

Employer-Sponsored Health Insurance Premiums and Income in U.S. Tax Data

Ithai Z. Lurie and Corbin L. Miller*

Office of Tax Analysis, U.S. Department of the Treasury

January 2022

Abstract

The distribution of employer-sponsored health insurance premiums in the United States is difficult to fully characterize with survey data. Using administrative tax data, we plot the population distribution of premiums and provide new evidence about the relationship between premiums and demographic characteristics, such as income. First, we demonstrate that the distribution of premiums measured in the tax data is comparable to survey data. Then we leverage detailed individual-level information for the population of tax filers to show that those with higher incomes are more likely to have an employer-sponsored insurance (ESI) policy and, conditional on having a policy, select more expensive plans. In fact, those in the top 10 percent of the income distribution account for 21 percent of aggregate ESI premiums. Even with these disparities though, inequality in ESI spending is less than half of that in income overall.

Keywords: Income, Health Insurance Premiums, ESI, Tax Data, Inequality

JEL Codes: D31, H51, I13, J32

* Direct correspondence to corbin.miller@treasury.gov. We would like to thank participants at the annual OTA Conference and NTA Annual Conference for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S. Department of the Treasury or the U.S. Government.

Introduction

Employers and employees together spend nearly \$1 trillion on Employer-Sponsored Health Insurance (ESI) premiums each year.¹ Most of these premiums are excluded from income and payroll taxes, resulting in an annual tax expenditure of about \$200 billion.² Even though growth in the cost of ESI has outstripped growth in wages – making health benefits an increasingly important percentage of employee compensation – little is known about the correlation between the cost of ESI and income due to data limitations. At the same time, income inequality has also been growing and there is an active debate about the best ways to measure income using administrative tax data (Auten and Splinter, 2018; Piketty, Saez, and Zucman, 2018). ESI premiums have been a blind spot in these studies because nontaxable benefits are not usually measured as carefully in administrative tax data. Current studies either impute health benefits from survey data, such as the Current Population Survey (CPS) or Medical Expenditure Panel Survey – Insurance Component (MEPS-IC), or rely on premium information in tax records with important limitations. Using newly linked administrative tax information about ESI coverage and premium costs this paper helps fill this gap in knowledge.

Starting in 2015, firms with more than 50 employees must file with the Internal Revenue Service (IRS) Form 1095-C for all full-time employees. This form reports ESI offers as well as monthly coverage for each covered person if the firm is self-insured. If the firm uses a third-party insurer, then that insurer must submit to the IRS coverage information on Form 1095-B. Each form

¹ See National Health Expenditure Accounts. Table 24 Employer-Sponsored Private Health Insurance. 2020. <https://www.cms.gov/files/zip/nhe-tables.zip> accessed on 02/19/20

² Office of Tax Analysis. Tax Expenditures. 2015. <https://www.treasury.gov/resource-center/tax-policy/Documents/Tax-Expenditures-FY2017.pdf> accessed on 02/19/20.

has identifiers that allow us to match policyholders and associated employers to create an individual-level administrative data set of ESI policies. Total spending on ESI premiums (i.e., the combined employer and employee portions) is reported on box 12 of IRS Form W-2 (Code DD). Starting in 2012, firms that provide ESI and submit at least 250 W-2s in the prior year are required to report this Code DD amount for each employee, although some smaller firms also report this information.³ This provides individual-level information on total ESI premium costs.

In this paper, we build on recent work that uses this administrative tax data to describe the distribution of health insurance coverage among the U.S. population (Lurie and Pearce, 2021). Rather than examine the source of coverage at the individual level, we focus on policy-level information. Specifically, we link ESI premium costs to information about the policy and policyholder to calculate an annualized premium that is comparable regardless of the fraction of the year the policy was held. Importantly, with the full population of tax returns, we have information that both reliably measures income and covers the population of ESI policyholders, unlike other existing survey-based data sources.

The only other study that explores the distribution of premiums in tax data compares Code DD values with insurance costs in the CPS (Larrimore and Splinter, 2019). They rely solely on total annual costs and do not link to the number of months of coverage or people covered. This is problematic since the cost of a single plan is dramatically less than a family plan. Even when restricting to single filers with no dependents – who are presumed to have a single coverage plan – it is difficult to interpret the dollar amounts without knowing the number of months covered.

³ Some firms include optional covered costs such as dental or vision insurance in the total cost of insurance reported as Code DD.

This approach also overlooks policyholders whose employer does not report ESI premium costs on the W-2.

We demonstrate that policies in the tax data align well with the best survey data (MEPS-IC) on features such as percent of single and non-single policies, average premiums, and age distribution of policyholders. Next, we show that the overall premium distribution in the tax data is remarkably similar to that in the MEPS-IC. Finally, we leverage the unique features of the tax data to show substantial inequality in ESI premiums not only due to differences in coverage but, conditional on having a policy, people with higher income buy more expensive policies. Our improved estimate of health compensation inequality shows that ESI premiums are more egalitarian than income or wages overall.

Data

We use IRS population data for tax year 2016 on ESI coverage (Form 1095-B and 1095-C), health care costs and wages (Form W-2), and total family income (Form 1040). Our unit of observation is a “policy”, which we define as a unique combination of a Social Security Number (SSN) and an Employer Identification Number (EIN) from a 1095 form. Hence, people might appear on multiple policies if they have multiple jobs during the year.

We do not observe the terms of the policies, such as deductibles and copays. However, we do observe monthly coverage for each person on the policy. Using this information, we define three types of plans based on the number of covered people: single, 2-person, and family plans (3 or more). We combine premium costs with coverage months to define the annualized premium for a policy as

$$\frac{\text{W-2 Premium (Code DD)}}{\text{\# of Covered Months}} \times 12.$$

Matching Process

Regulations are not uniform concerning which EIN employers should use in submitting different forms to the IRS (e.g., firms may have payroll operations for the entire firm in a subsidiary). As a result, the EIN on forms 1095B/C do not necessarily corresponds to the EIN on the W-2. Hence, matching each ESI policy to the corresponding W-2 with premium information is the primary challenge in using these data. We match about 68 million policies (70 percent) with a unique SSN-EIN in both datasets. However, as noted above, the same company can have multiple EINs and the EIN on a 1095 form may not match the EIN on a W-2 even for the same employer. For about 10 million policies, the policyholder has only one policy on file and one W-2, but the EINs do not match between the two forms. Despite this EIN mismatch, we also treat these records as matches.

There are several groups for which we either do not expect to match to a W-2 or have Code DD information and exclude them from our analysis. For example, the military is not required to report Code DD. Also, for multiemployer plans, it is unclear whether the W-2 issuer and union that sponsors the insurance are the same entity. Further, the match rate for these policies was extremely low, therefore we exclude them. About 5.5 million policies are held by someone without a W-2. These are mainly those who are not currently employed but retain health insurance coverage from a previous employer (e.g., retirees or individuals with COBRA coverage). Excluding these three groups leaves 84 million policies of which we match 78 million policies – or about 93 percent – to a W-2.

Results

Policy Characteristics

In Table 1 we show the average number of months of the year a policy covers, the average fraction of policies held for a full year, the average fraction of policies held by women, the fraction of

policies we matched, and the average premiums cost for the policies by the following characteristics: coverage source, policy type, policyholder's age, and firm size. Columns (1) and (2) show the count (in thousands) and fraction of policies in each category. The 97 million policies in our data are associated with 89 million unique policyholders.⁴ Similar to the MEPS-IC data, most plans are for single coverage. The age distribution of the policyholder is almost uniform between 27 and 64 years and most ESI policies are issued by firms with 1,000 or more employees. Column (3) shows the average number of months each policy is held, and Column (4) shows the fraction of policies held for the full year. Most notably, workers who are single or relatively younger are less likely to be covered by the same policy for a full year. This is consistent with higher rates of job switching among younger workers. Columns (5) and (6) show that 44 percent of policies are held by women and we match about two-thirds of policies to premium information. Given the regulation on which firms are required to provide code DD, it is not surprising that premium match rates increase at 250 employees. Match rates also increase with firm size. Column (7) shows the average premium by plan type. Overall, average premiums were \$6,098 for single plans, \$13,070 for 2-person plans, and \$17,171 for family plans.

Distribution of Annualized Premiums

Figure 1 compares average annualized premiums with the MEPS-IC by premium percentile. To match the MEPS-IC sampling frame and eliminate W-2 reporting selection by smaller firms, we restrict this comparison to large, private-sector firms. As expected, non-single premiums are larger

⁴ Compared to 90.07 million in the CPS (U.S. Census Bureau. Current population survey, annual social and economic supplement. 2017)

than single premiums across the distribution. The annualized premiums are similar in the MEPS-IC and tax data.

Figure 2 shows annualized ESI premiums by age, plan type, and gender. Premiums increase rapidly from age 20 to age 30. Premiums continue to increase at a slower pace from age 30 to age 60, after which they level off and become less precise due to policyholders reaching the age of retirement and Medicare eligibility and no longer receiving ESI coverage. For single plans, which we examine separately by gender, we see that women have slightly, but consistently, higher premiums than men. This difference – about 4.4 percent – remains despite provisions in the Affordable Care Act (ACA) that prohibit gender rating, although the difference is much smaller than the upwards of 50 percent difference before the ACA took effect (Pauly and Herring, 1999; Codispoti, Courtot, and Swedish, 2008).

We also find substantial geographic variation in premiums. Appendix Figures A1, A2, and A3 show county-level annualized premiums for single, 2-person, and family plans, respectively. There is clear regional variation as well as differences between counties within states and across some state borders.

Premiums and Income

Figure 3 compares income and premium percentiles by plan type. There is clearly a positive relationship between income and premiums with statistically significant slopes ranging from 0.2 for those with family plans to 0.3 for those with single plans. This suggests that those with higher incomes also have more expensive plans. Higher-income individuals could have relatively more expensive plans because they are more likely to have a wider range of plans available to them than those with lower-paying positions or they may select more generous plans among those available than people with lower incomes given the same options.

To formalize this, we plot Lorenz Curves and calculate Gini coefficients in Figure 4. We do this separately for 1040 income in panel A and W-2 wage earnings in panel B. The further the curve is from the 45-degree line (perfect equality), the more inequality there is in the distribution. This shows that those with higher incomes have a disproportionately higher share of income, premium spending, and policies. However, the share of income is clearly more inequitably distributed across the entire income range than the share of premiums or share of policies with Gini coefficients of 0.57, 0.40, and 0.25, respectively.⁵ This suggests that income is 42.5 percent more unequal than premium spending and over twice as unequal than the distribution of policies across income. The fact that there is more inequality in premium spending than policies held is consistent with higher-income individuals choosing more expensive plans or higher-income positions offering more generous health benefits.

We also calculate the share of premiums held by income percentile in Table A1. To properly relate premiums to the distribution of income, we impute premiums for policies that were not matched to a W-2 based on the distribution of observed premiums by plan type. Column (1) shows the fraction of tax returns with an ESI policy, Column (2) shows the share of policies held by families, we show the share of premiums in Column (3), and for reference we report the share of income in Column (4). The rate of ESI policy holding is high among those in the top income quartile and drops precipitously for those below the median. Those with incomes in the 95th to the 99th percentile account for 4 percent of the population but hold 5.6 percent of the aggregate number of policies. Families above median income have between 25 to 45 percent greater share of ESI

⁵ These numbers correspond to 1040 earnings, but the results for W-2 earnings are qualitatively similar.

policies compared to their share of the population. For example, the share of premiums in the top 10 percent of income is 21 percent. These columns clearly show that the amount of inequality in premiums is slightly higher than in policies, but the disparity is nowhere near as large as it is for income. Columns (5) through (8) show a similar skew at the individual-level for workers with W-2 wages.

Conclusion

We are the first to use a full population of employee-employer-linked administrative tax data in the United States to explore the distribution of ESI premiums. The tax data seems to align well with available survey data. For example, the number of individuals that have an ESI policy in their own name is similar between the CPS and our data. Furthermore, the MEPS-IC matches our data well on various dimensions especially in terms of the premium distribution.

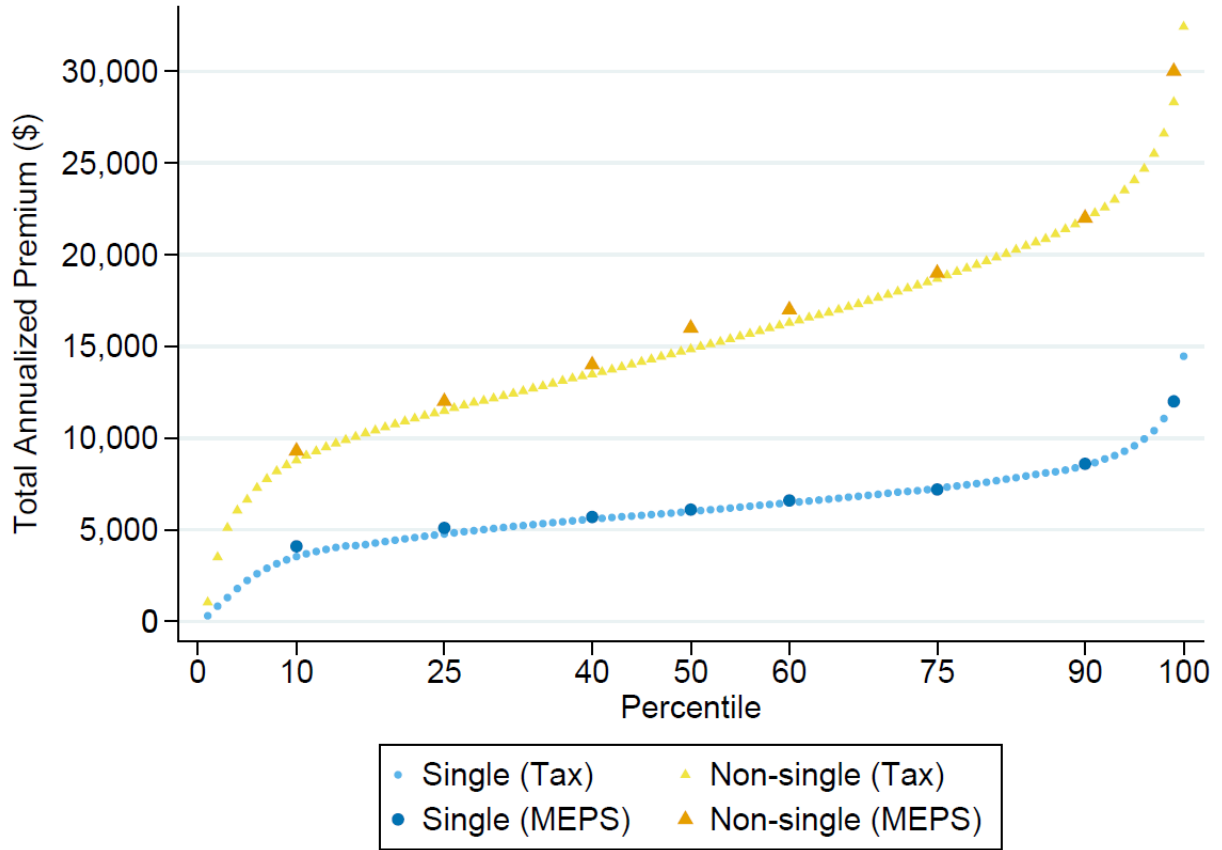
We also explore the relationship between income and ESI premiums, conditional on having ESI. Income and premiums are positively related across the entire distribution. The distribution of policies and premiums are skewed to those with higher incomes, but not nearly to the extent the distribution of income is skewed. This suggests that if studies measuring income inequality assume that health insurance premium spending follows the same distribution as income, then they will overstate the amount of overall income inequality. On the other hand, if these studies assume a uniform distribution of premium spending, then they will understate the total amount of income inequality.

References

- Auten, G., & Splinter, D. (2018). Income Inequality in the United States: Using Tax Data to Measure Long-term Trends.
- Larrimore, J., & Splinter, D. (2019). How much does health insurance cost? Comparison of premiums in administrative and survey data. *Economics Letters*, 174, 132-135.
- Lurie, I. Z., & Pearce, J. (2021). Health Insurance Coverage in Tax and Survey Data. *American Journal of Health Economics*, 7(2), 164-184.
- Pauly, M. V., & Herring, B. (1999). Pooling health insurance risks. *American Enterprise Institute*.
- Piketty, T., Saez, E., & Zucman, G. (2018). Distributional national accounts: methods and estimates for the United States. *The Quarterly Journal of Economics*, 133(2), 553-609.
- Codispoti, L., Courtot, B., & Swedish, J. (2008). Nowhere to Turn: How the Individual Health Insurance Market Fails Women. *National Women's Law Center*.

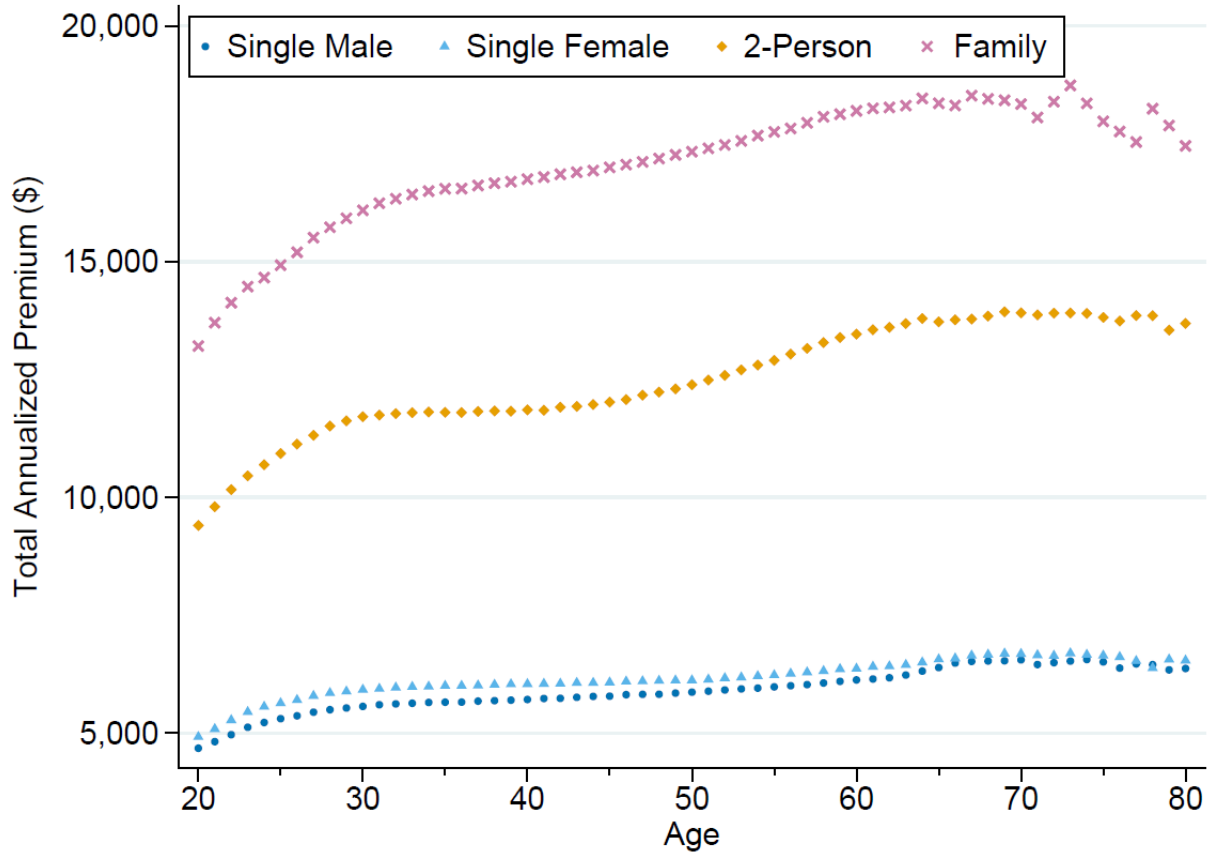
Figures

Figure 1: ESI Premium Distributions Are Similar in Tax Data and MEPS-IC



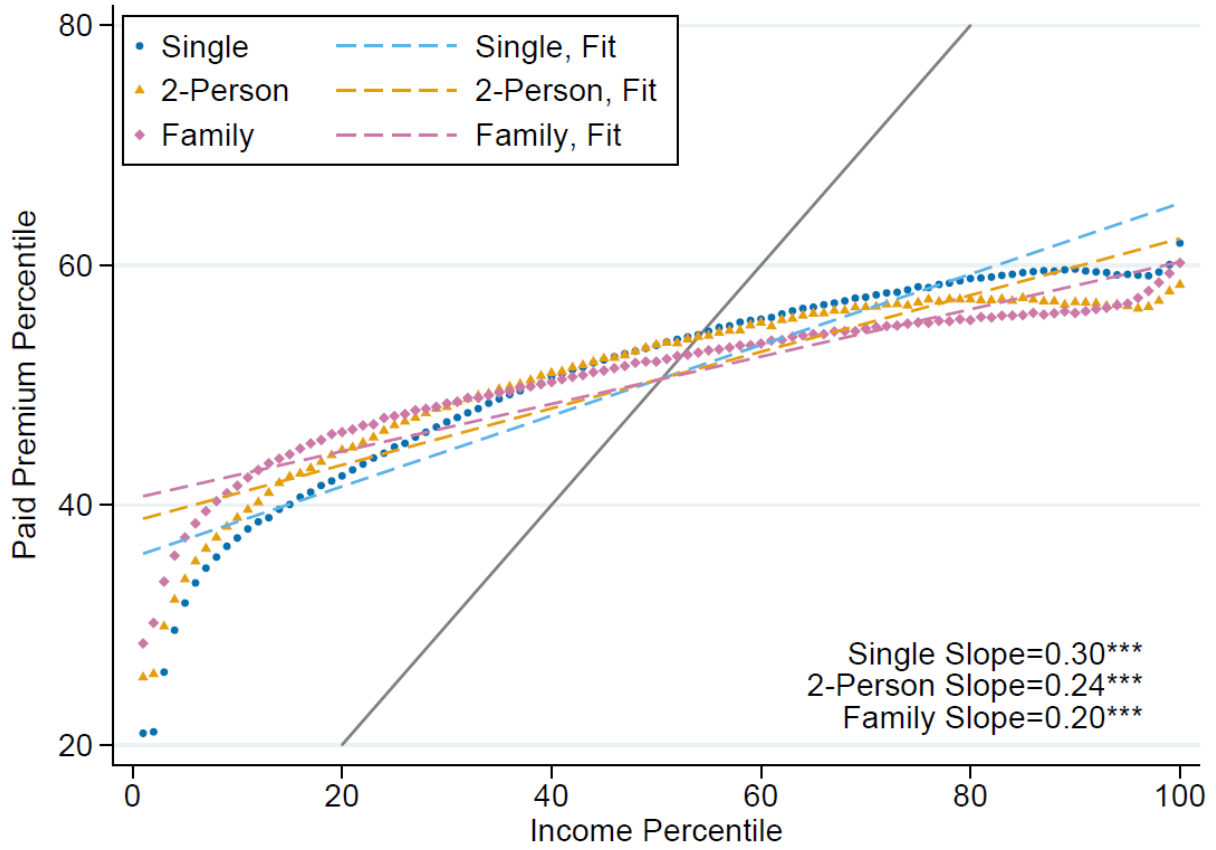
Notes: This figure displays the average annualized ESI premium for each percentile bin in the tax data along with the 10th, 25th, 40th, 50th, 75th, 90th, and 99th percentiles of premiums in the MEPS-IC.

Figure 2: ESI Premiums by Age, Plan Type, and Gender



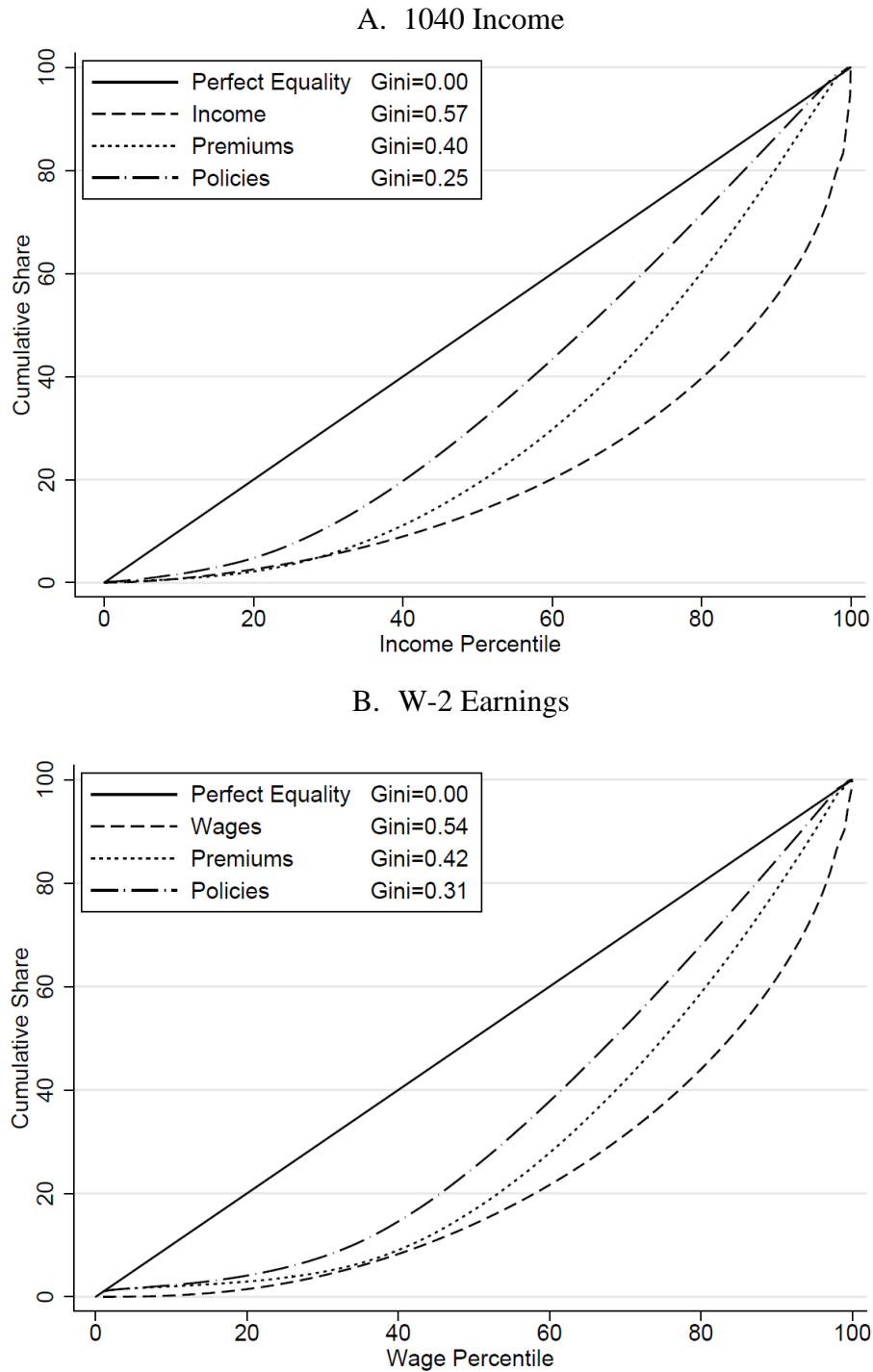
Notes: This figure shows the average annualized ESI premium at each age separately for family plans (3 or more covered people), 2-person plans, and single plans (also split by gender).

Figure 3: Income and Premium Percentiles by Plan Type



Notes: The percentile of actual premiums paid during 2016 (as opposed to annualized premiums) is plotted against income percentile. The gray line represents what a 1-to-1 relationship between the two percentiles would be, for reference. Slopes were calculated using OLS with robust standard errors. *** corresponds to a significance level of $p < .01$.

Figure 4: Lorenz Curves of Income and ESI Premium Inequality



Notes: This figure portrays the inequality in health insurance policies and premiums by earnings (Form 1040 income in panel A and wage income in panel B) percentile as Lorenz curves. The solid black line displays the curve that would result if policies and premiums were equally distributed across income. The Gini coefficient provides the ratio of the area between the line of perfect equality and the Lorenz curve to the entire area under the line.

Tables

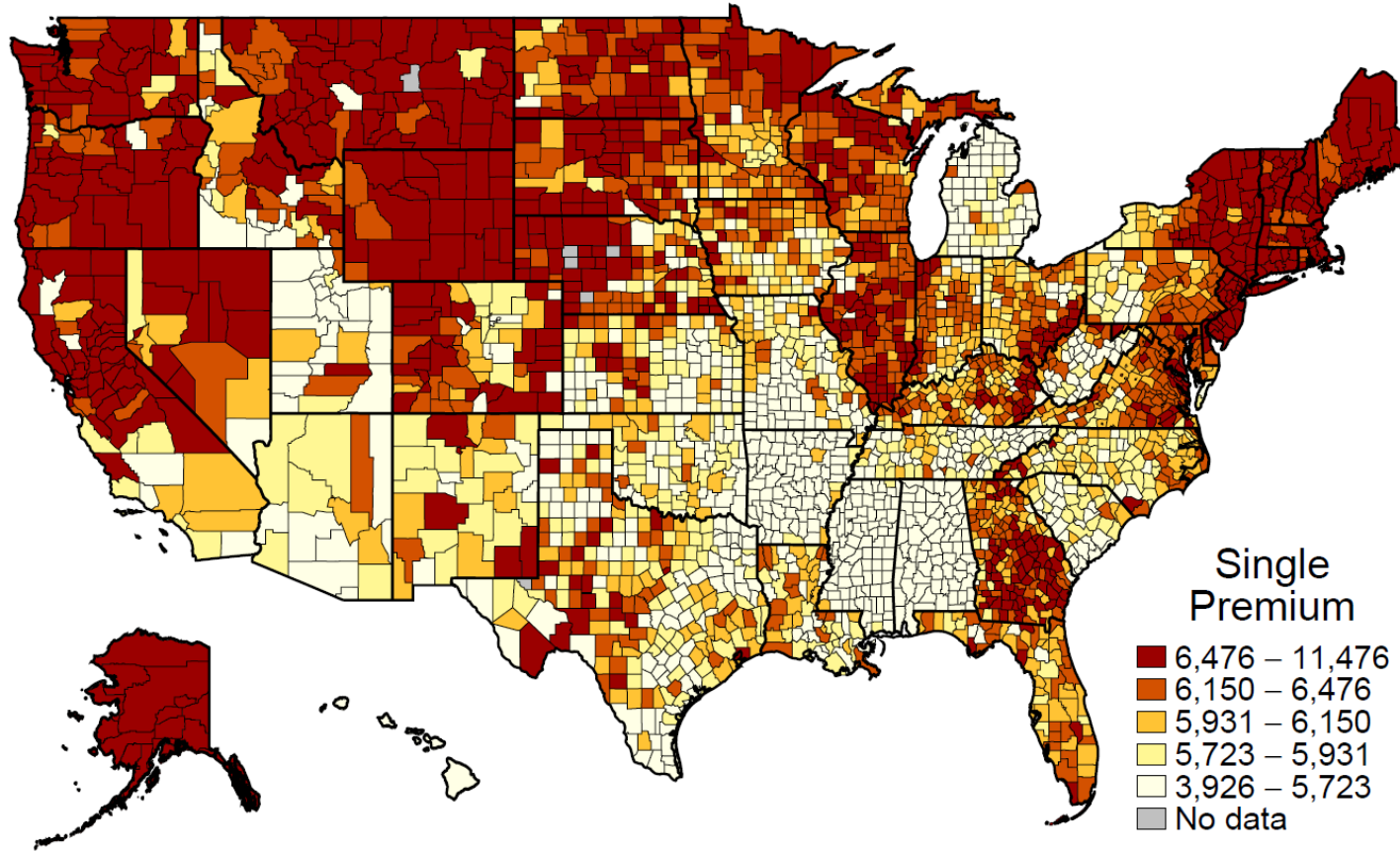
Table 1: Characteristics of ESI Policies in the Tax Data

	(1) N (1,000s)	(2) Fraction	(3) Covered Months	(4) Full-Year Coverage	(5) Female Held	(6) Matched Premiums	(7) Average Premiums
Coverage Source							
Self-Insured	50,378	0.52	9.87	0.67	0.46	0.75	
Third Party	38,726	0.40	9.29	0.59	0.45	0.37	
Other	8,099	0.08	10.36	0.75	0.22	0.07	
Policy Type							
Single	54,276	0.56	9.30	0.59	0.50	0.52	6,098
2-Person	18,276	0.19	10.15	0.72	0.39	0.52	13,070
Family	24,651	0.25	10.17	0.72	0.35	0.62	17,171
Policyholder Age							
<27	8,308	0.09	7.76	0.38	0.42	0.51	
27-34	18,441	0.19	8.91	0.53	0.45	0.58	
35-44	20,265	0.21	9.63	0.64	0.43	0.61	
45-54	21,435	0.22	10.05	0.70	0.44	0.61	
55-64	19,653	0.20	10.42	0.75	0.46	0.53	
65+	8,683	0.09	10.75	0.80	0.38	0.23	
Firm Size							
<10	3,549	0.04	10.34	0.74	0.44	0.03	
10-24	3,002	0.03	9.22	0.58	0.40	0.09	
25-99	8,057	0.08	8.97	0.54	0.38	0.17	
100-249	7,867	0.08	9.05	0.56	0.41	0.34	
250-499	6,732	0.07	9.37	0.60	0.44	0.64	
500-999	7,199	0.07	9.60	0.63	0.46	0.73	
1,000+	52,036	0.54	9.98	0.69	0.47	0.75	
No W-2	8,761	0.09	9.35	0.61	0.32	0.00	
Total/Average	97,203	1.00	9.68	0.65	0.44	0.65	

Notes: Each of the categories in Coverage Source, Policy Type, Policyholder Age, and Firm Size are mutually exclusive and collectively exhaustive, with the following exception: we omit 417,462 (0.4%) policies from the age categories because we did not have information on the policyholder's date of birth. Very few premiums are matched for "other" policies because these include military, union, and SHOP plans and of these three we only try to match SHOP. Union and military policies are associated with traditionally male-dominated occupations resulting in a lower fraction of female policyholders.

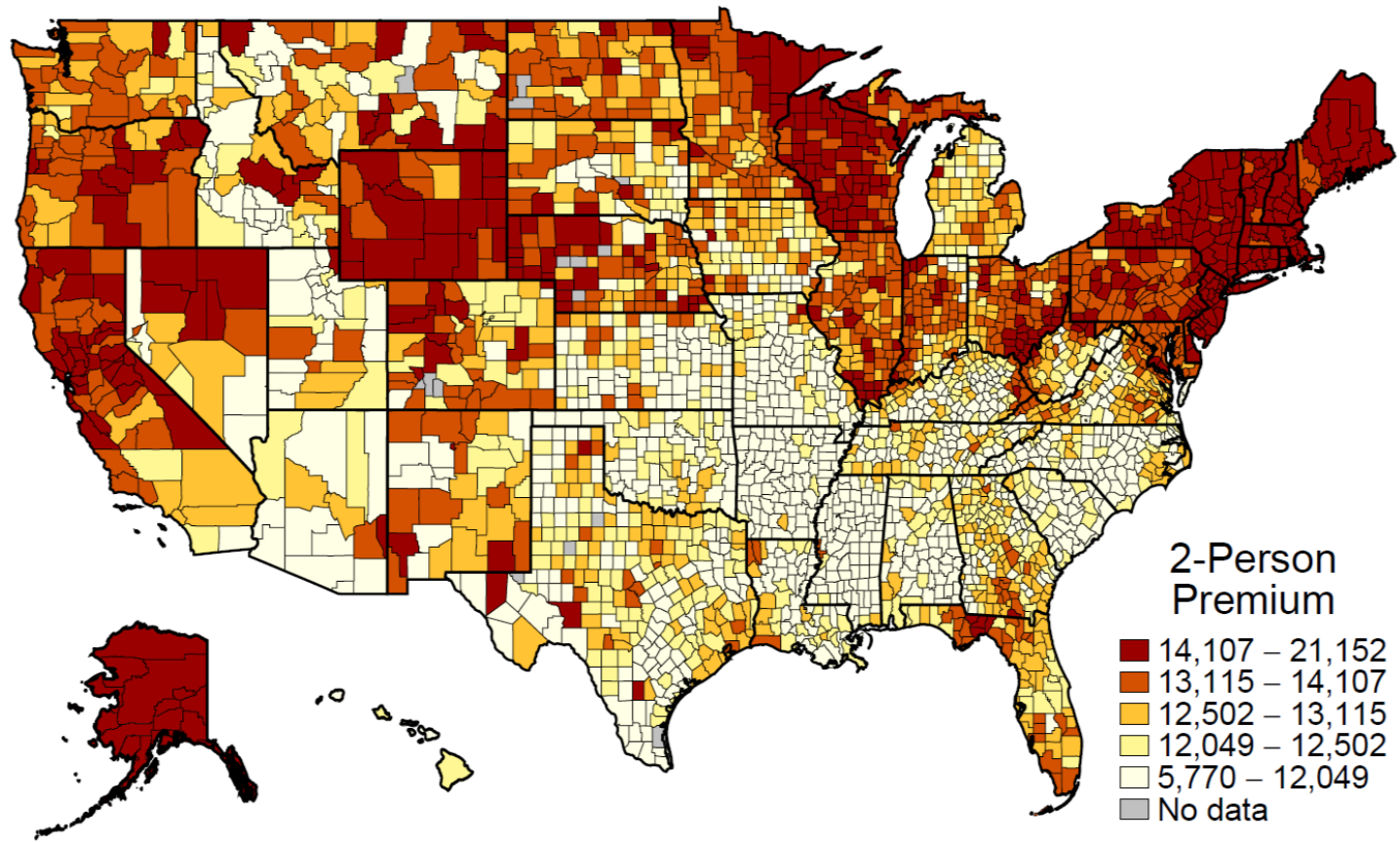
Online Appendix

Figure A1: ESI Premiums for Single Plans by County



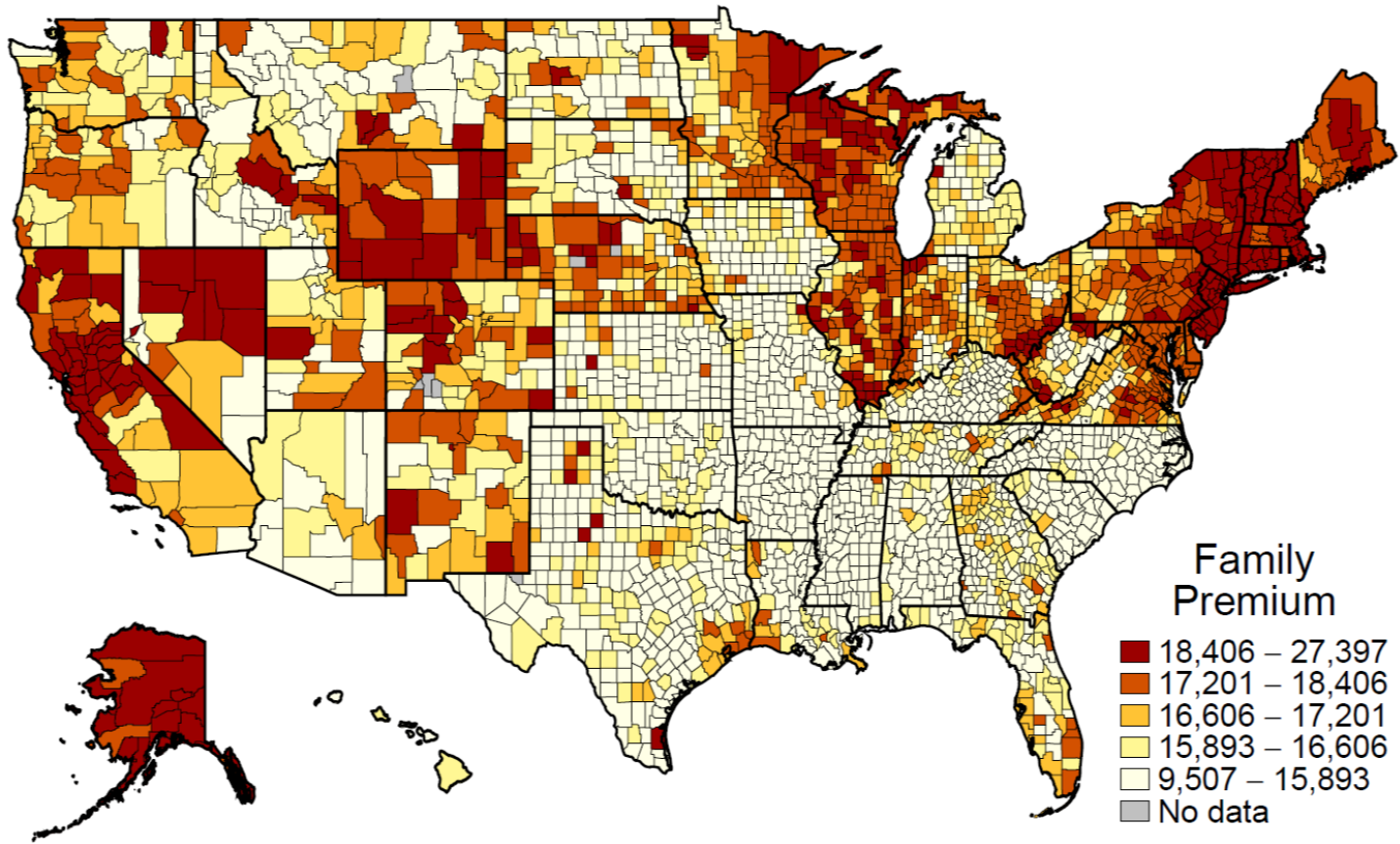
Notes: The average annualized ESI premium for single plans are shown for each county. Each shade (other than gray for insufficient data) corresponds to a quintile of county-level average premiums.

Figure A2: ESI Premiums for 2-Person Plans by County



Notes: The average annualized ESI premium for 2-person plans are shown for each county. Each shade (other than gray for insufficient data) corresponds to a quintile of county-level average premiums.

Figure A3: ESI Premiums for Family Plans by County



Notes: The average annualized ESI premium for family plans are shown for each county. Each shade (other than gray for insufficient data) corresponds to a quintile of county-level average premiums.

Table A1: Share of ESI Premiums by Income Percentile

Income Percentile Range	Household Income (Tax Units)				Wage Income (Individuals)			
	(1) Returns w/ ESI	(2) Policy Share	(3) Premium Share	(4) Income Share	(5) Workers w/ ESI	(6) Policy Share	(7) Premium Share	(8) Income Share
99.99+	82.9%	0.01%	0.02%	5.7%	93.6%	0.02%	0.03%	1.90%
99.9-99.99	78.3%	0.12%	0.18%	5.0%	90.8%	0.15%	0.25%	2.91%
99.5-99.9	76.5%	0.53%	0.79%	6.2%	88.3%	0.66%	1.06%	4.80%
99-99.5	77.6%	0.67%	1.02%	4.0%	86.3%	0.81%	1.23%	3.60%
95-99	81.0%	5.60%	8.43%	15.4%	83.9%	6.31%	8.83%	15.36%
90-95	83.3%	7.20%	10.55%	11.3%	82.7%	7.77%	10.19%	11.97%
75-90	81.3%	21.09%	27.60%	21.5%	79.2%	22.31%	26.22%	23.48%
50-75	72.0%	31.12%	29.92%	19.7%	67.4%	31.65%	29.91%	22.59%
25-50	48.9%	21.13%	14.18%	9.8%	36.2%	16.98%	11.42%	11.06%
0-25	14.4%	6.21%	3.10%	1.3%	10.3%	4.83%	3.26%	2.35%
Unmeasured		6.32%	4.21%		77.6%	8.50%	7.58%	

Notes: Household Income is the AGI reported on a tax return plus tax-exempt interest. Household Income is missing for those who did not file a tax return in 2016 and W-2 wages are missing for those who did not receive a W-2 (e.g., retirees or those who are unemployed with COBRA coverage). We found \$893 billion in total ESI premiums. Relevant to Columns (1) through (4), there are 146 million tax returns and \$10.4 trillion in total Household Income. For Columns (5) through (8), there are 177 million people with a W-2 and \$7.8 trillion in total wages.