Medicare Part D low-income subsidy	MMR
Medicaid Add-on	Medicaid add-on reimbursement code	MMR

Exhibit 1: (continued) Variable Definitions

Item	Definition	Data source
Mental health diagnoses	[3]1 claim with ICD-9 diagnosis code for mental illness within 365 days before or after the index prescription date. Only subjects with diagnosis of schizophrenia were retained in the study sample. Other diagnoses are not exclusive categories.	Claims
Schizophrenia	295.XX	Claims
Bipolar depression	296.4X, 296.5X, 296.6X, 296.8X, 296.7X	Claims
Senile dementia	290.XX	Claims
Major depression	296.2X, 296.82, 296.3X, 311.XX	Claims
Substance abuse	303.X, 304.X	Claims
Other psychosis	298.XX	Claims
Non-psychotic disorders	300.XX, 301.XX, 302.XX, 306.XX, 309.XX	Claims
Health screenings	Procedure claim code within 180 days prior to index prescription date	Claims
Diabetes screening	Diabetes screening (ICD-9 V771.XX), HbA1c test (CPT 83036, 86037), glucose tolerance test (CPT 82947, 82950, 82951)	Claims
Dyslipidemia screening	CPT 80061, 82465, 83718, 83721, 84478	Claims
Concomitant treatments		
Anti-diabetic drugs	AHFS codes 24-08-XX-XX, 24-24-XX-XX, 24-32-XX-XX	
Anti-lipemic drugs	AHFS codes 24-06-XX-XX, 56-24-XX-XX, 88-08-XX-XX, 80-00-XX-XX	
Baseline co-morbidities	Claim with associated diagnosis code or prescription drug event within 365 days prior to index date	Claims/PDE

Exhibit 1: (continued) Variable Definitions

Item	Definition	Data source
Obesity	ICD-9 278.XX	Claims
Diabetes	ICD-9 250.XX or anti-diabetic drugs (see above)	Claims/PDE
Dyslipidemia	ICD-9 272.XX or anti-lipemic drugs (see above)	Claims/PDE
Stroke/transient ischemic attack	ICD-0 433.1, 435.9, 435.8, 435.0, 434.91, 438.*, 431.*, 434.01, 434.11	Claims
Area poverty	Percent of residents in zip code living below the federal poverty line in 2010.	SAIPE

Notes:

AHFS = American Hospital Formulary Service (therapeutic class codes)

PDE = prescription drug event

MMR = monthly membership report

ICD-9 = International Classification of Disease 9th Edition (diagnosis & procedure codes)

CPT = Current Procedural Terminology (procedure codes)

SAIPE = Small Area Income and Poverty Estimates

References

Alexander, G.C., S.A. Gallagher, A. Mascola, R.M. Moloney, and R.S. Stafford. 2011. "Increasing Off-Label Use of Antipsychotic Medications in the United States, 1995–2008." *Pharmacoepidemiology and Drug Safety* 20(2): 177–84.

American Diabetes Association. 2004. "Consensus Development Conference on Antipsychotic Drugs and Obesity and Diabetes." *Diabetes Care* 27(2): 596–601.

Angrist, J.D., G.W. Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434): 444–54.

Baiocchi, M., J. Cheng, and D.S. Small. 2014. "Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference." *Statistics in Medicine* 33: 2297–340.

Brookhart, M.A., and S. Schneeweiss. 2007. "Preference-Based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results." *International Journal of Biostatistics* 3(1), Article 14. DOI: 10.2202/1557-4679.1072.

Brookhart, M.A., P.S. Wang, D.H. Solomon, and S. Schneeweiss. 2006. "Evaluating Short-Term Drug Effects Using a Physician-Specific Prescribing Preference as an Instrumental Variable." *Epidemiology* 17: 268–75.

Brooks, J.M., E.A. Chrischilles, S.D. Scott, and S.S. Chen-Hardee. 2003. "Was Breast Conserving Surgery Underutilized for Early Stage Breast Cancer? Instrumental Variables Evidence for Stage II Patients from Iowa." *Health Services Research* 38(6, pt. 1): 1385–402.

Brown, J.R., and A. Finkelstein. 2009. "The Private Market for Long-Term Care Insurance in the United States: A Review of the Evidence." *Journal of Risk and Insurance*

76(1): 5–29.

———————. 2008. "The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market." *American Economic Review* 98(3): 1083–1102.

———————. 2007. "Why Is the Market for Long-Term Care Insurance So Small?" *Journal of Public Economics* 91(10): 1967–91.

Brown, J.R., N.B. Coe, and A. Finkelstein. 2007. "Medicaid Crowd-Out of Private Long-Term Care Insurance Demand: Evidence from the Health and Retirement Survey."' *Tax Policy and the Economy* 76(1): 1–34.

Bureau of the Census, Population Division. 2014. "Table 1. Projections of the Population and Components of Change for the United States: 2015 to 2060 (NP2014-T1)." Computer file. [Release date Dec. 2014]. Washington, D.C.: Bureau of the Census. Available at http://www.census.gov/population/projections/data/national/2014/summarytables.html [last visited 6 Sept. 2016].

Casey, D.E. 2005. "Metabolic Issues and Cardiovascular Disease in Patients with Psychiatric Disorders." *American Journal of Medicine* 118(S2): 15S–22S.

Center for Retirement Research at Boston College. 2016. "National Retirement Risk Index." Available at http://crr.bc.edu/special-projects/national-retirement-risk-index/ [last visited 6 Sept. 2016].

Cole, J.A., H. Norman, L.B. Weatherby, and A.M. Walker. 2006. "Drug Copayment and Adherence in Chronic Heart Failure: Effect on Cost and Outcomes." *Pharmacotherapy* 26: 1157–64.

Congressional Budget Office. 2013. Rising demand for long-term services and supports for elderly people [Internet]. Washington (DC): CBO; Jun [cited 2016 Jun 16]. Available from: https://www.cbo.gov/sites/default/files/113th-congress-2013-2014/reports/44363-LTC.pdf [last visited 6 Sept. 2016].

Cornell, P., D.C. Grabowski, M. Cohen, X. Shi, and D.G. Stevenson. 2016. "Medical Un-

derwriting in Long-Term Care Insurance: Market Conditions Limit Options for Higher-Risk Consumers." *Health Affairs* (Millwood, VA) 35(8): 1494-503

Currie, J., and J. Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy* 104(6): 1263–96.

Cutler, D.M. 1996. "Why don't markets insure long-term risk?" Harvard University and National Bureau of Economic Research. Available at http://scholar.harvard.edu/files/cutler/files/ltc_rev.pdf?m=1360040872 [last visited 6 Sept. 2016].

Favreault, M.M., H. Gleckman, and R.W. Johnson. 2015. "Financing Long-Term Services and Supports: Options Reflect Trade-Offs For Older Americans and Federal Spending." *Health Affairs* (Millwood, VA) 34(12): 2181-91.

Feenberg, D.R. 2013. Internet TAXSIM. Version 9. Computer-program interface. Last updated 3 Apr. Available at http://users.nber.org/~taxsim/taxsim-calc9/index.html [last visited 6 Sept. 2016].

Feenberg, D.R., and E. Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12(1): 189–94.

Fisherkeller, K. 2015a. *Group Long-Term Care Insurance Survey—U.S.: Annual Review 2014*. Windsor, Ct.: Life Insurance and Market Research Association International, Inc.

———————. 2015b. *Individual Long-Term Care Insurance—U.S.: Annual Review 2014*. Windsor, Ct.: Life Insurance and Market Research Association International, Inc.

Frank, R.G., and R.J. Zeckhauser. 2007. "Custom-Made versus Ready-to-Wear Treatments: Behavioral Propensities in Physicians' Choices." *Journal of Health Economics* 26(6): 1101–227.

Garabedian, L.F., P. Chu, S. Toh, A.M. Zaslavsky, and S.B. Soumerai. 2014. "Potential Bias

of Instrumental Variable Analyses for Observational Comparative Effectiveness Research." *Annals of Internal Medicine* 161: 131–38.

Gareri, P., C. Segura-García, V.G.L. Manfredi, A. Bruni, P. Ciambrone, G. Cerminara, G. De Sarro, and P. De Fazio. 2014. "Use of Atypical Antipsychotics in the Elderly: a Clinical Review." *Journal of Clinical Interventions in Aging* 9: 1363–73.

Genworth Financial. 2015. *Genworth 2015 Cost of Care Survey: Home Care Providers, Adult Day Health Care Facilities, Assisted Living Facilities And Nursing Homes*. N.p.: Genworth Financial. Available at https://www.genworth.com/dam/Americas/US/PDFs/Consumer/corporate/130568_040115_gnw.pdf [last visited 6 Sept. 2016].

———————. 2013. *TrueView Underwriting Quick Reference Guide*. N.p.: Genworth Financial. Available at https://www.genworth.com/dam/Americas/US/PDFs/Producer/Forms/151328_030113_gnw.pdf [last visited 6 Sept. 2016].

Goda, G.S. 2011. "The Impact of State Tax Subsidies for Private Long-term Care Insurance on Coverage and Medicaid Expenditures." *Journal of Public Economics* 95(7–8): 744-57.

Greenhalgh-Stanley, N. 2012. "Can the Government Incentivize the Purchase of Private Long-Term Care Insurance? Evidence from the Long-Term Care Partnership Program." Center for Retirement Research at Boston College Working Papers. Available at http://crr.bc.edu/wp-content/uploads/2012/09/wp_2012-14.pdf [last visited 6 Sept. 2016].

Health and Retirement Study. 2012. Public-use dataset. Produced and distributed by the University of Michigan, with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Hebert, P.L., L.S. Geiss, E.F. Tierney, M.M. Engelgau, B.P. Yawn, and A.M. McBean. 1999. "Identifying Persons with Diabetes Using Medicare Claims Data." *American Journal of Medical Quality* 14(6): 270–77.

Hendren, N. 2013. "Private Information and Insurance Rejections." *Econometrica* 81(5): 1713–62.

Imbens, G.W., and D.B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.

Johnson, R.W., S.G. Schaner, D. Toohey, and C.E. Uccello. 2007. *Modeling the Decision to Purchase Long-Term Care Insurance.* Prepared for the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Office of Disability, Aging and Long-Term Care Policy. Jan. Washington, DC: The Urban Institute. Available at https://aspe.hhs.gov/basic-report/modeling-decision-purchase-private-long-term-care-insurance [last visited 6 Sept. 2016].

Koppes L.L., J.M. Dekker, H.F. Hendriks, L.M. Bouter, and R.J. Heine. 2005. "Moderate Alcohol Consumption Lowers the Risk of Type 2 Diabetes: A Meta-analysis of Prospective Observational Studies." *Diabetes Care* 28(3): 719–25.

Landrum, M.B., and J.Z. Ayanian. 2001. "Causal Effect of Ambulatory Specialty Care on Mortality Following Myocardial Infarction: A Comparison of Propensity Score and Instrumental Variable Analyses." *Health Services and Outcomes Research Methodology* 2(3): 221–45.

Lang, I., R.B. Wallace, F.A., Huppert, and D. Melzer. 2007. "Moderate Alcohol Consumption in Older Adults Is Associated with Better Cognition and Well-Being Than Abstinence." *Age and Ageing* 36(3): 256–61.

LifePlans. 2012. *Who Buys Long-Term Care Insurance in 2010–2011? A Twenty Year Study of Buyers and Non-Buyers (In the Individual Market).* Washington, DC: America's Health Insurance Plans. Available at http://ltcinsurancece.com/wp-content/uploads/2015/07/Who-Buys-LTC-Insurance-2010-2011.pdf [last visited 6 Sept. 2016].

———————. 2010. "Appendix Ja: A Profile of Declined Long-Term Care Insurance Ap-

plicants: A View of Selected Socio-Demographic Characteristics." Draft. 27 Dec. 26 pp. Appendix to *A Report on the Actuarial, Marketing, and Legal Analysis of the CLASS Program.* Prepared for the Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Office of Disability, Aging, and Long-Term Care Policy. Waltham, MA: LifePlans. Available at https://aspe.hhs.gov/report/report-actuarial-marketing-and-legal-analyses-class-program [last visited 6 Sept. 2016].

Lin, H., and J. Prince. 2013. "The Impact of The Partnership Long-Term Care Insurance Program on Private Coverage." *Journal of Health Economics* 32(6): 1205–1213.

McClellan, M., B.J. McNeil, and J.P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in The Elderly Reduce Mortality? Analysis Using Instrumental Variables." *JAMA* 272(11): 859–66.

Mor, V., J. Zinn, J. Angelelli, J.M. Teno, and S.C. Miller. 2004. "Driven To Tiers: Socioeconomic and Racial Disparities in the Quality of Nursing Home Care." *Milbank Quarterly* 82(2): 227–56.

Murtaugh, C.M., P. Kemper, and B.C. Spillman. 1995. "Risky Business: Long-Term Care Insurance Underwriting." *Inquiry* 32(3): 271–84.

National Association of Insurance Commissioners. 2014. "Model Regulation Service—4th quarter 2014: Long-Term Care Insurance Model Regulation." Available at http://www.naic.org/store/free/MDL-641.pdf [last visited 6 Sept. 2016].

———————. 2009. *A Shopper's Guide to Long-Term Care Insurance*. Kansas City, MO: NAIC. Available at http://www.nhli.org/webinars/2012/AARP_NHLI_ShoppersGuideToLongTermCareInsurance.pdf [last visited 6 Sept. 2016].

O'Shaughnessy, C.V. 2014. "National Spending for Long-Term Services and Supports (LTSS), 2012." National Health Policy Forum. 27 Mar. 8 pp. Washington, DC: The George Washington University. Available at http://www.nhpf.org/library/details.cfm/2783 [last visited 6 Sept. 2016].

Pincus, H.A., T.L. Tanielian, S.C. Marcus, M. Olfson, D.A. Zarin, J. Thompson, and J.M. Zito. 1998. "Prescribing Trends in Psychotropic Medications." *JAMA* 279: 526–31.

Pizer, S.D. 2016. "Falsification Testing of Instrumental Variables Methods for Comparative Effectiveness Research." *Health Services Research* 51(2): 790–97.

RAND Center for the Study of Aging. 2015. RAND HRS Data, Version O. Produced with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA: RAND.

Rassen, J.A., M.A. Brookhart, R.J. Glynn, M.A. Mittleman, and S. Schneeweiss. 2009. "Instrumental Variables II: Instrumental Variable Application—in 25 Variations, the Physician Prescribing Preference Generally Was Strong and Reduced Covariate Imbalance." *Journal of Clinical Epidemiology* 62(12): 1233–41.

Rector, T.S., S.L. Wickstrom, M. Shah, N. Thomas Greeenlee, P. Rheault, J. Rogowski, V. Freedman, J. Adams, and J.J. Escarce. 2004. "Specificity and Sensitivity of Claims-Based Algorithms For Identifying Members of Medicare+Choice Health Plans That Have Chronic Medical Conditions." *Health Services Research* 39(6, pt. 1): 1839–57.

Royston, P. 2004. "Multiple imputation of missing values." *Stata Journal* 4(3): 227–41.

Sanghavi, P., A.B. Jena, J.P. Newhouse, and A.M. Zaslavsky. 2015. "Outcomes of Basic versus Advanced Life Support for Out-of-Hospital Medical Emergencies." *Annals of Internal Medicine* 163(9): 681–90.

Schneeweiss, S. 2007. "Developments in Post-Marketing Comparative Effectiveness Research." *Clinical Pharmacology and Therapeutics* 82(2): 143–156.

Schneeweiss, S., D.H. Solomon, P.S. Wang, J. Rassen, and M.A. Brookhart. 2006. "Simultaneous Assessment of Short-Term Gastrointestinal Benefits and Cardiovascular Risks of Selective Cyclooxygenase 2 Inhibitors and Nonselective Nonsteroidal Anti-inflammatory Drugs: An Instrumental Variable Analysis." *Arthritis and Rheuma-*

tology 54: 3390–98.

Senate Commission on Long-Term Care. 2013. *Report to the Congress*. 113th Congr., 1st Sess., 30 Sept. Report prepared as required by Section 643 of the American Taxpayer Relief Act of 2012 (P.L. 112-240), by Bruce Chernof, Mark Warshawsky, et al. Washington, D.C.: United States Government Printing Office. Available at http://www.gpo.gov/fdsys/pkg/GPO-LTCCOMMISSION/pdf/GPO-LTCCOMMISSION.pdf [last visited 6 Sept. 2016].

Shin, J., N. Choi, S. Jung, J. Lee, J.S. Kwon, and B. Park. 2013. "Risk of Ischemic Stroke with the Use of Risperidone, Quetiapine and Olanzapine in Elderly Patients: A Population-Based, Case-Crossover Study." *Journal of Psychopharmacology* 27(7): 638–644.

Smedley, B.D., A.Y. Stith, and A.R. Nelson, eds. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press.

Staiger, D., and J.H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65(3): 557–86.

Stevenson, D.G., M.A. Cohen, E.J. Tell, and B. Burwell. 2010. "The Complementarity of Public and Private Long-Term Care Coverage." *Health Affairs* (Millwood, VA) 29(1): 96–101.

Stewart, K.A., D.C. Grabowski, and D.N. Lakdawalla. 2009. "Annual Expenditures For Nursing Home Care: Private and Public Payer Price Growth, 1977 To 2004." *Med Care* 47(3): 295–301.

Swanson, S.A., and M.A. Hernán. 2013. "Commentary: How To Report Instrumental Variable Analyses (Suggestions Welcome)." *Epidemiology* 24(3): 370–374.

Taub, A.A.L., A. Kolotilin, R.S. Gibbons, and E.R. Berndt. 2011. "The Diversity of Concentrated Prescribing Behavior: an Application to Antipsychotics." NBER Work-

ing Paper Series, no. 16823. Cambridge, MA: National Bureau of Economic Research. Available online at http://www.nber.org/papers/w16823 [last visited 6 Sept. 2016].

Temkin-Greener, H., D.B. Mukamel, and M.R. Meiners. 2000-2001. "Long-Term Care Insurance Underwriting: Understanding Eventual Claims Experience." *Inquiry* 37(4): 348–58.

Tracer, Z., and L. Davison. 2014. "Genworth CEO Sees Tough Turnaround from $844 Million Loss." Bloomberg [serial on the Internet]. 5 Nov.; updated 6 Nov. Available at http://www.bloomberg.com/news/articles/2014-11-05/genworth-posts-844-million-loss-on-long-term-care-shortfall-1- [last visited 6 Sept. 2016].

Verdoux, H., M. Tournier, and B. Bégaud. 2010. "Antipsychotic Prescribing Trends: A Review of Pharmaco-epidemiological Studies." *Acta psychiatrica Scandinavica* 121(1): 4–10.

Wang, P.S., S. Schneeweiss, J. Avorn, M.A. Fischer, H. Mogun, D.H. Solomon, and M.A. Brookhart. 2005. "Risk of Death in Elderly Users of Conventional vs. Atypical Antipsychotic Medications." *New England Journal of Medicine* 353(22): 2335–41.