

Working Paper Series
Congressional Budget Office
Washington, D.C.

Data and Methods for Constructing Synthetic Firms in CBO's Health Insurance Simulation Model, HISIM2

Lucas Goodman
Department of the Treasury, Office of Tax Analysis
lucas.goodman@treasury.gov

Ben Hopkins
Congressional Budget Office
ben.hopkins@cbo.gov

Alex Minicozzi
Congressional Budget Office
alexandra.minicozzi@cbo.gov

Eamon Molloy
Congressional Budget Office
eamon.molloy@cbo.gov

Working Paper 2021-15

December 2021

To enhance the transparency of the work of the Congressional Budget Office and to encourage external review of that work, CBO's working paper series includes papers that provide technical descriptions of official CBO analyses as well as papers that represent independent research by CBO analysts. Papers in this series are available at <http://go.usa.gov/xUzd7>.

The views expressed in this paper do not necessarily represent the official position of the Department of the Treasury.

We thank Karen Stockley, Bilal Habib, and Chapin White for helpful comments and suggestions and Casey Labrack and Erik O'Donoghue for help in, respectively, designing and reviewing the graphics. We also thank Ithai Lurie of the Department of the Treasury, Jessica Banthin of the Urban Institute, and Federico Girosi of RAND Corporation for their reviews. Karen Stockley, Ru Ding, and Daria Pelech provided invaluable assistance in assembling data from the Health Care Cost Institute. We gratefully acknowledge editorial assistance provided by John Skeen.

www.cbo.gov/publication/57431

Abstract

This paper discusses how the Congressional Budget Office constructs “synthetic firms”—businesses composed of artificial groups of workers—used in the agency’s health insurance simulation model, or HISIM2, which underlies projections of insurance coverage. Detailed information on a worker’s firm and coworkers is important for modelling whether that worker is offered employment-based insurance, because a worker’s access to such insurance depends on the collective effect of how much employees at the firm demand insurance and how much they will cost to insure. Unfortunately, no complete data source has both rich information on the entire nonelderly population and detailed data on those who are employed. Consequently, CBO uses nationally representative data from the Current Population Survey and supplements those data by building synthetic firms that mimic the composition of actual firms in other data sets. This working paper expands on a slide deck released in April 2020, which detailed both how the agency constructs synthetic firms and how HISIM2 models a firm’s decision about offering health insurance. Specifically, the paper describes how the agency constructs synthetic firms, which are the model’s decisionmaking units for whether or not to offer insurance. Since the publication of that slide deck, CBO has improved its approach, adding data on marital status and health care spending to its method of selecting coworkers for firms. The new data and methods better capture the observed heterogeneity in workforces’ demand for employment-based insurance and thus, in theory, more accurately allow for modelling the insurance coverage effects of a wide variety of policies affecting employers’ offer decisions. The paper includes summary statistics from the newly created data sources that should help others create synthetic firms that are representative of actual firms.

Keywords: health insurance simulation model, HISIM2, health insurance, health coverage, employment-based insurance, premiums, Current Population Survey

JEL Classification: H51, I13

Contents

Introduction.....	1
The Link Between an Individual Worker’s Coverage Decision and His or Her Coworkers.....	1
Data Linking Workers to Their Coworkers	3
Overview of CBO’s Current Method of Constructing Synthetic Firms	4
Data	5
Tax-Filing Data Set.....	5
Health Care Cost Institute Data Set	11
Methodology.....	12
Step 1: Assigning Firms a Type.....	14
Step 2: Specifying the Composition of Synthetic Firms.....	16
Step 3: Populating Synthetic Firms.....	19
Assessment of the Quality of the Synthetic Firms.....	21
Comparison of Workforces in the CPS and Tax Data	21
Distribution of Traits Within Firms	24
Simulated Change in Demand for an Offer at Synthetic Firms for Two Policies	39
Appendix A: Data Sources for the Tax-Filing Data Set	44

Introduction

The Congressional Budget Office uses a health insurance simulation model, HISIM2, to generate estimates of coverage, premiums, and budgetary costs for the nonelderly.¹ Nonelderly workers decide what type of coverage to enroll in on the basis of availability and net premiums,² among other factors. The most prevalent form of coverage, employment-based insurance, receives a tax preference, which, on average, increases with income.³ CBO and the staff of the Joint Committee on Taxation (JCT) estimate that in 2021, about 159 million people under age 65 in the United States (59%) obtained health insurance through their employer or that of a family member.⁴ Whether or not a firm offers employment-based insurance to its employees can have a large effect on coverage and costs, because employees take up employment-based coverage at very high rates when it is offered. Employees without an employment-based offer are more likely to be uninsured, and the per-enrollee budgetary cost to subsidize coverage varies substantially across types of coverage. While most large firms offer health insurance and still would in response to a wide range of policies, some small or even medium-sized firms could change their decision to offer health insurance in response to policy changes.⁵ CBO estimates that small and medium-sized firms (those with fewer than 100 employees) employed 52 million Americans under age 65 (33% of total employment) in 2021. Therefore, it is important that simulation models accurately measure the availability of and net premiums for employment-based insurance for individual workers under different proposals.

The Link Between an Individual Worker's Coverage Decision and His or Her Coworkers

In theory, an individual worker's access to employment-based insurance depends on how much the person and his or her coworkers cost to insure and how much they demand insurance. On average, larger firms pay less, per-employee, for their workers to be insured primarily because

¹ For general information on HISIM2, see Congressional Budget Office, "HISIM2: The Health Insurance Simulation Model Used in Preparing CBO's July 2021 Baseline Budget Projections" (July 2021), www.cbo.gov/publication/57205.

² The gross premium for health insurance is the total amount that an insurer receives to insure individuals covered by the plan. The net premium subtracts subsidies, like the tax exclusion for employment-based insurance or premium tax credits in the marketplaces established by the Affordable Care Act, from the gross premium.

³ Matthew Rae and others, *Tax Subsidies for Private Health Insurance* (The Henry J. Kaiser Family Foundation, October 2014), <https://tinyurl.com/emcph7y6>.

⁴ Congressional Budget Office, "Federal Subsidies for Health Insurance Coverage for People Under Age 65: CBO and JCT's July 2021 Projections" (July 2021), <https://go.usa.gov/xMSEh> (PDF, 451 KB).

⁵ CBO's assumptions regarding the elasticity of firms' offer decisions are based in part on Jonathan Gruber and Michael Lettau, "How Elastic Is the Firm's Demand for Health Insurance," *Journal of Public Economics*, vol. 88, no. 7–8 (July 2004), pp. 1273–1293, [https://doi.org/10.1016/S0047-2727\(02\)00191-3](https://doi.org/10.1016/S0047-2727(02)00191-3).

they spread their administrative costs over more enrollees.⁶ If employment-based insurance is available, the gross premium depends upon the collective health care spending for workers in firms with more than 50 workers (the large-group market).⁷ For example, all else being equal, insurers that expect enrollment of workers who will utilize expensive health care services will set a higher gross premium for insurance (because everyone is in a pool that is charged a group rate).

Employers offer health insurance to attract workers who demand it. That demand depends on the net premium, attractiveness, and availability of employment-based insurance relative to alternate health insurance options that the workers might choose. Public policies can affect whether a firm offers employment-based insurance and, if so, which type of plans to offer, through the policies' collective effect on the premium, attractiveness, and availability of insurance options for the firm's workforce.⁸ For example, a proposal that offers free alternate insurance only to people with low income (expanding eligibility for Medicaid, for instance) could affect coverage for a higher-income worker depending on the other employees at that worker's firm. A higher-income worker at a firm with mostly lower-income coworkers (who would be newly eligible for the free insurance) might lose his or her offer of employment-based coverage because the policy caused the collective demand for employment-based insurance to decline substantially. Conversely, a higher-income worker at a firm with mostly higher-wage workers (ineligible for the free alternative) would likely retain his or her employment-based insurance offer because the policy caused little change in collective demand. For a model to accurately predict whether higher-income workers would retain their offer of employment-based coverage in response to such a policy, it must account for the diversity of earnings distributions among firms.

Given that the net premium, attractiveness, and availability of employment-based insurance depend on the firm's collective demand for insurance, a simulation model that predicts an individual worker's coverage decisions would ideally be based on nationally representative data that links individual workers to their coworkers. Unfortunately, no single data source has both rich information on the entire nonelderly population—including those who do and do not file a tax return—and detailed data on coworkers. CBO uses nationally representative survey data in HISIM2 from the Current Population Survey and supplements those data by building “synthetic

⁶ Jean Marie Abraham and others, “Access to Health Insurance at Small Establishments: What Can We Learn From Analyzing Other Fringe Benefits?” *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, vol. 46, no. 3 (2009), pp. 253–273, https://doi.org/10.5034%2Finquiryjml_46.03.253.

⁷ The price of insurance in the small-group market (in which employers have fewer than 50 workers) is based on guidelines put in place by the Affordable Care Act and cannot vary with the medical claims history of workers at a firm.

⁸ In CBO's model, firms make decisions by maximizing an objective function that depends on their employees' willingness to pay for health insurance, the tax benefit of offering employment-based insurance, the switching costs incurred by altering offer decisions from the previous year, and firms' preferences that are unobserved. For more information on firms' decisionmaking in the agency's model, see Congressional Budget Office, “How CBO Models Firms' Behavior in HISIM2 in Its Baseline Budget Projections as of March 6, 2020” (March 2020), www.cbo.gov/publication/56303.

firms” that mimic the composition of actual firms in other data sets for those traits strongly correlated with demand for employment-based insurance and health care spending.

Though many characteristics of an individual can affect the net premium, attractiveness, and availability of health insurance options, CBO focuses on age, earnings, eligibility for employment-based insurance, marital status, health care spending, and state of residence when building synthetic firms. Age indirectly affects health care spending⁹ and directly affects insurance premiums in the nongroup and small-group markets.¹⁰ Earnings indirectly affect subsidies for insurance such that, on average, higher earners are eligible for lower net premiums for employment-based insurance and face higher net premiums for nongroup insurance. Eligibility to take up an offer of employment-based insurance directly affects the availability of the insurance option and indirectly affects its gross premium by determining whose health care spending determines the insurer’s cost. Marital status indirectly affects subsidies in the nongroup market because family size is used in the calculation of the ratio of income to the federal poverty level (FPL), on which the subsidy formulas are based.¹¹ Health care spending for the worker directly affects the gross premium of insurance in the large-group market for employment-based insurance, which depends on the insurer’s cost. Lastly, state of residence directly affects the availability of highly subsidized alternate insurance (because Medicaid eligibility rules differ by state) and the value of the tax exclusion for employment-based insurance (because income tax rates differ by state) and also indirectly affects premiums in the individual and small-group markets (because regulation of those markets varies by state).¹²

Data Linking Workers to Their Coworkers

In setting up its health insurance simulation model, CBO relied on two data sets to understand how workers’ and coworkers’ traits (namely, age, earnings, marital status, health care spending, and state of residence) are related. CBO and JCT collaborated to create the first data set, which was based primarily on employers’ and employees’ tax filings and provided reliable information on the joint distribution of employees’ age, wages, marital status, residence, and eligibility for employment-based insurance by size of firm and whether a firm offered health insurance.¹³ To create the tax-filing data set, CBO and JCT matched employers’ tax returns with employees’

⁹ Dale H. Yamamoto, *Health Care Costs—From Birth to Death?* Report 2013-1 (Health Care Cost Institute, June 2013), <https://tinyurl.com/xs5a593w>.

¹⁰ Centers for Medicare & Medicaid Services, “Market Rating Reforms” (June 2, 2017), <https://go.usa.gov/xs3VB>.

¹¹ Internal Revenue Service, “Questions and Answers on the Premium Tax Credit” (accessed May 17, 2021), <https://go.usa.gov/xs3Vn>.

¹² While CBO considers state of residence when constructing synthetic firms, the agency’s health insurance simulation model is not representative at the state level and cannot be used to generate state-level predictions on the effects of changes in state or federal policies.

¹³ Previous versions of CBO’s health insurance simulation model relied on more limited data sets. See Congressional Budget Office, *CBO’s Health Insurance Simulation Model: A Technical Description* (October 2007), www.cbo.gov/publication/19224.

wage and salary information returns. Supplemental administrative sources provided additional information on employers and employees, such as the type of industry and employees' age. Employers' and employees' tax returns identified some, but not all, employers that offered insurance and employees who were eligible to receive an offer. For the employers without tax records identifying that they offered insurance, CBO imputed that action to some of them on the basis of information from the Medical Expenditure Panel Survey's Insurance Component (MEPS-IC)—a survey of employers' health insurance offerings. Relying on data from the Medical Expenditure Panel Survey's Household Component (MEPS-HC)—a survey of households about health insurance coverage and health care spending—CBO also imputed whether employees were eligible to receive an employment-based offer.

That extremely large and rich data source, unfortunately, did not include health care spending. So CBO created a second data set using data held by the Health Care Cost Institute (HCCI) on health care spending for a sample of policyholders of employment-based plans (that is, spending both by the policyholders and on their behalf). That data set provided information on the correlation of spending for policyholders who are employed by the same firm.

Overview of CBO's Current Method of Constructing Synthetic Firms

CBO uses synthetic firms in HISIM2 to simulate changes in employment-based coverage under various proposals. The agency creates those firms using individual-level data on workers from the Current Population Survey (CPS). The individual workers drawn from the CPS are matched with coworkers, who are randomly drawn from among other CPS workers. The resulting firm is called "synthetic" because it consists of artificial, rather than actual, coworkers.

Synthetic firms are constructed in three steps. First, CBO and JCT classify employers in the tax-filing data set into different types of firms that are characterized by distinct joint distributions of employees' age and earnings. For example, one type of synthetic firm might tend to have young, low-income coworkers, while another type might have older, high-income coworkers.

Second, the agency specifies the desired composition of workers' synthetic firms by selecting the age, earnings, eligibility for employment-based insurance, marital status, health care spending, and states of residence of their coworkers on the basis of estimated distributions of those characteristics from the aforementioned data sets. The estimated distribution of coworkers' health care spending conditioned on workers' health care spending, for example, determines the likelihood that the desired composition of a synthetic firm includes many coworkers who are representative in terms of being relatively healthy or unhealthy. At this stage, CBO specifies the characteristics of each coworker in a synthetic firm but does not assign specific CPS respondents to the firm. As part of the process, each synthetic firm is assigned one of the types identified in the first step. The desired workforce composition for synthetic firms from this second step determines the probabilities that workers are selected as synthetic coworkers.

Third, to populate each synthetic firm, CBO selects workers in the CPS in firms in the same size category and with the same “offer status” (offering health insurance or not offering it) who have the appropriate traits, accounting for the fact that the CPS does not include workers with all possible combinations of traits. Larger firms offer insurance at much higher rates, and workers sort into firms on the basis of their demand for insurance,¹⁴ so by conditioning the selection probability on the size of the firm and whether or not it offers health insurance, CBO ensures that synthetic coworkers will be similar along dimensions affecting demand other than the traits directly chosen (namely, age, earnings, etc.).

Those large and detailed firm-level data sets (both the tax data and the claims data) reveal substantial heterogeneity among firms across dimensions affecting workers’ demand for insurance. By relying on those data, CBO constructs synthetic firms that are substantially more similar to real-world firms than were the ones created using the agency’s prior approach. The improvements have considerably enhanced the agency’s ability to model the effects of policies on employers’ offer decisions. Moreover, by publishing summary statistics from the newly created data sources, CBO hopes to aid others in creating synthetic firms that are more representative of actual firms.

Data

To construct synthetic firms that mimic actual firms, CBO and JCT first needed to characterize the composition of actual firms. Although the workers and coworkers in synthetic firms are drawn from the CPS, that data set does not include information on the composition of actual firms. Therefore, CBO and JCT relied on other data sets: For information on the age, earnings, marital status, state of residence, and eligibility to purchase health insurance for the employees of actual firms, CBO and JCT created a data set based on employers’ and employees’ tax filings in 2017. For information on the health care spending of and for employees of actual firms, CBO used commercial claims data held by HCCI.

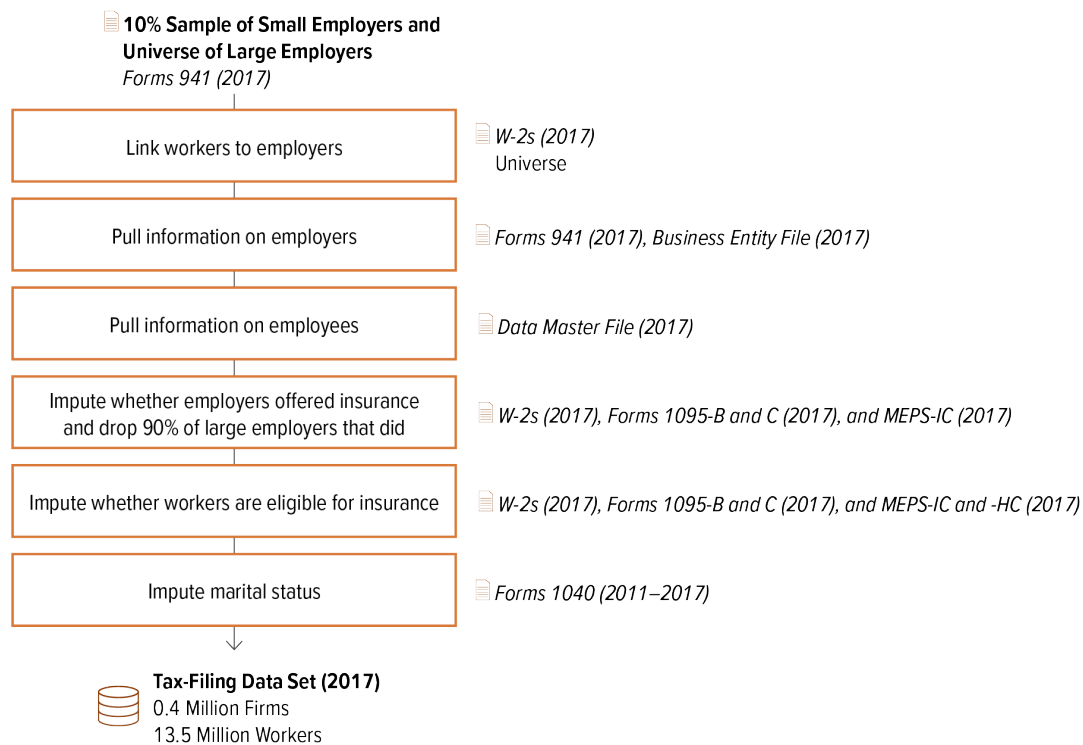
Tax-Filing Data Set

CBO and JCT created a data set based on employers’ and employees’ tax filings that, for the employees of actual firms in 2017, provided information on their age, earnings, marital status, state of residence, and eligibility to purchase health insurance (see Figure 1). The first step in creating the data set was linking employers’ tax returns, employees’ wage and salary information returns, and other relevant administrative data (see Appendix A). After combining those data, the agencies excluded records that were incomplete, erroneous, or irrelevant to the purpose of constructing synthetic firms. Next, the agencies used survey data from the Medical Expenditure Panel Survey to guide the imputation of two key variables that were not fully observed in the

¹⁴ Alan C. Monheit and Jessica Primoff Vistnes, “Health Insurance Enrollment Decisions: Preferences for Coverage, Worker Sorting, and Insurance Take-Up,” *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, vol. 45, no. 2 (2008), pp 153–167, https://doi.org/10.5034%2Finquiryjrnl_45.02.153.

Figure 1

Assembly of the Tax-Filing Data Set



Data source: Congressional Budget Office.

MEPS-HC = Household Component of the Medical Expenditure Panel Survey; MEPS-IC = Insurance Component of the Medical Expenditure Panel Survey.

administrative data: whether or not employers offered health insurance to their employees and whether or not employees were eligible for employment-based insurance if their employer offered it. Finally, the agencies used individual income tax returns to link employees with their spouses and to determine whether spouses were offered health insurance by their employer.

Linking Administrative Data Sets. To create the tax-filing data set, CBO and JCT began by drawing a sample of employers' tax returns (Forms 941)¹⁵ and linking them with employees' wage and salary information returns (W-2s). The agencies drew a 10 percent sample from the universe of employers with fewer than 100 employees and retained 100 percent of employers

¹⁵ Most employers are required to report a Form 941 on a quarterly basis. The only exceptions include seasonal employers for quarters in which they have paid no wages, employers of household employees, and employers of farm employees.

with 100 employees or more that filed a Form 941 in 2017. The agencies then used employer identification numbers to link Forms 941 to W-2s.

CBO and JCT at first retained all large employers to end up with an accurate characterization of the small subset of such employers that did not offer health insurance. Eventually, after imputing whether employers offered insurance, the agencies drew a 10 percent sample from among those that offered insurance—the vast majority—and retained all large employers that did not offer insurance. Basing the tax-filing data set on a 10 percent sample curbed computational time while still providing a data set on 0.4 million employers and 13.5 million employees with which to analyze the relationship between workers and coworkers.

CBO and JCT used tax filings and additional administrative data to obtain information on important traits of employers and employees. To determine an employer's size and industry, the agencies used Form 941 and the Internal Revenue Service's Business Entity File, respectively. To determine a worker's earnings and state of residence, the agencies used the W-2; to determine his or her age and sex, they used the Social Security Administration's Data Master File.

Excluding Incomplete, Erroneous, and Irrelevant Records. CBO and JCT dropped Forms 941 and W-2s that had no information on a key variable or information that was probably incorrect. Specifically, the agencies dropped from the sample Forms 941 with missing or invalid industry codes and employers that reported average compensation amounting to less than \$500 per employee per year.¹⁶ CBO and JCT also discarded W-2s that did not record a wage, had a missing or invalid taxpayer identification number, or were issued for individuals without an age provided or an age above 105. Those restrictions collectively caused the agencies to drop firms that employed 8 million people, relative to 158 million in the original universe.¹⁷

CBO and JCT also dropped W-2s that were not relevant for creating synthetic firms because of the scope and income reporting of the CPS. When creating synthetic firms, CBO classifies CPS respondents into earnings categories on the basis of their earnings from their longest-held job in the previous year. For those tax filers who received multiple W-2s, CBO and JCT use the earnings from the W-2 that reported the higher earnings and discard the rest. If the agencies had kept W-2s from secondary jobs in its tax-filing data set but matched CPS respondents to synthetic firms on the basis of their earnings from their main job, then CPS respondents with relatively low earnings in their main jobs would have been overrepresented in synthetic firms. Because the CPS does not survey individuals living outside the 50 states or the District of

¹⁶ Firms with under \$500 in reported average compensation were dropped in order to address an occasional data entry error in which the same number (sometimes with a decimal moved) was reported on the "wages and compensation" line of Form 941 (Line 2) and the "employee count" line (Line 1), leading to implausible employee counts for such firms.

¹⁷ The 158 million figure is computed after first dropping firms with reported average compensation under \$500 per employee.

Columbia, the agencies also dropped W-2s that did not record one of those areas as the employee's residence.

In addition, CBO and JCT dropped W-2s that recorded less than \$5,000 in earnings. If the agencies had retained those W-2s, the earnings of workers in the tax-filing data set would have been highly skewed toward low earnings relative to the amounts reported in the CPS. Because the CPS asks respondents in March of the current year about their employment in the previous year, it is probably missing some information on short-term employment that respondents do not remember to report. Many of the excluded W-2s were probably issued for new, part-time, seasonal, or temporary employees, who employers may deem ineligible for health insurance.¹⁸ Because ineligible employees do not affect an employer's decision to offer health insurance in CBO's analysis, excluding W-2s for those types of employees from the tax-filing data set has a limited effect on simulations of employers' decisions. For the same reason, CBO and JCT also excluded W-2s for employees under age 18.

Finally, CBO and JCT dropped Forms 941 associated with one W-2 or none. Those records were not relevant for constructing synthetic firms because they provide no information on coworkers. A Form 941 can have one W-2 or none for several reasons. Some Forms 941 are filed by entities that do not have any employees. Others are filed by self-employed individuals without employees.¹⁹ Finally, some Forms 941 are associated with one W-2 or none because some of the W-2s they were associated with were dropped for one of the reasons cited.

Imputing Whether Employers Offered Health Insurance. CBO and JCT used employees' tax filings to identify employers that offered health insurance. If an employer was associated with any employee who received a W-2, 1095-B, or 1095-C indicating that the employee was offered employment-based coverage, the agencies assigned that firm as having offered health insurance to employees it deems eligible. W-2s often report employers' contributions to health insurance premiums for employees who purchased employment-based coverage. For individuals who enrolled in minimum essential coverage,²⁰ insurers issue 1095-Bs, which indicate whether those individuals were enrolled in employment-based plans. Finally, for their full-time employees,

¹⁸ Nondiscrimination rules restrict employers' ability to offer health insurance to some employees but not others. However, those rules do not apply to certain types employees, including ones with less than three years of service, part-time employees, seasonal employees, and employees under the age of 25. For additional information on nondiscrimination rules, see Jones Day, *Fully Insured Group Health Plans May No Longer Provide Discriminatory Benefits* (November 2010), <https://tinyurl.com/5c9brw8e>.

¹⁹ Roughly half of the Forms 941 that the agencies dropped because they had only one associated W-2 were filings for S corporations whose sole employee was the owner of the firm reporting his or her wages.

²⁰ Minimum essential coverage is defined as any insurance plan that meets the shared responsibility provisions of the Affordable Care Act. Most employment-based coverage satisfies the definition of minimum essential coverage. Notable exclusions include stand-alone vision or dental plans, workers' compensation coverage, and coverage for a specific disease or condition. For more information, see Internal Revenue Service, "Questions and Answers on the Individual Shared Responsibility Provision" (April 13, 2021), <https://go.usa.gov/xMgds>.

large employers issue 1095-Cs, which indicate whether those employees were offered employment-based insurance.

The share of employers that CBO and JCT could determine offered health insurance on the basis of administrative data fell short of the share that offered health insurance according to the MEPS-IC, which, as a large survey of employers, is well-suited to measuring the number and types of employers that offer insurance. The differences between administrative data and the MEPS-IC were minor for large employers but pronounced for small ones.

There are many reasons why the tax filings for an employee who was offered health insurance may not reflect that fact. For one, reporting is not always required. Notably, employers that issue fewer than 250 W-2s or purchased health insurance through the Small Business Health Options Program or a multiemployer plan are not required to report their contributions to their employees' health insurance premiums on W-2s.²¹ Furthermore, employers and insurers do not always report what they are required to. For example, insurers that provide non-self-insured employer plans are required in most cases to report the identification number of sponsoring employers on Part II of the employees' Forms 1095-B but sometimes do not.²²

Consequently, in the tax-filing data set, CBO and JCT adjusted the number of employers offering health insurance upward. Specifically, when, on the basis of tax filings, the share of employers of a particular size and industry²³ that offered health insurance was less than the share estimated to do so according to the MEPS-IC, CBO and JCT randomly reclassified some employers of that type as having offered health insurance—to match the share in the MEPS-IC. The agencies reclassified 8.0 percent, 8.2 percent, and 1.1 percent of employers with fewer than 25 employees, 25 to 99 employees, and 100 or more employees, respectively (see Figure 2).

Imputing Whether Employees Were Eligible to Purchase Health Insurance. Employers that offer health insurance often have policies barring some employees from enrolling in it. Eligibility

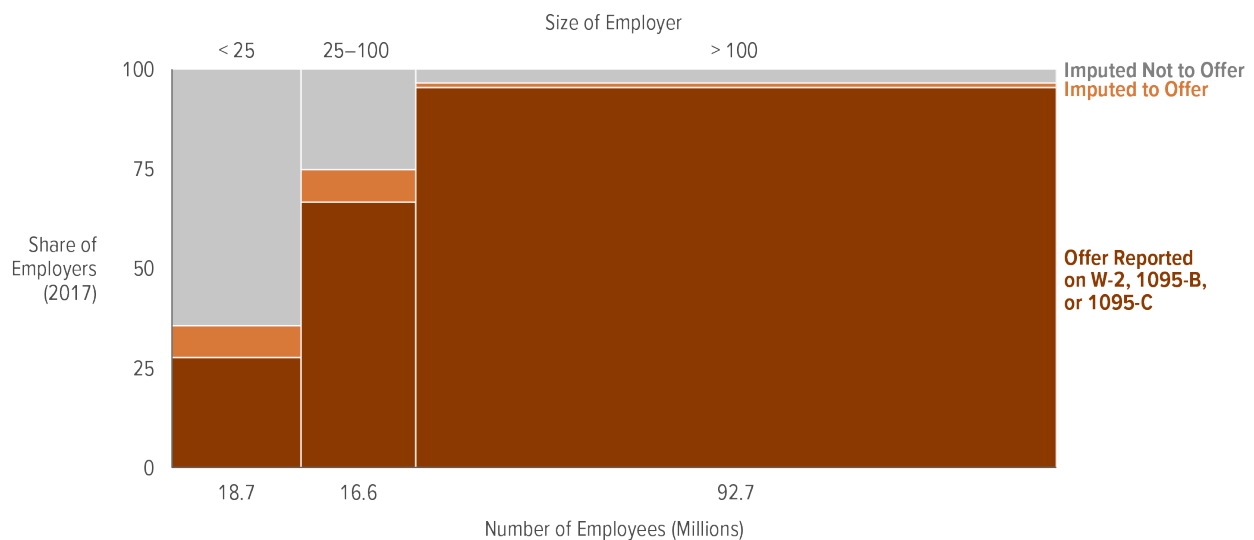
²¹ Internal Revenue Service, “Form W-2 Reporting of Employer-Sponsored Health Coverage” (June 24, 2021), <https://go.usa.gov/xecfr>.

²² See Internal Revenue Service, “Who Must File” and “Specific Instructions for Form 1095-B” in *Instructions for Forms 1094-B and 1095-B (2020)* (October 15, 2020), www.irs.gov/instructions/i109495b#idm139700478017104.

²³ For the sake of that imputation, firms were classified into the following 10 industry groups on the basis of their two-digit NAICS (North American Industry Classification System) codes: “Agriculture, Forestry, Fishing and Hunting” (11), “Mining and Manufacturing” (21, 31, 32, and 33), “Construction” (23), “Utilities, transportation, and warehousing” (22, 48, and 49), “Wholesale trade” (42), “Financial services, insurance, real estate rental and leasing, and management of companies and enterprises” (52, 53, and 55), “Retail trade” (44 and 45), “Information, professional, scientific and technical services, educational services, and health care and social assistance” (51, 54, 61, and 62), “Other services” (56, 71, 72, and 81), and “State and local governments” (92).

Figure 2

Share of Employers Imputed to Offer Health Insurance in the Tax-Filing Data Set



Data source: Congressional Budget Office.

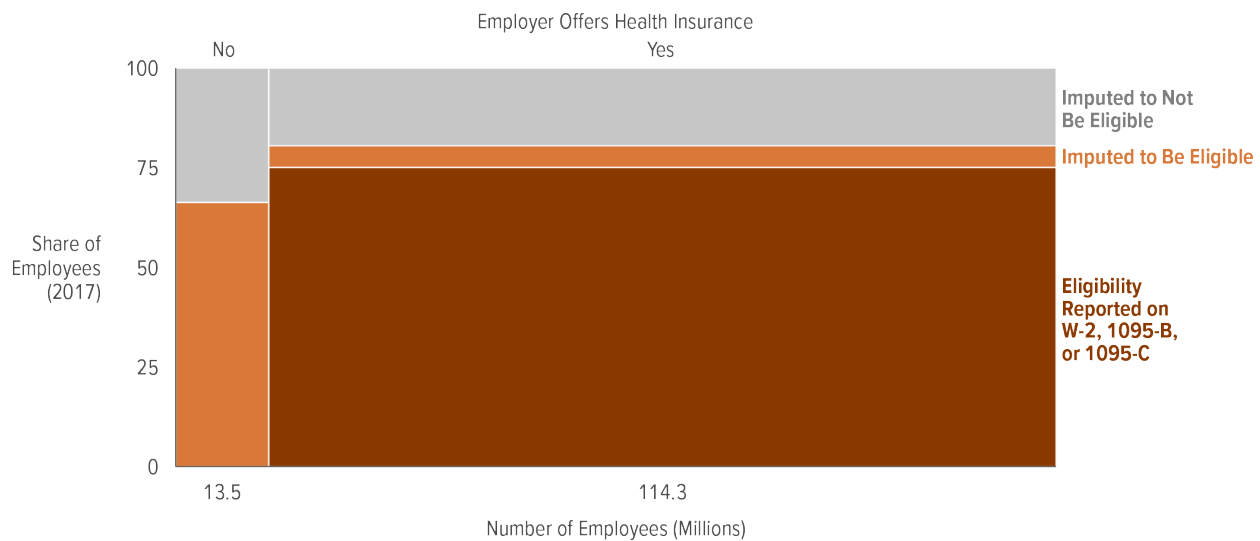
may be based on, for example, a person’s tenure, weekly hours, or status as a permanent (as opposed to temporary or seasonal) employee.

CBO and JCT used tax filings to identify many employees who were eligible to purchase employment-based insurance. If employees were issued a W-2 or 1095-B indicating that they purchased employment-based insurance or if they received a 1095-C indicating that they were offered it, CBO and JCT inferred that they were, in fact, eligible. However, because of the reporting issues described earlier, the share of employees who CBO and JCT determined were eligible to purchase health insurance on the basis of tax filings fell short of the share who actually were on the basis of the MEPS-IC (see Figure 3). Furthermore, because synthetic firms in HISIM2 may change their decision to offer insurance, it was also necessary for the agencies to identify employees who were not offered insurance but would have been eligible to purchase it if they had been. For obvious reasons, however, it was not possible to identify those employees on the basis of tax filings.

Consequently, as CBO and JCT did for the number of employers offering insurance, the agencies adjusted the share of employees eligible to purchase employment-based insurance upward in the tax-filing data set. Specifically, the agencies used data from the MEPS-HC to estimate the probability that a person working at a firm offering insurance was eligible to purchase it on the basis of the person’s age, marital status, and sex; the size of the employer; the employer’s industry; and whether or not the employer offered employment-based insurance. The agencies adjusted those estimates so that the predicted shares of workers who were eligible to purchase employment-based coverage at firms of different sizes were consistent with estimates from the

Figure 3

Share of Employees Imputed to Be Eligible for Employment-Based Insurance in the Tax-Filing Data Set



Data source: Congressional Budget Office.

MEPS-IC. Finally, CBO and JCT randomly reclassified additional workers in the tax-filing data set to be eligible to purchase employment-based coverage until the shares of workers who were eligible in the tax-filing data set equaled the estimated shares.

Inferring Marital Status. To infer a worker’s marital status, CBO and JCT drew on the filing statuses reported on individual income tax returns (Form 1040) between 2011 and 2017. If an employee’s filing status was either “Married Filing Jointly” or “Married Filing Separately” in the most recent year in which he or she filed an individual tax return, and the spouse in that filing year remained alive in 2017, the agencies inferred that the employee was married in 2017. Drawing on individual tax returns over a range of years allowed the agencies to identify the marital status of employees who did not file a tax return in 2017. For married employees, CBO and JCT used the taxpayer identification number of their spouse to determine whether the spouse was employed and, if so, whether the spouse was reported to have been offered health insurance by his or her employer (or whether the offer was imputed).

Health Care Cost Institute Data Set

For information on the health care spending for the employees of actual firms, CBO analyzed claims for approximately 45 million small- and large-market enrollees in 2017 from three large national insurers (UnitedHealthcare, Humana, and Aetna) held by HCCI. Those enrollees—of whom 24 million were policyholders—represented about one-quarter of all enrollees in employment-based insurance. The data included inpatient, outpatient, pharmacy, and physician claims by full- and partial-year policyholders and dependents. The data also included

information on enrollees' age, a group identifier that allowed for enrollees to be grouped by firm, and an indicator for whether the group was in the small- or large-group market.²⁴

CBO excluded records from the data set for several reasons. The agency excluded all dependents' claims; they could not be linked to policyholders and were therefore not useful for determining the composition of synthetic firms.²⁵ CBO also excluded policyholders who were enrolled for less than six months because their spending could not be reliably annualized. In addition, the agency excluded enrollees who were over age 65 and had more than half of their spending covered by a secondary payer, on the presumption that they were retirees rather than current employees. Finally, the agency excluded policyholders at firms with a single employee, with negative total claims during the year, and who were younger than 18. The resulting data set included information on 19.1 million policyholders with 0.3 million group IDs (identifications).

CBO made two adjustments to spending in the data set. First, after dropping policyholders who were enrolled for less than six months, CBO annualized the spending of the remaining partial-year policyholders by multiplying their average monthly spending by 12. Second, CBO inflated the spending for enrollees in plans for whom insurers had no information on prescription drug spending using the ratio of total spending to nondrug spending among enrollees of self-insured large-group plans for whom insurers did have information on prescription drug spending.²⁶ That adjustment was based on the inference that policyholders of self-insured large-group plans without prescription drug benefits would have had the same spending on prescription drug benefits if they had been enrolled in plans with those benefits as those who were enrolled.

Methodology

CBO focused on three goals when developing its method for constructing synthetic firms. First, the traits used to select coworkers for synthetic firms should be strongly correlated with demand for employment-based health insurance. Second, the method should produce realistic distributions of those traits within firms. Third, the method should produce realistic heterogeneity among firms with respect to those traits.

CBO's method for constructing synthetic firms focuses on traits that are strongly related to demand for employment-based insurance and for which the agency could obtain high-quality

²⁴ For more information on the fields included in the data set, see Health Care Cost Institute, *Legacy HCCI 1.0 Data Dictionary* (2019), <https://healthcostinstitute.org/research/research-resources> (PDF, 205 KB).

²⁵ While CBO does not use the health care spending for policyholders' dependents to build synthetic firms, HISIM2 does account for that spending when imputing employment-based insurance premiums and modelling coverage decisions.

²⁶ The HCCI data do not have information on the prescription drug spending for policyholders at firms that use an external pharmacy benefits manager to administer benefits or, less frequently, at firms that do not offer prescription drug benefits.

data. Some traits, like age or earnings, are related to demand because they are used to define policies that determine the premiums, attractiveness, and availability of different forms of health insurance. Other traits, like expected health care spending (by an employee and on his or her behalf), help determine the relative attractiveness of health insurance plans with different premiums and benefits. For example, the after-subsidy amount that a worker pays to purchase a policy in the health insurance marketplaces established by the Affordable Care Act (ACA) is formulaically determined by the ages of the people who would be covered and the ratio of their household's modified adjusted gross income relative to the FPL.²⁷ If a worker can purchase a policy through an ACA marketplace that would cover most of his or her household's expected spending at an affordable out-of-pocket cost, employment-based insurance becomes less attractive.

The extent to which employers respond to changes in policies depends in turn on how their workers, given their traits, would be affected. For instance, an employer with mostly high-income employees is unlikely to change its health insurance benefits in response to changes in the subsidy design for the ACA marketplaces. Even if subsidies were extended to households with higher incomes, those subsidies would probably be worth less than the value of the tax exclusion for higher-income employees. However, an employer might stop offering health insurance if it employed many lower-income workers who could access larger subsidies in the ACA marketplace if they were not offered affordable employment-based insurance.²⁸ Therefore, to make realistic predictions about whether a particular worker would gain or lose access to employment-based insurance in response to such a policy change, CBO's method for constructing synthetic firms must produce realistic distributions of traits within the firms. In addition, to make realistic predictions of the total number of workers who would gain or lose access to employment-based insurance, the method must yield realistic heterogeneity among firms with respect to those traits.

In practice, CBO constructs synthetic firms for each of the 86,667 workers in the CPS in three steps. First, CBO and JCT classify employers in the tax-filing data set into different types of firms that are characterized by distinct joint distributions of employees' age and earnings.

²⁷ Under the ACA, insurers may charge marketplace enrollees different premiums based on age, tobacco use, family size, and geography. Someone who is eligible to receive a subsidy through an ACA marketplace and selects the benchmark plan is required to pay up to a designated amount of his or her income toward premiums, and the subsidy covers the remainder. (A benchmark plan is the second-lowest-cost silver plan available in the marketplace in any given rating area.) The designated amount someone is required to pay increases with the ratio of the household's modified adjusted gross income relative to the FPL. The person can also apply up to that amount toward the purchase of another plan and receive a subsidy that would have applied for the benchmark plan.

²⁸ A person who is offered affordable employment-based coverage that meets a minimum value or who is offered coverage as a family member of a person with such an offer is not eligible for premium tax credits in the ACA marketplaces. For definitions of affordability and minimum value and information on special rules and exceptions, see Center on Budget and Policy Priorities, *Employer Coverage & Premium Tax Credit Eligibility Guide* (December 2019), <https://tinyurl.com/467k5tw9>.

Second, CBO specifies the desired composition of workers' synthetic firms by selecting the age, earnings, marital status, and state of residence of their coworkers on the basis of estimated distributions of those characteristics from the tax-filing data set and their health care spending on the basis of an estimated distribution of such spending from the HCCI data set. Third, CBO selects CPS respondents with the appropriate traits to populate each synthetic firm and imputes their eligibility for employment-based insurance. Whereas in the previous step, CBO specified the desired characteristics for each coworker in a synthetic firm, in this step the agency identifies CPS respondents that exactly or closely match those characteristics and assigns them to the firm. The agency's method for populating synthetic firms adjusts the distribution of coworkers' marital status, states of residence, and health care spending to account for the fact that the CPS does not include respondents with every combination of traits.

Step 1: Assigning Firms a Type

To replicate the real-world heterogeneity among firms, CBO has each synthetic firm mimic one of the many types of actual firms identified in the tax-filing data set. In the real world, young, low-earning workers, for example, can work at firms with very different distributions of coworkers depending on whether they are, say, a clerk at a law firm or a hotel. Consequently, among firms of the same size and offer status in the tax-filing data set, CBO and JCT identified types that differ from one another in terms of the age, earnings, marital status, and eligibility of the people employed. While the agencies did not use industry to distinguish firm types because the CPS sample size is insufficient to allow for doing so, the characteristics used to identify different types of firms capture differences in workforces in different industries that might influence the offer status. When constructing a worker's synthetic firm, CBO randomly selects one of the identified types for it to mimic.

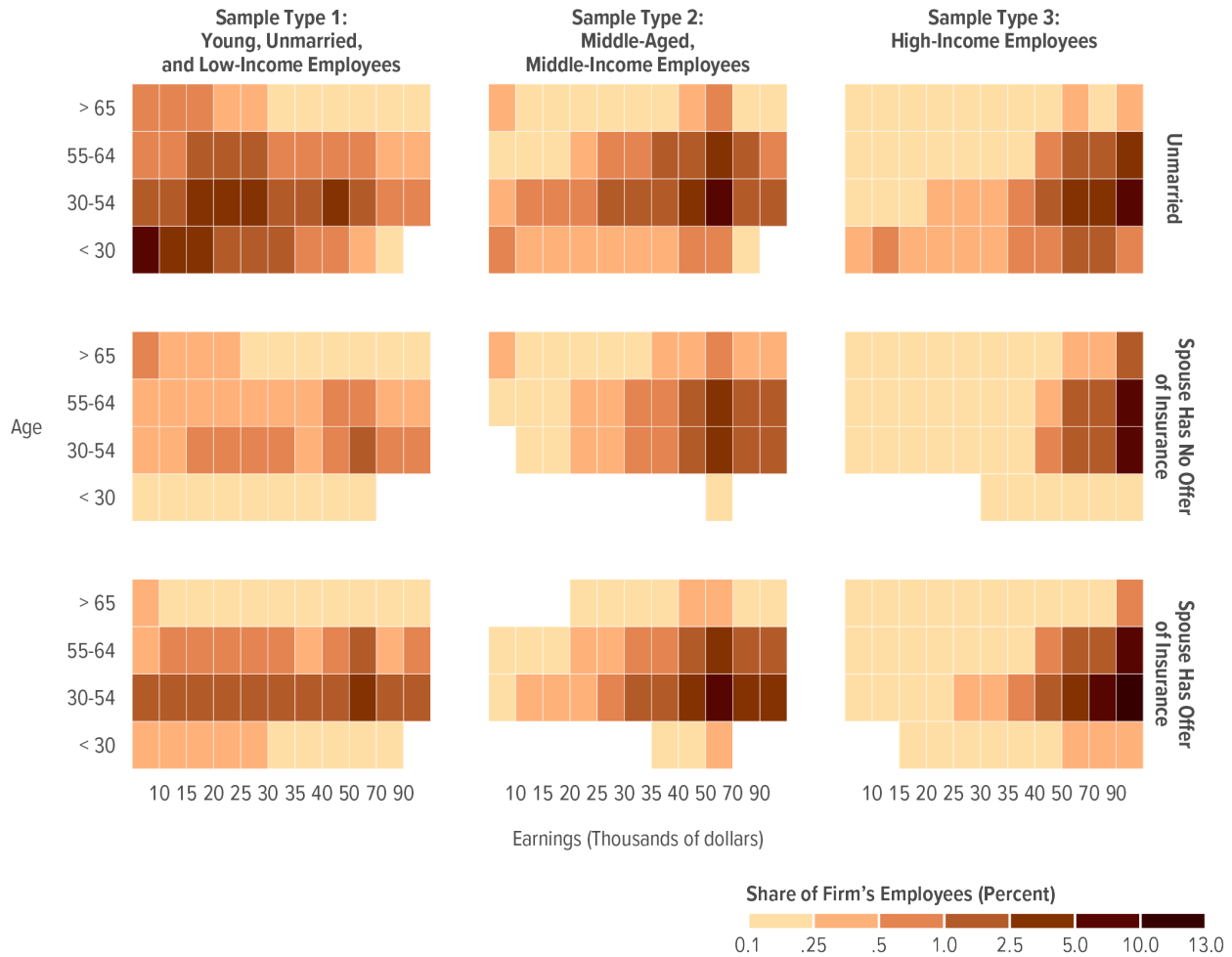
When CBO and JCT analyzed the tax-filing data set, they used the k -means clustering algorithm to identify different types of firms. That popular machine learning algorithm allowed the agencies to identify groups of firms that were distinct in the sense that firms within a cluster had similar distributions of employees' ages and earnings, but firms in other clusters had dissimilar distributions.²⁹ Then, for each type of firm, the agencies calculated the joint distributions of employees' ages, earnings, marital status, and eligibility that were subsequently used to determine the desired composition of synthetic firms and to impute the eligibility of coworkers.

Those distributions capture differences among firms that are relevant for modeling policies related to health insurance, as in this example: Three sample types of firms consist of firms with many young, unmarried, low-income coworkers; with many middle-aged, middle-income

²⁹ Specifically, the agencies computed a discrete grid of each firm's joint cumulative distribution function (CDF) of the earnings and ages of its workers. The k -means clustering algorithm was applied to that CDF grid.

Figure 4

Distribution of Employees' Ages, Earnings, and Marital Status (as Related to Insurance) at Three Sample Types of Firms



Data source: Congressional Budget Office.

employees; and with many high-income employees (see Figure 4). Under a policy that expanded Medicaid in all states, for instance, more employees would become eligible for Medicaid at firms of the first sample type than the second or third; consequently, firms of the first sample type would be relatively more likely to stop offering insurance in response to the policy.

Before identifying types, the agencies sorted firms by their size and whether or not they offered health insurance to their employees. When applying the *k*-means clustering algorithm, the user specifies how many types to identify. Classifying employers into many small groups results in synthetic firms that closely resemble actual firms, but it also introduces a risk of inadvertently disclosing the composition of individual firms. CBO and JCT balanced those considerations by identifying many different types of firms but not so many that more than a handful of types

Table 1

Total Number of Types of Firms Identified

	Number
Firms That Offer Health Insurance	
Small firms (Fewer than 25 employees)	28
Medium-sized firms (25 to 99 employees)	40
Large firms (100 employees or more)	117
Firms That Do Not Offer Health Insurance	
Small firms (Fewer than 25 employees)	33
Medium-sized firms (25 to 99 employees)	9
Large firms (100 employees or more)	14
Total	241

Data sources: Congressional Budget Office and the staff of the Joint Committee on Taxation.

consisted of fewer than 25 firms. Types that did consist of fewer than 25 firms were not used to create synthetic firms. Because there are more medium-sized and large firms that offer health insurance than do not, CBO and JCT identified more types of the former than the latter.

All told, CBO and JCT identified 241 types of firms in the tax-filing data set (see Table 1).

Step 2: Specifying the Composition of Synthetic Firms

Before populating workers' synthetic firms, CBO first selects the traits of each coworker in the firms. To ensure that synthetic firms mimic actual firms, CBO quantifies the relationships among a specific set of traits of individuals and their coworkers in the tax-filing and HCCI data sets and imposes those relationships on synthetic firms. The agency assigns individuals' synthetic firms to correspond to one of the types of firms identified in the previous step. Then, the agency specifies the age,³⁰ earnings,³¹ and marital status³² of individuals' coworkers on the basis of their own age, earnings, and marital status, as well as the size of their firm,³³ whether or not they were offered health insurance, and the type of firm they were assigned. CBO specifies the states that individuals' coworkers reside in on the basis of their own state of residence as well as the size of their firm and whether or not they were offered health insurance. The agency specifies the health

³⁰ Age is categorized as 18 to 29, 30 to 54, 55 to 64, or 65 and older.

³¹ Earnings are categorized as less than \$10,000, \$10,000 to \$14,999, \$15,000 to \$19,999, \$20,000 to \$24,999, \$25,000 to \$29,000, \$30,000 to \$34,999, \$35,000 to \$39,999, \$40,000 to \$49,999, \$50,000 to \$69,999, \$70,000 to \$89,999, or \$90,000 or more.

³² Marital status is categorized as unmarried, married to someone who does not have an offer of employment-based insurance, and married to someone who does have an offer of employment-based insurance.

³³ Firms are categorized as small (fewer than 25 employees), medium-sized (25 to 99 employees), or large (100 employees or more).

care spending³⁴ for individuals' coworkers on the basis of their own age and health care spending, the size of their firm, whether or not they were offered health insurance, and the coworkers' previously determined age.

Specifying Coworkers' Age, Earnings, and Marital Status. CBO specifies coworkers' age, earnings, and marital status because those traits are closely related to the demand for employment-based insurance:

- Age directly determines eligibility for some types of coverage, like Medicare³⁵ or marketplace insurance through a parent,³⁶ and determines premiums in the small-group and nongroup markets.³⁷ It is correlated with people's expected expenditures on health care,³⁸ their adjusted gross income relative to the FPL, and the age and number of people that they live with.
- Earnings³⁹ are the largest component of adjusted gross income for most households⁴⁰ and are therefore directly related to the value of the tax exclusion for employment-based insurance;⁴¹ eligibility for some types of coverage, like Medicaid;⁴² and subsidies in the nongroup market.⁴³ Earnings are also correlated with demand for insurance for other reasons, like the affordability of premiums, eligibility for financial assistance from providers due to an inability to pay, and—through their correlation with wealth—the desire for asset protection.⁴⁴

³⁴ Health care spending is categorized relative to the distribution of it for employment-based insurance policyholders as below the 25th percentile, between the 25th and 50th percentiles, between the 50th and 75th percentiles, between the 75th and 90th percentiles, between the 90th and 95th percentiles, or above the 95th percentile.

³⁵ Department of Health and Human Services, "Who Is Eligible for Medicare?" (September 11, 2014), <https://go.usa.gov/xeaAV>.

³⁶ Healthcare.gov, "How to Get or Stay on a Parent's Plan" (accessed September 17, 2021), www.healthcare.gov/young-adults/children-under-26/.

³⁷ Centers for Medicare & Medicaid Services, "Market Rating Reforms" (June 2, 2017), <https://go.usa.gov/xearQ>.

³⁸ Dale H. Yamamoto, *Health Care Costs—From Birth to Death?* Report 2013-1 (Health Care Cost Institute, June 2013), <https://tinyurl.com/xs5a593w>.

³⁹ In HISIM2, earnings are defined as wage and salary earnings from the job the worker held for the longest time over the past 12 months.

⁴⁰ Internal Revenue Service, *Statistics of Income—2018 Individual Income Tax Returns* (September 2020), www.irs.gov/pub/irs-pdf/p1304.pdf (6.09 MB).

⁴¹ Matthew Rae and others, *Tax Subsidies for Private Health Insurance* (The Henry J. Kaiser Family Foundation, October 2014), <https://tinyurl.com/emcph7y6>.

⁴² Tricia Brooks and others, *Medicaid and CHIP Eligibility and Enrollment Policies as of January 2021: Findings From a 50-State Survey* (The Henry J. Kaiser Family Foundation, March 2021), <https://tinyurl.com/29tjzn28>.

⁴³ Internal Revenue Service, "Questions and Answers on the Premium Tax Credit" (May 17, 2021), <https://go.usa.gov/xea6G>.

⁴⁴ Pamela J. Farley and Gail R. Wilensky, "Household Wealth and Health Insurance as Protection Against Medical Risks" in Martin David and Timothy Smeeding, eds., *Horizontal Equity, Uncertainty, and Economic Well-Being* (University of Chicago Press, 1985), pp. 323–358, <https://doi.org/10.7208/9780226137285-015>.

- Marital status is related to household composition. In addition, people who are married to someone with an offer of employment-based insurance are not willing to pay as much for insurance offered directly to them.

CBO assigns those traits jointly because they are closely related and are all observed in the tax-filing data set. Young coworkers, for example, are more likely to have low income or be unmarried than older coworkers. If CBO assigned coworkers' age independently of their earnings and marital status, the distribution of the three traits among coworkers in the synthetic firms would be unrealistic.

Specifying the Residence of an Individual's Coworkers. CBO specifies the residence of coworkers in synthetic firms so that the firms' incentives to offer health insurance reflect differences among states in terms of their relevant policies and demographics. State policies directly affect demand for employment-based insurance by setting income tax rates (which determine the value of the tax exclusion), eligibility policies for public programs like Medicaid, and the rules governing how insurers set premiums in the small-group and nongroup markets, among other things. Meanwhile, an individual's residence may be correlated with a variety of factors that are indirectly related to the demand for employment-based insurance, like family size.

CBO first probabilistically assigns individuals to work at firms where nearly all employees live in the same state (single-state firms) or firms where employees are distributed across two or more states (multistate firms) on the basis of the share of firms that were multistate firms in the tax-filing data set (see Table 2). The difference matters. One state's policies, such as Medicaid expansion decisions, do not strongly influence multistate firms. Moreover, multistate firms are more likely to be large and to offer health insurance.

CBO and JCT classify firms in the tax-filing data set as single-state firms if 75 percent or more of their employees live in a single state and classify firms as multistate when that is not the case. If an individual is assigned to work for a single-state firm, all of his or her coworkers are specified to reside in that individual's state. If an individual is assigned to work in a multistate firm, the coworkers are specified to reside in either the individual's state, other states in the individual's census division (Northeast, Southeast, Midwest, or West), or in one of the three other census divisions.

Table 2

Share of Workers Employed by Multistate Firms

	Share
Firms That Offer Health Insurance	
Small firms (Fewer than 25 employees)	9.8%
Medium-sized firms (25 to 99 employees)	13.4%
Large firms (100 employees or more)	47.1%
Firms That Do Not Offer Health Insurance	
Small firms (Fewer than 25 employees)	6.4%
Medium-sized firms (25 to 99 employees)	7.4%
Large firms (100 employees or more)	15.6%
Total	35.4%

Data sources: Congressional Budget Office and the staff of the Joint Committee on Taxation.

Specifying the Health Care Spending for an Individual's Coworkers. CBO specifies the health care spending for an individual's coworkers because it determines premiums at large firms and is related to demand for health insurance. In most states, premiums for firms with more than 50 employees reflect the expected cost of insuring enrollees in those plans. In addition, individuals with higher health care spending benefit more from enrolling in generous health insurance plans than those with lower health care spending.

The HCCI data set revealed a modest relationship between the health care spending for workers and their coworkers. Especially at large firms, workers with health care spending below the 25th percentile of spending for all policyholders in the HCCI data set are somewhat more likely than others to have coworkers with similarly low spending. By creating synthetic firms that reflect that relationship, CBO generates premiums that are realistically related to the health care spending by and for individuals.

CBO randomly assigns an individual's coworkers' health care spending categories on the basis of the individual's age and health care spending, the coworkers' ages, and the size of the firm and whether or not it offers employment-based insurance. Health care spending for coworkers at firms with fewer than 25 employees is assigned on the basis of the distribution of such spending for policyholders in the small-group market. At firms with 25 employees or more, it is assigned on the basis of the same distribution in the large-group market. Because the HCCI data set includes only information on employment-based insurance policyholders, CBO does not use the HCCI data to assign the health care spending of coworkers at firms that do not offer employment-based insurance. For firms that do not offer health insurance, selection of coworkers is independent of the coworkers' health care spending.

Step 3: Populating Synthetic Firms

After specifying the composition of synthetic firms, CBO populates them with people drawn from the CPS. To populate a synthetic firm around the individual worker, the agencies randomly

draw from the sample in the CPS people with the desired age, earnings, marital status, residence, and health care spending categories who work for firms of the same size that similarly did or did not offer health insurance. The agency then imputes coworkers' eligibility for employment-based insurance on the basis of their age, earnings, and marital status, as well as the size of their firm and whether or not they were offered health insurance.

When populating synthetic firms, CBO selects coworkers from the set of people in the CPS who work for firms of the same size and with the same offer status as the firm for the individuals around whom the synthetic firms are being constructed. Workers' decisions about where to work are partly determined by their demand for employment-based insurance. In examining the CPS, CBO found that workers who choose to be employed by the same type of firm tend to be similar along dimensions other than the traits selected in the agency's method for constructing firms but that may nonetheless be important in determining whether or not a firm offers insurance.

The probability that a particular person in the CPS is sampled is proportional to his or her CPS sample weight. CBO caps the size of synthetic firms at 300 employees. That cap reduces the time HISIM2 requires for policy simulations but is sufficient to model pooling of health care costs at larger firms and match the targeted composition of a worker's firm type.

CBO uses two methods to deal with the fact that the CPS does not include respondents with every possible combination of traits. While increasing the number of traits used to select coworkers yields more realistic synthetic firms, it also increases the chance that synthetic firms include coworkers with a combination of traits that no CPS respondents have. CBO's first method for dealing with this issue is to prioritize traits in the following order: (1) firm size and offer status, (2) age and earnings, (3) marital status, (4) health care spending, and (5) state of residence. For example, CBO will sometimes populate a synthetic firm with a coworker who has the right age but the wrong state, but never the opposite. The agency subjectively ranked traits on the basis of several factors, like how much each trait probably affects demand for employment-based insurance, the strength of the empirical relationship for the trait between worker and coworkers, and the quality of the data used to estimate that relationship. The agency's second method for dealing with this issue is to group marital status, health care spending, and residence categories. For example, if the CPS contains no respondents with a certain combination of traits above the 95th percentile of health care spending, CBO looks to draw a coworker with that combination of traits above the 90th percentile of such spending.

CBO imputes whether coworkers are eligible for employment-based insurance, because in its analysis, employers do not consider the preferences of ineligible employees when deciding whether to offer health insurance to employees. Firms often restrict their part-time, temporary, seasonal, and new employees from purchasing into group policies. Imputing whether synthetic firms' employees are eligible for employment-based insurance improves HISIM2's simulations

of employers' behavior by making employers less sensitive to policies affecting young, low-income employees, who tend to be ineligible for employment-based insurance.

CBO imputes coworkers' eligibility for employment-based insurance after populating synthetic firms with CPS respondents because it is not possible to conclusively identify CPS respondents who were not eligible for employment-based insurance. On the one hand, CPS respondents who were employed by firms that did not offer employment-based insurance could not say whether they would have been eligible for employment-based insurance had their firm offered it. On the other hand, CPS respondents who were employed by firms that did offer employment-based insurance often misreport their eligibility. To impute eligibility, CBO probabilistically assigns a coworker to be eligible on the basis of the share of employees in the tax-filing data set who were eligible for employment-based insurance and had the same age, earnings, and marital status and worked for firms of the same size that similarly did or did not offer health insurance.

Assessment of the Quality of the Synthetic Firms

CBO and JCT used three main groups of analyses to evaluate the data and methods used to construct HISIM2's synthetic firms. First, the agencies compared the distribution of traits used to select synthetic coworkers (age, earnings, marital status, state of residence, and firm size) among workers in HISIM2 and workers in the tax-filing data. Because those characteristics are used to build firm types and choose coworkers for synthetic firms, it is important to verify that, at an aggregate level, the two data sources agree on the overall composition of the employed. Second, the agencies calculated summary statistics of the distribution of coworkers' traits across firms based on a worker's own traits. Those statistics were first calculated for the administrative tax data to more easily understand the numerous firm types produced by the clustering algorithm. The same statistics were then calculated for the synthetic firms in HISIM2's data to allow the agencies to evaluate whether those relationships were successfully replicated. Third, the agencies used simulated changes in employees' demand for an offer of health insurance from their employers resulting from two policy changes. Specifically, the agencies compared the effects of those policies on the preferences of a single worker at a firm and the preferences of the entire synthetic firm's workforce.

Comparison of Workforces in the CPS and Tax Data

The following set of figures demonstrates that the distributions of age, earnings, marital status, and firm size for workers in HISIM2 in 2017 closely match those distributions in the 2017 tax-filing data set. The small differences are generally attributable to sampling variation and reporting error in the CPS or small differences in the concepts measured between survey and administrative data. Ensuring that the traits of workers in HISIM2 closely match those in the tax-filing data set is important to ensure that HISIM2's synthetic firms are similar to actual firms. Because, as outlined in the methodology section, a worker's type of firm is chosen on the basis

of that worker's own traits, matching the overall distribution of workers' traits in the tax data also ensures that HISIM2 will also closely match in terms of the distribution of firm types.

Because of the different methods used to collect and process the underlying data, workers' traits observed in HISIM2 and the tax-filing data might not necessarily be distributed similarly. In HISIM2, the traits are largely based on the CPS, whereas in the tax-filing data set that CBO and JCT created, they rely on administrative records imposing sample exclusions largely based on W-2 earnings. Variation in the timing of the measurement (e.g., partway through the year, averaged over the year, or at the end of the year), reporting error, or sample restrictions may cause the data sources to have dissimilar distributions. To address some of the shortcomings of survey data, CBO makes several modifications to CPS data to better match administrative records as part of the agency's standard process of preparing input data for HISIM2.⁴⁵ Two of those modifications are particularly relevant to the creation of synthetic firms. First, the agency adjusts the firm size reported by workers to match the distribution found in the MEPS-IC. Second, the agency adjusts both dollar amounts and sample weights to better match the wage distribution found in the administrative tax data. The results in this paper rely on HISIM2 data after those modifications have been made to provide a better sense of how the model's data compare with the administrative tax data. Additionally, for consistency with the sample restrictions placed on the tax data, the HISIM2 data used to construct figures and tables for this paper also exclude self-employed workers and those reporting less than \$5,000 in earnings from their longest-held job in the prior year.⁴⁶ Because the administrative tax data keep only the highest-wage W-2 for each worker, CBO and JCT use the CPS measure of annual earnings from the longest-held job (rather than total annual wages) to compare wages in HISIM2 with those in the tax data.⁴⁷ Finally, to match the CPS, which, for confidentiality reasons, caps earnings values at \$200,000 for individuals earning more than that amount, the agencies also cap earnings in the tax data at that amount.

The firm size, age, wage, and marital status distributions of workers in HISIM2 and the tax data all closely match. Figure 5 presents simple comparisons of estimated probability distributions of

⁴⁵ For details about how CBO processes CPS data, see Jessica Banthin and others, *Sources and Preparations of Data Used in HISIM2—CBO's Health Insurance Simulation Model*, Working Paper 2019-04 (Congressional Budget Office, April 2019), www.cbo.gov/publication/55087.

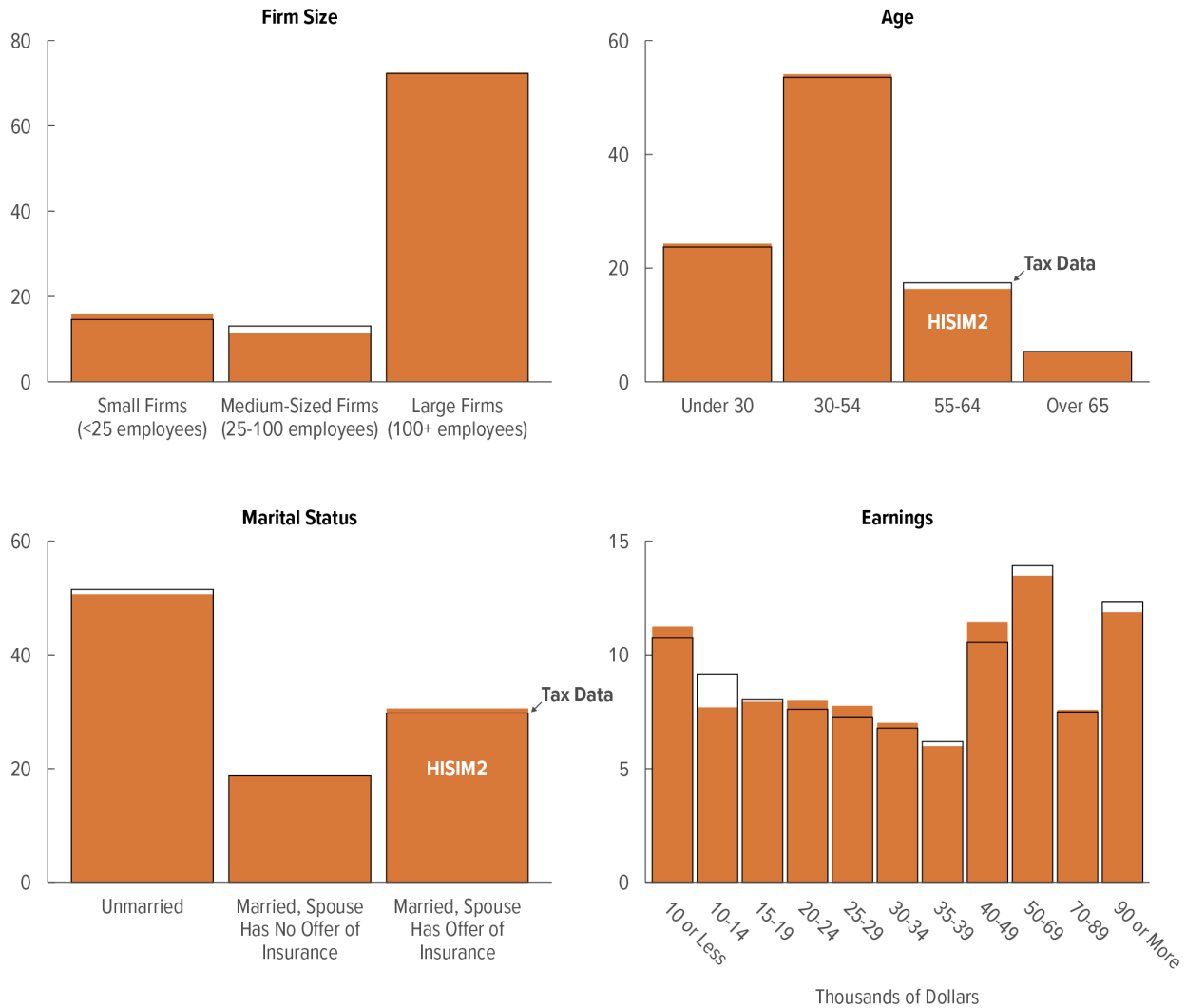
⁴⁶ Overall, 16.2% of workers in the CPS are excluded from this analysis: the 9.6% of workers in the CPS who are self-employed and the 6.6% who are not self-employed but who report less than \$5,000 in earnings from their main job. CBO does build synthetic firms for some self-employed workers, those who are imputed to have at least one coworker, using the same methods outlined in the methodology section and treating them as if they have similar coworkers to those who are not self-employed. The agency also builds synthetic firms for workers with less than \$5,000 in earnings from their main job, treating them as if they have coworkers who are similar to the coworkers of workers earning between \$5,000 and \$10,000.

⁴⁷ The specific CPS variable used is "ERN_VAL." For consistency, CBO applies to that measure the same multiplicative dollar amount modifications that the agency applies to annual total wages.

Figure 5

Comparison of Key Characteristics of Workers in HISIM2 and Tax Data, 2017

Percentage of Workers



Data source: Congressional Budget Office.

HISIM2 = CBO's health insurance simulation model.

those four characteristics. The relatively small differences in firm size are likely owing to different concepts used by both data sources.⁴⁸ The small differences in the age and wage distributions could result from sampling variation or reporting error in the CPS or from the

⁴⁸ In the administrative tax data, the firm size measure is based on the number of employees that were on a firm's payroll in a given quarter. However, HISIM2's CPS data are adjusted to match the firm size distribution from the MEPS-IC, which asks firms to report the number of workers they employ in a "typical pay period." Consequently, a measure of the number of people employed at any point in a quarter would be expected to be slightly higher than the number employed in a typical pay period.

exclusion of certain W-2s in the administrative tax data. Finally, while the distributions of marital status are broadly similar, the small differences could be caused by the discrepancy between a person's tax-filing status (single versus married filing jointly or separately) and how the person reports his or her marital status in a household survey.⁴⁹

Distribution of Traits Within Firms

The next set of figures demonstrates that CBO's methods successfully replicated actual firms, meeting the second and third goals listed in the Methodology section: producing realistic distributions of coworkers' traits within firms and realistic heterogeneity among firms in terms of those key traits.

Figure 6 shows the distribution of coworkers' average ages by the worker's age in HISIM2 and the tax data. Those statistics were created by first calculating the average age of coworkers at each firm (excluding the worker the firm is built around), then calculating the probability density function of that firm-level average age for all workers within a given age and offer status category. As shown in the figure, in the tax data, a clear positive correlation exists between a worker's age and the age of her coworkers. For instance, for a worker between 30 and 54 years old, the average age of his or her coworkers is higher than it is for a worker younger than age 30. In addition, there is also more cross-firm variation in small firms than in large firms (which can be seen in the height of the density plots). Despite some minor differences, HISIM2's synthetic firms replicate those relationships between workers' and coworkers' ages found in the tax data, including differences in cross-firm variation among firms of different sizes.

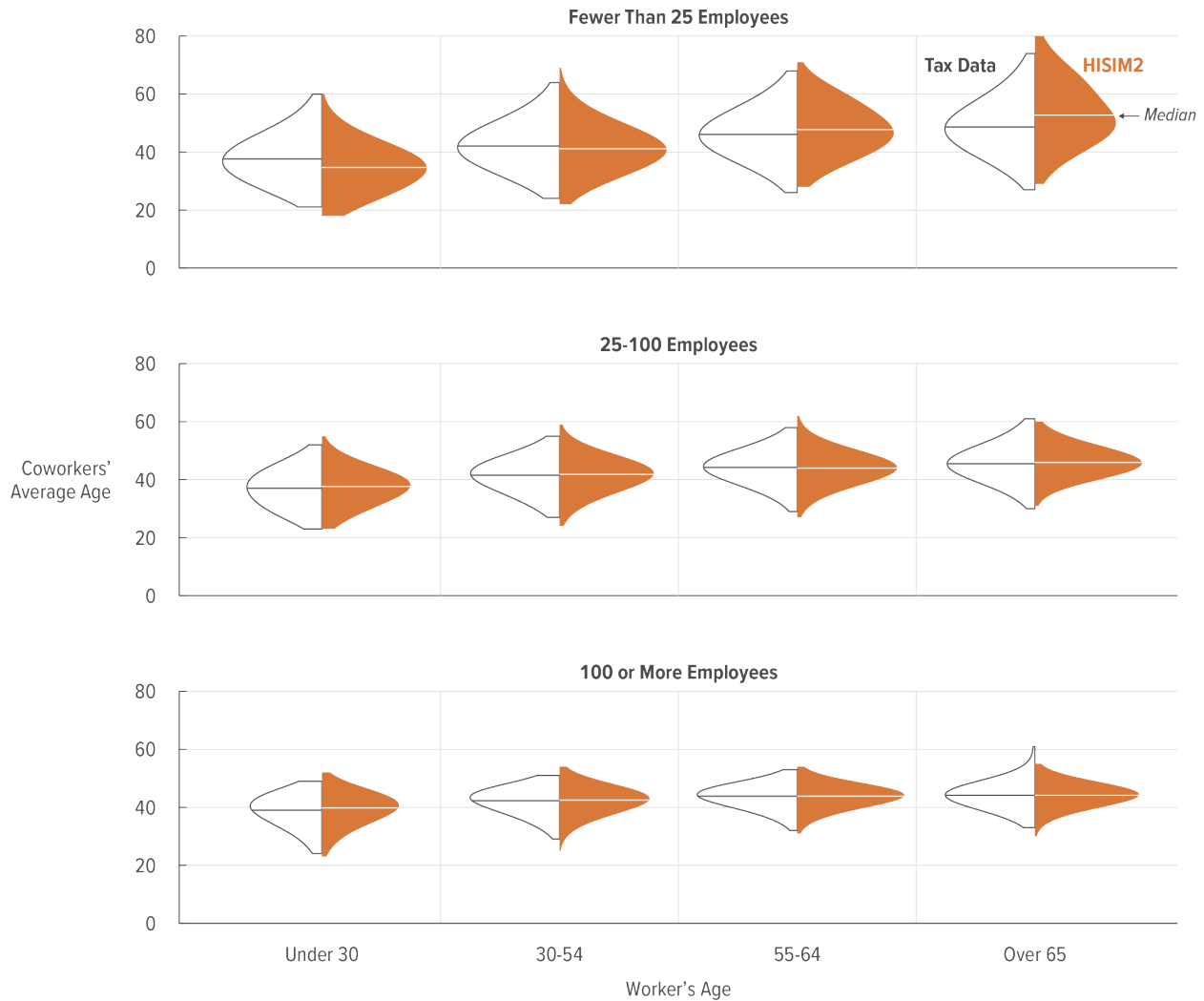
Similarly, Figure 7 shows the distribution of average coworkers' earnings conditioned on a worker's earnings, by firm size. As with age, in the tax data, there is a positive correlation between a worker's wages and the median wages of his coworkers; a higher-wage worker tends to have higher-wage coworkers. Additionally, the median earnings of coworkers are slightly higher at large firms than they are at small ones. However, unlike the situation for coworkers' ages, there is substantial cross-firm variation in coworkers' average wages in both small and large firms. Finally, coworkers' average wages appear to vary more for higher-wage workers than for lower-wage workers. Altogether, HISIM2's synthetic firms largely duplicate the distributions of coworkers' average earnings found in the tax data.

Figure 8 (on page 27) shows the relationship between a worker's and his or her coworkers' marital status as it relates to having an offer of insurance, in both HISIM2 and the administrative tax data. Workers tend to sort into firms by marital status; for unmarried workers, the majority of their coworkers are unmarried, and for married workers, the majority of their coworkers are married. Additionally, among married workers, those whose spouses have an offer of insurance

⁴⁹ CBO and JCT also confirmed that the distribution of states of residence also matches closely between the two data sources; the share of the national workforce that resided in each state differed by no more than 1 percentage point.

Figure 6

Distribution of Coworkers' Average Age, by Worker's Age, in Different-Sized Firms in HISIM2 and Tax Data, 2017



Data source: Congressional Budget Office.

To maintain the confidentiality of workers and firms in the tax-filing data set, the medians are calculated from rounded-value probability distribution functions.

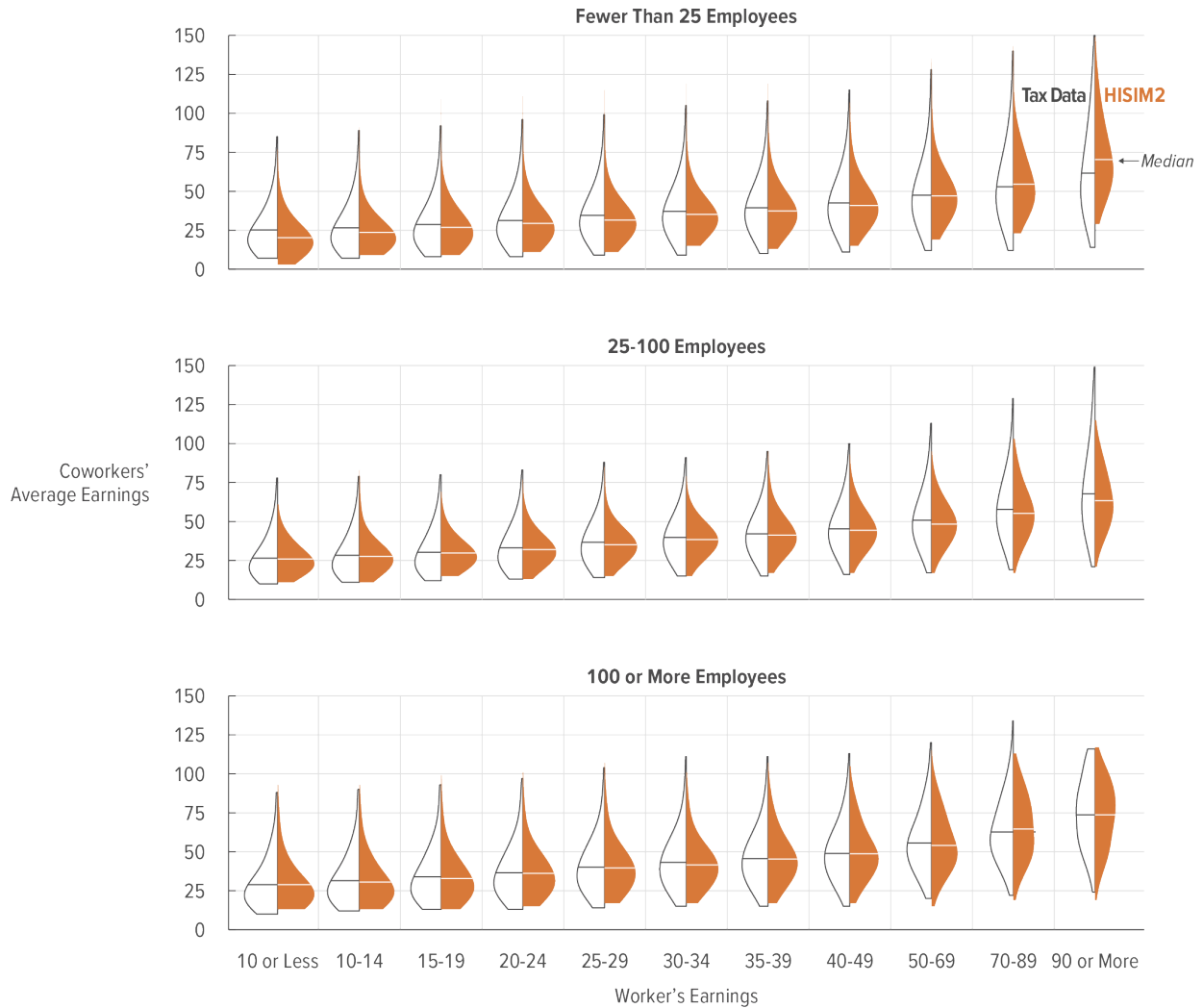
HISIM2 = CBO's health insurance simulation model.

are more likely to work with people in the same situation than is the case for those whose spouses do not have an offer of insurance. Though there are some small differences, the synthetic firms created for HISIM2 replicate those relationships in the tax data.

Figure 7

Distribution of Coworkers' Average Earnings, by Worker's Earnings, in Different-Sized Firms in HISIM2 and Tax Data, 2017

Thousands of Dollars



Data source: Congressional Budget Office.

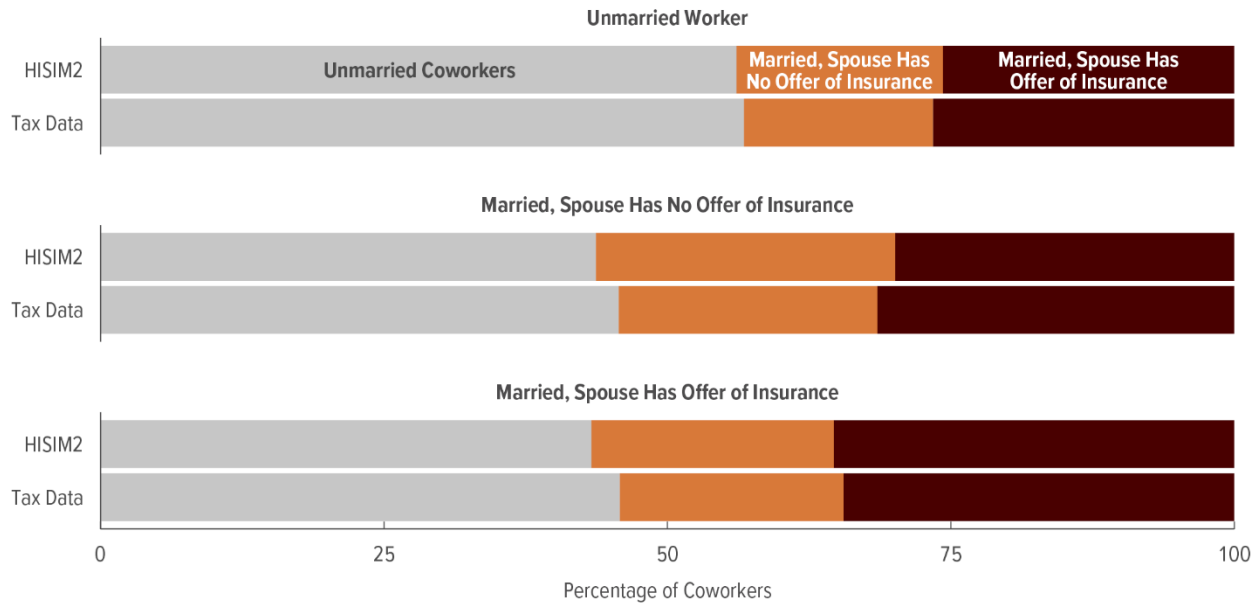
To maintain the confidentiality of workers and firms in the tax-filing data set, the medians are calculated from rounded-value probability distribution functions.

Earnings in the HISIM2 data are those from a worker's longest-held job, and those in the tax-filing data set are from a worker's highest-paying job.

HISIM2 = CBO's health insurance simulation model.

Figure 8

Coworkers' Marital Status as Related to Insurance, by Worker's Status, in HISIM2 and Tax Data, 2017



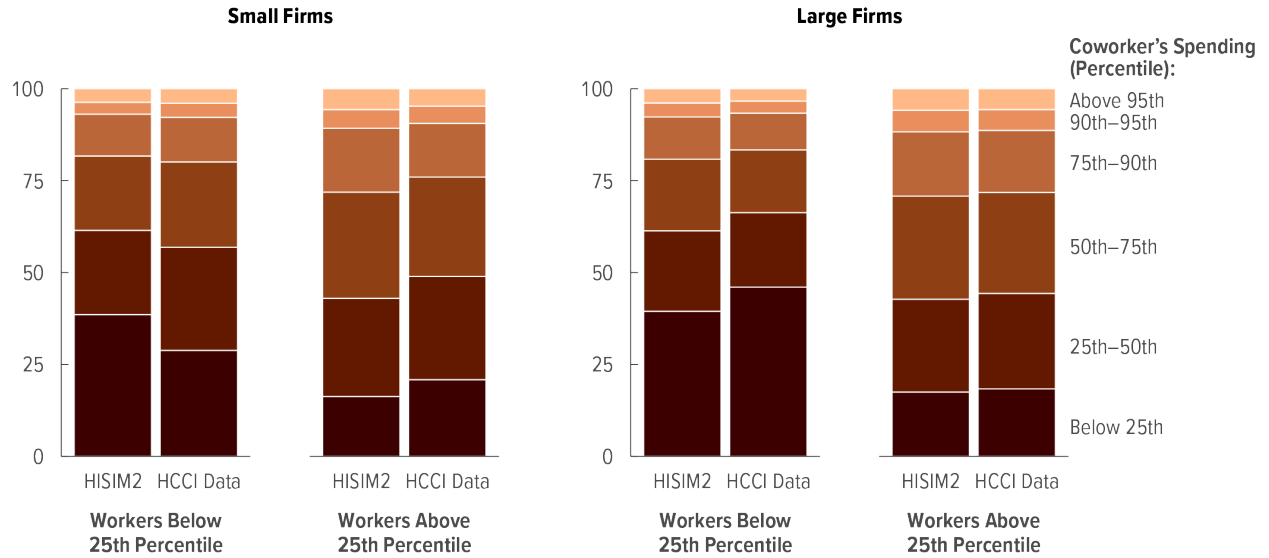
Data source: Congressional Budget Office.

HISIM2 = CBO's health insurance simulation model.

Figure 9 shows the distribution of health care spending for coworkers conditioned on spending for a worker at small and large firms that offered insurance, both in HISIM2 and HCCI data. In the HCCI data, at large firms, workers with spending in the bottom 25 percent are markedly more likely to work with similarly healthy coworkers than are workers with spending in the top 75 percent. For other types of workers, there is a more modest correlation between the worker's and coworkers' health care spending. Differences between the distributions in HISIM2 and the HCCI data result from the limited size of the CPS sample. Because a coworker fitting the exact assigned characteristics is not always available, CBO prioritizes selecting coworkers on the basis of their age, earnings, and marital status over their health care spending, which causes those discrepancies.

Figure 9

Coworkers' Health Care Spending, by Worker's Spending, at Firms That Offered Insurance, in HISIM2 and HCCI Data, 2017



Data source: Congressional Budget Office.

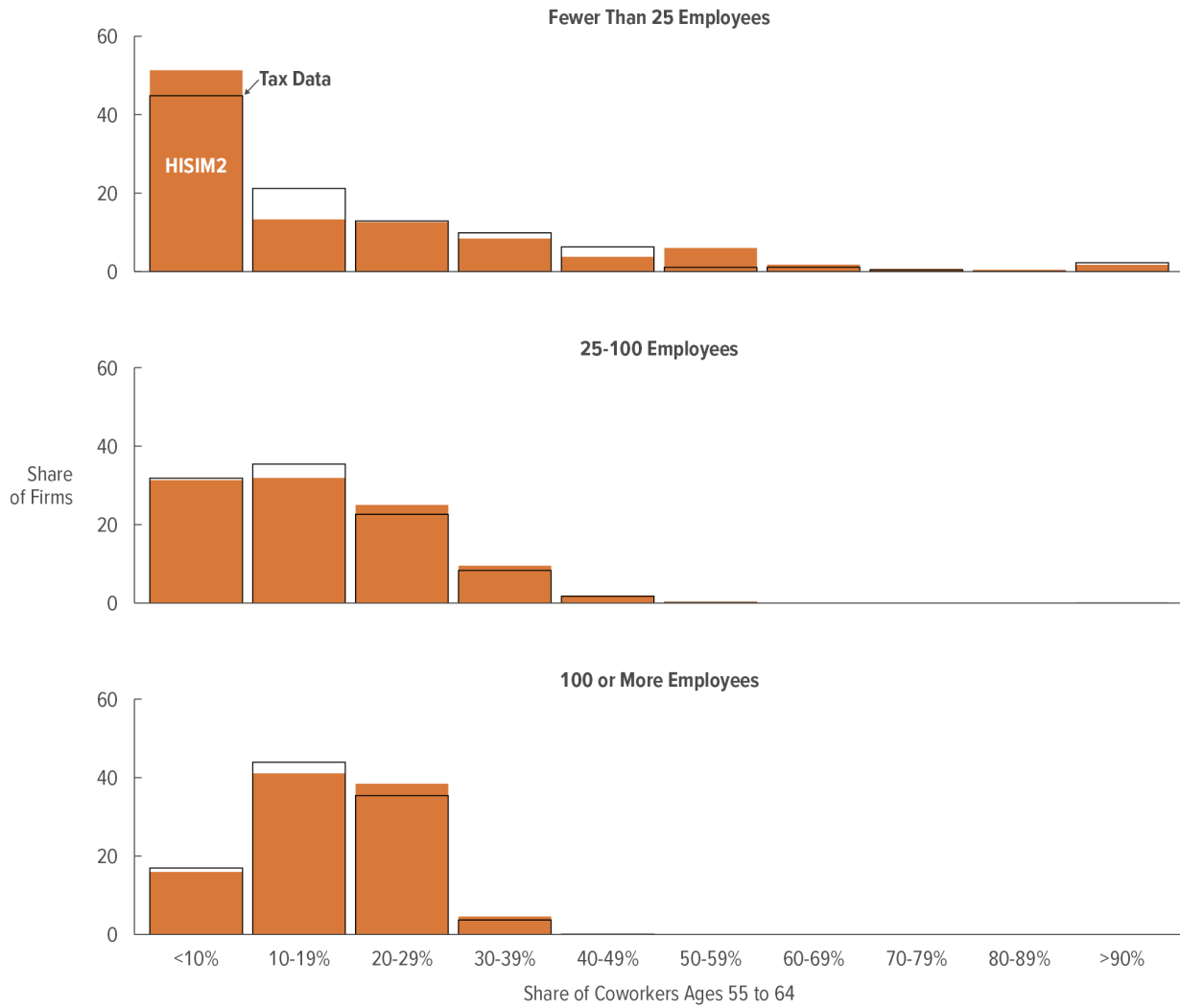
HCCI = Health Care Cost Institute; HISIM2 = CBO's health insurance simulation model.

The next set of figures also demonstrates the heterogeneity of coworkers' characteristics among firms. Each graph shows the share of workers at firms with a given share of coworkers that, in 2017, had a specific characteristic potentially substantially influencing demand for employment-based insurance: being between 55 and 64 years old, being single, and earning less than \$40,000. The figures demonstrate the substantial variation in coworkers' characteristics that affect the demand for employment-based insurance and the ability of HISIM2's synthetic firms to replicate that variation.

Figures 10a-c present the distributions of those three different characteristics of coworkers. The first panel shows the worker-weighted distribution of the share of coworkers ages 55 to 64 in both HISIM2 and the administrative tax data. Compared with those percentages for medium-sized and large firms, a greater share of small firms has less than 10 percent of their workforce in that age group. Additionally, there are some small firms where a majority of coworkers are ages 55 to 64, while such medium-sized and large firms are virtually nonexistent. The second figure of the group shows the distribution of the share of single coworkers. Again, there is more variation among small firms; at about 10 percent of them, fewer than 10 percent of the coworkers are single, and at about 10 percent, more than 90 percent of the coworkers are single. Finally, the third figure of the group gives the share of coworkers with less than \$40,000 in earnings from their main job, roughly the median individual level. About a third of small firms have workforces with more than 90 percent of workers earning less than that amount. However, at large firms, workforces are distributed somewhat evenly across the 10 categories of earnings. The similar distributions of these characteristics in both the administrative tax data and HISIM2 further demonstrate that the synthetic firms successfully mimic the variation observed among real firms.

Figure 10a

Share of Coworkers Ages 55 to 64 in Different-Sized Firms in HISIM2 and Tax Data, 2017



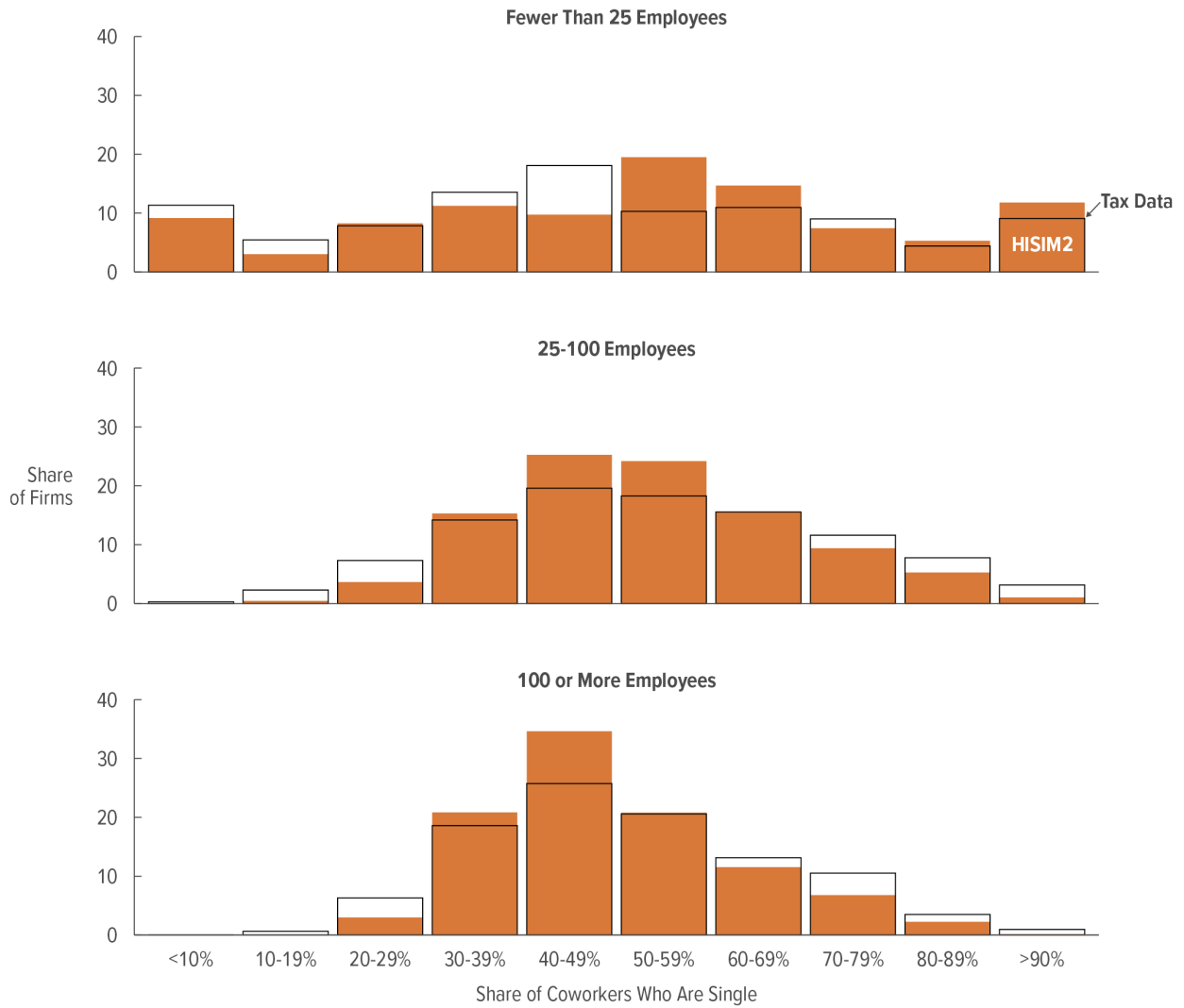
Data source: Congressional Budget Office.

Within each panel, the HISIM2 and Tax Data bars respectively sum to 100.

HISIM2 = CBO's health insurance simulation model.

Figure 10b

Share of Coworkers Who Are Single in Different-Sized Firms in HISIM2 and Tax Data, 2017



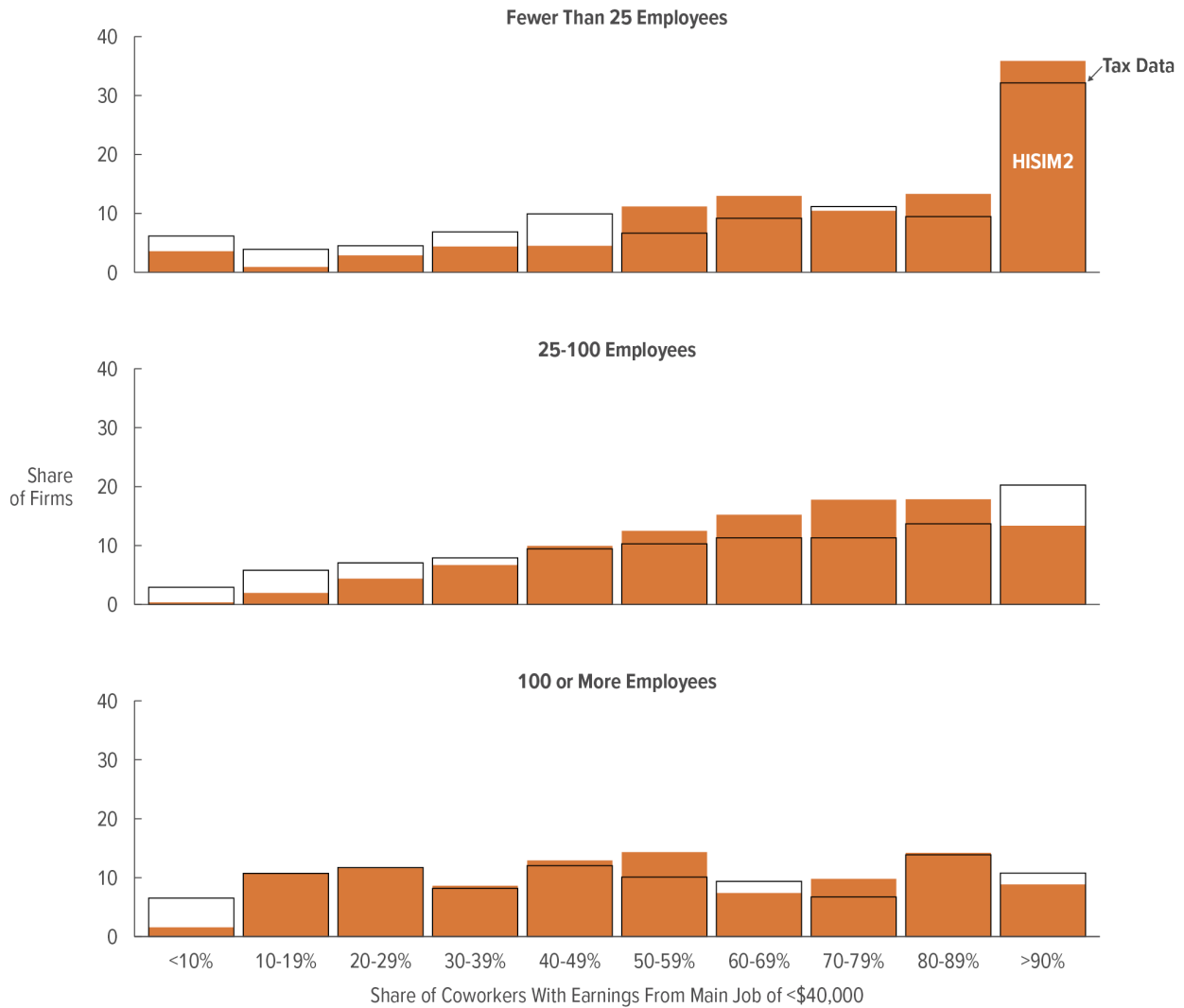
Data source: Congressional Budget Office.

Within each panel, the HISIM2 and Tax Data bars respectively sum to 100.

HISIM2 = CBO's health insurance simulation model.

Figure 10c

Share of Coworkers With Earnings of Less Than \$40,000 in Different-Sized Firms in HISIM2 and Tax Data, 2017



Data source: Congressional Budget Office.

Within each panel, the HISIM2 and Tax Data bars respectively sum to 100.

Earnings in the HISIM2 data are those from a worker's longest-held job, and those in the tax-filing data set are from a worker's highest-paying job.

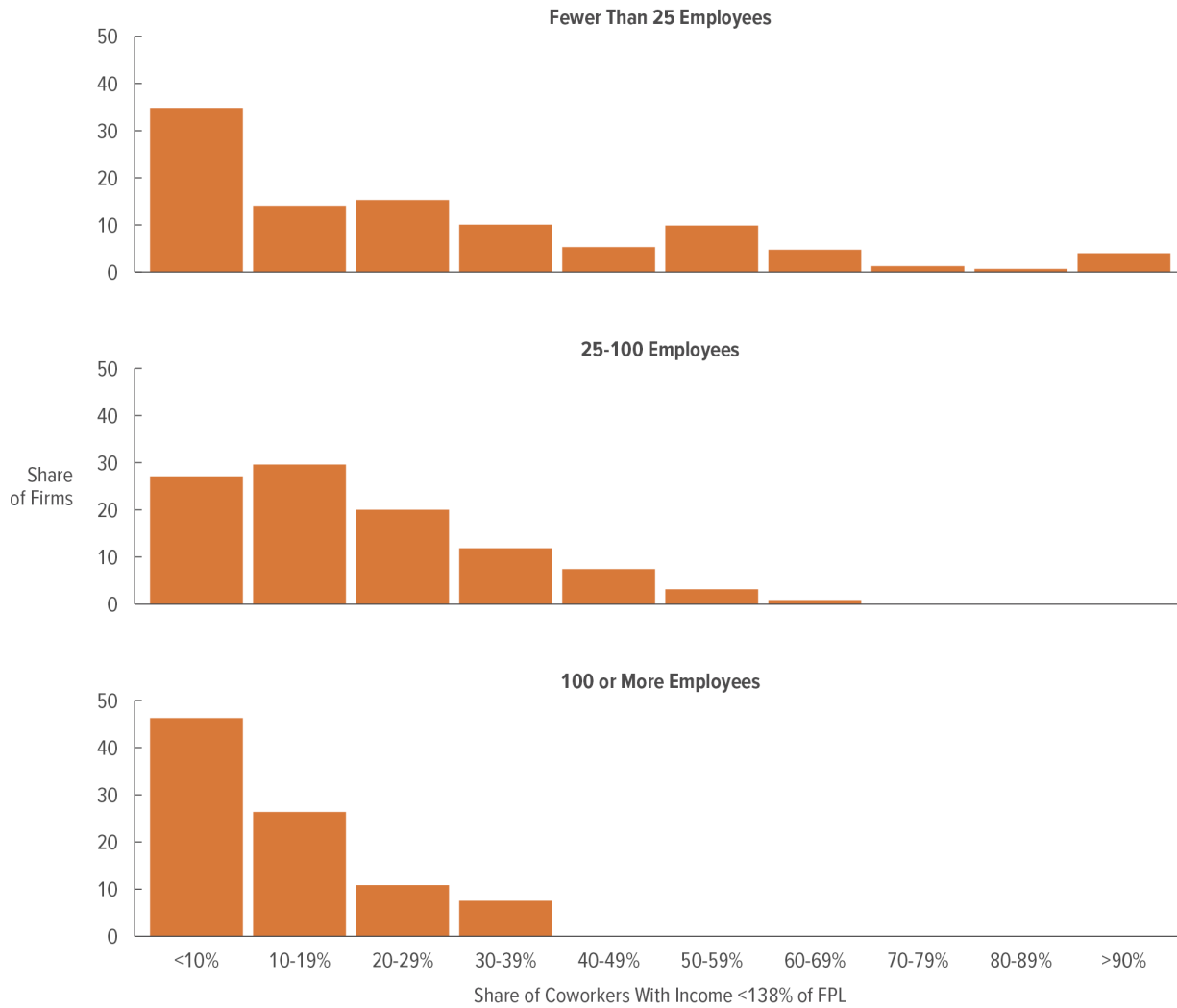
HISIM2 = CBO's health insurance simulation model.

Figures 11a-e use HISIM2's data to display the share of coworkers with characteristics not readily available in the administrative tax data. The first three figures of the group show the shares of coworkers in three different income groups (expressed as a percentage of the FPL) that are particularly important for eligibility for publicly subsidized insurance coverage. In general, there are few large firms with high shares of coworkers in the income groups associated with Medicaid eligibility (<138% of the FPL) or generous nongroup subsidies (138-250% of the FPL). The fourth figure shows the share of coworkers in the Medicaid coverage gap, which means they have income less than the FPL and live in a state that has not expanded Medicaid.⁵⁰ While most firms have very few employees that fall in that coverage gap, a small share of small firms have large percentages of their workforce that would be eligible for Medicaid if all states expanded eligibility. Finally, the fifth figure shows the distribution of the share of a firm's workforce that has greater than the 75th percentile of health care spending, as calculated across all workers in that firm size category. As expected for large firms, the share of coworkers with health care spending exceeding the overall 75th percentile is closely clustered around 25 percent, meaning that the health care spending distributions of the firms are roughly similar to the overall distribution for all workers with employment-based coverage. However, there is substantially more variation in that share for small firms, implying considerably more variation among firms in both the demand for and costs of health insurance.

⁵⁰ Individuals living in a state that has not expanded Medicaid eligibility with income between 100% and 138% of the FPL are typically eligible for subsidized nongroup coverage.

Figure 11a

Share of Coworkers With Income Below 138 Percent of the FPL in Different-Sized Firms in HISIM2, 2017



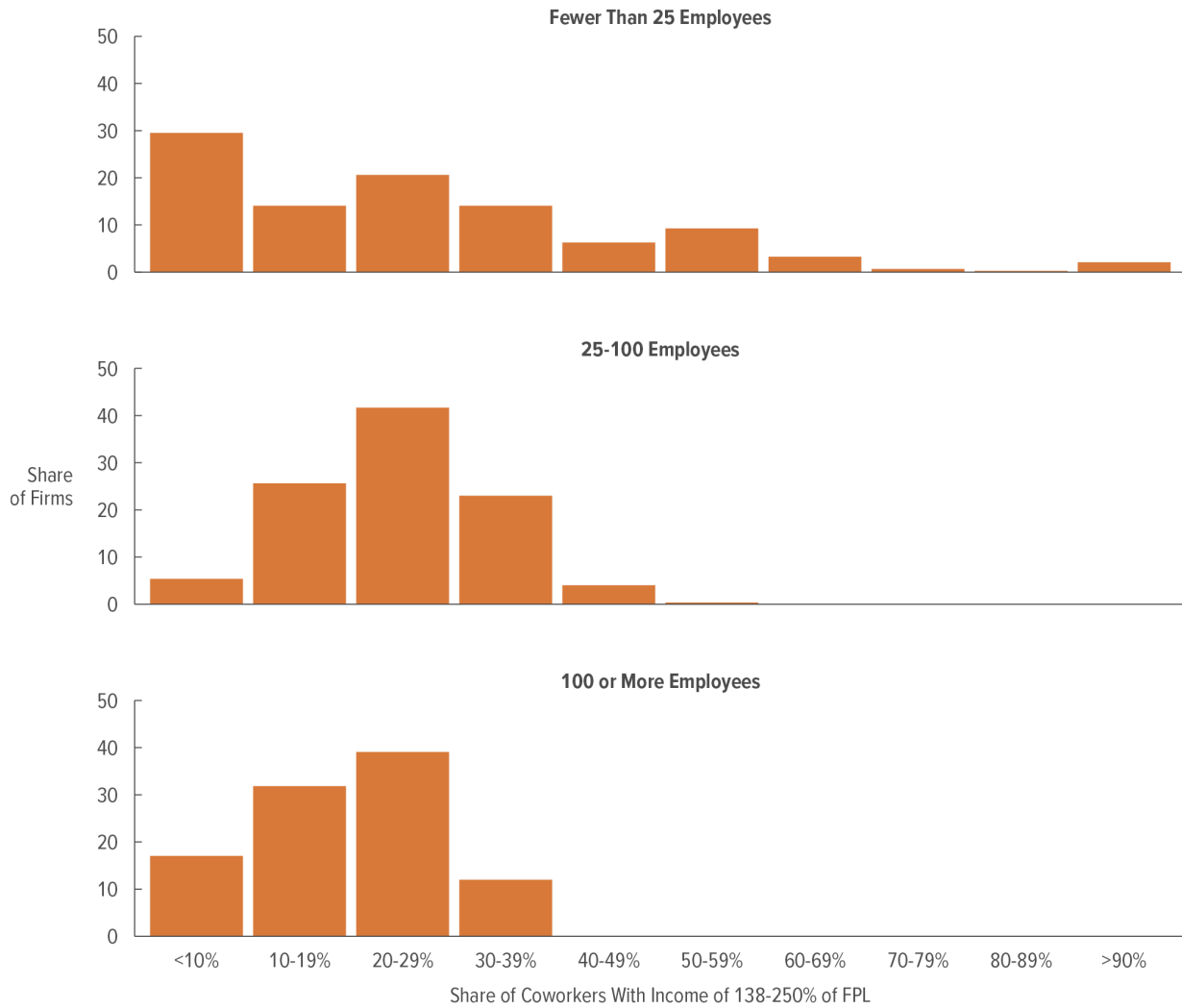
Data source: Congressional Budget Office.

Within each panel, the bars sum to 100.

FPL = federal poverty level; HISIM2 = CBO's health insurance simulation model.

Figure 11b

Share of Coworkers With Income Between 138 Percent and 250 Percent of the FPL in Different-Sized Firms in HISIM2, 2017



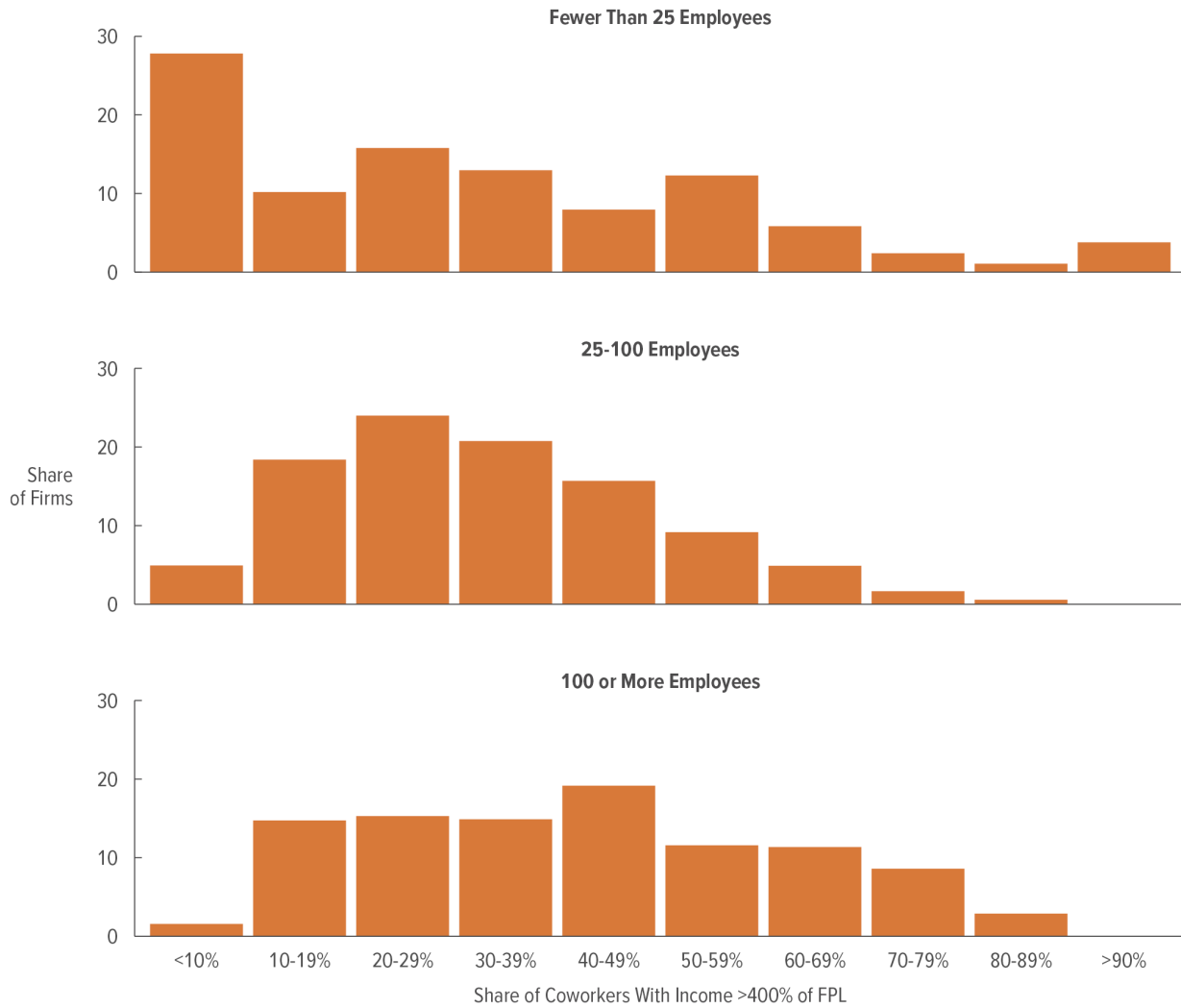
Data source: Congressional Budget Office.

Within each panel, the bars sum to 100.

FPL = federal poverty level; HISIM2 = CBO's health insurance simulation model.

Figure 11c

Share of Coworkers With Income Greater Than 400 Percent of the FPL in Different-Sized Firms in HISIM2, 2017



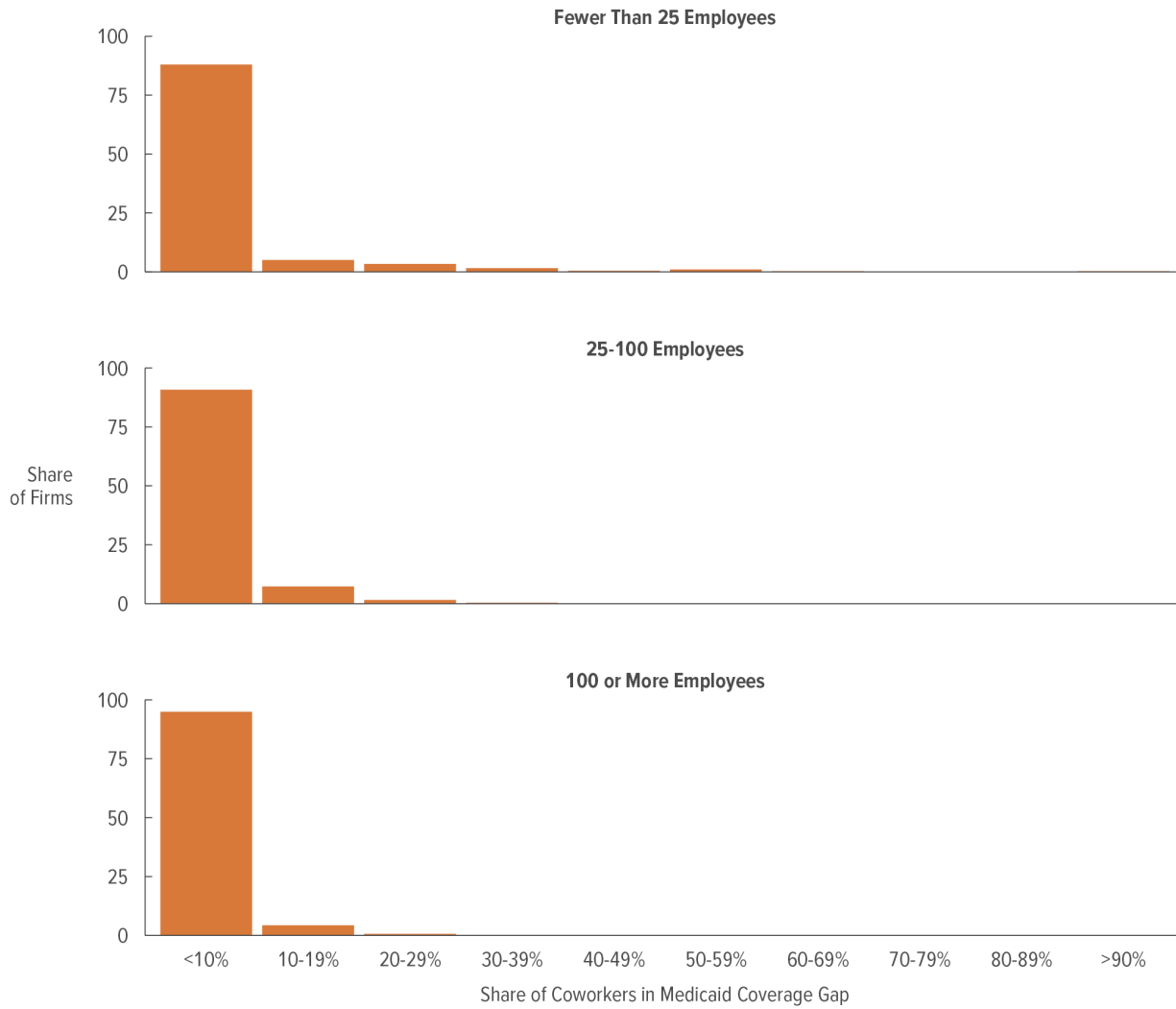
Data source: Congressional Budget Office.

Within each panel, the bars sum to 100.

FPL = federal poverty level; HISIM2 = CBO's health insurance simulation model.

Figure 11d

Share of Coworkers in the Coverage Gap for Medicaid in Different-Sized Firms in HISIM2, 2017



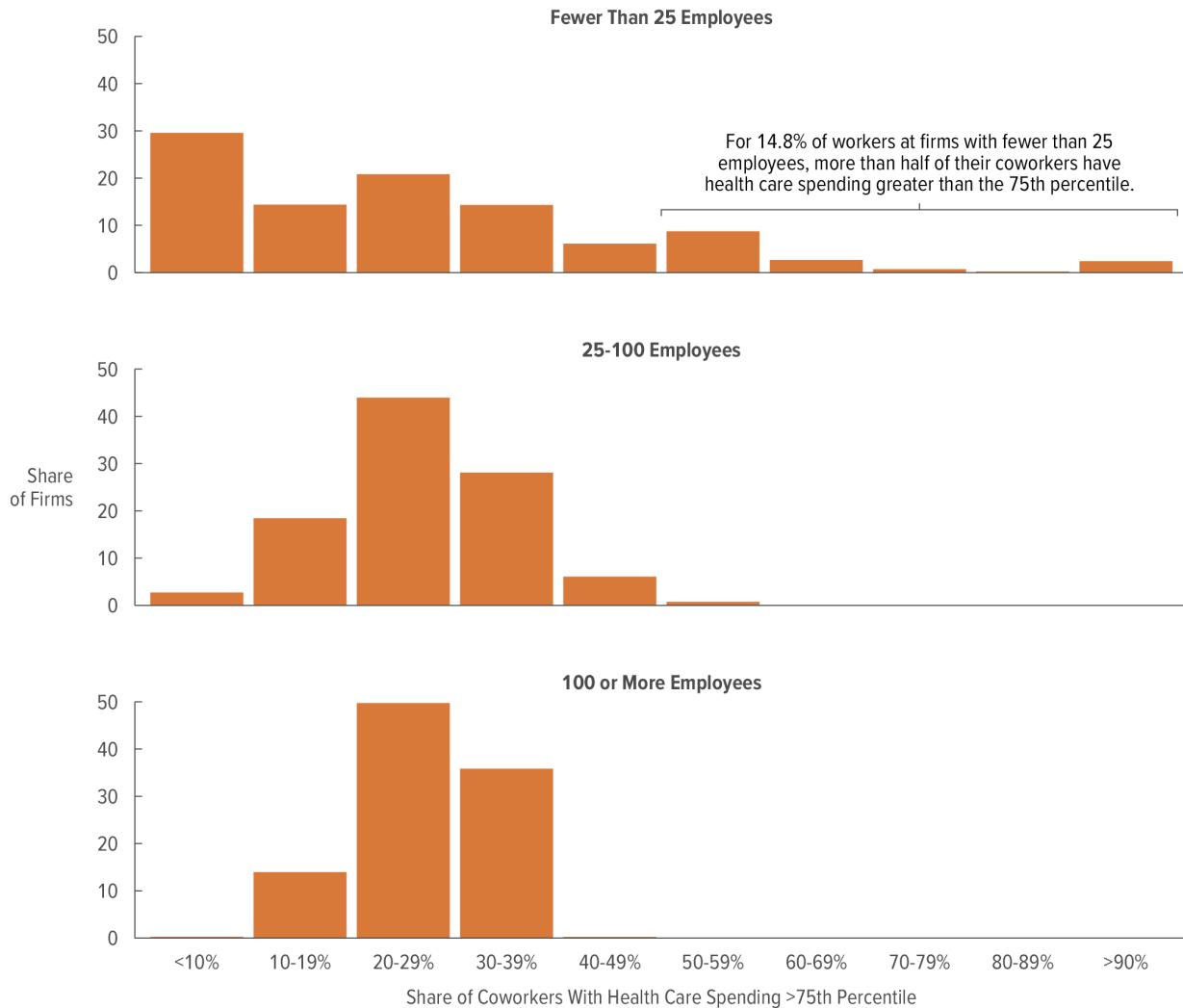
Data source: Congressional Budget Office.

Within each panel, the bars sum to 100.

HISIM2 = CBO's health insurance simulation model.

Figure 11e

Share of Coworkers With Health Care Spending Greater Than the 75th Percentile, in Different-Sized Firms in HISIM2, 2017



Data source: Congressional Budget Office.

The 75th percentile of health care spending is calculated across all workers in that firm size category.

Within each panel, the bars sum to 100.

HISIM2 = CBO's health insurance simulation model.

All in all, HISIM2's synthetic firms largely replicate the variation in the workforces observed in the tax-filing data. However, summary statistics of those synthetic firms do not exactly match those from the tax data for several reasons. First and probably most significantly, for some combinations of coworkers' traits, there are few or no available respondents in the CPS data meeting those criteria, which does not allow for perfectly filling in targeted distributions in HISIM2. As explained in the Methodology section, CBO uses a set of rules to prioritize traits to

choose coworkers when no coworker with the exact set of traits is available, but that approach of course yields a variation from the tax data. Second, although, as demonstrated above, the workforce characteristics of HISIM2 and the tax data are similar, they are not identical. Differences in the share of workers in a given group for firm size, offer status, age, and earnings can result in small differences in firms' characteristics, on average, once the shares are aggregated. In particular, given the dissimilarities between firms that offer health insurance and those that do not, differences between HISIM2 and the tax data in the share of firms offering coverage within subgroups (e.g., at firms of a certain size) result in overall discrepancies between HISIM2's synthetic firms and those in the tax data. Third, by definition, clustering firms eliminates a small amount of within-cluster variation. Finally, there is some small amount of random variation in the age or wages of specific coworkers chosen from a given age or wage group, which is defined by a range (e.g., ages 30 to 54 or wages of \$50,000 to \$69,999).

Simulated Change in Demand for an Offer at Synthetic Firms for Two Policies

The final set of analyses uses two policy tests—expanding Medicaid in all states and changing the age rating in the nongroup market—to demonstrate the strengths of modelling employers' behavior using a synthetic firm approach.

For the first policy test, CBO estimated the effect of having all remaining states expand eligibility for Medicaid to childless adults with income at or below 138 percent of the FPL, using the enhanced federal funding available under the ACA. The agency then calculated the change in employees' willingness to pay for employment-based insurance as a result of the policy using two different modelling methods.⁵¹ The first is a naïve approach that considers the change in willingness to pay for a single employee at the firm, the worker a synthetic firm is built around. The second uses HISIM2's current method of averaging the willingness to pay across all employees at the firm. Although comparing those worker-level and firm-level reactions is to some extent simplistic, it demonstrates the usefulness of a synthetic firm approach that mirrors real-world variation in the composition of firms. Modeling firms' behavior in offering insurance or not offering it using only a single worker may result in firms that are too responsive to a narrowly targeted policy that substantially affects only a small number of workers. Using a synthetic firm approach without sufficient variation in firm types may result in firms that are too unresponsive to such a policy that substantially affects only a small number of firms. By mimicking the composition of actual firms in HISIM2's synthetic firms, CBO is, in theory, able

⁵¹ Willingness to pay for employment-based health insurance is estimated as the difference in a worker's expected utility with that offer and his or her utility without the offer. In other words, it represents the maximum amount that worker would be willing to pay for that insurance offer. For more information on how CBO specifies expected utility, see Congressional Budget Office, "HISIM2: The Health Insurance Simulation Model Used in Preparing CBO's July 2021 Baseline Budget Projections" (July 2021), www.cbo.gov/publication/57205, and for more information on how the agency models firms' behavior, see Congressional Budget Office, "How CBO Models Firms' Behavior in HISIM2 in Its Baseline Budget Projections as of March 6, 2020" (March 2020), www.cbo.gov/publication/56303.

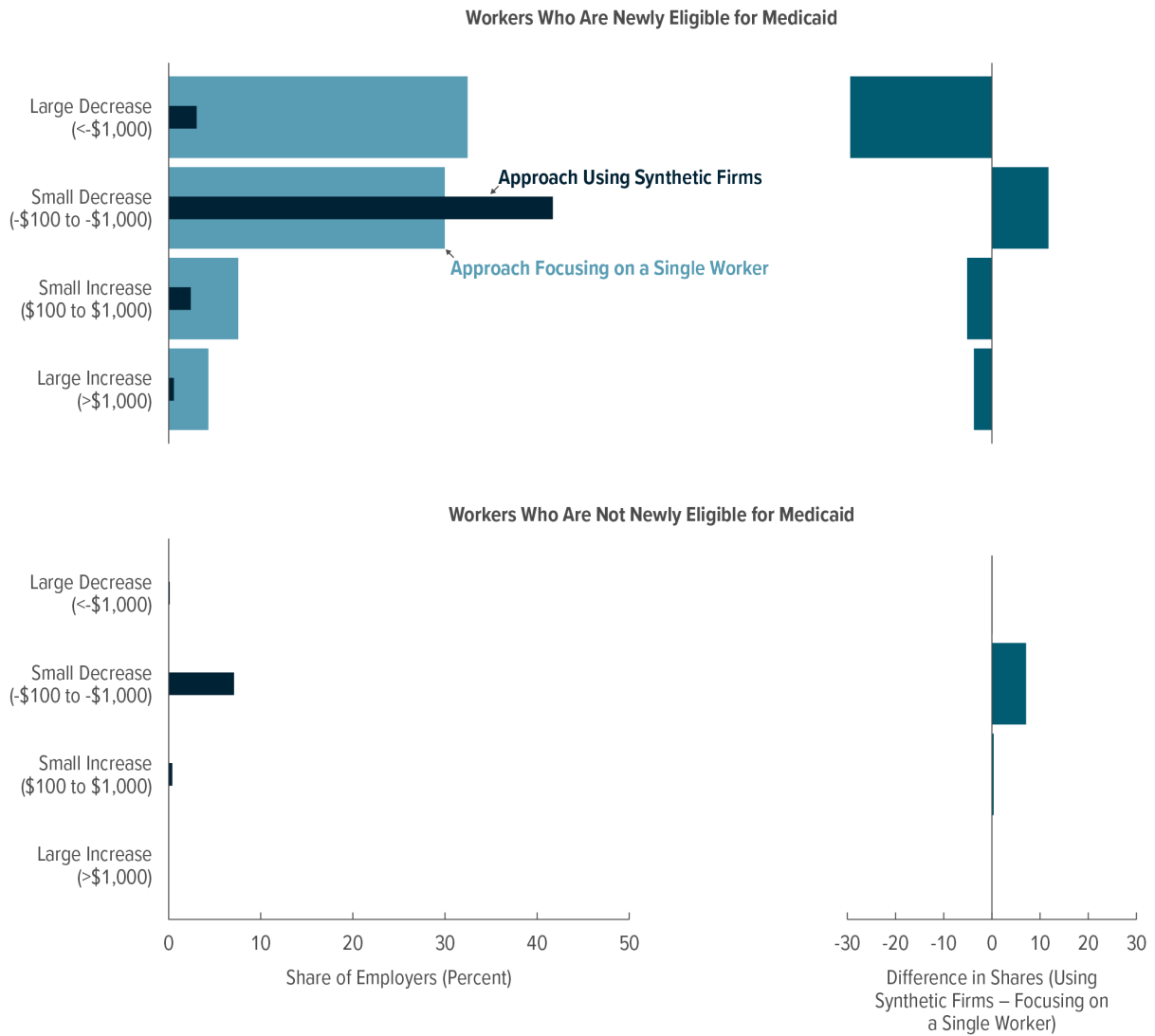
to more accurately model the broad array of effects a policy might have on the diverse set of firms that constitute the labor market.

Figure 12 displays the change in willingness to pay resulting from expanding Medicaid as specified using the naïve approach focusing on a single worker and HISIM2's synthetic firm approach. The effects of the policy are particularly concentrated among workers newly eligible for Medicaid (namely, those residing in nonexpansion states with income below 138 percent of the FPL). Under the approach focusing on a single worker, willingness to pay decreases substantially for a large share of such workers because they would now be able to enroll themselves and their families in Medicaid coverage. In HISIM2, a change in willingness to pay of that size would place substantial pressure on a firm that offered insurance to cease doing so. However, evaluating the average change in willingness to pay using all workers in a synthetic firm results in a significantly different distribution. While many workers in nonexpansion states with income below 138 percent of the FPL would experience large decreases in willingness to pay ($-\$5,000$ to $-\$1,000$), the average change at their firms would typically be more modest ($-\$1,000$ to $-\$100$) or essentially zero ($-\100 to $\$100$).⁵² Conversely, while there is almost no effect of the policy on the individual-level willingness to pay for those not newly eligible, many of those people work at firms with coworkers who are affected, which also would lead to a modest decrease in their firm-level willingness to pay. Overall, the naïve approach to evaluating the policy to expand Medicaid leads to the conclusion that the policy would have a large effect on a small group of firms, while the synthetic firm approach reveals that the policy would have a more modest effect on a larger group of firms.

⁵² A small share of workers experience an increase in willingness to pay following Medicaid's expansion. Under current law, those workers have income between 100% and 138% of the FPL and live in a nonexpansion state, making them eligible for generous marketplace subsidies if their employer does not offer affordable coverage. As a result, their willingness to pay for employment-based insurance is negative (i.e., they would like to pay their employer not to offer insurance) because an employer's affordable offer would block them from that subsidized marketplace coverage. If their state expanded Medicaid eligibility, they would no longer be eligible for those subsidies but could enroll in Medicaid regardless of whether their employer offered insurance. Consequently, their willingness to pay for that employment-based insurance shifts from strongly negative to weakly positive.

Figure 12

Change in Willingness to Pay for Insurance Following an Expansion of Medicaid, Using an Approach Focusing on an Individual Worker and One Using Synthetic Firms



Data source: Congressional Budget Office.

The dollar values in parentheses indicate how much less or more a worker would be willing to pay for employment-based insurance than the person currently pays.

The shares of employers where the change in willingness to pay is between -\$100 and \$100 are not shown. They can be derived by subtracting the shares of employers for which there is a large decrease, small decrease, small increase, or large increase from 100%. In the case of workers who are not newly eligible, for example, the change in willingness to pay is between -\$100 and \$100 for 99% of employers under an approach focusing on an individual worker.

Shown in Figure 13, the second policy test uses an identical approach to study the effects of a change that would allow the price of the most expensive age-rated nongroup premium in an ACA marketplace to be as much as five times that of the least expensive, whereas under current law, the price can be just three times as much. While the effects would vary by state, CBO predicts that, in 2022, such a policy would typically result in lower reference premiums for enrollees younger than 47 and higher reference premiums for older enrollees.

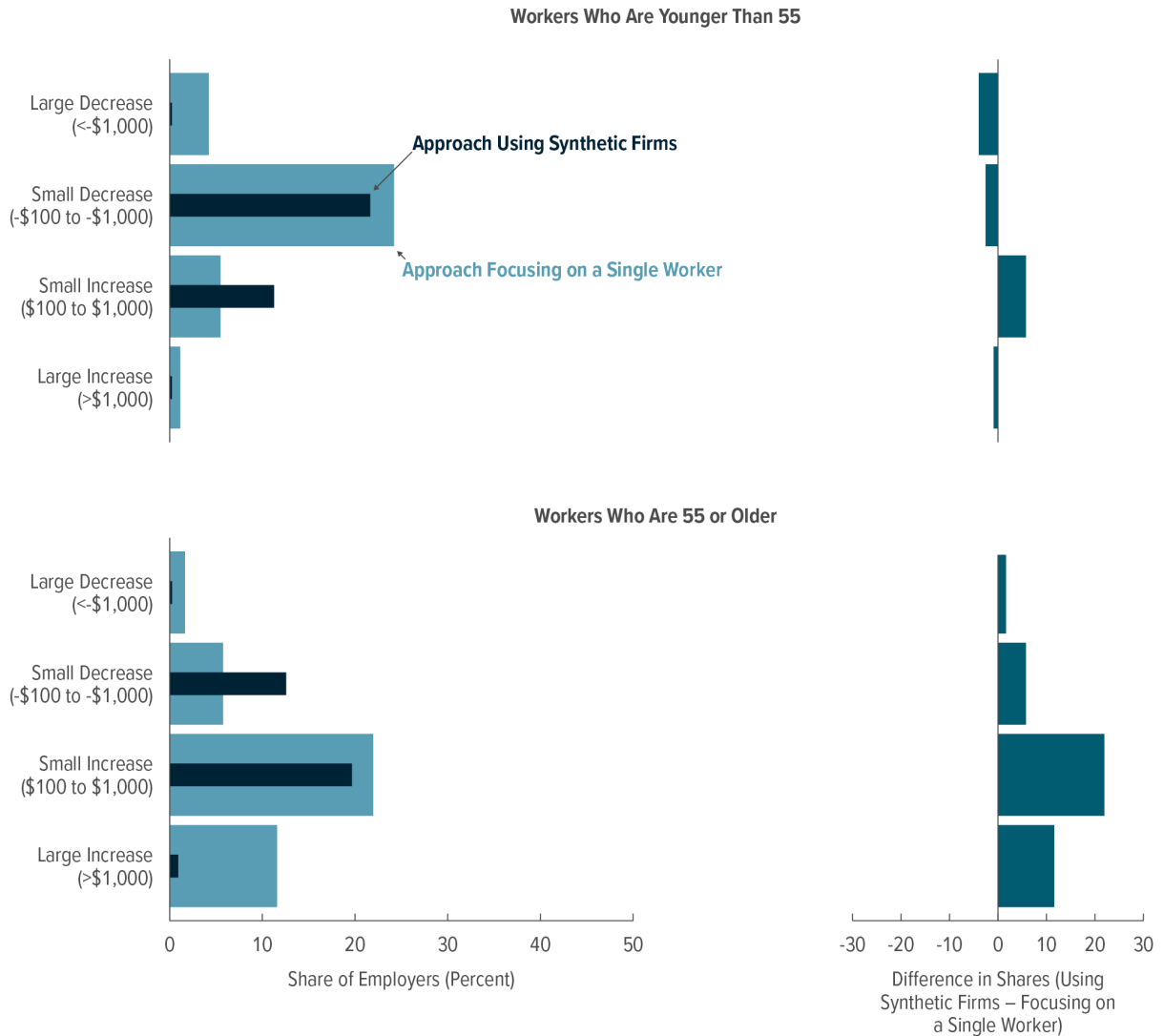
Under an approach that focuses on a single worker, the effects of the change on his or her willingness to pay are heterogeneous. Many younger workers would be less willing to pay for employment-based insurance because purchasing coverage in the nongroup market would now be cheaper.⁵³ Conversely, many older workers would face higher premiums in the nongroup market and would be more willing to pay for employment-based insurance.

However, under an approach that uses synthetic firms, the effects on the average firm-level willingness to pay are much more homogenous. The majority of firms would see little net change in willingness to pay, as the effects of lower nongroup premiums for younger workers and higher premiums for older workers would largely cancel each other out. For those firms that do experience a nontrivial change in willingness to pay, it is overwhelmingly likely to be a small one, in the range of \$100 to \$1,000. Further, more younger workers experience a small increase in willingness to pay driven by their older coworkers' facing higher premiums, while, conversely, more older workers experience a decrease in willingness to pay due to the effects on their younger coworkers. While some individual workers' willingness to pay would change substantially under the policy, most firms would see little change in the average willingness to pay because their employees are not concentrated in any specific age group.

⁵³ A worker's willingness to pay for employment-based insurance depends on the expected health care spending of his or her entire health insurance unit. As a result, a young worker's willingness might increase under the policy being analyzed if his or her health insurance unit included older members.

Figure 13

Change in Willingness to Pay for Insurance Following a Policy Allowing More Expensive Nongroup Premiums, Using an Approach Focusing on an Individual Worker and One Using Synthetic Firms



Data source: Congressional Budget Office.

The shares of employers where the change in willingness to pay is between -\$100 and \$100 are not shown. They can be derived by subtracting the shares of employers for which there is a large decrease, small decrease, small increase, or large increase from 100%.

Appendix A: Data Sources for the Tax-Filing Data Set

Data Source	Description	Use
Social Security Administration Data Master File	The database of Social Security numbers maintained by the Social Security Administration, combined with applications for individual taxpayer identification numbers compiled by the Internal Revenue Service	To identify employees' ages
Internal Revenue Service Entity File	A data set constructed by the Internal Revenue Service, combined from business-level tax filings and auditors' examinations. It contains a variety of information, including industry and filing requirements.	To identify employers' industries
Form 1040	The standard individual income tax form	To infer employees' marital status
Form 1095-B	A form sent by health insurers to enrollees to verify health insurance coverage	To identify employers that offer and employees who are eligible for employment-based insurance
Form 1095-C	A form sent by applicable large employers to employees to report offers of health insurance coverage and enrollment in self-insured plans	To identify employers that offer and employees who are eligible for employment-based insurance
Form 941	A form filed quarterly by employers to calculate their payroll tax liability, report their payroll tax and individual income tax withholding deposits, and claim certain credits	To identify employers and their number of employees
Form W-2	A form, sent by employers annually to their employees, that reports wage payments and certain other payments from employers to employees, including employers' contributions to health insurance premiums	To connect employees with employers, to identify employees' earnings, and to identify employers that offer and employees who are eligible for employment-based insurance