

A New Measure of the Likely Authorization Status of Noncitizens in the American Community Surveys: Methodology and Descriptive Results for 2010¹

Robert Warren, Vicky Virgin, Jose Pacas, and John R. Warren

Abstract

We describe a method for inferring the authorization status of foreign-born U.S. residents in the 2010-2013 American Community Surveys (ACS). Based on this method, we construct a new variable that reflects likely authorization status. The variable can be cross-tabulated with any variable now available in the ACS and at whatever level of geographic aggregation desired (e.g., state, county, city, PUMA). Our method overcomes conceptual problems that characterized previous efforts (as described, for example, by Van Hook et al, 2015). The resulting variable should be useful for academic research (for example, on the spatial dispersion or personal attributes of unauthorized people) and for policy and service planning at the regional, state, and local levels.

Introduction

The variable described here was created by the authors as part of a project carried out by the Center for Migration Studies (CMS) with the financial support of the MacArthur Foundation. For convenience, the text and tables below often refer to the data under discussion as “CMS data,” or “CMS database.” This draft was derived primarily from Warren (2014).

Over the past few decades, demographers have made considerable progress in estimating the size of the unauthorized resident population. Passel, Van Hook, and Bean (2004) and Van Hook et al. (2015) provide detailed descriptions of the various methods that evolved beginning in about the mid-1980s. The most widely accepted current estimates rely on residual methods of estimation. Essentially, the residual method involves subtracting estimates of the legally foreign-born resident population from census or survey data for the total foreign-born population; the difference, or residual, is assumed to be the population counted in the census or survey. Usually, adjustments are made to account for under-enumeration.

¹ We are grateful for the valuable comments and suggestions from an advisory group that included Sarah Bohn, Laura Hill, Enrico Marcelli, Rob Paral, Jeffrey Passel, and Audrey Singer. The authors are solely responsible for any errors or omissions.

The residual method has established the contours of the population: the total population is approximately 11 million; just over half are from Mexico; many of the rest are from Central and South American countries; and they are concentrated in the top 10 states, the most populous being California and Texas. The best evidence indicates that unauthorized immigration increased rapidly in most states in the late 1990s to 2000, and then dropped steadily after 2001 (Warren and Warren 2013). The total population has been at or near zero growth during the past five years. It is also known that approximately 40 percent of the total unauthorized resident population entered with temporary visas and overstayed or otherwise violated the terms of their admission (Warren 1997).

With so much information available, one might ask why additional information about unauthorized immigrants is needed. Although significant progress *has* been made over the past few decades, it is increasingly clear that the residual method does not provide sufficiently detailed characteristics of the population needed for many purposes. Until recently, most of the information that has been available about unauthorized immigrants has been produced by a handful of statisticians and disseminated in the form of reports that show selected data points at the national and sometimes the state level. In addition to the limited scope of the residual estimates, the quality of the data and assumptions underlying the residual method are becoming more tenuous each year.²

² For example, in the estimates published by the Pew Research Center and the Department of Homeland Security (DHS) in the past few years, legal immigrants admitted more than *30 years* ago are “survived” to current dates based on emigration rates—assumed to be applicable to all countries and states—that are increasingly out of date. In addition, the lack of data on internal migration of *legal* immigrants affects the accuracy of the state estimates.

The limitations inherent in the residual method have led to efforts in recent years to derive individual-level estimates by assigning probable legal status to survey respondents. Passel and Clark (1998) were the first to provide population estimates of immigrants by legal status. The Pew Research Center has used this approach to estimate the social and demographic characteristics of unauthorized residents using data from the Current Population Survey (CPS) sponsored by the US Census Bureau and the US Bureau of Labor Statistics (Passel and Cohn 2008), and, more recently, from the ACS (Passel and Cohn 2014). Capps et al. (2013) combined procedures developed by Passel and Clark, information on legal status from the relatively small US Census Bureau Survey of Income and Program Participation (SIPP), and multiple imputation techniques³ to assign legal status to respondents in the ACS. In both cases, the results were controlled to a residual estimate of the national population. In contrast, we employed a technique of controlling the initial estimates—the number counted in the micro data collected by the ACS in 2010—to separate independent population controls for 145 countries or areas. A detailed comparison of the methods used to construct prior estimates is beyond the scope of this paper. However, the analysis of the data indicates that the technique represents a significant improvement in this type of approach.

Overview of the methodology

The estimates are based partly on the reported characteristics of non-US citizens (henceforth, noncitizens) in the micro data of the 2010 ACS. The ACS is an annual statistical survey covering approximately one percent of the total US population. The survey gathers information previously obtained in the decennial census—the ACS questionnaires are very similar to the 2000 census

³ See Van Hook et al. (2015) for an evaluation of the efficacy of using multiple imputation techniques to estimate legal status in surveys.

long form. The survey provides detailed social and economic data for all states, as well as all cities, counties, metropolitan areas, and population groups of 100,000 people or more.

It is assumed that nearly all of the unauthorized resident population contained in the ACS data is comprised of noncitizens that entered the United States after 1981. Very few who entered before 1982 would still be residing in the country as unauthorized residents in 2010 because: (1) a large percentage of those who entered before 1982 obtained legal status under the Immigration Reform and Control Act of 1986 (IRCA);⁴ and, (2) the remainder have had about 30 years in which to leave the unauthorized resident population—that is, to adjust to legal status, be removed, leave voluntarily, or die.

The methodology involved three major steps: (1) applying a series of edits, referred to here as “logical edits,”⁵ to identify as many legal residents as possible based on responses in the survey; (2) deriving separate population controls, for 145 countries or areas, for unauthorized residents counted in the 2010 ACS; and (3) using those population controls to make final selections of individual respondents in the ACS to be classified as unauthorized residents. The following discussion focuses exclusively on *noncitizens that arrived after 1981* unless noted otherwise.

⁴ The Immigration Reform and Control Act of 1986 (IRCA) went into effect in 1987. Two main groups were eligible for legalization, each with their own residency requirements: legalization applicants who continuously resided in the United States since before January 1, 1982 and Special Agricultural Workers (SAWs) who had 60 days of seasonal agricultural work experience in qualifying crops from May 1985 to May 1986. About 1.6 million legalization applicants and 1.1 million SAW applicants were approved.

⁵ The term “logical edit” refers to the process of determining probable legal status by examining survey data. Respondents were assigned to the legal category if they worked in occupations that generally require legal status, were legal temporary migrants, were immediate relatives of US citizens, received public benefits, were from countries where most arrivals would be refugees, or were age 60 or older at entry. This method was developed by Jeffrey Passel and Rebecca Clark (1998), extended by Passel, Bean and Van Hook (2004), and refined by Passel and others in recent years at the Pew Research Center.

Although this discussion focuses on 2010, annual estimates were also derived for 2011 to 2013. The methods used to derive the estimates for all four years are described in the Appendix.

A total of 19.9 million noncitizens that arrived after 1981 were counted in the 2010 ACS.⁶ About 10.9 million, or 55 percent, were unauthorized residents (Warren and Warren 2013). Based on the logical edits, about 7.3 million of the 19.9 million noncitizens were moved into the legal category, leaving 12.6 million as “likely unauthorized.” Finally, we selected 10.9 million of the remaining 12.6 million to be unauthorized residents using the country-by-country population controls, as described in the Appendix. The technique of controlling the likely unauthorized population to 145 country controls produced estimates by age, year of entry, and state of residence that are very similar to the best available residual estimates.

The final step in the methodology was to adjust the estimates for under-enumeration. As described in detail below, the data was adjusted using the same methodology used to construct residual estimates by state (Warren and Warren 2013). The most recent entrants are assumed to have the highest undercount rates (about 12 percent), and the undercount rate drops steadily with length of residence, falling to two percent for those who entered in 1982. The estimated undercount rate for the total population is about 7.5 percent. The Appendix shows additional details.

For most countries the logical edits produce numbers that are fairly close to the control totals. However, the logical edits do not shift *all* of the legal residents out of the likely unauthorized

⁶ Natives of Cuba are omitted because in most cases they are eligible to apply for legal status under the Cuban Adjustment Act.

category. As a result, a small percentage of those remaining in the database actually are legal residents. However, a sizeable proportion of these legal residents had previous experience as unauthorized residents. Overall, the database consists of an estimated 89 percent unauthorized residents, but adding in those with previous unauthorized experience raises the total to 93 percent. The Appendix shows how these percentages were computed and provides more detail.

Sources of data

The estimates for each country used as population controls for the 2010 estimates are based partly on administrative data from the Department of Homeland Security (DHS), estimates and assumptions consistent with Warren and Warren (2013), and a variety of other sources, including foreign-born data from the 2000 Census, legalization applications by country from IRCA, and estimates of nonimmigrant overstays, by country, derived by DHS annually from 1988 to 1992. The Appendix provides a more complete description of the data and methods used to derive the population controls for 145 countries or areas for 2010 as well as the annual estimates for 2011 to 2013.

Public-use samples of individual survey records from the ACS were tabulated to provide the data used in the estimation process. The public-use file is a representative one percent sample of the entire US (including about three million individual records for 2010) obtained from the Integrated Public-Use Microdata Series, or IPUMS (Ruggles et al. 2010).

Validation of the results

Comparisons with other estimates

Comparisons with other estimates address the issue of “face validity”— that is, do the estimates generated by this method conform generally to the accepted parameters of the unauthorized resident population? The first comparisons, shown in Table 1, are not a test of the overall database; they are a comparison of our population controls, by country, with DHS’ residual-based estimates. This section later shows how estimates compiled from this database compare to other widely-accepted residual estimates.

Based on the comparisons shown, the estimates described in this paper are consistent with the best current estimates of the unauthorized resident population. Although all methodological approaches have inherent errors, the close correspondence between estimates derived from such disparate approaches indicates that they are measuring approximately the same population.

Population controls compared to DHS estimates

We chose to control the final results to estimates by country of origin, rather than states, because residual estimates for states are subject to error due to the lack of information about internal migration of legal residents.⁷ Unfortunately, estimates by country of origin were available for only a few countries at the beginning of this project, thus it was necessary to derive estimates for more than 100 countries to be used as population controls.

⁷ In the residual estimates for states, net out-migration of legal residents from entry to the survey date will produce underestimates of the population; net in-migration will produce overestimates. To produce accurate estimates for California, for example, it would be necessary to know the cumulative net migration of legal immigrants in every year from 1980 to the survey date. Residual estimates by country of origin are unaffected by the lack of data on internal migration.

Table 1 shows a comparison between our population controls and DHS' estimates for their top ten countries in 2010. The two sets of estimates are generally consistent with each other. The estimates for these ten countries are important because they make up more than 80 percent of the unauthorized resident population.

Table 1. Comparison of CMS Population Controls with DHS Estimates: 2010

Numbers in thousands; rounded independently

Country	Unauthorized population		Percent of total population	
	CMS 2010 (1)	DHS 2010 (2)	CMS 2010 (3)	DHS 2010 (4)
All countries	11,725	11,590	100%	100%
Mexico	6,600	6,830	56%	59%
El Salvador	615	670	5%	6%
Guatemala	465	520	4%	4%
Honduras	315	380	3%	3%
Philippines	245	290	2%	3%
India	365	270	3%	2%
China	205	220	3%	3%
Korea	325	300	2%	2%
Ecuador	175	210	1%	2%
Vietnam	110	190	1%	2%
All other	2,300	1,710	20%	15%

Source: CMS population controls, see text; Baker and Rytina (DHS) 2013, Table 3, revised estimates for 2010 based on the 2010 ACS data.

These estimates compared to published estimates for 2010

This section compares the estimates compiled from the database described here for 2010 with estimates for 2010 derived using the residual method. Note that in Tables 2, 3, and 4 below our estimates are compilations of micro data, with the exception of the estimate for Mexico in Table 2. The estimate for Mexico is a control total computed as described in the Appendix.

Table 2. Comparison of CMS data with Residual Estimates of Unauthorized Residents, for Selected Areas: 2010

Numbers in thousands; rounded independently

Area	<i>Estimated unauthorized resident population in 2010</i>			
	CMS (1)	Warren and Warren (2013) (2)	DHS (2013) (3)	Pew Research Center (2013) (4)
US total	11,725	11,725	11,570	11,400
California	2,915	2,935	2,910	2,500
Texas	1,720	1,610	1,780	1,650
Florida	785	1,000	730	900
New York	915	705	690	800
Illinois	595	585	550	500
New Jersey	485	410	440	500
Georgia	395	395	430	n/a
North Carolina	335	370	390	n/a
Arizona	300	340	350	n/a
Washington	240	260	260	n/a
Other states	3,040	3,115	3,040	n/a
Mexico	6,600	n/a	6,830	6,150
Pct. from Mexico	56%	n/a	59%	54%

Source: See references shown in column headings.

The estimates shown in Columns 2, 3, and 4 of Table 2 are based on estimates derived by Warren and Warren (2013), DHS (Baker and Rytina 2013), and Pew Research Center (Passel and Cohn 2013), respectively. The estimates for 2010 shown in Table 2 are generally consistent with those estimates, which were derived using the residual method. Some differences among the estimates are to be expected considering the diverse data and methods used to construct the estimates. The important point is that the estimates described here for 2010 “fit” well with the best available residual estimates.

Age distribution

The age distribution of the unauthorized resident population derived here for 2010 is similar to the published DHS estimates for 2012 (Table 3). In both sets of estimates, about 83 percent of the population is under age 45. The DHS estimates have relatively fewer under age 25 (22 percent compared to 27 percent) and more in the 25 to 44 age range (61 percent compared to 56 percent). The percentages in the 45 to 54 and the 55 and over age ranges are identical (Table 3).

Table 3. Age Distribution of the Unauthorized Resident Population Derived from the CMS Estimates for 2010 Compared to DHS Estimates for 2012

Numbers in thousands; rounded independently

Age	<u>Estimated unauthorized population</u>		<u>Percent of the total population</u>	
	CMS 2010 (1)	DHS 2012 (2)	CMS 2010 (3)	DHS 2012 (4)
Total	11,725	11,430	100%	100%
Under 18	1,415	1,120	12%	10%
18-24	1,785	1,410	15%	12%
25-34	3,610	3,660	31%	32%
35-44	2,945	3,320	25%	29%
45-54	1,420	1,400	12%	12%
55+	550	520	5%	5%

Source: 2010 CMS, see text; 2012 DHS, from Baker and Rytina (2013), Table 5.

Year of entry

Comparable estimates of unauthorized immigration by period of entry are not readily available because of differences⁸ in data and methodology, and because the distributions by period of entry can change in just a few years. Nevertheless, the figures shown in Table 4 are roughly comparable for the time periods shown.

⁸ For example, the estimates in Table 4, column 3 were derived from CPS data controlled to estimates based on updated 2000 Census data, while the data in Columns 2 and 4 were derived from ACS data controlled to 2010 Census data. Note also that data are not shown for the 2005 to 2009 period for Pew 2009 and DHS 2012 because (a) the Pew estimates refer to March 2009, so the period is incomplete; and (b) the source of the DHS 2012 estimates, Table 1, shows only the 2005 to 2011 period.

Table 4. Estimated Unauthorized Resident Population, by Period of Entry: 1980 to 2009
Rounded to thousands

Period of entry	CMS 2010 (1)	Warren 2010 (2)	Pew 2009 (3)	DHS 2012 (4)
2005 to 2009	3,100	2,500	n/a	n/a
2000 to 2004	3,600	3,500	3,500	3,250
1990 to 1999	3,700	n/a	4,500	4,640
1980 to 1989	1,100	n/a	1,400	2,000

Source of Columns: (1) CMS 2010, see text; (2) Warren 2010 from Warren and Warren (2013); (3) Pew 2009 from Passel and Cohn (2010), Table 5; (4) DHS 2012 from Baker and Rytina (2013), Table 1.

Estimates for states

Table 5 shows our estimates for every state and the District of Columbia along with estimates derived by Warren and Warren (2013) using the residual method. The two distributions of unauthorized residents are remarkably similar considering the fundamental differences in the methods of estimation. Note that the estimates for each state shown in Table 5 were *not* controlled to marginal state controls; they are compilations from the database.

Table 5. Estimated Unauthorized Resident Population Derived from the CMS Database Compared to Estimates Derived Using the Residual Method: 2010

In thousands; states might not add to total because of rounding

State of Residence	CMS estimates	Warren and Warren	State of residence	CMS estimates	Warren and Warren
US total	11,725	11,725	Missouri	60	70
Alabama	90	95	Montana	1	1
Alaska	10	5	Nebraska	35	40
Arizona	300	340	Nevada	185	180
Arkansas	60	70	New Hampshire	10	2
California	2,915	2,935	New Jersey	485	410
Colorado	180	195	New Mexico	80	85
Connecticut	115	110	New York	915	705
Delaware	20	20	North Carolina	335	370
DC	15	25	North Dakota	1	z
Florida	785	1,000	Ohio	95	100
Georgia	395	395	Oklahoma	90	95
Hawaii	40	15	Oregon	130	130
Idaho	35	30	Pennsylvania	150	150
Illinois	595	585	Rhode Island	25	30
Indiana	105	110	South Carolina	95	105
Iowa	45	45	South Dakota	5	1
Kansas	80	75	Tennessee	120	135
Kentucky	50	50	Texas	1,720	1,610
Louisiana	55	60	Utah	90	100
Maine	2	2	Vermont	2	z
Maryland	240	225	Virginia	250	265
Massachusetts	170	195	Washington	240	260
Michigan	95	90	West Virginia	2	1
Minnesota	95	100	Wisconsin	80	80
Mississippi	25	25	Wyoming	5	5

Source: CMS estimates compiled from CMS database; Warren and Warren estimates from Warren and Warren (2013).
z rounds to zero.

Unauthorized resident population in 2013

The main purpose of this article is to describe the construction of the database. However, this article includes the estimates shown in Tables 6 and 7 below to illustrate the kinds of information that can be compiled. Table 6 shows the states that have the highest number of unauthorized residents from the countries listed under the state. For example, even though many states have

large numbers from Mexico, California has the *largest* population from Mexico, about 30 percent of the total number in the United States (Table 6). As might be expected, Illinois has the highest number of unauthorized residents from Poland (44 percent).

Table 6. States with the Largest Numbers of Unauthorized Residents, by Country of Origin: 2013

(Numbers in thousands)

Minimum of 4,500 unauthorized residents

State and Country	Unauthorized pop.		Pct. of total (3)=2/1	State and Country	Unauthorized pop.		Pct. of total (6)=5/4
	In the US (1)	In the State (2)			In the US (4)	In the State (5)	
California				Florida			
Mexico	6,088	1,806	30%	Haiti	122	81	66%
El Salvador	588	151	26%	Colombia	133	51	38%
China	359	103	29%	Jamaica	104	36	35%
Philippines	246	100	41%	Venezuela	57	29	51%
Korea	187	62	33%	Nicaragua	55	25	45%
Iran	16	9	56%	Argentina	38	11	29%
Egypt	25	7	28%	Uruguay	16	8	50%
Armenia	6	5	83%				
Fiji	8	5	63%	New York			
Indonesia	11	5	45%	Dominican Rep.	194	78	40%
				Ecuador	144	63	44%
Illinois				Jamaica	104	31	30%
Poland	66	29	44%	Trinidad & Tob.	30	15	50%
				Guyana	25	15	60%
Virginia				Bangladesh	26	8	31%
Bolivia	23	12	52%				

Source: CMS data; see text.

Table 7 shows estimates of the unauthorized resident population for three major sending areas in the Western Hemisphere, cross-classified by the top six states of residence of the unauthorized resident population for each year from 2010 to 2013. The annual estimates for large sending areas and the top six states indicate that, at least at this level of aggregation, the methodology produces stable results from year to year. The total population declined each year from 2010 to 2013, but the rate of decline slowed each year. The population from Mexico fell by about a half

million from 2010 to 2013, a reduction of about eight percent. Population decline was the largest for South American countries; the population dropped from 820,000 in 2010 to 640,000 in 2012, a decline of 180,000, or 22 percent, in just two years. The population from Central America was largely unchanged from 2010 to 2013.

Table 7. Annual Estimates of the Unauthorized Resident Population from All Countries, Mexico, Central America, and South America, for the Top 6 States: 2010 to 2013

Numbers in thousands

Ranked on state estimates for all countries in 2013

Area and State	2010 (1)	2011 (2)	2012 (3)	2013 (4)	Change, 2010 to 2013	
					Number (5) = 4-1	Percent (6) = 5/1
All countries	11,725	11,315	11,110	11,010	-715	-6%
California	2,915	2,825	2,765	2,665	-250	-9%
Texas	1,720	1,710	1,650	1,725	10	z
New York	915	865	855	820	-100	-11%
Florida	785	755	725	705	-85	-11%
Illinois	595	545	540	530	-65	-11%
New Jersey	485	465	475	475	-10	-2%
All other states	4,305	4,150	4,100	4,090	-220	-5%
Total Mexico	6,600	6,360	6,115	6,090	-515	-8%
California	2,015	1,930	1,850	1,805	-210	-10%
Texas	1,330	1,320	1,250	1,330	z	z
New York	200	190	175	180	-20	-11%
Florida	175	200	165	155	-25	-14%
Illinois	395	390	385	355	-40	-10%
New Jersey	110	95	90	100	-10	-9%
All other states	2,375	2,230	2,195	2,165	-210	-9%
Total Central America	1,505	1,495	1,515	1,520	15	1%
California	340	330	345	325	-15	-5%
Texas	190	195	195	195	10	4%
New York	115	110	135	125	10	8%
Florida	150	145	150	160	10	6%
Illinois	20	15	15	15	-5	-30%
New Jersey	75	75	80	80	5	9%
All other states	615	620	600	620	5	1%
Total South America	820	715	640	640	-180	-22%
California	55	45	40	40	-15	-28%
Texas	35	25	20	30	-5	-14%
New York	150	140	135	125	-25	-17%
Florida	210	170	150	155	-55	-26%
Illinois	15	15	10	15	z	4%
New Jersey	115	100	95	95	-20	-19%
All other states	240	220	185	180	-55	-23%

Source: DHS database; see text for methods of estimation.

z rounds to zero.

Discussion

Uses of the data

This data should be useful for many purposes, from planning and evaluating changes in US immigration policy to assisting states, counties, cities, and private entities that serve immigrants. Although much attention has been focused on President Obama's executive actions to defer deportation of millions of unauthorized residents, the information in this database will be particularly valuable to service providers at the local level.⁹ For example, the data tool will provide information about unauthorized residents at the local level for such variables as age, gender, years in the US, languages spoken, ability to speak English, educational attainment, occupation, poverty status, and legal status of parents and children.

Local area data

Organizations that provide services to immigrants will need the following kinds of information available from this database: the number of unauthorized residents in the area, their countries of origin, their ability to speak English, which languages they speak, how many children they have, what percent are living below the poverty level, and (beginning with the 2013 estimates) how many have access to a computer or the internet. To make sure that such local-area data are viable, the estimates were evaluated by comparing them to the best and most detailed information available.

⁹ In this context, "local level" refers to areas that have approximately 100,000 or more total residents.

Strengths and limitations of the data

There are several strengths of the data described here:

1. Controlling the data (after the logical edits) for 2010 to control totals for 145 countries or areas is a new and important step in constructing estimates of the unauthorized. The estimates produced by applying this technique are consistent with other information available about the unauthorized resident population.
2. Unauthorized residents constitute a high percentage of the population included in the database, as described in the Appendix, making it an excellent resource for researchers, policymakers, and NGOs service providers.
3. The large sample size of the ACS makes it possible to compile data for unauthorized residents for geographic areas with population sizes of 100,000 or more.
4. The individual-level data available from the ACS makes it possible to measure the social and economic conditions of the unauthorized resident population for such variables as income, education, health care coverage, and language usage.
5. The methods used to construct the database for 2010 and 2011 can be used for several years to produce similar annual estimates based on ACS data (see Appendix).

Potential limitations include:

1. Although the database contains a high percentage of unauthorized residents, inevitably some legally resident noncitizens are also included. However, as explained below, this variation in percent unauthorized by country of origin should have little effect on estimates of unauthorized residents compiled for most purposes.

2. Although the logical edits have been reviewed carefully, misclassification errors are inevitable. In cases where *some* error would occur regardless of the choice for a logical edit, we made the choice that was likely to introduce the least amount of error. For example, all noncitizens who were aged 60 and over at entry are assumed to be legal residents. While a small number *could* have been unauthorized, the available evidence indicates that very few unauthorized residents arrive at ages 60 and over.
3. The data are based on a sample, and sampling variability must be taken into account, especially as the size of the population or the geographic area under study decreases.
4. If comprehensive legislative reform were to be enacted, the logical edits described here would have to be reviewed and revised. It is possible that the characteristics specified in the legalization program could be used to modify the methodology so that estimates could continue to be made.

Privacy of respondents

A primary consideration in this project was the protection of the privacy of all ACS respondents. In most cases, access to the information about unauthorized immigrants will be limited to the creation of tables through the CMS website. In rare cases, the ACS micro data used to compile tables might be made available, but only upon request, after consultation with expert advisors, and after the establishment of a confidentiality (non-sharing) agreement.

From a demographic perspective, the possibility of identifying individuals or groups of unauthorized immigrants from the database is extremely unlikely for three primary reasons:

1. The annual ACS is an approximately one in 100 sample, and these estimates are restricted to areas with population sizes of 100,000 or more. In a population that large, individuals or small groups of unauthorized residents are likely to be indistinguishable from legal foreign-born residents.
2. Almost three quarters (73 percent) of the total foreign-born population in the United States are legal residents. The legally resident foreign-born population outnumbered unauthorized residents in *every* state and the District of Columbia, making it extremely unlikely that individuals or groups of unauthorized residents can be distinguished from legal foreign-born residents.¹⁰ In the 10 states with the lowest numbers of unauthorized residents, more than 90 percent of the foreign-born population consists of legal residents.
3. As with the residual estimates of unauthorized residents, this database contains a small percentage of legal residents, making it unlikely that any particular population group is composed exclusively of unauthorized residents.

Sampling variability

Estimates of sampling variability will be available, along with the estimates of the unauthorized resident population, at the CMS website: <http://cmsny.org>. A detailed discussion of the accuracy of 2010 PUMS data, including measures used to assure confidentiality, weighting and estimation, measuring sampling error, and design factors, is available from US Census Bureau (2010) at:

http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2010AccuracyPUMS.pdf.

¹⁰ For each state, the source of data for unauthorized residents is Warren and Warren 2013, and the source of data for total foreign-born population by state is the 2010 ACS data from IPUMS.

Conclusion

This data base makes it possible for a broad cross-section of users to tabulate detailed geographic, social and economic data on the unauthorized and compare that data to variables collected in the ACS. Historically, most of the information that has been available about unauthorized immigrants has been produced by a handful of statisticians and disseminated in the form of reports that show selected data points at the national and sometimes the state level. These estimates are highly consistent with the most widely-accepted estimates derived using the residual estimation methods while providing detailed information for each figure at population units as small as 100,000 persons.

Increasing access to data on legal and unauthorized residents has significant advantages for policymakers, government entities, and NGOs that legislate, craft, implement and evaluate programs that serve immigrants. In the context of a legalization program, detailed, sub-state data will allow stakeholders to determine: (1) how to allocate resources, including staffing; (2) which populations would likely need assistance with fees and would struggle to meet income requirements; (3) where to expand charitable legal service capacity and English language classes; (4) how to conduct outreach, in what languages and where; and (5) which immigrants might have other relief available.

References

- Baker, Bryan C. and Nancy Rytina. 2013. *Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2012*. Washington, DC: Office of Immigration Statistics, Policy Directorate, US Department of Homeland Security.
http://www.dhs.gov/sites/default/files/publications/ois_ill_pe_2012_2.pdf.
- Bean, Frank D., Barry Edmonson, and Jeffrey S. Passel, eds. 1990. *Undocumented Migration to the United States: IRCA and the Experience of the 1980s*. Washington, DC: Urban Institute Press.
- Capps, Randy, Bachmeier, James, Fix, Michael, and Van Hook, Jennifer. 2013. *A Demographic, Socioeconomic, and Health Coverage Profile of Unauthorized Immigrants in the United States*. Washington, DC: Migration Policy Institute.
- GAO (US Government Accountability Office).1995. *Illegal Immigration: INS Overstay Estimation Methods Need Improvement*. GAO/PEMD-95-20.
<http://www.gao.gov/products/PEMD-95-20>.
- Jasso, G., Massey, D. S., Rosenzweig, M. R. and Smith, J. P. 2008. “From Illegal to Legal: Estimating Previous Illegal Experience among New Legal Immigrants to the United States.” *International Migration Review* 42: 803–43.
- Passel, Jeffrey S. and Rebecca L. Clark. 1998. *Immigrants in New York: Their Legal Status, Incomes and Taxes*. Washington, DC: Urban Institute.
- Passel, Jeffrey S., Jennifer Van Hook, and Frank D. Bean. 2004. *Estimates of the Legal and Unauthorized Foreign-born Population for the United States and Selected States, Based on Census 2000*. Washington, DC: US Census Bureau.
- Passel, Jeffrey S. and D’Vera Cohn. 2008. *Trends in Unauthorized Immigration: Undocumented Inflow Now Trails Legal Inflow*. Washington, DC: Pew Research Center.
- . 2009. *A Portrait of Unauthorized Immigrants in the United States*. Washington, DC: Pew Research Center.
- . 2010. *US Unauthorized Immigration Flows Are Down Sharply Since Mid-Decade*. Washington, DC: Pew Research Center.
- . 2013. *Population Decline of Unauthorized Immigrants Stalls, May Have Reversed*. Washington, DC: Pew Research Center.
- . 2014. *Unauthorized Immigrant Totals Rise in 7 States, Fall in 14*. Washington, DC: Pew Research Center.

Riosmena, Fernando. 2013. "At the Edge of US Immigration's 'Halt of Folly': Data, Information, and Research Needs in the Event of Legalization." *Journal on Migration and Human Security* 1(4): 148-62. <http://dx.doi.org/10.14240/jmhs.v1i4.18>

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota.

US Census Bureau. 2010. PUMS Accuracy of the Data (2010), available at: http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2010AccuracyPUMS.pdf.

Van Hook, Jennifer, James Bachmeier, Donna Coffman, Ofer Harel. Forthcoming. "Can We Spin Straw into Gold? An Evaluation of Immigrant Legal Status Imputation Approaches." *Demography*, February 2015.

Warren, Robert and Warren, J. R. 2013. "Unauthorized Immigration to the United States: Annual Estimates and Components of Change, by State, 1990 to 2010." *International Migration Review* 47: 296-329.

Warren, Robert. 2014. "Democratizing Data about Unauthorized Residents in the United States: Estimates and Public-use Data, 2010 to 2013," *Journal on Migration and Human Security*. <http://jmhs.cmsny.org/index.php/jmhs/article/view/38>

Warren, Robert. 1997. "Estimates of the Undocumented Immigrant Population Residing in the United States: October 1996," Joint Statistical Meetings, Anaheim, CA, August 13.

Appendix. Detailed methodology

This section first describes the methods used to derive estimates of the unauthorized resident population *counted* in the 2010 ACS. The estimates are based partly on the characteristics of noncitizens reported in the micro data of the 2010 ACS. It is assumed that nearly all of the unauthorized immigrant population is in the ACS data for noncitizens who entered the US after 1981 because (a) pre-1982 entrants could have legalized under IRCA and (b) those who did not do so have had 28 years in which to leave the unauthorized resident population.¹¹ The final step, described last, was an adjustment for undercount.

Step 1. Logical edits

The first step examined individual characteristics reported in the survey, comparing them to other data and assumptions, to identify those who are likely to be legal residents. The responses of all noncitizens that arrived after 1981 were examined to determine whether they were in any of the following categories:

- a) Occupations that require legal status. Examples include lawyers, judges, police, firefighters, and pilots. Others in this group included government workers, veterans, active members and employees of the Armed Forces, and their dependents.
- b) Legal temporary migrants. Noncitizens in this category, often referred to as nonimmigrants,¹² were selected based on their date of entry and on certain educational

¹¹ Choosing 1982 as the starting date is an example of making the assumption that is likely to produce the least error. Including 1980 and 1981 entrants, as others have done, probably would have added a small number of unauthorized residents, but it most likely would have added an even greater number of legal noncitizens to the database, especially since the ACS data for entrants in 1980 are subject to heaping.

¹² Nonimmigrants are non-US citizens who are admitted legally to the United States for specified, temporary periods. Examples include tourists, students, foreign government officials, and temporary workers and their families.

and occupational characteristics. Those who are likely to be enumerated in censuses or surveys are included; tourists and other short-term visitors are not.

- c) Immediate relatives of US citizens. Noncitizens who were immediate relatives of US citizens were assumed to be legal residents *except* noncitizens from Mexico, El Salvador, Guatemala, and Honduras who arrived after 2000.¹³
- d) Receiving public benefits. Noncitizens were classified as legal if they reported receiving public benefits that are restricted to legal residents. Examples include Medicare, Social Security, Medicaid (for certain recipients¹⁴), Supplemental Security Income (SSI), and Temporary Assistance for Needy Families (TANF).
- e) Age 60 or older at entry. Respondents who were 60 years old or older *in the year they entered the US* were considered to be legal residents in the 2010 ACS. This assumption is based on work by Passel and Cohn (2010) and on estimates of unauthorized residents by age published by DHS (2013). In the DHS estimates, 4.5 percent of all unauthorized residents were 55 years and older; in the estimates derived here, 4.7 percent are 55 years and older.
- f) From “refugee countries.” If the number of refugees was *double* the number of new arrivals from a country in a particular year (based on DHS data), the arrivals for that country/year were assumed to be legal residents in the 2010 ACS.

¹³ This exception was made because unauthorized immigrants who entered the US after 2000 from these countries are more likely than others to be subject to the three- and ten-year bars, which prohibit applicants from returning to the United States if they entered the US illegally (that is, they entered without inspection, most likely across the southern border). Note that noncitizens that entered the US after 2000 from these four countries are not *automatically* assumed to be unauthorized residents; they could be classified as legal residents as the result of any of the logical edits.

¹⁴ The following noncitizen Medicaid recipients were classified as legal residents: men age 19 and over; and women age 19 and over who had not had a child in the previous year.

As a result of the logical edits, a total of 7.3 million noncitizens who entered the US after 1981 were judged to be legal residents. After the 7.3 million were shifted into the legal category, about 12.6 million remained in the “likely unauthorized” category. At that point, about 86 percent (10.9 million of 12.6 million) of the noncitizens remaining in the database were unauthorized residents.

Step 2. Population controls for each country

The next major step in the methodology was to derive independent population controls by country of origin. For this project, estimates of the unauthorized resident population from 145 countries¹⁵ or areas of origin were needed; unfortunately, such estimates were available for only a few countries. A variety of demographic techniques were used, including residual estimation procedures, to derive a complete set of estimates. The residual method alone could not produce acceptable estimates for every country primarily because detailed information about emigration—which can vary considerably by country—is not available. Therefore, other relevant data, described below, were used to ensure that the estimate for each country fell within a plausible range. The total number of 10,850,000 counted in the 2010 ACS is based on estimates by Warren and Warren (2013).¹⁶

Provisional estimates were derived for 145 countries using the residual method, as illustrated below in Table A-1. Then, an extensive analysis and evaluation of the estimates for each country

¹⁵ Hereinafter, to simplify the description, the terms “145 countries,” or “each country,” or “all of the countries” are used to refer to the individual countries *or areas* for which estimates of the unauthorized resident population were derived. The countries or areas are listed in IPUMS for 2010 (country codes 15000 to 71042).

¹⁶ See Warren and Warren 2013 for the methodology used to estimate the number of unauthorized residents counted in the 2010 ACS. The estimate of 10,850,000 *counted* in the 2010 ACS was derived by following the same procedures described in that report but without any adjustment for undercount.

was undertaken. The provisional estimates were adjusted so that they fell within certain demographic guidelines. For example, the final estimate for a country could not exceed the number derived in Step 1 for that country. Also, estimates of the percent of unauthorized to foreign-born in 2000 and 2010, by country, were used to calibrate the estimates. Finally, two other sets of data were used as *indirect* indicators of the general magnitude of the unauthorized resident population from various countries and regions of the world: (1) the country-by-country distribution of the three million applicants under IRCA; and (2) DHS' annual estimates of nonimmigrant overstays for every country from 1988 to 1992 (see Bean, Edmonson and Passel 1990; GAO 1995).

Table A-1. Computation of Provisional Residual Estimates for 145 countries and Areas.

A. Foreign-born population, entered annually, 1986¹⁷ to 2009, from 145 countries.

Computation of legally resident population.

1. Legal permanent residents (LPRs) and refugees admitted annually from 1986 to 2009 from 145 countries (from DHS annual Statistical Yearbooks)
2. Emigration of LPRs and refugees. Rates from Warren and Warren (2013): 3.3 percent in the first year in the US; annual rate drops by five percent for each additional year in the US
3. Deaths to LPRs and refugees. Rate in initial year from Warren and Warren (2013): 5.7 per 1,000 in the first year in the US; annual death rate increases by one percent for each additional year in the US (the increase accounts for ageing of the population).
4. Undercount of *legal* residents. Rates from Warren and Warren (2013): five percent in the first year in the US; rate drops by 10 percent for each additional year in the US
5. Nonimmigrants. Data for 2010 from 145 countries (from 2010 DHS Yearbook) for temporary legal residents likely to be counted in censuses and surveys.

B. Legally resident foreign-born population, entered annually, 1986 to 2009, from 145 countries. $B = 1-2-3-4+5$

C. Provisional estimates of unauthorized residents from 145 countries, counted in the 2010 ACS. $C = A - B$

¹⁷ The provisional estimates covered the 1986 to 2009 period, rather than the 1982 to 2010 period, because detailed data for LPRs by country of origin were not readily available from the DHS website for years prior to 1986. The steps taken to convert the provisional estimates to final estimates generated estimates for the entire unauthorized resident population.

Step 3. Selection of unauthorized residents counted in the ACS

Step 3 selected the individual respondents to be included in the database. After applying the logical edits to compile the number of likely unauthorized residents, by country, a random process was used to select the appropriate number, i.e., the population control, for each country. For Mexico, for example, 8.1 million noncitizens arrived after 1981; about 1.6 million were judged to be legal residents based on the logical edits (Step 1), leaving 6.4 million as possible unauthorized residents. The control total for Mexico was 6.1 million (Step 2). In Step 3, about 6.1 million, or 95.6 percent, of the 6.4 million were randomly selected to be in the database. Those not selected were classified as legal residents. Overall, a total of 10.9 million as unauthorized residents were selected; the remaining 9.1 million noncitizens who arrived after 1981 were classified as legal residents. After this selection process was completed, about 89 percent of the total consisted of unauthorized residents.¹⁸

Table A-2 illustrates the estimation procedure for South American countries. The first three columns in the table represent the first major step in the methodology. The procedure began with total noncitizens (column 1) and then identified legal noncitizens based on the logical edits (column 2). Column 3 shows the number remaining as likely unauthorized residents after the logical edits. The population controls for each country are shown in column 4. Note that in the unlikely event that the logical edits had identified *all* of the legal noncitizens, the numbers in columns 3 and 4 would have been identical.

¹⁸ Table A-5, Items A, B, and C, illustrates how this percent was computed for each country.

Table A-2. Selection of Unauthorized Residents from South American Countries to be Included in the 2010 CMS Estimates (data not adjusted for undercount)

Numbers in thousands; rounded independently

Country of birth	Noncitizens, entered 1982 to 2010			In the 2010 CMS estimates	
	Total (1)	Legal based on the logical edits (2)	Remainder after the logical edits (3)=(1)-(2)	Selected based on population controls (4)	Percent of column 3 selected (5)=(4)/(3)
Total	1,449	539	910	755	n/a
Argentina	86	36	50	40	79%
Bolivia	40	13	27	21	79%
Brazil	240	95	145	130	90%
Chile	40	16	24	19	80%
Colombia	317	125	192	155	81%
Ecuador	259	68	190	160	84%
Guyana	67	31	36	28	77%
Paraguay	8	4	4	3	80%
Peru	230	90	140	115	82%
Uruguay	34	8	27	22	82%
Venezuela	121	50	72	60	84%
Other S. America	5	2	3	2	77%

Source: See text.

The numbers included in the database were derived by randomly selecting (from the micro data) the numbers shown in column 4 from those in column 3 (Table A-2). For Brazil, for example, 130,000 (column 4) of the 145,000 in column 3 were selected to be in the database. To summarize the procedures for estimating the number *counted* in 2010: the number of noncitizens who entered from each country after 1981 was compiled; the procedures described above (logical edits) were used to move some of them into the legal category; and the remainder was controlled to country-by-country population totals. The process was repeated for 145 countries or areas.

Annual estimates for 2011 to 2013

The method of constructing the database for unauthorized residents in 2011 is illustrated in Table A-3. Note that Columns 1, 2, and 3 were compiled from 2011 ACS data in the same way as Columns 1, 2, and 3 in Table A-2 above. The percents in Column 4 of Table A-3 are the same percents shown in Table A-2, Column 5. In essence, it is assumed that the logical edits shift the same *percent* of the noncitizen population into the legal category in 2011 as they did in 2010; that is, they produce the same percent unauthorized in 2011 as they did in 2010.¹⁹

Table A-3. Selection of Unauthorized Residents from South American Countries for the CMS Database in 2011 (data not adjusted for undercount)

Numbers in thousands; rounded independently

Country of birth	Noncitizens, entered 1982 to 2011			In the 2011 CMS database	
	Total	Legal based on logical edits	Remainder after the logical edits	Percent of column 3 that are unauthorized	Estimated unauthorized population in 2011
	(1)	(2)	(3)=(1)-(2)	(4)	(5)=(3)*(4)
Total	1,343	550	793	n/a	660
Argentina	75	34	41	79%	33
Bolivia	40	12	27	79%	22
Brazil	221	97	124	90%	111
Chile	46	23	23	80%	18
Colombia	306	139	167	81%	136
Ecuador	233	67	166	84%	140
Guyana	66	35	31	77%	24
Paraguay	8	2	6	80%	4
Peru	202	80	122	82%	100
Uruguay	24	8	16	82%	13
Venezuela	120	51	70	84%	58
Other S. America	4	2	2	77%	1

Source: See text.

¹⁹ Under this assumption, the estimated unauthorized resident population varies from year to year in response to changes in the logically-edited (i.e., mostly unauthorized) noncitizen population.

The estimates of unauthorized residents in 2011 (Table A-3, column 5) were derived by multiplying the percent unauthorized in Column 4 by the likely unauthorized population, that is, the number of noncitizens remaining after the logical edits (Column 3). Then, the numbers in Column 5 were randomly selected from the figures in Column 3. The same estimation procedures used for 2011 were used to derive estimates for 2012 and 2013.

An important aspect of the estimation procedure for 2011 and subsequent years is that it does not rely on administrative data or residual estimates. In addition to avoiding the errors inherent in that data, estimates for the most recent year of ACS data can be derived using this estimation procedure as soon as the public-use data become available.

Adjustment for undercount

The estimates were adjusted for undercount using the approach described by Warren and Warren (2013). The undercount rate is relatively higher for recent arrivals than for those who have been in the US the longest. Unauthorized residents who arrived in the five years before the survey were adjusted by about 12 percent; those who entered before 1986 were adjusted by about 2 percent. The undercount rate used here dropped by about 25 to 30 percent for each additional five years of residence in the US²⁰ The adjustments were made by re-weighting the microdata based on year of arrival from 1982 to the survey date. The result for 2010 is a total population of 11,725,000, the total number estimated by Warren and Warren (2013). The estimated undercount rate for the total population is 7.5 percent.

²⁰ Specifically, for the 2010 estimates, those who entered in 2010 and 2009 were adjusted for undercount by 13.2 percent; that rate dropped by 7 percent *each year*, falling to 1.9 percent for those who entered in 1982.

Overall percent unauthorized

As shown in Table A-2 above (Column 5), for most countries the logical edits produce numbers that are fairly close to the control totals. However, a small percentage of those selected for the database are erroneously classified as unauthorized residents. Many of these legal residents who are misclassified as unauthorized actually had previous experience as unauthorized residents (see Table A-4). Therefore, they probably would have characteristics similar to unauthorized residents.

Data from the 1996 DHS longitudinal study of legal immigrants (Jasso et al. 2008) indicates that a sizeable proportion of those who attain LPR status have spent time in an unauthorized status at some time in the past. As shown in Table A-4, almost one-third of those who became legal permanent residents in fiscal year 1996 had previous unauthorized experience in the United States.

Table A-4. Estimated Percent of Legal Permanent Residents who had Previous Unauthorized Experience in the United States: 1996 Cohort of LPRs

<u>Continent of birth</u>	<u>Percent</u>	<u>Country of birth</u>	<u>Percent</u>
All countries	31.7	Mexico	73.7
North America	58.4	El Salvador	65.4
South America	43.2	Dominican Republic	45.2
Africa	22.8	Jamaica	33.1
Europe	15.8	Poland	32.1
Asia	12.9	Philippines	13.5
		China	11.0
		India	3.6

Source: Jasso et al. 2008, Table 4.

For the 1996 cohort of LPRs, the percent with previous unauthorized experience varied by country and continent of origin. For Mexico and El Salvador, the percents were 74 percent and

65 percent, respectively (Table A-4). For North and South America combined, approximately half of the 1996 cohort of LPRs had previous unauthorized experience. The percents from Africa, Europe, and Asia, ranging from 13 to 23 percent, were lower than those from the Western Hemisphere.

Table A-5 shows the computation of the percent unauthorized, as well as the percent that had previous unauthorized experience, for El Salvador in 2010. Based on random selection, about 87 percent of the 570,000 unauthorized residents from El Salvador in the database for 2010 are likely to be unauthorized residents (Item C). Adding the number who probably were legal but had previous unauthorized experience (Item G), the percent of Salvadorans in the database who were unauthorized or had previous unauthorized experience increases to 96 percent (Item I). Replicating Table A-5 for each country and summing across all countries yields a total of 89 percent unauthorized; adding legal residents who had previous unauthorized experience to those estimated to be unauthorized yields an overall total of 93 percent.

Table A-5. Estimation of the Percent of Unauthorized Residents from El Salvador who were Unauthorized or had Previous Unauthorized Experience: 2010 ACS

Numbers in thousands; rounded independently

Item	Category	Estimate
A.	Noncitizens from El Salvador, entered 1982-2010, <i>after</i> logical edits	653,000
B.	Estimated unauthorized resident population from El Salvador, 2010	570,000
C.	Percent of A above who are unauthorized $C = B/A$	87%
D.	Probable unauthorized, based on random selection $D = B \times C$	497,000
E.	Probable legally resident, based on random selection $E = B - D$	73,000
F.	Percent of legal residents with previous unauthorized experience	65%
G.	Probable legal residents with previous unauthorized experience $G = E \times F$	48,000
H.	Unauthorized or with previous unauthorized experience $H = D + G$	544,000
I.	Pct. unauthorized <i>or</i> with previous unauthorized experience $I = H/B \times 100$	96%

Source: Item F from Jasso et al. 2008; other estimates calculated as shown or as described in the text.

The percent unauthorized varies by country; the countries with the largest numbers of unauthorized residents tend to have the highest percent unauthorized. This variation in percent unauthorized by country of origin should have little effect on estimates of unauthorized residents compiled for most purposes. For example, estimates for each state will have a high percentage of unauthorized residents because all of the states have relatively large numbers of unauthorized residents from Mexico and Central America (countries with high percentages of unauthorized residents) and relatively few from Europe and Canada (countries with lower percentages of unauthorized residents).

While clearly a limitation, this statistic—about 89 percent unauthorized—is comparable to the percent unauthorized in the residual estimates and in more recent estimates that derive legal status from ACS microdata. The estimates published in recent years by the Pew Research Center, DHS, Warren and Warren (2013), and Capps, et al. (2013) are, strictly speaking, "mostly unauthorized" because data are not available for some categories of legal residents. That is, some categories of legal or quasi-legal residents, such as recent asylees and parolees, aliens who have filed for (but not officially received) LPR status, and recipients of temporary protected status, are included as unauthorized residents in the residual estimates. Passel and Cohn (2009) noted that their residual estimate of 11.9 million unauthorized residents in 2008 included "immigrants from certain countries holding temporary protected status (TPS) or people who have filed for asylum status but whose claims are unresolved. This group may account for as much as 10 percent of the unauthorized estimate."