



# Comparing the Effectiveness of Assignment Strategies for Estimating Likely Undocumented Status in Secondary Data Sources for Latino and Asian Immigrants

Annie Ro<sup>1</sup> · Jennifer Van Hook<sup>2</sup>

Received: 23 October 2020 / Accepted: 15 May 2021 / Published online: 26 May 2021  
© The Author(s) 2021

## Abstract

Researchers are increasingly interested in the role of undocumented status in immigrant economic, social, and health outcomes. A major obstacle to this work is that detailed immigration status is not widely collected in representative data sources. Some secondary data sources collect enough information to identify immigrants without a green card (non-LPRs), and researchers take different approaches to assign undocumented status to immigrants within this population. These approaches have not been compared to one another, nor do we know if they work equally well for Latino and Asian immigrants. In this research note, we test the validity of several assignment strategies using the 2001, 2004, and 2008 panels of the restricted version of the Survey of Income and Program Participation (SIPP) to measure differences in health-related outcomes (e.g., health insurance coverage and self-rated health) by immigration status. We compare results when immigration status is directly measured using the detailed information in the SIPP to several strategies to assign undocumented status among non-LPRs. The probabilistic approach produced the smallest biases, but Asian immigrants had larger biases compared to Latinos across all strategies.

**Keywords** Undocumented immigrants · Asian immigrants · Simulations · Health insurance · Survey of income and program participation

---

✉ Annie Ro  
annie.ro@uci.edu

Jennifer Van Hook  
jxv21@psu.edu

<sup>1</sup> Department of Health, Society, and Behavior, UC Irvine Program in Public Health, 653 E. Peltason Drive AIRB 2030, Irvine, CA 92697, USA

<sup>2</sup> Pennsylvania State University, 706 Oswald Tower, University Park, PA 16802, USA

## Introduction

There is strong interest in understanding the outcomes and costs of undocumented status among immigrants, particularly in health. While the body of empirical evidence is growing (Hamilton et al., 2019; Vargas Bustamante et al., 2012), there are only a limited number of large-scale representative data sources with detailed immigration status information. Yet these surveys, such as the restricted version of the California Health Interview Survey (CHIS) or the public-use version of the Survey of Income and Program Participation (SIPP), only distinguish between legal permanent residents (LPRs) and non-LPRs among non-citizens. To identify undocumented immigrants, researchers either assume that the majority of non-LPRs are undocumented (Pourat et al., 2003) or use an imputation method for assigning undocumented status among non-LPRs (Vargas Bustamante et al., 2014). Moreover, we do not know whether race/ethnic-specific trends can be studied with these methodologies and data sources. There is especially no guidance for researchers on whether Asian undocumented immigrants, the second-largest and fastest-growing racial/ethnic undocumented group, can be reasonably identified in these data sources.

Researchers primarily employ two imputation approaches to distinguish undocumented immigrants from other authorized immigrants. The first, the logical edit method, is more common and uses characteristics such as country of origin, educational attainment, or occupational status to identify likely undocumented immigrants from non-citizens (Borjas, 2017), as well as from non-LPRs who are lawfully present, such as refugees who have not yet adjusted to LPR status or students on a non-immigrant visa (Vargas Bustamante et al., 2014). One potential problem is that the logical edit method assumes undocumented immigrants have certain demographic and occupational patterns, yet we know that Asian and Latino populations differ widely on such characteristics (Acevedo-Garcia et al., 2010).

Another strategy, statistical assignment, constructs regression prediction equations from surveys that collect undocumented status information and apply them to other surveys without detailed migration measures. Until now, it has not been possible to use statistical assignment methods to distinguish between undocumented immigrants and other non-LPRs because, to our knowledge, no public use survey data distinguish undocumented immigrants from other non-LPRs. This has made it difficult to obtain the appropriate data on which to estimate the regression prediction equations. The validity of the statistical assignment approach has been established under certain data conditions, but only among the aggregated group of undocumented immigrants, which is majority Latino (Van Hook et al., 2015).

Different approaches may contribute to inconsistent conclusions on the role of immigration status because of differences in who is identified as undocumented. We also do not know whether the different approaches work equally well across different racial/ethnic groups. This paper measures the biases of different assignment approaches when estimating differences in health insurance coverage and self-rated health by immigration status for both Latino and Asian immigrants using the SIPP. The restricted version of the SIPP is one of the only representative

data sources with detailed immigration information that allows researchers to distinguish between legal non-immigrants (LNIs; refugees/asylees, diplomatic, student, business, or tourist visa holders) and undocumented immigrants among non-LPRs. Our findings offer recommendations for researchers looking to use available secondary data sources to study undocumented immigrants.

## Methods

### Data

The SIPP is a longitudinal study of the U.S. civilian non-institutionalized population that collects information on income, labor force participation, social program participation, and general demographic characteristics (Census, 2016). Each panel is followed for 3–4 years and is interviewed for up to 13 waves. We use the 2001, 2004, and 2008 panels and the third and fifth waves of each panel, as these contain the migration and health items. Each panel interviews 14,000 to 52,000 households; we only included foreign-born Asian ( $n > 6500$ ) and Latino ( $n > 12,000$ ) adults in these panels and waves (exact sizes withheld for disclosure rules).

All respondents were asked their country of birth. Foreign-born respondents were then asked their citizenship status, whether they had LPR status upon arrival, and if not, whether they subsequently adjusted to LPR status. For users of public-use data, this information can be used to distinguish naturalized citizens, LPRs, and non-LPRs. As noted above, the problem is that non-LPRs include immigrants with a mixture of statuses and not only undocumented immigrants. However, the restricted data identify non-LPRs who were granted refugee status or granted asylum, admitted as LNIs, and all others.<sup>1</sup> With information from the restricted data, we categorized immigrants who entered as refugees/asylees and non-immigrants (e.g., diplomatic, student, business, or tourist visa), have not adjusted their status, and have under 6 years of US duration as LNIs. We classified all other non-LPRs as undocumented.

To handle missingness in the sequence of migration questions, which can go as high as 15%, we used multiple imputation by chained equations (MICE). We first averaged across ten imputed datasets to create a single dataset that was only missing migration information. We conducted MICE again for the migration information only and created ten imputed datasets. Each of the simulations (described below) was conducted on a random selection of one of the ten datasets.

---

<sup>1</sup> The restricted SIPP data also provide additional detail about the entry status of new arrival LPRs, distinguishing among immediate family or relative sponsored permanent residents, employment-based permanent residents, and other permanent residents.

## Assignment Approaches

We compared the results using the detailed migration information in the restricted SIPP with results that would be obtained by four approaches for assigning the undocumented status among non-LPRs: (1) the combination approach; (2) the logical edit; (3) the probabilistic approach; and (4) random assignment.

For the combination approach, we simply used non-LPR status to approximate undocumented immigrants. This assumes that non-LPRs are composed primarily of undocumented immigrants and that the presence of LNIs has a negligible impact on the results.

For the logical edit method, we took the population of non-LPRs and used similar criteria as Bachmeier et al. (2014) (demographic characteristics, occupational status, industry, and less than five years duration in the United States) to identify individuals who were likely to be on temporary employment or student visas (i.e., LNIs). For instance, we assumed high-tech H1B holders (temporary work visas) were currently employed, had been in the United States for under 5 years, were college-educated, worked for a private, for-profit employer, and were in occupations such as computer software engineers, network systems and data communications analyst, or computer systems analysts and scientists (for full criteria, see Appendix 1). For refugees, we assigned individuals a probability of being a refugee based on the proportion of individuals from the same country of birth who entered the United States the same year (Bachmeier et al., 2014). Those not assigned as LNI or refugee were classified as undocumented.

For the probabilistic method, we calculated the probability of being undocumented as a function of gender, age, poverty, education, year of immigration, marital status, household size, home ownership, employment status, region, disability status, self-rated health, and interactions between insurance coverage and gender, age, year of immigration, and marital status. We included health insurance and self-rated health, as the probabilistic method has the smallest bias when the dependent variables are incorporated in the model that is estimating immigration status (Van Hook et al., 2015). We interacted insurance status by other predictors to account for possible differences in coverage by these characteristics between LNIs and undocumented immigrants. This probability was generated from a regression model estimated for the non-LPRs in the restricted SIPP data in which detailed immigration status is known. We used the coefficients to calculate predicted probabilities in the simulated data in which detailed immigration status was unknown. The coefficients from this model are provided in Appendix 2. If a non-LPR immigrant had a higher probability than a random draw between 0 and 1, they were assigned as undocumented; the remaining were assigned as LNI.

Finally, we randomly assigned non-LPR immigrants as LNI or undocumented to compare the assignment methods to statuses assigned by chance. If an assignment performs worse than random assignment, it means that it is no better than if LNIs and undocumented immigrants were randomly assigned into different categories.

## Simulations

We compared the associations of immigration status with current health insurance and self-rated health between those obtained from the detailed migration information in the restricted SIPP, on the one hand, and those using the assigned method, on the other, utilizing the Monte Carlo simulation approach detailed by Van Hook et al. (2015). We conducted simulations separately for Asians and Latinos. First, we randomly chose one of the ten imputed datasets and drew 300,000 cases with replacement from the weighted Asian or Latino samples of the combined SIPP panels and randomly divided it into two equal groups. Second, we assigned immigration status. For the first group (the “donor” sample), we utilized the detailed migration information in the restricted SIPP to categorize individuals into one of four immigration statuses: naturalized citizens, green card holders, LNIs, and undocumented. For the second group (the “target” sample), we masked the SIPP data on detailed immigration status, and applied the combination approach, logical edit, the probabilistic approach, or random assignment to separate LNIs from undocumented immigrants. Third, we estimated a logistic regression with current health insurance or poor/fair self-rated health as the outcome and four immigration categories as the primary independent variable, controlling for age, gender, education, year of survey, marital status, and region. We repeated this sequence 500 times and calculated the average coefficients and standard errors of the coefficients. We calculated the relative bias (difference between target and donor coefficient, divided by the donor coefficient) between the coefficients of the target and donor samples. We consider a relative bias within 10% of the estimate obtained from the migration information in the restricted SIPP to be unbiased (Van Hook et al, 2015). A positive relative bias means the estimate from the imputation approach overestimates differences compared to the donor sample and a negative relative bias means the imputation approach underestimates differences.

## Results

Table 1 shows the demographic characteristics. LPR was the modal status category for Latinos (40.5%) and the majority of Asians were citizens (58.2%). Only 6.6% and 23.1% Latino LNIs and undocumented had current health insurance, respectively. In contrast, 21% and 76.1% of Asian LNIs and undocumented had current health insurance. Nearly 8% of undocumented Latinos had fair/poor self-rated health compared to 5.4% of undocumented Asians.

Table 2 compares assignment strategies for estimating differences in current health insurance coverage by immigration status. The first column is the estimate (expressed in logged odds) using the detailed migration information in the restricted SIPP, obtained from the donor sample. Compared to Asian citizens, Asian LNIs had the lowest odds of having health insurance ( $\beta = -1.61$ ; OR=0.2), followed by LPRs ( $\beta = -0.79$ ; OR=0.45) and undocumented immigrants ( $\beta = -0.65$ ; OR=0.53). Of the four imputation methods, the probabilistic approach had the lowest relative bias. The estimate for LNIs was 4% higher compared to the donor estimate

Table 1 Sample descriptives (weighted means) from 2001, 2004, 2008 SIPP, for all analytic variables

	Latinos ( <i>n</i> > 12,000)				Asians ( <i>n</i> > 6500)			
	FB citizens	FB LPR	FB LNI	FB undoc	FB citizens	FB LPR	FB LNI	FB undoc
Mean age	33.3%	40.5%	0.4%	25.8%	58.2%	30.9%	0.4%	10.4%
Male	46.7	38.2	30.4	32.7	47.7	40.1	34	33.5
Married	49.6%	53.2%	51.6%	57.9%	45.3%	45.6%	41.0%	51.4%
Years of education	63.1%	57.3%	52.2%	48.7%	67.5%	65.9%	53.9%	56.8%
Less than HS	36.1%	54.7%	34.9%	58.5%	11.6%	15.8%	45.2%	10.3%
High school graduate	29.1%	26.5%	26.9%	27.8%	19.0%	20.3%	26.9%	12.7%
Some college	22.8%	13.8%	27.6%	9.3%	24.7%	20.7%	15.4%	14.9%
College graduate	12.0%	4.9%	10.7%	4.5%	44.7%	43.1%	12.5%	62.1%
Region								
Northeast/midwest	20.5%	17.9%	9.3%	17.7%	32.1%	37.6%	46.0%	38.6%
South	36.2%	36.1%	51.7%	35.1%	19.2%	22.0%	16.7%	26.3%
West	43.4%	46.0%	39.0%	47.2%	48.6%	40.4%	37.3%	35.1%
Household size	3.8	4.5	4.6	4.9	3.5	3.6	5.45	3.1
Number of minors in household	0.88	1.23	0.46	1.11	0.68	0.83	1.19	0.62
Employed	67.0%	70.0%	58.2%	69.0%	70.7%	69.2%	56.4%	61.0%
Income to Poverty	2.9	2.1	1.74	1.9	4.7	4.4	2.07	4
Over 15 year US residence	66.2%	41.9%	12.7%	17.0%	65.1%	26.3%	41.5%	7.8%
Currently insured	69.0%	42.4%	6.6%	23.1%	85.8%	71.6%	21.0%	76.1%
Poor health	17.0%	12.1%	7.0%	7.7%	12.1%	9.5%	18.6%	5.4%

**Table 2** Simulated coefficients of migration status in models predicting insurance coverage by assignment method

	Donor sample			Random assignment			Combining			Logical edit			Probabilistic		
	Coef	SE		Coef	SE	Rel. Bias	Coef	SE	Rel. Bias	Coef	SE	Rel. Bias	Coef	SE	Rel. Bias
<i>Asians</i>															
Naturalized															
LPRs	-0.79	0.02		-0.79	0.02	-0.01	-0.80	0.02	0.01	-0.79	0.02	-0.01	-0.79	0.02	0.00
LNIs	-1.61	0.08		-0.71	0.09	-0.56				0.46	0.05	-1.29	-1.68	0.08	0.04
Undocumented	-0.65	0.02		-0.70	0.02	0.08	-0.70	0.02	0.09	-1.03	0.03	0.59	-0.64	0.02	-0.02
<i>Latinos</i>															
Naturalized															
LPRs	-0.80	0.01		-0.80	0.01	0.00	-0.81	0.01	0.00	-0.80	0.01	0.00	-0.80	0.01	0.00
LNIs	-2.39	0.14		-1.50	0.13	-0.37				0.58	0.03	-1.24	-2.31	0.15	-0.04
Undocumented	-1.47	0.02		-1.48	0.02	0.01	-1.48	0.02	0.01	-1.62	0.02	0.10	-1.47	0.02	0.00

and the estimate for undocumented immigrants was 2% lower. Combining the LNI and undocumented into a single group produced a coefficient 9% higher than the one from the donor sample. The approach with the largest bias was the logical edit. The estimate for LNIs was actually in the opposite direction ( $\beta=0.46$ ), erroneously suggesting that LNIs had higher odds for having health insurance compared to citizens. The estimate produced for the undocumented by the logical edit approach was higher by 59%. The logical edit performed even worse than random assignment into the LNI and undocumented categories.

For Latinos, all groups had lower odds of being insured compared to naturalized citizens. LNIs had the lowest odds of having current health insurance compared to US citizens ( $\beta=-2.39$ ;  $OR=0.09$ ), followed by the undocumented ( $\beta=-1.47$ ;  $OR=0.23$ ) and LPRs ( $\beta=-0.80$ ;  $OR=0.49$ ). As with Asians, the probabilistic approach produced the smallest bias for Latino immigrants. The estimate for the logged odds of current health insurance of LNIs relative to US citizens was 4% higher compared to the donor sample and less than 1% higher for the undocumented. Combining LNIs and undocumented immigrants overestimated the donor sample-derived effect by only 1%. The logical edit produced the largest biases. While the coefficient for undocumented immigrants was only 10% higher, the relative bias was for LNIs 124% and was in the opposite direction of the coefficient from the donor sample.

Table 3 compares the assignment strategies estimating differences in fair/poor self-rated health by immigration status. In the donor sample, all Asian groups had higher odds of reported fair/poor SRH compared to naturalized citizens. LNIs had the highest odds ( $\beta=1.25$ ;  $OR=3.50$ ), then the LPRs ( $\beta=0.13$ ;  $OR=1.14$ ) and undocumented ( $\beta=0.09$ ;  $OR=1.09$ ). The probabilistic approach had the smallest relative biases for Asians; the coefficient for LNIs was 4% higher than the estimate from the donor sample and 6% higher for undocumented immigrants. The logical edit approach underestimated the estimate from the donor sample for LNIs by 79% and overestimated it by 114% for undocumented. Combining LNIs and undocumented produced the highest bias, this approach overestimated the logged odds of fair/poor health rated health by 133% percent. This bias was even higher than random assignment.

For Latinos, all groups had lower odds for fair/poor health rated health than naturalized citizens. The probabilistic approach had the lowest bias, the estimates for LNIs and undocumented immigrants were 3% off from the estimate from the donor sample. The logical edit produced a small bias for undocumented immigrants (around 1%), but a sizeable bias for LNIs (87%). Combining LNIs and undocumented produced a small bias; this approach underestimated the effect from the donor sample by 7%.

## Discussion

This paper compared three assignment strategies (combining LNIs/undocumented, the logical edit, and probabilistic) to assign undocumented status within non-LPRs across two health outcomes: current health insurance and fair/poor self-rated health



**Table 3** Simulated coefficients of migration status in models predicting fair/poor self-rated health by assignment method

	Donor sample			Random assignment			Combining			Logical edit			Probabilistic		
	Coef	SE	Rel. Bias	Coef	SE	Rel. Bias	Coef	SE	Rel. Bias	Coef	SE	Rel. Bias	Coef	SE	Rel. Bias
<i>Asians</i>															
Naturalized															
LPRs	0.13	0.02		0.13	0.02	-0.03	0.14	0.02	0.02	0.13	0.02	0.00	0.14	0.02	0.03
LNIs	1.25	0.10		0.18	0.16	-0.86				0.26	0.05	-0.79	1.30	0.10	0.04
Undocumented	0.09	0.04		0.18	0.04	1.16	0.20	0.04	1.33	0.18	0.06	1.14	0.09	0.04	0.06
<i>Latinos</i>															
Naturalized															
LPRs	-0.05	0.02		-0.05	0.02	0.00	-0.04	0.02	-0.10	-0.05	0.02	-0.04	-0.05	0.02	0.00
LNIs	-0.16	0.20		-0.22	0.20	0.40				-0.02	0.05	-0.87	-0.16	0.20	0.03
Undocumented	-0.21	0.03		-0.20	0.03	-0.03	-0.19	0.03	-0.07	-0.21	0.03	0.01	-0.20	0.03	-0.03

among Asian and Latino immigrants. All three approaches produced unbiased estimates for the Latino undocumented population for both health outcomes. This offers some assurance that estimates for the undocumented population when only LPR status is known should be comparable despite using different approaches. While estimating the Latino LNI population is more problematic, their small proportion suggests that this group may be too small to impact overall results.

The estimates were more biased for Asian immigrants, likely because the share of LNIs among Asians non-LPRs is higher. While the probabilistic approach produced unbiased estimates for undocumented immigrants for both outcomes, we caution that the coefficients included in the prediction equation may not be available in all datasets (e.g., detailed occupation, homeownership). The combination approach produced an unbiased estimate for undocumented health insurance, but not for self-rated health.

In general, our results did not provide strong support for the logical edit approach as a preferred imputation approach. This was especially true for Asian immigrants; the logical edit approach produced biases that were no better than random assignment for either outcome. We acknowledge, however, that our logical edit approach did not include spousal immigration status, which has been used by other researchers (Borjas, 2017) and may improve validity. However, our work aligns with others who have found logical edit approaches to either mis-categorize undocumented immigrants (Sohn & Pebley, 2020) or have diverging estimates from other approaches (Altman et al., 2020; Spence et al., 2020). The logical edit approach may be subject to bias because the proxies for the statuses are themselves associated with certain outcomes. For instance, using an indicator such as duration in the US for undocumented immigrants can introduce bias if duration is also associated with a health outcome.

There are some limitations to our findings. Our data may not reflect newer trends in immigration policy. More recent SIPP panels do not allow for detailed immigration status coding, however. While meant to be nationally representative, the SIPP oversamples low-income adults to assess their participation in federal social service programs and we acknowledge that our descriptive data on undocumented immigrants vary from other sources, such as the American Community Survey. While general response rates ranged between 19 and 30%, undocumented immigrants may also have been less likely to participate in the SIPP. Yet other work has found that the detailed migration questions in the SIPP, which are presumably riskier for undocumented immigrants, do not have a higher level of missing compared to others (Bachmeier et al., 2014).

As interest in undocumented immigrants and health grows, researchers are looking to find data sources with detailed migration information to investigate their questions. This paper suggests that estimates for Latino undocumented immigrants are robust across different estimation strategies. For researchers using this type of data, combining LNI and undocumented immigrants to approximate the population of Latino undocumented immigrants should be acceptable. Our results underscore the difficulty in studying Asian undocumented immigrants in representative survey data. We suggest that future research on this population use on primary data sources to ensure accurate categorization of immigration status.

## Appendix 1. Criteria for Temporary Visa Holders and Refugees

### Temporary Visa Holders

\* Among non-LPRs only; limited to those under 6 years of residence, unless otherwise noted.

#### A Visas (Diplomats)

Occupational classification codes: General and operations managers; Managers all others.

Industry code: National security and international affairs.

#### F Visas (Students)

Age: 18–64.

Hours worked: part-time.

Educational status: Currently enrolled student; college or above.

Occupational classification: Architects; surveyors, cartographers, and photogrammetrists; all engineer type; dentists; dieticians/nutritionists, audiologists, podiatrists, audiologists, veterinarians,

Industry code: Research, development, and testing services; educational services.

#### J Visas (Exchange Visitors)

Employed.

Not a private for-profit employee.

Occupational classification: Architects; surveyors, cartographers, and photogrammetrists; all engineer type; veterinarians, dentist; physical therapist; audiologist; occupational therapist; physical therapist; health diagnosing and treating practitioners; radiation therapist; speech-language pathologist; medical scientists.

Industry: Scientific research and development services; college and universities; hospitals; nursing care facilities; other health care services; outpatient care centers.

Less than three years of duration.

#### H1 Visas (Nurses)

Employed.

Occupation: Licenses practical and licensed vocational nurses.

Industry: hospitals; nursing care facilities; other health care services; outpatient care centers.

Less than 3 years of US duration.

### **H1B Visas (High-Tech Workers)**

Employed.

Private, for-profit employee.

College degree.

Occupation: Accountants, auditors; budget analyst; engineers all types; scientists all types; economics, sociologists, post-secondary teachers; reporters; physicians/surgeons; technicians all types.

### **G1 Visa (International Organization)**

Employed.

Industry: National security and international affairs.

Occupation: Librarian, lawyer, secretary/administrative assistant; word processor/typist; desktop publisher; office machine operator; office/administrative support; answering service; telephone operator; computer operator; driver/chauffeurs.

### **R1 Visa (Religious Worker)**

Employed.

Less than 3 years of duration.

Occupation: clergy; directors, religious activity and education; religious workers.

### **O1 and P1 (Athletes and Entertainers)**

Employed.

Less than 3 years of duration.

Occupation: artists and related workers; designers; actors; producers and directors; dancers and choreographers; musicians, singers, and related workers; entertainer and performers, sports and related workers.

### **J1 Visas (HS exchange Students and Au Pairs)**

Unmarried (for HS exchange students only).

Between 14 and 20 (for HS exchange students only).

Currently enrolled in high school (for HS exchange students only).

No children (for au pairs only).

Occupation: Childcare workers, personal and home care aides (for au pairs only).  
Not head of household.

## Refugees

Probability of being refugee based on panel year and country of origin. Individuals granted asylum by country, divided by the total number of permanent visas granted in the same period for that country. Data from the Department of Homeland Security Immigration Yearbook.

### 2001 (Period Between 1999 and 2003)

- \* Afghanistan:  
Refugee probability: 100%
- \* Armenia.  
Refugee probability: 2.4%
- \* Cambodia.  
Refugee probability: 0.3%
- \* China.  
Refugee probability: 0.01%
- \* Colombia.  
Refugee probability: 0.2%
- \* Cuba.  
Refugee probability: 10.6%
- \* Indonesia.  
Refugee probability: 0.8%
- \* Laos.  
Refugee probability: 2.4%
- \* Pakistan.  
Refugee probability: 0.04%
- \* Vietnam.  
Refugee probability: 7.7%

### 2004 (Period Between 2002 and 2006)

- \* Afghanistan:  
Refugee probability: 42.4%
- \* Armenia.  
Refugee probability: 2.6%
- \* Cambodia.  
Refugee probability: 0.2%

- \*Colombia.  
Refugee probability: 1.0%
- \*Cuba.  
Refugee probability: 10.5%
- \* Indonesia.  
Refugee probability: 0.4%
- \* Laos.  
Refugee probability: 100%
- \* Pakistan.  
Refugee probability: 0.1%
- \* Vietnam.  
Refugee probability: 1.2%

### 2008 (Period Between 2006 and 2010)

- \* Afghanistan:  
Refugee probability: 19.2%
- \* Armenia.  
Refugee probability: 0.6%
- \* Cambodia.  
Refugee probability: 0.3%
- \*Colombia.  
Refugee probability: 0.3%
- \*Cuba.  
Refugee probability: 10.1%
- \* Indonesia.  
Refugee probability: 0.1%
- \* Laos.  
Refugee probability: 10%
- \* Pakistan.  
Refugee probability: 0.2%
- \* Vietnam.  
Refugee probability: 0.8%

### Appendix 2. Logged Odds of Being Undocumented among non-LPRs, 2001/2004/2008 SIPP

	Asians		Latinos	
	Coef	SE	Coef	SE
Male	-0.176	0.0992	-0.095	0.0356
Age	0.085	0.0185	0.005	0.0077

	Asians		Latinos	
	Coef	SE	Coef	SE
Age squared	-0.0001	0.0002	0.001	0.0001
Income to poverty	0.096	0.0208	0.0319	0.0127
Education	0.21	0.009	-0.111	0.0044
Insurance Cov * Age	-0.089	0.0066	-0.0863	0.0041
Insurance Cov * Female	2.598	0.2479	3.802	0.1769
Insurance Cov * Male	4.589	0.256	2.688	0.1553
Insurance Cov * Years in US	0.187	0.0278	0.2055	0.0224
Years in US	0.38	0.0181	0.3239	0.0062
Married	-0.153	0.0803	-0.4071	0.0333
Number in HH	-0.471	0.0163	0.0391	0.007
Own home	1.012	0.1221	-0.4442	0.0392
Employed	-1.58	0.0853	0.1705	0.0351
Region				
NE	Ref			
Midwest	X	x	x	x
South	x	x	x	x
West	x	x	x	x
Disabled	-0.147	0.1182	0.0958	0.0662
Fair/poor health	-2.479	0.1171	-0.0308	0.0687
Engineer/construction	x	x	x	x
Constant	-7.24	0.4707	6.742	0.2156

x = cell sizes too small to disclose

**Author Contributions** Dr. Ro conceptualized the project, contributed to the data analysis plan, conducted data analysis, interpreted data analysis, and wrote the manuscript. Dr. Van Hook contributed to the data analysis plan, interpreted data analysis, and wrote the manuscript.

**Funding** This research was supported by the Russell Sage Foundation (1809-08530).

**Data Availability** The authors do not have the authority to share the dataset. The restricted SIPP is available to qualified researchers with approved projects the US Census Federal Research Data Centers.

**Declarations**

**Conflict of Interest** Dr. Ro and Dr. Van Hook do not have any conflicts of interest or competing interests to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acevedo-Garcia, D., Bates, L. M., Osypuk, T. L., & McArdle, N. (2010). The effect of immigrant generation and duration on self-rated health among US adults 2003–2007. *Social Science & Medicine*, 71(6), 1161–1172. <https://doi.org/10.1016/j.socscimed.2010.05.034>
- Altman, C. E., Spence, C., Hamilton, C., & Bachmeier, J. D. (2020). Health insurance coverage: Logical versus survey identification of the foreign-born. *Journal of Immigrant and Minority Health*. <https://doi.org/10.1007/s10903-020-01045-y>
- Bachmeier, J. D., Van Hook, J., & Bean, F. D. (2014). Can we measure immigrants' legal status? Lessons from two US surveys. *International Migration Review*. <https://doi.org/10.1111/imre.12059>
- Borjas, G. J. (2017). The labor supply of undocumented immigrants. *Labour Economics*. <https://doi.org/10.1016/j.labeco.2017.02.004>
- Census. (2016). *SIPP introduction and history*. <http://www.census.gov/programs-surveys/sipp/about/sipp-introduction-history.html>. Accessed 5 Mar 2021
- Hamilton, E. R., Hale, J. M., & Savinar, R. (2019). Immigrant legal status and health: Legal status disparities in chronic conditions and musculoskeletal pain among Mexican-Born farm workers in the United States. *Demography*. <https://doi.org/10.1007/s13524-018-0746-8>
- Pourat, N., Lessard, G., Lulejan, L., Becerra, L., & Chakraborty, R. (2003). *Demographics, health, and access to care of immigrant children in California: Identifying barriers to staying healthy*. University of California, Los Angeles; Center for Health Policy Research, Los Angeles, CA. Available at: [http://www.healthpolicy.ucla.edu/pubs/files/NILC\\_FS\\_032003.pdf](http://www.healthpolicy.ucla.edu/pubs/files/NILC_FS_032003.pdf)
- Sohn, H., & Pebley, A. (2020). *New approaches to estimating immigrant documentation status in survey data*. University of California, Los Angeles; California Center for Population Research. Available at <http://papers.cpr.ucla.edu/index.php/pwp/article/view/1237/616>
- Spence, C., Bachmeier, J. D., Altman, C. E., & Hamilton, C. (2020). The association between legal status and poverty among immigrants: A methodological caution. *Demography*, 57(6), 2327–2335. <https://doi.org/10.1007/s13524-020-00933-0>
- Van Hook, J., Bachmeier, J. D., Coffman, D. L., & Harel, O. (2015). Can we spin straw into gold? an evaluation of immigrant legal status imputation approaches. *Demography*. <https://doi.org/10.1007/s13524-014-0358-x>
- Vargas Bustamante, A., Chen, J., Fang, H., Rizzo, J. A., & Ortega, A. N. (2014). Identifying health insurance predictors and the main reported reasons for being uninsured among US immigrants by legal authorization status. *The International Journal of Health Planning and Management*, 29(1), e83–e96. <https://doi.org/10.1002/hpm.2214>
- Vargas Bustamante, A., Fang, H., Garza, J., Carter-Pokras, O., Wallace, S. P., Rizzo, J. A., & Ortega, A. N. (2012). Variations in healthcare access and utilization among Mexican immigrants: The role of documentation status. *Journal of Immigrant and Minority Health*. <https://doi.org/10.1007/s10903-010-9406-9>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.