HOOVER INSTITUTION
ECONOMICS WORKING PAPERS

A Technical Overview of the Collection of Health Expenditures and Insurance (CHEI) Data

Tom Church and Daniel Heil

Economics Working Paper 19115

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010

December 2, 2019

This documentation provides an overview of the Collection of Health Expenditures and Insurance (CHEI) data. The CHEI is composed of annual microdata projections derived from recent federal surveys. The data allow for projections of US health insurance enrollment, health expenditures, eligibility for federal health subsidies, tax expenditures for employer-sponsored health insurance (ESI), and several other health-related data. This technical overview provides a description of the data sources and a description of the major imputations and calculations used to construct the datasets.

The authors are research fellows at the Hoover Institution and they welcome comments, questions, and suggestions. They may be reached at tvchurch@stanford.edu and dlheil@stanford.edu.

A Technical Overview of the Collection for Health Expenditures and Insurance

This documentation details the construction and design of the Collection of Health Expenditures and Insurance data (CHEI). The CHEI is composed of annual microdata projections derived from recent federal surveys. The data allow for projections of US health insurance enrollment, health expenditures, eligibility for federal health subsidies, tax expenditures for employer-sponsored health insurance (ESI), and several other health-related data. The CHEI is mainly intended to estimate health-related data for non-seniors, but the data include senior populations. CHEI data may be used to estimate the long-term fiscal effects of proposed reforms to federal healthcare programs.[1] Among other uses, the CHEI data is used to construct cost estimates for policy options in the Hoover Institution's America Off Balance Budget Calculator.[2]

CHEI data is intended to closely match the health statistics in Congressional Budget Office (CBO) reports. Consequently, it follows the methodology used to build CBO's Health Insurance Simulation Model (HISIM2).[3] Due to data limitations and the limited purpose of the data we make several simplifying assumptions relative to CBO's methodology; these are discussed below.

This document begins with an overview of the data used. Section II then explains the major imputations and calculations made to construct the initial-year dataset. Section III explains how the initial-year dataset is used to project future years. We conclude with a discussion of major limitations with the data and a discussion of planned additions.

I. Data Overview

The primary dataset used for the CHEI data is the Annual Social and Economic Supplement of the Current Population Survey (CPS).[4] The CPS is a household survey conducted in March of each year. It includes demographic, income, and health variables for the previous calendar year. Currently, CHEI data use the 2018 CPS with most data corresponding to the 2017 calendar year. We use CPS demographic and income variables to build the initial-year dataset.[5]

Demographic adjustments are made to ensure variables correspond to the previous calendar year. For example, we adjust the ages of approximately one quarter of the sample to account for respondents with birthdays between January and March. We also impute demographic information that affects eligibility for public health care programs. This includes imputations for

---

[1] We provide a general overview of the CHEI data here. This does not fully describe the methodology in building the dataset. The code is available upon request.

[2] The American Off Balance Budget Calculator assesses the long-term fiscal impact of various fiscal reforms to the US federal budget. The calculator is available at americaoffbalance.org/calculator.

[3] Congressional Budget Office (CBO) (2019). *Sources and Preparation of Data Used in HISIM2—CBO's Health Insurance Simulation Model: Working Paper 2019-04*. Available at https://www.cbo.gov/publication/55087.

[4] We extract the 2018 CPS microdata from the National Bureau of Economic Research (NBER). The raw data are available here https://data.nber.org/data/current-population-survey-data.html.

[5] CPS income variables suffer from widely known undercounts. This includes underestimates for self-employment income and investment income. CBO's HISIM2 data includes imputations for self-employment income and realized capital gains using tax records. Lacking recent filer microdata, we do not attempt these imputations. As a consequence, our data will likely underestimate tax liabilities, marginal tax rates, and tax expenditures for employer-sponsored health insurance to a small extent.

whether a respondent was pregnant during the previous calendar year and whether a respondent is lawfully present in the United States.[6]

We impute several additional variables to the original CPS data using other surveys and publicly available reports. We estimate initial tax liabilities using the National Bureau of Economic Research TAXSIM calculator.[7] We use the 2017 Medical Expenditure Panel Survey Household Component (MEPS-HC) to estimate health expenditures and health insurance status.[8] Employer-sponsored insurance premiums and related variables are primarily imputed using publicly available statistics from the MEPS Insurance Component (MEPS-IC). Finally, we use the Centers for Medicare and Medicaid Service's (CMS) Public Use Files for the Affordable Care Act's (ACA) Marketplace Exchange to estimate state-specific ACA premiums. We discuss our methods for imputing data from these sources for the initial-year dataset in section II.

After the initial-year dataset is completed, we construct annual projections. Sample weights of the initial-year dataset are adjusted to reflect US Census Bureau population and demographic projections.[9] We then make further adjustments to sample weights to reflect changes in health insurance status. Health spending and premiums follow National Health Expenditure (NHE) and CBO growth assumptions. Income-related variables are adjusted to reflect CBO's budget and economic outlooks. We explain the projection assumptions in section III.

II. Data Imputations

The CPS includes questions covering tax liabilities, health insurance status, insurance premiums, and health spending. These data, however, suffer from well-known undercounts and other issues that limit their usefulness. We thus supplement these data with other surveys and administrative sources. This section explains these imputations.

Tax Related Variables

The CHEI data estimates tax liabilities and marginal tax rates. Among other uses, we use estimates of filers' marginal tax rates to estimate the tax value of employer-sponsored insurance plans (ESI). We use tax liabilities to estimate the revenue and outlay effects of the ACA's premium subsidies.

Estimating tax liabilities begins with assigning all respondents to a tax unit composed of members of a family expected to file a single tax return. Each tax unit is assigned a taxpayer identification variable. Married respondents are assigned to the same tax unit; unmarried adults and dependents with incomes above IRS filing thresholds are assigned to their own tax unit. We then construct a series of income and demographic variables that affect tax liabilities.

---

[6] Consistent with CBO, we impute immigration status using the procedure in Borjas, George, (2017). "The labor supply of undocumented immigrants," *Labour Economics, 46*. Pages 1-13. Available at https://doi.org/10.1016/j.labeco.2017.02.004.

[7] We use the Stata package TAXSIM27 to complete this analysis.

[8] We use the 2017 MEPS-HC Full Year Consolidated Data file. Agency for Healthcare Research and Quality (August 2019). *MEPS HC-201 2017 Full Year Consolidated Data File.* Available at https://meps.ahrq.gov/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-201.

[9] We use the Stata raking package ipfraking for all weighting adjustments. See Kolenikov, Stanislav, (2014). "Calibrating survey data using iterative proportional fitting (raking)," *Stata Journal 14*. Pages 22-59.

We calculate tax variables in two ways. First, we use the National Bureau of Economic Research's TAXSIM calculator.[10] TAXSIM provides a detailed breakdown of federal and state tax liabilities, but is currently only capable of making estimations through 2023. Second, since we are projecting tax liabilities beyond 2023, we also calculate separate tax estimates using a more basic tax calculator created for the CHEI data. Our basic tax calculator calculates tax liabilities before accounting for tax credits. It uses separate thresholds and deduction amounts for single filers, joint filers, head-of-household filers, and filers who are claimed as a dependent on another tax return.

Beyond our omission of tax credits, the biggest difference between our tax calculator and TAXSIM is the treatment of state taxes. TAXSIM calculates state income taxes paid and assumes filers will itemize their deductions in cases where a taxpayer's state and local tax deductions would exceed the filer's standard deduction.[11] In contrast, our calculator does not include state income taxes and assumes all filers will claim the standard deduction. These assumptions mean our tax calculator's estimates will overstate tax liabilities. Marginal tax rates will also be overestimated for filers who would fall to lower tax brackets after accounting for itemized deductions. Despite these limitations, our tax calculations generally match TAXSIM's estimates once we exclude state tax effects.[12]

Since our tax calculator does not calculate state taxes, we use TAXSIM's state marginal tax rates to calculate a respondent's tax savings from ESI coverage. Beyond 2023, we assume a filer's state marginal tax rate remains at its 2023 levels.[13] We discuss tax projections in more detail in section III.

Health Insurance Status Imputations

Several health insurance status variables are available in the CPS. These include enrollment in Medicare, Medicaid, military health care, private group health insurance, and private individual health insurance. For private health insurance, variables exist indicating whether a respondent is the policyholder of the plan or is a dependent on another respondent's plan. Generally, we use these CPS variables to determine enrollment status. The CPS health insurance variables, however, are inconsistent with other surveys and administrative sources.

To ensure enrollment totals match administrative sources, we adjust the CPS sample weights to reflect enrollment totals in MEPS-HC. We use a raking technique to adjust enrollment totals while maintaining original CPS distributions for sex, race, age, and poverty status.[14]

---

[10] We use the Stata package TAXSIM27 to run TAXSIM. For more details see Feenberg, Daniel R., and Elizabeth Coutts, (1993). "An introduction to the TAXSIM model," *Journal of Policy Analysis and Management, 12(1)*, pages 189-194. Available at http://users.nber.org/~taxsim/taxsim27/.

[11] Note, this is a smaller issue for tax years 2018 through 2026 because the Tax Cut and Jobs Act of 2017 temporarily limited state and local tax deductions to $10,000 and increased standard deductions above those thresholds.

[12] To test our tax calculator, we re-ran TAXSIM without assigning a state to any respondent. Our pre-credit tax liability numbers matched TAXSIM's estimates.

[13] State tax calculations in TAXSIM for 2018 and beyond use 2017 state tax laws. Specifically, TAXSIM authors state "State tax is 1977 through 2017 with 2018+ calculated using the 'real' value of the 2017 law."

[14] We use the original categorical CPS poverty variable (POVLL). It divides families into 14 poverty level groups ranging from below 50% of the federal poverty line to above 500% of the poverty line. The race variable has five categories: Hispanics; non-Hispanic whites, black, Asians; and an other category.

Beyond the weighting adjustments, we impute whether respondents who report non-group insurance coverage purchase their insurance through the ACA's Marketplace Exchange. Using MEPS-HC, we estimate the likelihood that someone with non-group coverage participates in the exchanges. Limiting the sample to non-senior respondents who report non-group coverage, we regress a variable indicating exchange participation on several covariates that exist in both the CPS and MEPS-HC surveys. The covariates include poverty status, health status, and several demographic variables. We then fit the regression output on non-senior CPS respondents who report individual coverage to predict the probability that they will participate in the exchange. We adjust probabilities to ensure members of a household reporting individual coverage have a similar likelihood for participating in the exchanges. Exchange enrollment is then assigned to those with probabilities exceeding age-specific thresholds. The age-specific thresholds ensure a similar age distribution to the actual 2017 exchange population.[15]

Health Expenditures

Expected health expenditures for the non-senior population are imputed using modified MEPS-HC data. Our methodology closely follows CBO's methodology with some simplifying assumptions that are explained below.

We rely on MEPS-HC data to estimate healthcare spending. Raw MEPS-HC spending totals, however, are inconsistent with the National Health Expenditure (NHE) data. Following CBO's methodology, we adjust MEPS-HC data to reconcile aggregate MEPS-HC spending estimates with NHE data.[16] We adjust the MEPS-HC sample weights to reflect known undercounts in Medicaid enrollment and high-cost recipients. The reweighting methodology maintains the initial MEPS-HC distributions for sex, race, and poverty status. Finally, we reconcile any remaining difference in spending aggregates with an across-the-board spending increase for all respondents.

Using the modified MEPS-HC data we then perform a series of regressions to estimate health spending for CPS respondents. The regressions estimate the probability a non-senior respondent with private insurance has health spending within specific spending buckets. We divide spending into 12 buckets: $0; $1 to $499; $500 to $999; $1,000 to $1,499; $1,500 to $1,999; $2,000 to $2,999; $3,000 to $3,999; $4,000 to $4,999; $5,000 to $9,999; $10,000 to $17,499; $17,500 to $25,000; and $25,000 and above. We estimate separate regressions for different age groups and sexes. Covariates used in some or all of these regressions include age, health status, disability

[15] We use the CMS Marketplace Open Enrollment Period Public Use Files to determine age distributions. See the 2017 OEP Snapshot Public Use Files available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Marketplace-Products/Plan_Selection_ZIP.

[16] The corrections follow the methodology described in Bernard, Selden, and Pylypchuck (2015) and Bernard et al., (2012). Bernard, Didem, Thomas M. Selden, and Yuriy O. Pylypchuk (2015). *Aligning the Medical Expenditure Panel Survey to Aggregate U.S. Benchmarks, 2010*. Available at https://meps.ahrq.gov/data_files/publications/workingpapers/wp_15002.pdf. Bernard, Didem, Cathy Cowen, Thomas Selden, Liming Cai, Aaron Catlin, and Stephen Heffler, (2012). "Reconciling medical expenditure estimates from the MEPS and NHEA, 2007." *Medicare & Medicaid Research Review, 2(4)*. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4006479/pdf/mmrr2012-002-04-a09.pdf

status, household size, family income as a percent of the family's poverty threshold, and an indicator variable for having group insurance coverage.[17]

We then fit these regression estimates on CPS data to estimate the probability a respondent under 65 has health expenditures within a particular bucket. Each CPS respondent is then assigned an expected spending amount for each bucket that is randomly selected between a bucket's lower and upper bounds. The expected amount spent for the above-$25,000 bucket is estimated by regressing total health expenditures on similar covariates for non-senior MEPS-HC respondents reporting spending above $25,000. We then fit these regression estimates on the CPS sample. To ensure adequate variance in the above $25,000 spending estimate, we add a random variable drawn from a normal distribution with zero mean and standard deviation equal to the root mean squared error of the regression.[18] Total expected health spending for a respondent is equal to the sum of products of the fitted probabilities and the expected spending amount in each spending bucket.

Adjustments to spending imputations are made to reflect state-level variations in per capita health spending. We use NHE's health expenditure by state of residence to calculate an index of per capita private health insurance spending for personal health care relative to the national average.[19] We then adjust total expected health spending so mean expected health spending by state matches this index.

In addition to total expected spending estimates, we use MEPS-HC to estimate the expected composition of spending in each spending bucket for each CPS respondent. This composition is divided into hospital spending, provider spending, and an other spending category. In addition, we estimate expected out-of-pocket spending as a share of total expected health spending by spending bucket.

Group Health Premiums

We use MEPS-IC data to impute group health premiums for non-federal workers. Unlike MEPS-HC data, however, we do not have access to MEPS-IC microdata; instead, we use publicly available summary tables.[20] As a consequence, the CHEI premium data have less variance than reported in actual data.

Group health insurance premiums vary by state, firm size (measured by the number of employees), and the type of insurance plan–*e.g.,* preferred provider organizations (PPOs), high-deductible health plans (HDHP), or Health Maintenance Organizations (HMO). We estimate firm size and assign a plan type to each CPS respondent who reports being the policyholder of a

---

[17] Our regression analysis is simpler than CBO's. CBO uses more precise spending buckets (16 in total) and includes a covariate for previous year spending. Nevertheless, our expected spending levels are broadly consistent with CBO's reported estimates.
[18] We set a minimum value of $25,000 for the fitted estimate.
[19] NHE state-level data is only available through 2014. We assume per capita state health spending relative to the national average remains constant since 2014. Data are available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsStateHealthAccountsResidence.
[20] We use the MEPS Insurance Component National-Level Summary Tables available here: https://meps.ahrq.gov/mepsweb/data_stats/quick_tables_search.jsp?component=2&subcomponent=1.

group plan. We explain our method for determining respondents' firm size below. Insurance plan types are randomly assigned to respondents to reflect enrollment totals reported in the *MEPS Insurance Component Chartbook 2017*.[21] We then assign self-only and family group premiums to each respondent using the premiums reported in MEPS-IC national-level tables.[22] These premiums are adjusted to reflect state-level differences in ESI premiums.[23] In cases where a respondent has only one covered dependent, we also estimate a self-plus-one plan premium equal to the arithmetic average of the respondent's assigned family and self-only plan premiums. Following similar procedures, we impute the share of premiums paid by employees.

Respondents in the private sector are assigned one of five firm sizes: 1 to 9 workers, 10 to 49 workers, 50 to 99 workers, 100 to 999 workers, and 1000 and above. The CPS includes a similar firm size variable, but this variable understates the number of workers employed by larger-size firms. To ensure consistency, we move some individuals from smaller firm sizes to larger employers. This is completed in an iterative process where a portion of respondents are randomly moved into larger employer groups until the share in each group is within 1 percent of the share reported in MEPS-IC data. State and local public employees are assigned to one of two firm sizes: 100 to 999 and 1000 and above.

We randomly assign all federal employees to one of ten insurance carriers, weighted by 2015 enrollment levels. In 2015 these ten insurance carriers covered 95 percent of federal employees, with 70 percent of federal employees enrolled in one of two federal employee fee-for-service (FFS) programs offered by Blue Cross and Blue Shield.[24] We use carrier-specific premium information from the Office of Management and Budget to impute nationwide fee-for-service premiums.[25] We use MEPS-IC data to impute premiums for federal workers who are covered by an HMO.

Affordable Care Act Premiums

We estimate ACA premiums, potential subsidies, and metal levels for each respondent enrolled in the exchanges. Each exchange enrollee is assigned an ACA premium based on their imputed ACA metal level (bronze, silver, gold, platinum, or catastrophic).[26] We also determine the

---

[21] Agency for Healthcare Research and Quality, (2018). *Medical Expenditure Panel Survey Insurance Component: 2017 Chartbook*. Available at https://meps.ahrq.gov/data_files/publications/cb22/cb22.pdf.

[22] MEPS-IC summary tables do not report HDHP premiums in the same way they report other plan types. We use the Kaiser Family Foundation's 2017 Employer Health Benefits Survey to estimate the percent difference in mean HDHP premiums and the mean premium of other group plans. HDHP premiums are then imputed using this difference and MEPS-IC premiums by firm size.

[23] State-level differences are estimated using the MEPS summary data. Often these data are censored or otherwise missing in MEPS-IC summary tables. We use available data from 2015 to 2017 to estimate the percent difference in self and family premiums in a particular state relative to the national average.

[24] Enrollment by carrier is available Bloom, Kirstin B. and Ada S. Cornell (2016). Federal Employees Health Benefits (FEHB) program: An overview," *Congressional Research Service, R43922*. Available at https://fas.org/sgp/crs/misc/R43922.pdf.

[25] When a nation-wide carrier offers multiple plans, we take the arithmetic average of the plans. FFS data for 2017 are available here: https://www.opm.gov/healthcare-insurance/healthcare/plan-information/plans/premiums/2017/ffs/non-postal.

[26] The Affordable Care Act exchanges feature four metal tiers that reflect the actuarial value of an insurance plan. Bronze plans must have an actuarial value of about 60%; silver, 70%; gold, 80%; and platinum, 90%. Catastrophic plans with less than 60% actuarial value are available to individuals under 30. Subsidies for exchange participants are based on the second-cheapest silver plan in their geographical area. Most exchange enrollees choose bronze or silver plans.

second-cheapest silver plan available to the respondent. Finally, we calculate each respondent's required contributions to determine the amount of ACA exchange premium subsidies, if any, the respondent would receive. Each of these imputations is explained below.

We first estimate average state ACA premium by age and metal level. Using CMS Health Insurance Exchange Public Use Files, we collect plan-specific premium information by age, metal level, and rating area.[27] We then rank each plan from cheapest to most expensive in each rating area. A state-level, population-weighted premium is then calculated for each rank, age, and metal level.[28]

We then impute a particular ACA metal level for each exchange participant, assigned in a manner that ensures enrollment by metal level matches administrative data.[29] Since metal level selection and poverty level are correlated, we include an adjustment to ensure metal level enrollment by poverty level reflects administrative data.[30] A priori, we also expect high-cost enrollees to choose more generous metal levels. Thus, within each poverty group we assign higher metal levels to respondents with higher expected health expenditures.

Participants are then randomly assigned a plan in their assigned metal level. Each exchange enrollee receives one of three premiums: (1) the premium from their state's cheapest plan in their metal level, (2) the second-cheapest plan in their metal level, or (3) a mean of more expensive plans in their metal level. We use public data from California's state exchange to estimate the share of participants who choose the cheapest, second cheapest, or more expensive plan within a particular metal level.[31] The data show that 47 percent of bronze plan participants choose the cheapest option and an additional 19 percent choose the second-cheapest option. Among silver plan participants, 32 percent choose the cheapest and 25 percent choose the second-cheapest plan.

We then calculate exchange subsidies. The exchange subsidy is calculated for each family with at least one exchange enrollee. The subsidy is equal to the sum of second-cheapest options for all exchange participants within a family minus the required health contribution for that family.

---

[27] Detailed ACA Marketplace Exchange plan information for states who participate in the federal exchange is available here: https://www.cms.gov/CCIIO/Resources/Data-Resources/marketplace-puf. Plan information for states with separate exchange programs is available here: https://www.cms.gov/CCIIO/Resources/Data-Resources/sbm-puf.

[28] Rating areas by county or zip code are available here: https://www.cms.gov/CCIIO/Programs-and-Initiatives/Health-Insurance-Market-Reforms/state-gra. We use the American Community Survey to determine population weights for each rating area.

[29] We use the CMS Marketplace Open Enrollment Period Public Use Files to determine metal level enrollment. See the 2017 OEP Snapshot Public Use Files available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Marketplace-Products/Plan_Selection_ZIP.

[30] This is partly due to ACA cost-sharing subsidies that effectively increase the actuarial value of some silver plans for eligible participants. These subsidies are available for participants with family incomes between 100% and 250% of federal poverty guidelines. As a consequence, eligible individuals largely choose silver plans. For example, among federal exchange participants, 76% of individuals with family incomes between 100% to 250% chose silver plans in 2019. Only 27% of participants at other poverty levels chose silver plans. See sheet (11) Metal Level by FPL in the state-level public use file, available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Marketplace-Products/2019_Open_Enrollment.

[31] Covered California enrollment data is available here https://hbex.coveredca.com/data-research/. The Covered California data is unlikely to be representative of other states' experiences, but we are unaware of any other data source available for this calculation.

Subsidies are capped at the total exchange premiums for the family.[32] Using imputations for pre-credit tax liability we estimate the share of a family's subsidy that lowers its tax liability and the share in excess of its tax liability. This allows us to estimate the aggregate revenue and outlay effects from ACA subsidies.

Data on individuals who purchase individual health insurance off the exchanges are not readily available. We use state-level ACA premium information to estimate premiums for off-exchange individual plans. We randomly assign all individuals who purchase an off-exchange plan a catastrophic, bronze, or silver premium.

III. Projections

We generate separate datafiles with health-related variables through 2049, the final year available in CBO's *2019 Long-Term Budget Outlook*. Each dataset is derived from the initial-year dataset with adjustments made to reflect income growth, price changes, and expected changes in national demographics. Generally, this process involves using projections from three sources (the US Census Bureau, CBO, and NHE) to change certain variables or adjust sample weights. Not all years are projected in an identical fashion. NHE projections are only available to 2027. CBO's detailed economic, premium, and insurance status projections are only available to 2029. Beyond these years we use simplified CBO assumptions derived from the current annual Long-Term Budget Outlook to adjust relevant variables.[33] We discuss the specific assumptions below.

Demographics and Population

We use the US Census Bureau's 2017 National Population Projections Datasets to adjust sample weights to reflect population and demographic changes.[34] Each year's projection is based on the previous year's sample weight adjusted to reflect Census projections. All sample weights are first increased by the projected growth in US population. We then use a raking scheme to adjust race, sex, and age distributions. The raking method preserves the 2017 poverty status distribution to ensure relative income distributions remain constant over time.

Income Variables

Income variables are adjusted using CBO income projections. We use different approaches to adjust income during early projection years (up to 2029) and later projection years (2030 to 2049). We discuss these approaches below.

For the projection years 2019 to 2029, we use CBO's 10-year economic forecasts to grow different income categories.[35] CBO includes aggregate projections for different types of income including wages and salaries, self-employment income, dividend income, and rental income. We adjust these income projections to calculate income per recipient using CPS estimates of the

---

[32] In some cases, bronze plans premiums may be less than a recipient's maximum subsidy.

[33] Congressional Budget Office (June 2019). *The 2019 Long-Term Budget Outlook.* Available at: https://www.cbo.gov/publication/55331.

[34] We use Table 1 of US Census Bureau, (2017). *2017 National Population Projections Datasets*. Available here https://www.census.gov/data/datasets/2017/demo/popproj/2017-popproj.html.

[35] We use calendar-year estimates from CBO's 10-year economic forecasts. CBO (August 2019). *An Update to the Economic Outlook: 2019 to 2029*. Available at https://www.cbo.gov/about/products/budget-economic-data#4.

number of individuals receiving each type of income per year. We then estimate a per enrollee growth rate for each year and adjust the relevant income type by our calculated growth rate. To account for expected faster income growth among top earners, we calculate wages and salaries separately for CPS respondents with 2017 wages below and above the Social Security taxable maximum. We use CBO's *2019 Long-Term Budget Outlook* to estimate the share of aggregate earnings below the taxable maximum.[36]

After 2029, we exclusively use data from CBO's *Long-Term Budget Outlook* to grow incomes. The available data allow us to calculate wages and salary growth. Aggregate income growth for other income categories is unavailable. We thus assume other income categories grow at CBO's projected nominal GDP growth rate.

We make a separate calculation for other income variables not included in CBO's 10-year economic forecasts. The lion's share of these variables is related to public assistance programs. We grow these variables at CBO's projected CPI-U growth rate (using CBO's *Long-Term Budget Outlook)*.[37] In addition, we grow imputed HHS poverty guidelines (used to determine ACA subsidy eligibility) by the projected growth in the CPI-U**.**

Tax Projections

As discussed in section II, TAXSIM only allows for tax calculations through 2023. We thus use a separate tax calculator for all projected years. Generally, the tax calculator works the same in all years. We index tax bracket thresholds, standard deductions, and other tax code provisions that are indexed to an estimate of the chained CPI. The Tax Cuts and Jobs Act of 2017 (TCJA) permanently changed the index measurement from the traditional CPI to a chained metric. We estimate the chained CPI as CBO's projected CPI-U minus .25 percentage points.[38]

There are significant changes to the tax code beginning in 2026 when most of the TCJA's individual income tax provisions expire. We return post-2025 tax rates to their pre-TCJA amounts. We also estimate the post-2025 personal exemptions, standard deductions, and tax bracket thresholds using their 2017 amounts indexed for growth in the chained CPI-U. In addition, the tax calculator includes calculations for the personal exemption phaseout, which was temporarily eliminated with TCJA.

As noted above, we do not attempt to estimate state tax rates; instead, we borrow TAXSIM's estimates for state income tax data. We assume filers' state marginal tax rates remain constant after 2023.

Healthcare Premiums

Private insurance healthcare premiums are adjusted to reflect expected increases in premiums for all private health insurance. Through 2027, we use NHE projections for per enrollee private

---

[36] We use the economic variables from CBO's June 2019 outlook. CBO (June 2019). *The 2019 Long-Term Budget Outlook*. Available at https://www.cbo.gov/about/products/budget-economic-data#1.

[37] This will likely understate Social Security income growth and retirement income growth as these income sources are generally linked to wage growth. Since the CHEI data is mainly focused on the non-senior population, this bias is unlikely to have a significant effect.

[38] McClelland, Rob (2013). "Differences Between the Traditional CPI and the Chained CPI." *Congressional Budget Office*, April 19, 2013. Available at: https://www.cbo.gov/publication/44088.

health insurance spending.[39] For 2028 and 2029, we use CBO's private insurance premium growth rate projection as reported in their 2019 report *Federal Subsidies for Health Insurance Coverage for People Under Age 65: 2019 to 2029.* Beyond 2029, we rely on CBO's private health insurance "excess cost factor" as reported in the *2019 Long-Term Budget Outlook.* The excess cost factor is an estimate of the growth in private insurance spending beyond the growth rate in real GDP per capita. CBO projects the excess cost factor to fall from 1.53 percentage points in 2030 to 1.0 percentage points in 2049. We use the same growth projections to determine the future employee-share of group-plan premiums.[40]

The CBO projects ACA Marketplace Exchange premiums will grow at different rates than other private insurance premiums. Through 2029, we use CBO estimates of the change in the benchmark premium (i.e. second-cheapest silver premiums) to grow all ACA plan premiums.[41] Beyond 2029, ACA premiums grow at the same rate as other private insurance premiums.[42]

Healthcare Spending

Expected healthcare spending is adjusted to reflect expected increases in health spending. Similar to the healthcare premium projections, through 2027 we use NHE projections on per enrollee private health insurance spending to grow expected health spending. We make separate calculations to grow expected spending by spending type (hospital, provider, and other) using NHE data for each category. Beyond 2027, we use CBO's estimates for the growth in private insurance premiums to grow all spending categories.

Insurance Status

Unlike CBO's HISIM2 model, the CHEI data do not directly attempt to simulate future health insurance enrollment. Instead, to ensure the data closely match CBO's public reports, we adjust sample weights and insurance status of a portion of the sample to match CBO's enrollment projections. We discuss these adjustments below.

CBO's *Federal Subsidies for Health Insurance Coverage for People Under Age 65: 2019 to 2029* projects the number of non-seniors by health insurance status (*e.g.,* enrolled in an ESI plan, exchange enrollment, or uninsured). We adjust each year's sample weights to reflect CBO's projected share of the non-senior population in each insurance category. We use a raking technique that preserves the projected year's age, race, sex, and poverty distributions. The weighting adjustments are generally minor and robust to various specifications.

The adjustments, however, fail to capture expected changes in the risk pool of exchange participants. Over the next 10 years, CBO projects mean ACA subsidies will rise faster than

[39] The NHE projections are available in Table 17 of the NHE Projections 2018-2027 available here: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected.

[40] This likely underestimates the growth in the employee share, which has recently grown faster than overall premiums.

[41] The benchmark premium growth assumptions are found in Figure A1 of CBO (2019). *Federal Subsidies for Health Insurance Coverage for People Under Age 65: 2019 to 2029*. Available at www.cbo.gov/publication/55085.

[42] This assumption should not materially affect our projections. CBO's projected growth rates for ACA benchmark premiums and all other private insurance premiums are nearly identical by the end of CBO's 10-year budget window.

ACA benchmark plans.[43] This increase may be due to changes in expected incomes of ACA participants or from changes in the ACA risk pool. For example, this could occur if the average age of exchange participants grows over time. In either case, the reweighting adjustments fail to capture this phenomenon. Thus, without further adjustments, CHEI data would underestimate future ACA subsidies. To better account for expected changes in the ACA risk pool, we determine the expected risk-neutral value of a respondent's exchange enrollment (accounting for the respondent's expected health spending, premium subsidies, and gross premium). We then remove respondents with the least expected value until our exchange enrollment projections match CBO's reported estimates. These changes are completed before the reweighting adjustments described above.

Detailed enrollment projections by CBO are only available through 2029. Beginning in 2029, we assume the share of the non-senior population in each insurance category remains constant.

IV. Conclusion

The CHEI data include 33 annual datasets (2017 through 2049) each with microdata on projected health spending, insurance premiums, and enrollment data. Beyond the normal caveats for any forecasts, the CHEI data suffer from several limitations. We have noted many of these limitations above. We discuss major limitations below.

The limitations include simplified growth assumptions for income. We have attempted to account for relative increases in wages and salaries for top earners, but otherwise we assume income growth rates are uniform across all CPS respondents. In addition, we make further simplifying income growth assumptions for projected years beyond CBO's 10-year budget window.

As discussed above, we do not attempt to predict insurance enrollment, but rather rely on CBO's public estimates for enrollment changes. This simplifying assumption allows us to emulate CBO's cost estimates, but it could lead to issues if there are significant demographic changes within a particular health insurance category. We addressed this potential issue for the ACA exchange population, but otherwise we assume no significant demographic changes within a particular health insurance category.

Finally, CBO's HISIM2 model uses MEPS-IC microdata (among other sources) to model firm behavior. This includes modeling which firms will offer health insurance, the plans and premiums chosen, and the employees who will be eligible for coverage. Lacking MEPS-IC microdata, we make no similar attempt to model firm behavior.

In the future, we plan to update CHEI data to reflect new data, include additional modules that determine respondents' eligibility for Medicaid, and make additional imputations to better estimate ESI coverage and firm decisions.

[43] The average ACA subsidies are found in the Supplement Material of CBO (2019). *Federal Subsidies for Health Insurance Coverage for People Under Age 65: 2019 to 2029*. Available at www.cbo.gov/publication/55085.