Kaiser Family Foundation ACA Eligibility Analysis: Technical Appendix C: Imputation of Offer of Employer-Sponsored Insurance

To impute the presence of an offer of employer-sponsored insurance (ESI) for each person in the sample, we followed the broader methods of our Immigration Status Imputation (described in Technical Appendix B), tailored to this characteristic of survey respondent. This approach uses the Survey of Income and Program Participation (SIPP) to develop a model that predicts whether each individual in the microdata has an offer of ESI Unlike the Immigration Status Imputation, this method does not control the model results to a second data source, since the source data set in this case (SIPP) can be thought of as the authoritative source for this statistic. Below we describe how we developed the regression model and applied it to the Current Population Survey. We also describe how the model may be applied to other data sets. The programming code, written using the statistical computing package R v.3.1.1, is available upon request for people interested in replicating this approach for their own analysis.

Data Sources

We used the sixth wave of the 2008 Survey of Income and Program Participation (SIPP) panel data to build the regression model. The SIPP Wave Six Topical Module dataset contains questions on employer-provided health benefits as well as a pathway of questions for those not covered about the workplace making an offers of ESI at the worker-level.[1]

The regression model is designed to be applied to other datasets in order to impute offer status. The code mentioned above includes programming to apply the model to either the SIPP Core file or the Current Population Survey (CPS) (for years 2007 on). Because the SIPP Core file and CPS contain different survey questions and variable specifications, we create unique regression models to apply the model to each dataset. For the updated analysis underlying *The Coverage Gap: Uninsured Poor Adults in States that Do Not Expand Medicaid*, we apply the regression model to the 2014 CPS-ASEC.

Construction of Regression Model

We use the SIPP Wave Six to create a binomial, dependent variable that identifies a respondent as a recipient of an offer of employer-sponsored insurance. The dependent variable is constructed at the worker-level based on the factors below, and then distributed to all non-qualifying relatives sharing the tax-filing unit of that worker (more detail on tax-filing unit construction can be found in Technical Appendix A):

1) Worker was not covered by employer-sponsored insurance,

2) Worker either indicated holding an ESI plan or eligibility to be covered that was then voluntarily declined.[2]

We use the following independent variables to predict offer status within tax-filing units:

1) Any public coverage,

2) Any nongroup coverage,

3) Highest earnings among all workers,

4) Number of full-time workers within tax-filing unit,

5) Number of part-time workers within tax-filing unit,

6) Age of oldest worker,

7) Presence of any worker at a large-firm (100+ employees),

8) Presence of any workers in the construction industry,

9) The unemployment rate within the state of residence for the year.

The regression model was sub-populated to remove respondents who could not have received an offer of ESI. People in tax-filing units without any employed persons and persons not in tax filing units could not be considered to have an offer of ESI. This imputation does not account for the affordability of the offer or whether it meets the minimum value test.[3]

Imputing Offer Status in Other Datasets

We use the SIPP regression results as our estimate of the total number of individuals with access to an Employer-Sponsored Health Plan. To generate the imputed offer status variable, we first calculated the probability that each tax-filing unit in the dataset was offered ESI based on the SIPP regression model. Next, we selected tax-filing units within the target dataset (the Current Population Survey) using the sampling probabilities resultant from our model. Since the imputation of documentation status (discussed in Technical Appendix B) required a multiply-imputed approach, this secondary imputation and subsequent tax-filing unit sampling was only conducted once per implicate, keeping the number of implicates to five.

To easily apply the regression model to other data sets, we created a function that applies this approach to a chosen data set. The function first loads the dataset of choice; then standardizes the data to match the independent variables from the SIPP regression model; and finally applies the multiple imputation to generate a variable for legal immigration status.

---

[1] http://www.census.gov/programs-surveys/sipp/tech-documentation/topical-modules/topical-modules-2008.html

[2] Indicators that disqualify workers from holding an offer include reporting not enrolled in health insurance through work because of being Ineligible, Denied coverage, or Other.

[3] For an explanation of affordability, see: http://kff.org/health-reform/fact-sheet/explaining-health-reform-how-will-the-affordable-care-act-affect-small-businesses-and-their-employees/