

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

December 2, 2021

Presentation to the  
International Microsimulation Association

Ben Hopkins  
Health Analysis Division

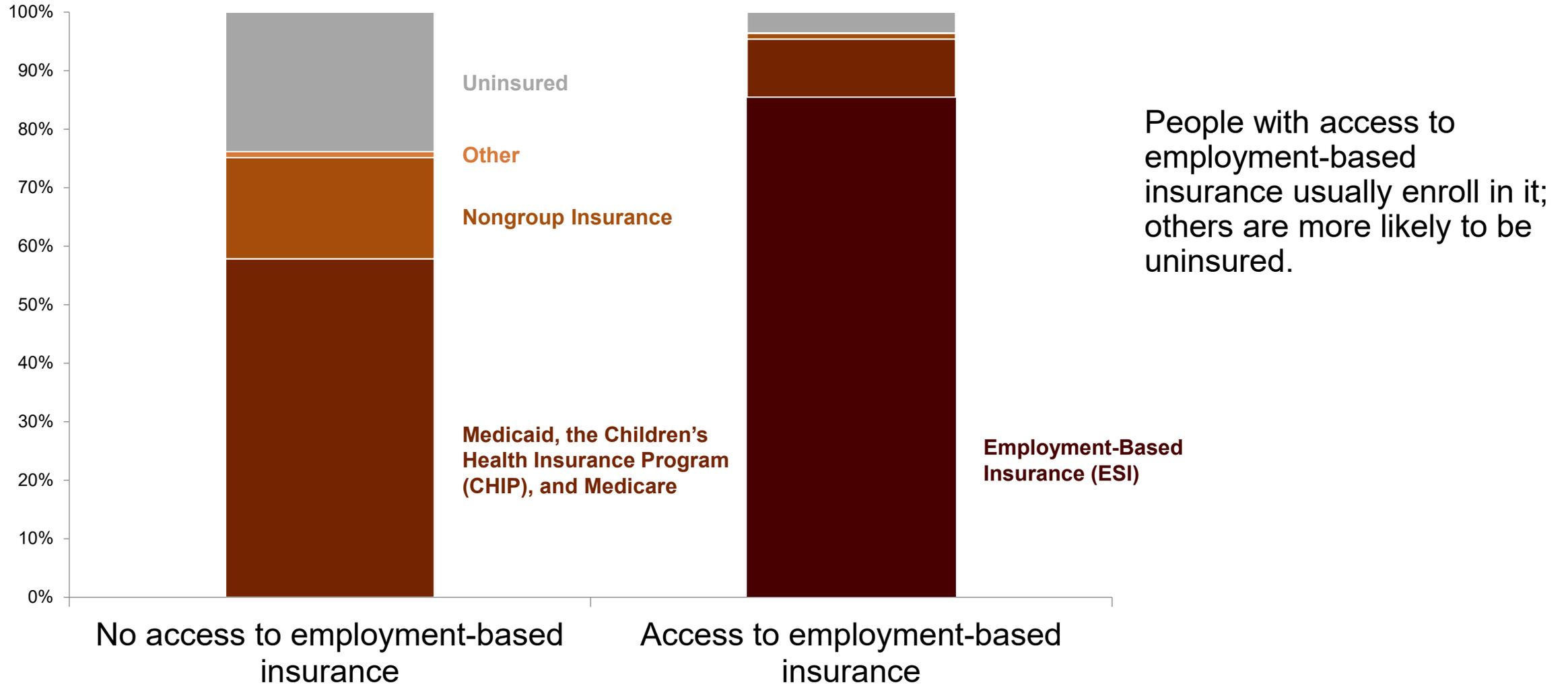
The U.S. Congressional Budget Office (CBO) produces independent analyses of budgetary and economic issues to support the Congressional budget process.

CBO uses a health insurance microsimulation model, HISIM2, to generate estimates of coverage, premiums, and budgetary costs for the nonelderly.

HISIM2 models **household decisions** between different sources of health insurance coverage and **employer decisions** to offer health insurance or not.

The model of whether employers offer their employees health insurance has a large effect on CBO's **coverage and budgetary projections**.

# Insurance Coverage of People Under Age 65, 2017



In theory, an individual worker's access to employment-based insurance (ESI) depends on how much the person and his or her coworkers cost to insure and how much they demand insurance.

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.

In firms with more than 50 workers, gross premiums depend upon the collective health care spending for workers.

HISIM2 is based on nationally representative, individual-level data from the Current Population Survey (CPS), which does not include any information on respondents' coworkers.<sup>1</sup>

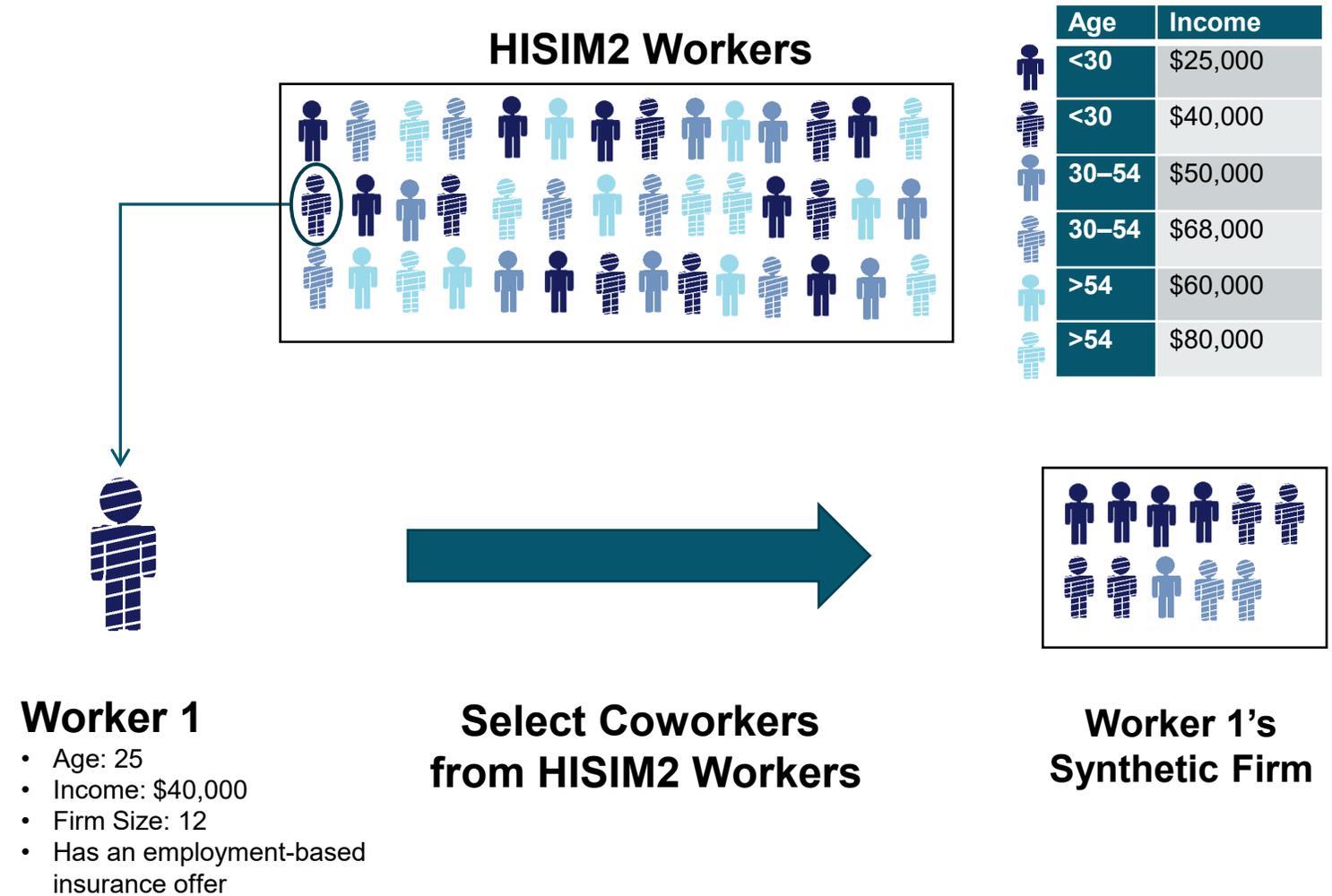
There are other data sets that link employees to firms and include information on employees' age, income, marital status, residence, eligibility for employment-based insurance, and health spending.<sup>2</sup>

CBO creates **synthetic firms** by randomly selecting coworkers for each CPS respondent. Those firms mimic the composition of actual firms in those other data sets.

<sup>1</sup> 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](http://www.cbo.gov/publication/55087).

<sup>2</sup> Those data sets lack sufficient information on all members within households, people who don't file tax returns, hours worked, lack national representation, and other factors that are essential to estimating insurance coverage.

# Simplified Example of How Synthetic Firms are Created in HISIM2



Each HISIM2 worker is given a synthetic firm composed of other HISIM2 workers.

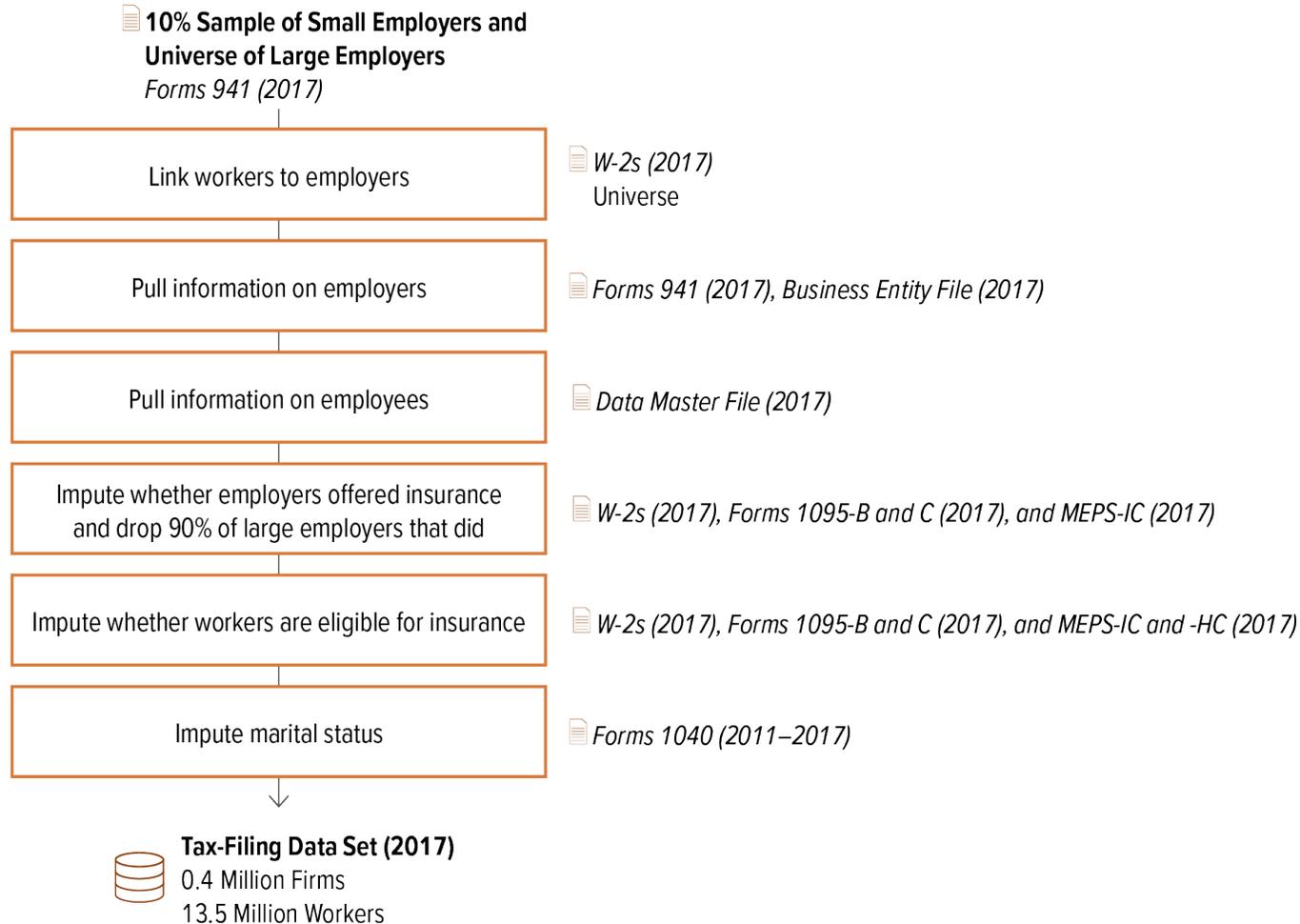
This graphic presents a simplified process. In HISIM2, coworkers are also selected on the basis of marital status, health spending, and state of residence.

# Data



CBO and the staff of the Joint Committee on Taxation (JCT) collaborated to create a novel data set with information on employee age, income, marital status, residence, and eligibility for employment-based insurance within and across firms. That data set is based on **employers' and employees' tax filings** and the Medical Expenditure Panel Survey Insurance Component (MEPS-IC) and Household Component (MEPS-HC).

# Assembly of the Tax-Filing Data Set



To create the tax-filing data set, CBO and JCT matched employers' tax returns (Form 941) with employees' wage and salary information returns (W-2).

Additional information about employees was provided by supplemental administrative sources (IRS Business Entity File, Social Security Administration's Death Master File, 1040s, 1095-Bs, and 1095-Cs).

CBO used data from a proprietary data set on the **health spending** of a sample of enrollees in employment-based insurance (ESI) plans to measure the correlation in the health spending of ESI policy holders that are employed by the same firm.



# Methods

CBO focused on three goals when developing its method for constructing synthetic firms.

1. The traits used to select coworkers for synthetic firms should be strongly correlated with demand for employment-based health insurance.
2. The method should produce realistic distributions of those traits within firms.
3. The method should produce realistic heterogeneity among firms with respect to those traits.

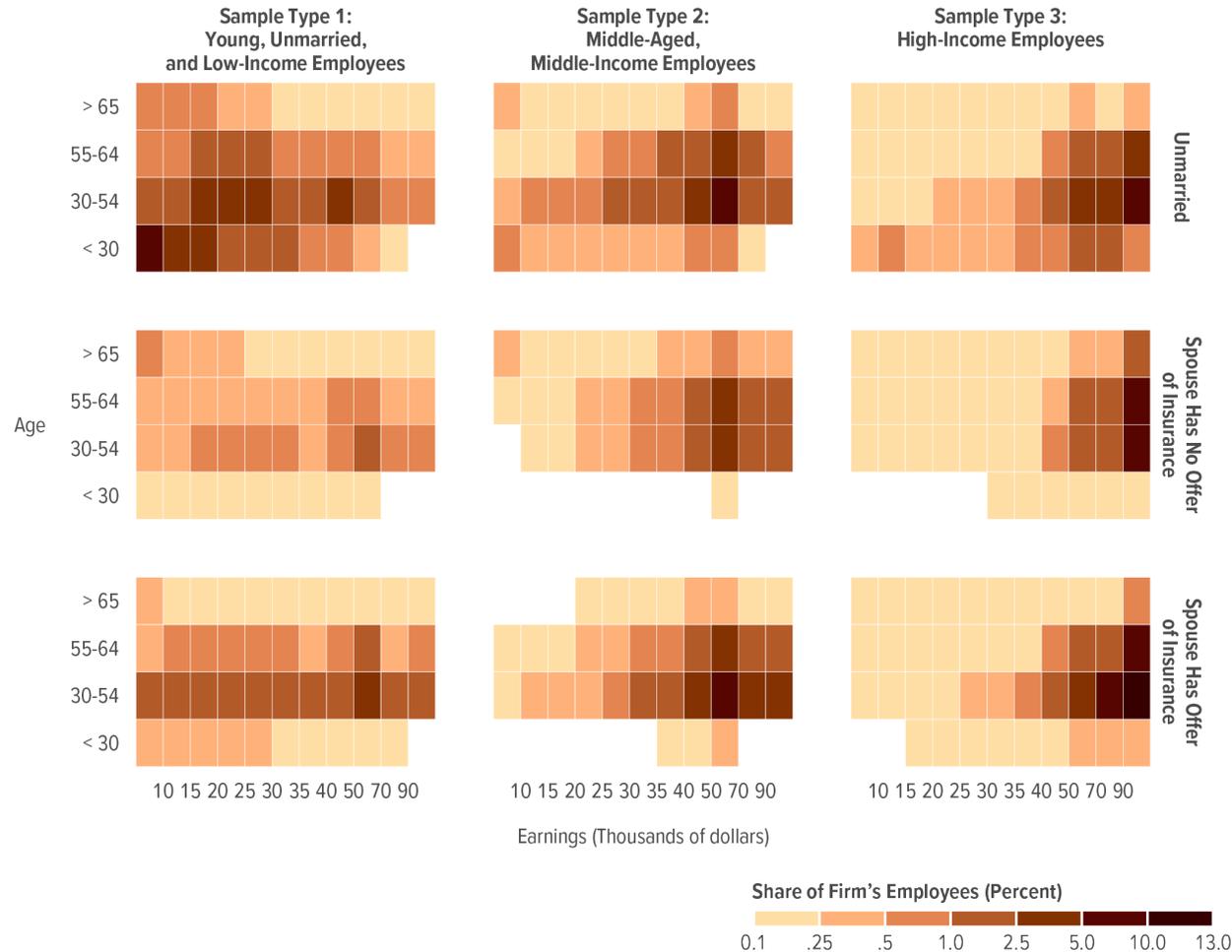
CBO used the following traits to create synthetic firms: **age, earnings, marital status, health spending, state of residence**, and **eligibility** to receive an offer of employment-based insurance.

To ensure that synthetic firms mimicked actual firms, CBO quantified the relationships among those traits of individuals and their coworkers in the tax-filing and health spending data sets and imposed those relationships on synthetic firms.

In practice, CBO constructed synthetic firms for each of worker in the CPS in three steps:

1. The agency classified employers in the tax-filing data set into different **types of firms**.
2. It specified the **desired composition** of workers' synthetic firms.
3. It selected CPS respondents with the **appropriate traits** to populate each synthetic firm.

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



CBO classified employers in the tax-filing data set into different types of firms that are characterized by **distinct joint distributions** of employees' age and earnings.

The agency used the **k-means clustering algorithm** to identify different types of firms.

Each synthetic firm mimics one of the many types of actual firms identified in the tax-filing data set

CBO imposed the relationships among a specific set of traits of individuals and their coworkers in the tax-filing and health spending data sets on synthetic firms.

For example, the agency specified the age, earnings, and marital status of individuals' coworkers on the basis of their own age, earnings, and marital status, as well as the size of their firm and whether or not they were offered health insurance according to the distribution of the same at actual firms of the same type in the tax-filing data set.

The agency assigned the health spending (based on the health spending data set) and state of residence of coworkers in a similar fashion.

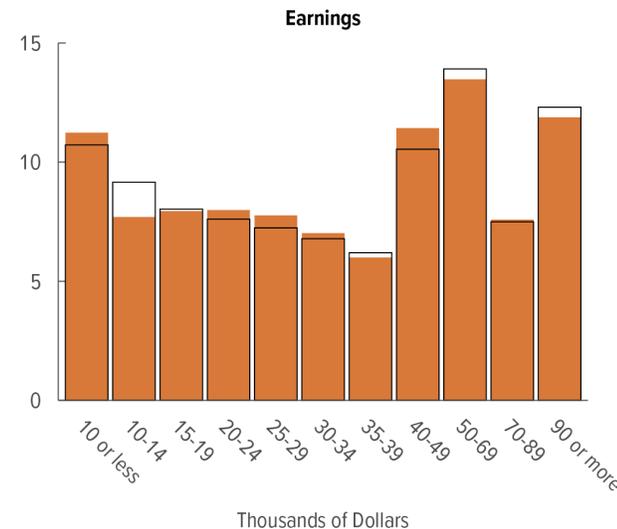
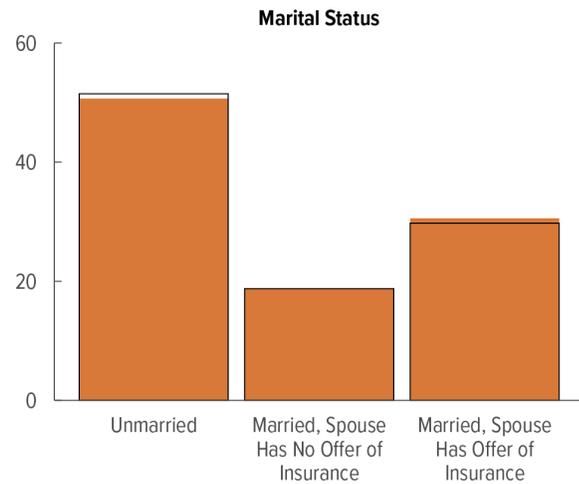
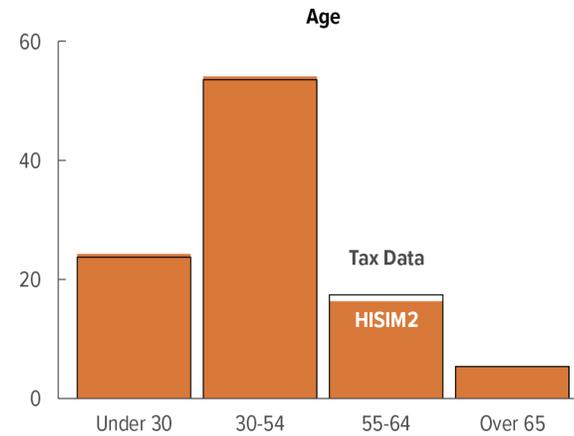
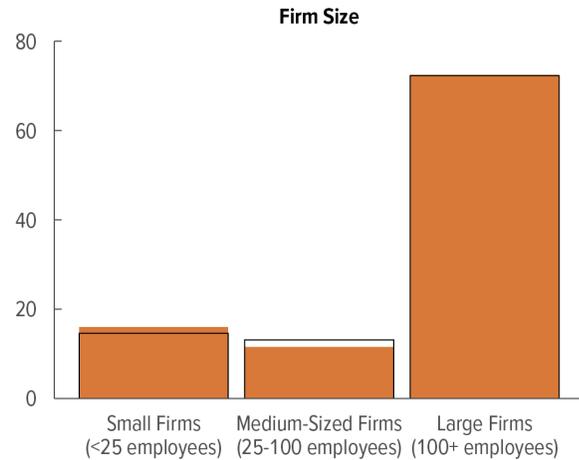
To populate a synthetic firm around the individual worker, CBO randomly drew people from the CPS who had the desired age, earnings, marital status, residence, and health care spending categories and who worked for firms of the same size that similarly did or did not offer health insurance.

# **Assessment of the Quality of HISIM2's Synthetic Firms**

CBO used three main groups of analysis to evaluate the data and methods used to construct HISIM2's synthetic firms.

1. The agency compared **the distribution of traits** used to select synthetic coworkers **among workers** in HISIM2 and workers in the tax-filing data.
2. It compared summary statistics of **the distribution of coworkers' traits across firms**, conditional on a worker's own traits in HISIM2 and workers in the tax-filing and health spending data sets.
3. It compared the **simulated effects of policies** on employees' demand for an offer of health insurance from their employers on the basis of (i) the preferences of a single worker at a firm and (ii) the preferences of the entire synthetic firm's workforce.

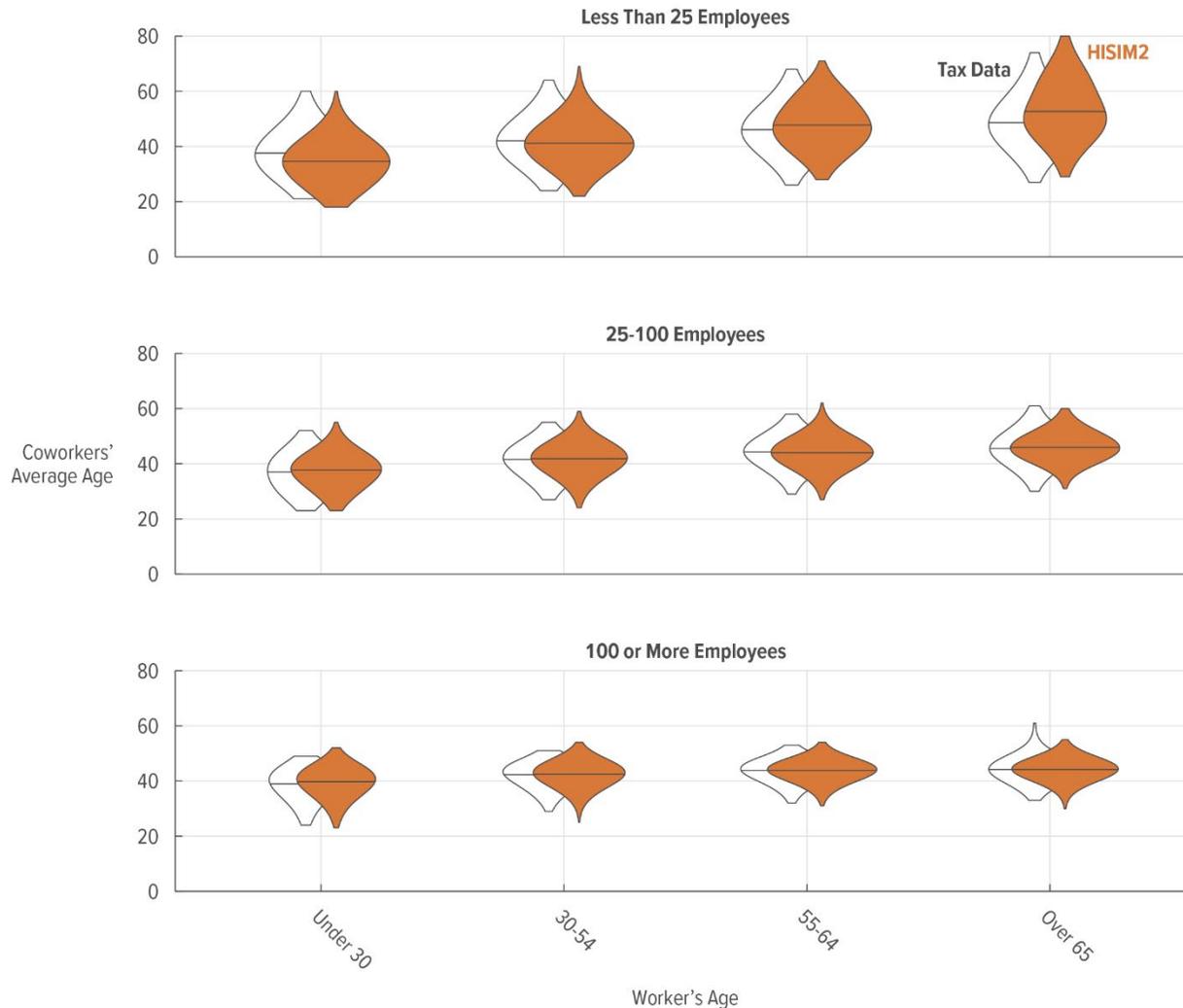
# Comparison of Key Characteristics of Workers in HISIM2 and Tax Data, 2017



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.

The distribution of firms and workers are broadly similar in HISIM2 and the tax-filing data set by age, wage, marital status, and firm size.

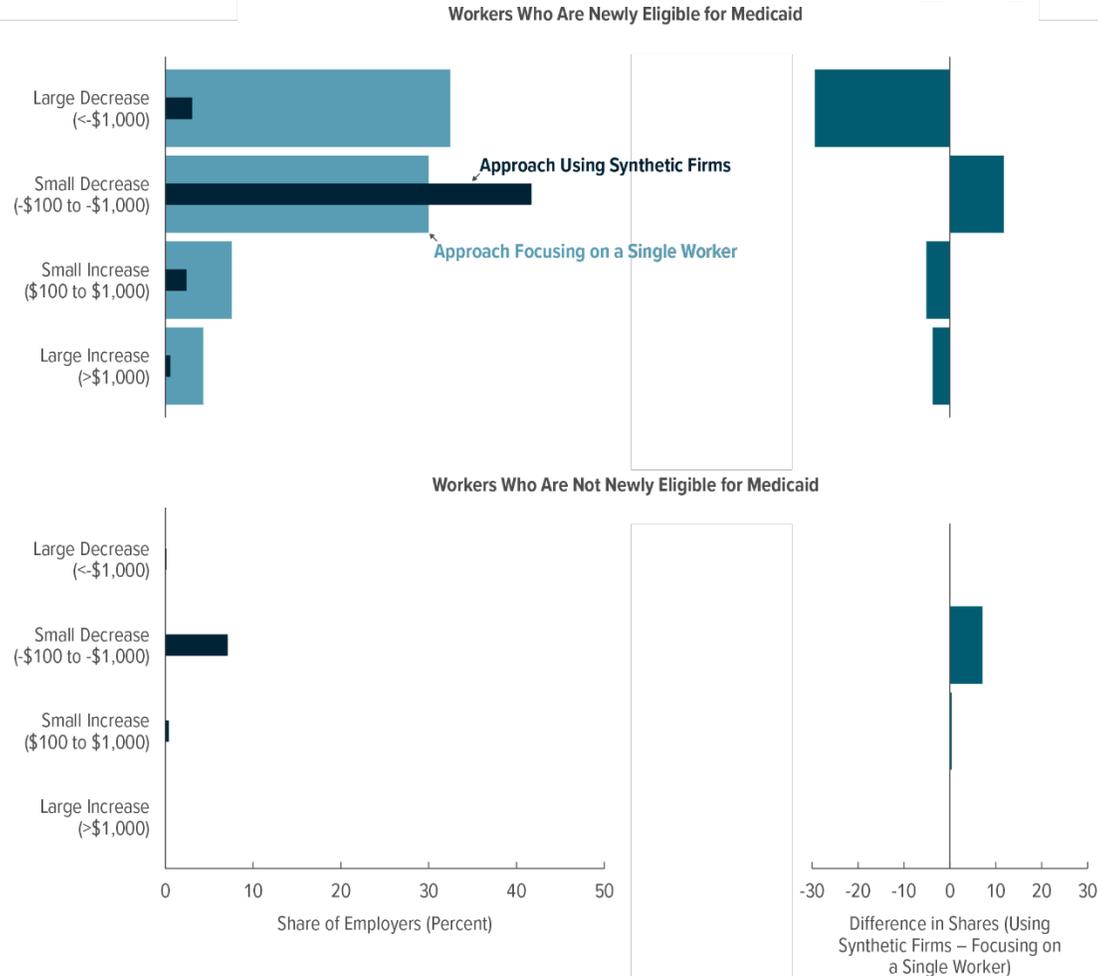
# Distribution of Coworkers' Average Age, by Worker's Age, in Different-Sized Firms in HISIM2 and Tax Data, 2017



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.

CBO conducted similar analyses for earnings, marital status, and health spending.

# 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



Modeling firms' behavior 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 a policy that substantially affects only a small number of firms.

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 in which 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 whom 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.