February 2024 | Appendix

KFF ACA Eligibility Analysis, Technical Appendix B: Immigration Status Imputation

To impute documentation status, we draw on the methods underlying the 2013 analysis by the State Health Access Data Assistance Center (SHADAC) and the recommendations made by Van Hook et. al..[1,2] This approach uses the 2023 KFF/LA Times Survey of Immigrants to develop a model that predicts immigration status for each person in the sample.[3] We apply the model to a second data source, controlling to state-level estimates of total undocumented population as well as the undocumented population in the labor force from the Pew Research Center.[4] Below we describe how we developed the regression model and applied it to the American Community Survey (ACS). We also describe how the model may be applied to other data sets. The programming code, written using the statistical computing package R v.4.3.1, is available upon request for people interested in replicating this approach for their own analysis.

Data Sources

We used the 2023 KFF/LA Times Survey of Immigrants l data to build the regression model. The 2023 Survey of Immigrants dataset contains questions on citizenship and legal status at the person level. The KFF/LA Time Survey of Immigrants[5] is a probability-based survey exploring the immigrant experience in the U.S. and draws on three different sampling frames including an address-based sample (ABS), a random digit dial (RDD) sample of pre-paid cell phone numbers, and callbacks to an RDD sample in which the individual did not speak English or Spanish. The survey includes interviews with 3,358 immigrant adults and was offered in ten different languages.

The regression model is designed to be applied to other datasets in order to impute legal immigration status in surveys that do not ask about migration status. The code mentioned above includes programming to apply the model to either the Survey of Income and Program Participation (SIPP) Core files, ACS, or the Current Population Survey (CPS). Because the SIPP Core file contains different survey questions and variable specifications from the ACS and CPS, we create unique regression models to apply the model to each dataset. For the analysis underlying this brief and other KFF estimates of eligibility for ACA coverage, we apply the regression model to the 2013 ACS and then each subsequent year of the ACS.

Due to underreporting of legal immigration status in survey datasets, in imputing immigration status we control to state and national-level estimates of the total undocumented population and also the undocumented population in the labor force from the Pew Research Center. Pew reports these estimates for all states and the District of Columbia.[6]

Construction of Regression Model

We use the 2023 Survey of Immigrants to create a binomial, dependent variable that identifies a respondent as a potential unauthorized immigrant. The dependent variable is constructed based on the following factors:

1) Respondent was not a United States (US) citizen,
2) Respondent did not have permanent resident status or a valid work or student visa, , and
3) Respondent does not have other indicators that imply legal status.[7]

We use the following independent variables to predict unauthorized immigrant status:

1) Year of US entry,
2) Job industry classification,
3) State of residence,
4) Household Income,
5) Ownership or rental of residence,
6) Number of occupants in the household (< or >= six occupants),
7) Whether all household occupants are related,
8) Health insurance coverage status,
9) Sex, and
10) Ethnicity.

The regression model was sub-populated to remove respondents who could not be considered unauthorized. People who could not be considered unauthorized include people who are US citizens or have other indicators that imply legal status.

Imputing Unauthorized Immigrants in Other Datasets

We use the Pew estimates as targets for the total number of unauthorized immigrants that the imputation generates. We first apply this strategy to the 2013 ACS, which contains health insurance information prior to the ACA's coverage expansions. We stratify the targets by state and the District of Columbia and by participation in the labor force. We impute immigration status within each of these 102 strata.[8]

To generate the imputed immigration status variable, we first calculated the probability that each person in the dataset was unauthorized based on the 2023 Survey of Immigrants regression model. Next, we isolated the dataset to each individual stratum described above. Within each stratum, we sampled the data using the probability of being unauthorized for each person. After sampling, we summed the person weights until reaching the Pew population estimate for each stratum. The records that fell within the Pew population estimate were considered to be unauthorized immigrants. We repeated the process of

sampling using the probability of being unauthorized and subsequently summing the person weights to reach Pew targets five times, creating five different unauthorized variables per record. These five imputed authorization status variables were then incorporated into a standard multiple imputation algorithm, closely matching the imputed variable analysis techniques used by the Centers for Disease Control and Prevention for the National Health Interview Survey.[9]

We used this first pass on the ACS 2013 to inform our sampling targets for the latest available microdata (ACS 2022). Looking at the results of our undocumented imputation on the ACS 2013, we calculated the share of undocumented immigrants lacking health insurance within each of those 102 strata prior to the ACA's coverage expansions and transferred that information into a new dimension of sampling strata for the ACS 2022. We split each of the 102 sampling strata used on the pre-ACA ACS 2013 into uninsured versus insured categories, resulting in 204 sampling strata for subsequent years. We then repeated our imputation on the ACS 2022 with the newly-divided strata, allowing for a small decline in the undocumented uninsured rate based off of the percent drop in the uninsured rate among citizens.[10]

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 2023 Survey of Immigrants regression model, and finally applies the multiple imputation to generate a variable for legal immigration status.

Endnotes

[1] State Health Access Data Assistance Center. 2013. “State Estimates of the Low-income Uninsured Not Eligible for the ACA Medicaid Expansion.” Issue Brief #35. Minneapolis, MN: University of Minnesota. Available at: http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2013/rwjf404825.

[2] Van Hook, J., Bachmeier, J., Coffman, D., and Harel, O. 2015. “Can We Spin Straw into Gold? An Evaluation of Immigrant Legal Status Imputation Approaches” *Demography*. 52(1):329-54.

[3] This data source is a change from previous KFF analyses, which used microdata from the 2008 Panel of the Survey of Income and Program Participation (SIPP)

[4] This data source is a change from previous KFF analyses, which used estimates from the Department of Homeland Security.

[5] More information about the survey methods is available at https://www.kff.org/report-section/understanding-the-u-s-immigrant-experience-the-2023-kff-la-times-survey-of-immigrants-methodology/

[6] Pew updates these estimates periodically. We use the most recent estimates available at the time of our analysis, and in some cases incorporate estimates received from correspondence with researchers at Pew prior to their publication - however we do not release these numbers ourselves. We draw on Pew directly for all published data and interpolate years missing from their trend. Our analysis uses the year applicable to the year for the data sets to which we apply the regression model. The most recent estimates as of the time of our analysis were: J Passel, D Cohn. *Mexicans decline to less than half the U.S. unauthorized immigrant population for the first time*. (Pew Research Center), June 2019. Available at: https://www.pewresearch.org/fact-tank/2019/06/12/us-unauthorized-immigrant-population-2017/.

[7] Indicators that imply legal status include: (i) respondent entered the US prior to 2000, or (ii) respondent is enrolled in Medicare or military health insurance.

[8] For more information, see SHADAC 2013, footnote 6. The table created for this function contains estimates of the undocumented across 2013-2022.

[9] For more detail, see documentation available at: National Health Interview Survey. *2022 Imputed Income Files.* Available at:https://www.cdc.gov/nchs/nhis/2022nhis.htm.

[10] As an example of this calculation, we found that approximately 66% of undocumented uninsured individuals did not have health coverage in 2013. We allow the undocumented rate to drop slightly after 2013. We base the percent drop in the uninsured rate among the undocumented on the drop for citizens (half the scale of the drop for citizens) each year until 2022, resulting in the final undocumented uninsured rate of 50% in calendar year 2022. Prior to implementing this new sampling dimension, we found unrealistic drops in the uninsured rate of the undocumented population that we largely attributed to our prediction model's inability to discern this group from legally-present non-citizens, many of whom are eligible for assistance under the ACA's coverage expansions. Although a few states have implemented programs that allow for coverage of the undocumented population, these programs are state-funded and relatively small in scale compared to the nationwide coverage expansions accompanying the ACA.