Demography (2024)
DOI 10.1215/00703370-11189687 © 2024 The Authors

ADVANCE PUBLICATION

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

A Test of the Validity of Imputed Legal Immigration Status

Marcelo Castillo, Alexandra Hill, and Thomas Hertz

ABSTRACT We evaluate the performance of a widely used technique for imputing the legal immigration status of U.S. immigrants in survey data—the logical imputation method. We validate this technique by implementing it in a nationally representative survey of U.S. farmworkers that includes a well-regarded measure of legal status. When using this measure as a benchmark, the imputation algorithm correctly identifies the legal status of 78% of farmworkers. Of all the variables included in the algorithm, we find that Medicaid participation poses the greatest challenge for accuracy. Using the American Community Survey, we show that increased Medicaid enrollments stemming from the implementation of the Affordable Care Act in 2014 led to sizable changes in the share of immigrants imputed as legal over time and across space. We explore the implications of these changes for two previous studies and conclude that including Medicaid criteria in the imputation algorithm can significantly impact research findings. We also provide tools to gauge the sensitivity of results.

KEYWORDS Undocumented immigration • Legal status • Imputation • Immigration

Introduction

Despite the political and administrative salience of the distinction between legal and undocumented immigration, no national household survey in the United States routinely measures immigrants' legal status. Clark and King (2008) characterized this omission as the most serious data-related obstacle to studying immigrant populations. To overcome this challenge, researchers have developed several techniques for imputing legal status in survey data. One of the most widely used is the logical imputation method, which consists of using a set of categorical markers (logical criteria) to assign immigrants as being likely legal or likely undocumented. These person-level imputations have been used to generate rich demographic, economic, and geographic descriptions of the U.S. undocumented population, study differences in labor market and other outcomes by legal status, and more.[1] However, relatively little work has

[1] Examples of descriptive work include Heinzel et al. (2021), Lopez et al. (2021), Passel and Cohn (2019), and Warren (2014). Examples of econometric work include Amuedo-Dorantes et al. (2018), Amuedo-Dorantes et al. (2020), Bohn et al. (2014), Bozick and Miller (2014), Flores (2010), Hall et al. (2010), Kaushal (2006), Orrenius and Zavodny (2016), Potochnick (2014), and Tamborini and Villarreal (2021).

ELECTRONIC SUPPLEMENTARY MATERIAL The online version of this article (https://doi.org/10.1215/00703370-11189687) contains supplementary material.

examined the validity of this imputation technique (notable exceptions include Van Hook et al. (2015) and Spence et al. (2020)).

In this article, we evaluate the performance of one specific imputation algorithm, developed in Borjas (2017) and refined in Borjas and Cassidy (2019). The most frequent criteria used by this algorithm to classify immigrants as legal are being a U.S. citizen, having arrived in the United States before 1980, having received any of several federal social services that are accessible only to citizens or lawful permanent residents (including Medicaid), and having a spouse who meets one of these criteria. We focus on this algorithm because it was designed to be easily reproducible in publicly available data, making it one of the most clearly documented, transparent, and widely used methods for imputing legal status.[2] Importantly, most of the criteria used in the algorithm are the same or similar to those applied in other well-established methods (see Table A1 in the online appendix; all tables and figures designated with an "A" appear in the appendix).

Our evaluation of the Borjas–Cassidy (BC) algorithm has three steps. In the first step, we test the algorithm's accuracy by applying it to the National Agricultural Workers Survey (NAWS), a survey dataset with a well-regarded measure of legal immigration status. We find that the BC algorithm correctly determines the legal status of 78% of farmworkers during 2000–2018. The algorithm performs relatively well for undocumented immigrants (88% are correctly classified) but is not as accurate at categorizing the documented (59% are correctly classified). This analysis also shows that the algorithm's ability to correctly categorize undocumented immigrants has decreased over time, while its ability to correctly identify legal immigrants has improved. We examine how different legal markers contribute to these changes in accuracy and find that the criteria for individual and spousal Medicaid participation are arguably the most problematic for three reasons: (1) they flag a large number of immigrants, (2) many undocumented immigrants misreport participating in Medicaid, and (3) changes in government policy influence Medicaid eligibility and thus participation rates. We show that the recent growth in Medicaid enrollments, which accelerated in 2014 after the implementation of the Affordable Care Act (ACA), is in great part responsible for the changes in accuracy rates uncovered in the NAWS.

In the second and third steps of our analysis, given Medicaid's distinct role in the BC algorithm, we turn our attention to exploring whether Medicaid inclusion affects estimates of parameters of the broader immigrant population. In the second step, we apply the BC algorithm to the American Community Survey (ACS), a data source that better reflects the U.S. immigrant population than the NAWS but lacks a measure of legal status. Using the ACS, we examine how Medicaid inclusion affects the size of the imputed legal immigrant population across space and over time. We find that for low-income populations these effects are large and increased substantially in response to the 2014 ACA Medicaid expansion. We estimate that the imputed shares of legal immigrants among "low-skill" workers, immigrants from Mexico, and noncitizens increased by about six percentage points in ACA expansion states. Because Medicaid expansion is unlikely to have impacted the true legal share,[3] we caution

[2] For examples, see Albert (2021), Amuedo-Dorantes and Antman (2017), Bhandari et al. (2021), Borjas and Slusky (2022), Docquier et al. (2019), and Lessem and Nakajima (2019).

[3] We offer evidence from the NAWS to support this claim in Figure A7.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

researchers against using logical imputation techniques that include Medicaid when examining trends in the imputed legal shares across states and over time, particularly when focusing on low-income immigrant groups.

Finally, in the third step, we replicate two studies applying the BC algorithm and demonstrate that the consequences of Medicaid inclusion can be severe, as in the case of Borjas and Cassidy (2019), or negligible, as in the case of Borjas (2017).[4] Accordingly, researchers may benefit from demonstrating the robustness of their research findings to Medicaid inclusion. Using ACS data, Borjas and Cassidy (2019) found that the wage gap between legal and undocumented workers declined between 2013 and 2016. Using a decomposition based on the regression anatomy theorem, we show why including Medicaid in the algorithm is problematic for this analysis: average wages are much lower for immigrants classified as legal solely by the Medicaid criteria than for other immigrants. As a result, even if rising Medicaid enrollments cause only modest increases in the share of immigrants imputed as legal, they can lead to large declines in estimates of the wage gap obtained using the BC algorithm. We demonstrate that ACA Medicaid expansion can explain most of the decline in the BC wage gap. Using Current Population Survey (CPS) data, Borjas (2017) found that labor supply elasticities are highest for native-born workers, lower for legal immigrants, and close to zero for undocumented immigrants. We find that these results are not sensitive to Medicaid inclusion because the labor supply elasticities of the Medicaid group and other immigrants are similar in magnitude, and the size of the Medicaid group is relatively small during the study period.

Background and Prior Research

The current best practice for estimating the size of the undocumented immigrant population uses the "residual method" pioneered by Warren and Passel (1987) (Passel et al. 2004; Van Hook et al. 2021). This approach starts with survey-based estimates of the size of the total immigrant population and then subtracts counts of the legal immigrant population computed using administrative data. While the residual method can tell us the size of the undocumented population, it has limited utility for researchers interested in comparing the behavior or characteristics of legal and undocumented immigrants. To deal with this issue, researchers have developed several techniques for assigning a legal status to individual immigrants in survey datasets. These techniques fall under two broad categories: statistical imputation and logical imputation, though they are sometimes used together (Capps et al. 2018). Statistical imputation consists of using the relationships between a set of predictors and legal status in a survey that includes a measure of legal status (the donor sample) to impute a legal status variable in a dataset that includes the same set of predictors but that lacks a measure of legal status (the target sample) (Van Hook et al. 2015).

Perhaps because of the complexity involved in statistical imputation, most researchers opt to use logical imputation approaches (Capps et al. 2018). First developed by

[4] Both papers are very careful to note that their results are only as reliable as the underlying status imputations and stress the importance of validating their procedure.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

Passel and Clark (1998), and later refined by Passel at the Pew Research Center (Pew), logical imputation categorizes individual immigrants in survey datasets, usually the ACS, the CPS, or the Survey of Income and Program Participation (SIPP), as likely legal or likely undocumented, on the basis of categorical markers associated with legal status (Van Hook et al. 2015). There are different versions of this method in the literature, but most use a similar set of categorical markers to impute legal status. For example, most versions classify individual immigrants as likely legal if they or their spouses report being a U.S. citizen or having received any of a number of federal social services (e.g., Medicaid) that are legally accessible only to citizens or lawful permanent residents (LPRs). Immigrants who are not flagged as likely legal by any categorical markers are classified as likely undocumented.

We focus on the BC algorithm, which is a simplified version of the Pew approach. George Borjas was granted access to Pew's CPS datasets for 2012 and 2013, which contain Pew's status imputation variable. After careful examination of the demographic characteristics of those imputed as undocumented in these files, he reverse-engineered the Pew imputation algorithm, and applied it to all CPS files since 1994, to study the labor supply of undocumented immigrants (Borjas 2017). The algorithm was later refined in Borjas and Cassidy (2019), which added an H-1B filter. A key strength of this algorithm is that it is simpler (it consists of fewer criteria) than most other logical imputation methods. It is also arguably one of the most clearly documented methods for imputing legal status. Hence the BC algorithm has become, to our knowledge, one of the most frequently used methods for imputing documentation status (e.g., Albert 2021; Amuedo-Dorantes and Antman 2017; Borjas and Slusky 2022; Docquier et al. 2019; Heinzel et al. 2021; Lessem and Nakajima 2019; Spence et al. 2020).

There is a modest literature on how different legal imputation approaches affect parameter estimates (e.g., Altman et al. 2023; Altman et al. 2021; Hamilton et al. 2022; Ro and Van Hook 2022; Sohn and Pebley 2020; Spence et al. 2020; Van Hook et al. 2015). Simulations by Van Hook et al. (2015) revealed that the logical approach can bias estimates of the link between legal status and health coverage. Spence et al. (2020) found similar biases in relation to poverty. Our study contributes to this body of work by using a novel dataset with a well-regarded legal status measure to track changes in imputation accuracy across space and over time, highlighting the role of specific legal markers. We devote particular attention to demonstrating theoretically and empirically how one legal marker, Medicaid, can have strong implications for parameter estimates. Crucial to the relevance of our analysis to other algorithms is that most logical imputation methods use Medicaid participation as a marker of legal status (see Table A1).

Validation Data: The National Agricultural Workers Survey

The NAWS is an ongoing, employer-based, repeated cross-sectional survey of employed crop workers, conducted by the U.S. Bureau of Labor Statistics. Launched in 1989, as mandated by the Immigration Reform and Control Act (IRCA) of 1986, the survey originally aimed to monitor potential shifts in farm labor supply due to IRCA's legalization of previously undocumented workers. Over time, it expanded in

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

scope and has informed significant research on the relationship between legal status and socioeconomic outcomes (Chung and Leigh 2015; Grzywacz et al. 2010; Hamilton et al. 2019; Pena 2010; Richards 2018). The survey is conducted in three seasonal waves each year, encompassing 1,500 to 3,500 individuals per fiscal year. We focus on the period from 2000 to 2018 because, during this period, the data include specific questions relevant to our study.

Because legal status is a sensitive issue, there are natural grounds for skepticism that truthful responses can be elicited from undocumented immigrants in surveys. We argue that the NAWS measure is, in fact, highly reliable for several reasons. First, compared with other well-regarded surveys that directly measure legal status, such as the SIPP, the NAWS has a lower nonresponse rate to immigration-related questions. According to Bachmeier et al. (2014), the NAWS had significantly lower nonresponse rates than the SIPP, resulting in a smaller percentage of immigrants whose legal status could not be determined (0.4% in the 2004–2005 NAWS vs. 13.6% in the 2004 SIPP).[5] This low nonresponse rate may reflect the fact that the NAWS is well-known and trusted in the farmworker community and is based on in-person interviews conducted by well-trained and highly experienced enumerators who are not federal government employees.

Second, the NAWS legal status measure is based on more than a dozen immigration-related questions that are checked for internal consistency. Individuals are asked about their immigration status, with response options including naturalized citizen, green card holder, and undocumented. This measure is then validated by ensuring consistency with responses to several other questions, such as involvement in immigration-related programs, and personal information, such as date of birth, country of birth, and date of entry into the United States.[6]

Finally, the survey finds an extraordinarily high undocumented share in subpopulations where we would expect undocumented workers to be overrepresented. For example, between 2000 and 2018, the NAWS included 3,424 male, noncitizen immigrant farmworkers who had been in the country less than two years and who lacked a high school degree. Of these, 99% self-reported as undocumented. Although we lack a direct validation method, such as cross-referencing with confidential administrative immigration records, we regard the NAWS legal status measure as one of the most precise available. For clarity of exposition, we treat it as corresponding to the true values of this variable.

An important limitation of the NAWS is that it is restricted to the farm sector, and this has important consequences for the portability of some parameter estimates (e.g., wage gap estimates). There are notable differences between foreign-born farmworkers and the broader foreign-born population in areas such as naturalization rates, education levels, geographic distribution, and country of birth. Between 2008 and 2018, NAWS data show that foreign-born farmworkers had lower naturalization rates (7% vs. 43%), a higher rate of workers with at most a high school education (96% vs. 52%), a larger proportion residing in California (47% vs. 25%), and a substantially greater percentage born in Mexico (92% vs. 32%) compared with the general

[5] The nonresponse rates for questions related to legal status in the SIPP ranged from 4.6% to 25.3% between 1996 and 2008, while in the NAWS it ranged from 0.5% to 5.5% between 1988 and 2009.

[6] See the online appendix for more details on this variable's construction.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

foreign-born population as measured in the ACS (see Tables A2 and A3). Moreover, a much larger share of the agricultural workforce is undocumented compared with the entire immigrant population. Using the SIPP and ACS, Capps et al. (2020) estimated that 23% of all immigrants in 2018 were undocumented, whereas NAWS data show a much higher 54% in the farm sector.

Given the unique characteristics of farmworkers, some measures of the accuracy of the BC algorithm in the NAWS may better reflect the algorithm's performance when applied to data on subpopulations with similar attributes (e.g., low-income workers in California) than when applied to nationally representative data on the foreign-born population. Nonetheless, we argue that some measures likely align in key instances. Importantly, we argue that *trends* in Medicaid's contribution to the BC algorithm's accuracy for farmworkers are similar to those in the broader population. Our analysis in the ACS provides evidence to support this claim. Below, we are careful to note which findings for farmworkers are likely applicable to the wider immigrant population and which are not.

One alternative to the NAWS is the SIPP, which has been used in previous studies to evaluate the performance of various imputation methods (e.g., Van Hook et al. 2015). The SIPP is the only survey representative of the entire U.S. population that includes questions on legal status. However, recent changes in its data collection methods make the SIPP ill-suited for the analysis at hand. Public SIPP data from 1996–2008, used in most studies that use these data to measure legal status, include information on citizenship, entry status into the United States (permanent or nonpermanent resident), and status transitions from nonpermanent to permanent residency. Legal status is generally inferred by categorizing citizens, permanent arrivals, and those who have transitioned to permanent residence as likely legal immigrants. A key limitation of the SIPP is that from 2014 onward, it ceased to include status transition information, thus reducing its usefulness for inferring legal status in recent years (Heinzel et al. 2021). Given that we devote substantial attention to how the 2014 ACA implementation affected the algorithm's performance, the SIPP is of limited usefulness for our study.[7]

The BC Legal Status Imputation Algorithm

The components of the BC algorithm in the NAWS are listed in Table 1. Column 1 shows the number of immigrant farmworkers who are flagged as legal by each marker, and column 2 shows the percentage of these workers who self-report as legal. Columns 3 and 4 show analogous statistics from applying the markers sequentially so that each individual is categorized as legal only by a single marker. The rationale behind most of these criteria is straightforward, but they are also clearly imperfect indicators of legal status. For example, Borjas and Cassidy (2019) assumed that all

[7] Compared with the NAWS, the SIPP has other potential drawbacks, including a higher nonresponse rate, an undercounting of the undocumented population (Capps et al. 2018), and the fact that its legal status measure misclassifies legal non-LPRs as undocumented. To our knowledge, no other U.S. survey with a legal status measure (e.g., Los Angeles Family and Neighborhood Survey) enables studying changes in the BC algorithm's accuracy around ACA implementation.

Table 1 BC legal markers in the National Agricultural Workers Survey

BC Criterion	*N* (1)	% Correct (2)	*N* (3)	% Correct (4)
1. Citizen	2,123	100.0	2,123	100.0
2. Citizen—Spouse	1,641	62.6	1,437	58.1
3. Pre-1980 Arrival	4,529	89.4	3,197	86.1
4. Pre-1980 Arrival—Spouse	1,376	90.8	372	79.1
5. Cuban-born	27	89.8	25	89.0
6. Social Security or SSI[a]	668	82.5	230	68.8
7. Veteran/Armed Forces[a]	14	25.4	5	7.8
8. Public Housing[a]	231	49.5	167	40.0
9. Medicaid	2,697	68.7	1,545	55.9
10. Medicaid—Spouse	2,958	58.1	712	21.4

Notes: Statistics in columns 3 and 4 are estimated only for the subset of workers not classified as legal by previous criteria. In other words, we apply each criterion in order, so that each person is categorized as legal by a single legal marker. In columns 2 and 4, we use sample weights to compute the share of workers correctly classified. The sample includes all immigrant farmworkers interviewed between 2000 and 2018, omitting 1.25% for whom documentation information was missing, for a final count of $N = 32{,}715$. We omit criteria that are not applicable to the NAWS sample—working in the government sector, having an occupation that requires special licensing, or being classified as a likely H-1B visa holder—and one criterion not reported in the NAWS—having a Cuban-born spouse.

[a] Household-level variable in the NAWS.

immigrants married to U.S. citizens are legal immigrants; in fact, more than one third of such workers in the NAWS are undocumented.[8]

The shares correctly classified for the two Medicaid-based criteria are also low. In the NAWS, 44% of the more than 1,500 immigrants who are flagged by this question alone are undocumented. The misclassification rate rises to 78% when classifying based solely on the spouse's Medicaid status. Undocumented immigrants' Medicaid participation, despite federal restrictions, requires explanation. One possibility is that Medicaid payments are made for undocumented individuals' emergency services, which could result in misreported full-benefit receipts (KFF 2023). Another factor could be that immigrants who receive free clinic services may mistakenly report Medicaid participation (Capps et al. 2013). Moreover, in California, undocumented immigrants can access a limited-scope Medicaid plan that covers pregnancies and emergency care, which is funded solely by the state (California Department of Health Care Services 2017).

Importantly, the two Medicaid criteria classify a relatively large number of immigrants as legal— these criteria classify more than 21% of those imputed as legal in the NAWS, and the criteria are similarly important in the ACS (Heinzel et al. 2021).[9] Given that policy shifts can affect Medicaid eligibility and participation rates, the significance of Medicaid in the algorithm, combined with its high misclassification

[8] This aligns with findings from the SIPP and ACS indicating that in 2018, 15% of undocumented immigrants were married to U.S. citizens and another 6% to LPRs (Capps et al. 2020).

[9] Medicaid is relatively more important in the NAWS than in the ACS (see Table A2), potentially owing to lower earnings and naturalization rates among farmworker immigrants compared with the general immigrant population.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

rates, suggests potential issues with its use in logical imputation. For example, the expansion of the ACA in 2014 led to a surge in Medicaid enrollments within expansion states. This increase could result in the reclassification of many new enrollees' imputed status, switching them from undocumented to legal. As we will show, this inconsistency in status assignment over time may complicate the interpretation of trends in some parameter estimates during the ACA expansion period.

Accuracy of the BC Algorithm in the NAWS

To evaluate the accuracy of logical imputation methods, we adapt a simple classification error framework related to the work of Aigner (1973). We then document changes in the accuracy rates of the BC algorithm over time in the NAWS and show how different legal markers contribute to these changes. Our analysis suggests that Medicaid criteria are particularly problematic for the algorithm—they are the least accurate criteria and their impact on accuracy rates changes considerably over time.

Framework

Let L be an indicator variable equal to 1 if the person truly has legal status and 0 if truly undocumented. Let $L(Z) = \max(Z_1, Z_2, \ldots, Z_n) = L^Z$ be an indicator equal to 1 if the person is imputed as legal and 0 otherwise when we use a vector $\mathbf{Z}$, composed of n "legal status marker" indicators in the classification algorithm. Classification error, u^Z, in the imputed legal status variable L^Z is the difference between L^Z and true legal status, L, so that $L^Z = L + u^Z$. The true probabilities that an individual is legal or undocumented are $\Pr(L=1) = P_L$ and $\Pr(L=0) = P_U$, respectively, which equal the share of immigrants that are truly legal and undocumented. The probabilities that an individual is categorized as legal or undocumented by algorithm Z are $\Pr(L^Z=1) = P_L^Z$ and $\Pr(L^Z=0) = P_U^Z$, respectively, which equal the share of immigrants that are imputed as legal and undocumented, regardless of their true status.

We define the *legal accuracy rate* of algorithm Z as the probability that a legal immigrant is correctly classified as legal, $\Pr(L^Z=1 \mid L=1) = A_L^Z$, and the *undocumented accuracy rate* as the probability that an undocumented immigrant is correctly classified as undocumented, $\Pr(L^Z=0 \mid L=0) = A_U^Z$. The *overall accuracy rate*, a standard summary measure of accuracy, of algorithm Z is the total share correctly imputed, which can be written as

$$A^Z = A_L^Z P_L + A_U^Z P_U. \tag{1}$$

For convenience, we denote the algorithm including all BC legal markers as the "BC" or "full algorithm" and index relevant variables and parameters with the letter F. The algorithm that includes all BC markers except for the two Medicaid criteria is referred to as the "base algorithm" and relevant variables and parameters are indexed with the letter B.

Table 2 shows the joint distribution of L and L^F in terms of the accuracy rates A_L^F and A_U^F along with our estimates of their values in the NAWS sample. Overall, the

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

Table 2 Joint distribution of true and imputed legal status

		Imputed Legal Status, BC Algorithm		
		$L^F = 1$	$L^F = 0$	$f(L)$
True Legal Status	$L = 1$	$A_L^F P_L = .21$	$(1 - A_L^F)P_L = .14$	$P_L = .35$
	$L = 0$	$(1 - A_U^F)P_U = .08$	$A_U^F P_U = .57$	$P_U = .65$
	$f(L^F)$	$P_L^F = .29$	$P_U^F = .71$	1
$A_L^F = .59$		$A_U^F = .88$	$A^F = A_L^F P_L + A_U^F P_U = .78$	

Notes: The values in each cell are estimated using data on all immigrant farmworkers in the NAWS interviewed between 2000 and 2018 and applying the NAWS sampling weights. The bottom row shows the BC legal, undocumented, and overall accuracy rates.

algorithm correctly classifies 78% of immigrants in our sample, with a higher accuracy rate for undocumented ($A_U^F = .88$) than for documented immigrants ($A_L^F = .59$). A relatively high undocumented accuracy is expected and likely mirrored when the algorithm is applied in the ACS, given that immigrants are by default assumed to be undocumented unless flagged by a legal marker, and these markers are chosen specifically for their low probability of applying to undocumented individuals. Legal accuracy rates, however, are likely much higher in the ACS because naturalized citizens' share of legal immigrants is higher in the general population compared with farmworkers.[10]

Figure 1 shows the legal, undocumented, and overall accuracy rates of the full algorithm over time. In the first half of the sample, 2000–2009, the overall accuracy rate is stable. Beginning in 2010, the overall accuracy rate steadily declines, falling 10 percentage points by 2018. The decline in the undocumented accuracy rate follows a similar trajectory. Nearly all (95%) undocumented workers are correctly classified at the beginning of the period, but this declines by nearly 20 percentage points by 2018 (reaching a 76% accuracy rate). Legal accuracy increases somewhat over time, partially counteracting the decreasing accuracy for undocumented workers.

Decomposing BC Accuracy by Imputation Categories

To show how different legal markers contribute to the BC algorithm's decline in overall accuracy, we decompose accuracy rates by the source of classification. In this initial analysis, we emphasize the distinct impact of Medicaid criteria on the algorithm's accuracy in the NAWS.

Most immigrants in the NAWS who are imputed as legal are classified by three categorical markers—U.S. citizenship, entering before 1980, and Medicaid. We focus on accuracy rates associated with these three legal markers. We group all other "small" criteria together with citizenship into a single category, which we refer to as "citizenship." One way to designate which categorical marker is responsible for assigning an individual as being documented is to apply the markers sequentially. To

[10] Capps et al. (2020) estimated that, in 2018, more than half of legal immigrants were naturalized citizens. Naturalized citizens are never more than a quarter of all legal immigrants in the NAWS during 2000–2018.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

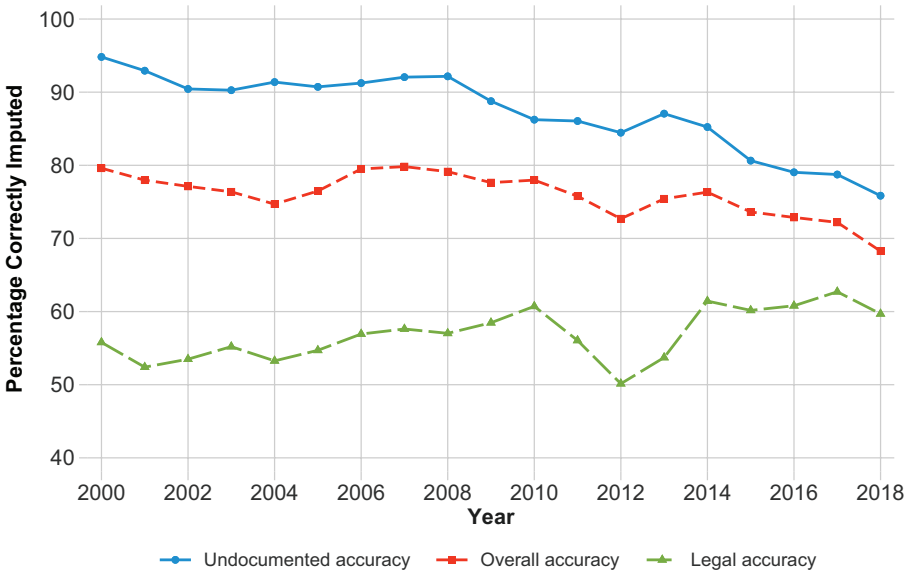


Fig. 1 Legal, undocumented, and overall accuracy rates in the NAWS. Accuracy rates are calculated as two-year moving averages across interview years and applying the NAWS sampling weights. The overall accuracy rate is estimated as the share of the immigrant population whose imputed immigration status matches their self-reported status. The undocumented accuracy rate is the share of undocumented workers correctly classified by the BC algorithm. The legal accuracy rate is the share of legal workers correctly classified by the BC algorithm.

mimic the order in which the markers are listed in Borjas and Cassidy (2019), and in which they are applied by Heinzel et al. (2021), we apply the citizenship criteria first, the 1980 criteria second, and the Medicaid criteria last. In Figure A4, we show that our main results regarding trends in the Medicaid effects on BC accuracy are robust to the order in which we apply these markers.

Following this ordering of legal markers and letting $j = \{\text{Citizen}, 1980, \text{Medicaid}\}$ index the source of classification, we can express overall accuracy as

$$A^F = P_L A_L^F + P_U A_U^F = P_L\left(\sum_j A_L^j\right) + \left(P_U - P_U\left(\sum_j (1 - A_U^j)\right)\right), \tag{2}$$

where A_L^j is the share of truly legal imputed as legal due to marker j, and $(1 - A_U^j)$ is the share of truly undocumented incorrectly imputed as legal because of marker j. For example, A_L^M and $(1 - A_U^M)$ are the shares of legal and undocumented workers who receive or have a spouse who receives Medicaid, but are not citizens, did not arrive in the United States before 1980, and were not flagged by any of the other "small" criteria. We refer to A_L^j and $-(1 - A_U^j)$ as the contribution of marker j to the legal and undocumented accuracy rates of the BC algorithm, and $A_L^j P_L - (1 - A_U^j) P_U$ as the contribution of marker j to overall BC accuracy.

Figure 2 illustrates the undocumented share of immigrant farmworkers in the NAWS (panel a) and the contribution of the three markers to the BC algorithm's overall (panel b), legal (panel c), and undocumented (panel d) accuracy rates. Panel a shows that, in all years, more than half of all farmworkers are estimated to be undocumented. The undocumented share peaks in the early 2010s (67%) and then declines to 52% by 2018.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d5290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

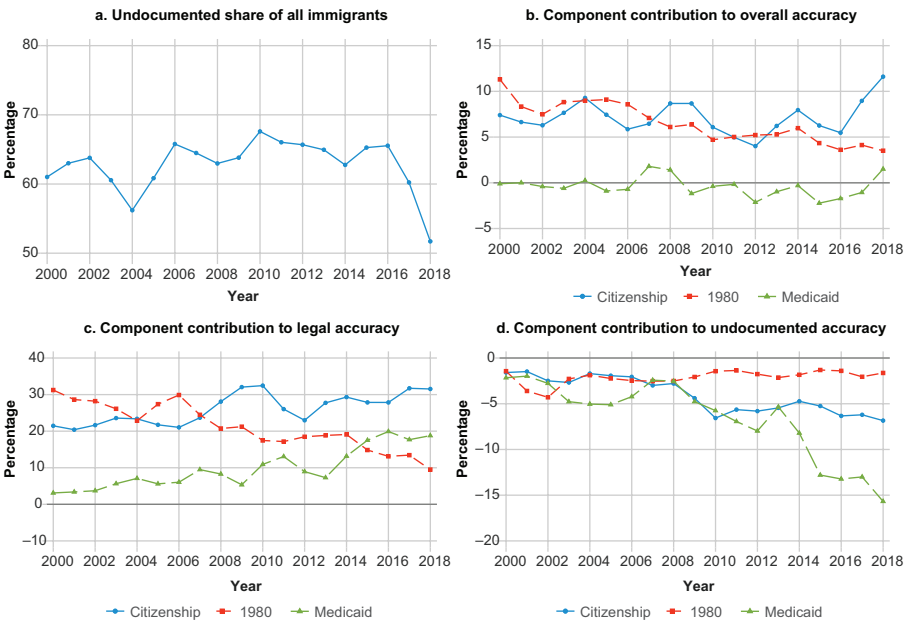


Fig. 2 Component contribution to accuracy rates. All accuracy rates are calculated as two-year moving averages across interview years and applying the NAWS sampling weights. Panel a shows the share undocumented, panel b shows each component's contribution to overall accuracy, and panels c and d display their contributions to legal and undocumented accuracy rates of the BC algorithm, respectively. The overall accuracy rate is the share of all immigrants correctly classified by the BC algorithm. Legal and undocumented accuracy rates are the shares of legal and undocumented workers that are correctly classified by the BC algorithm.

Panel b shows that citizenship criteria have a positive and mostly stable contribution to overall accuracy, the 1980 criteria have a positive but declining contribution, and the Medicaid criteria have small and often negative contributions. Taken as a whole, panels a and b suggest that a large portion of the algorithm's overall accuracy is due to the relatively large share of undocumented workers in our sample. The post-2010 decline in overall accuracy results from a decreasing share of undocumented workers and the diminished contribution of the 1980 criteria. Some of this decline in the contribution of the 1980 criteria likely stems from the fact that the pool of pre-1980 births shrinks over time as people age. This trend in accuracy is also likely to be present in the ACS, where there is also a notable decline in the share of immigrants deemed legal by the 1980 criteria.[11]

Unlike citizenship and 1980 criteria, Medicaid criteria tend to have small and sometimes negative impacts on overall accuracy. Panels c and d of Figure 2, which show component contributions to BC legal and undocumented accuracy rates, help explain why this is the case. Over the study period, the effect of Medicaid on legal accuracy increased by approximately 16 percentage points and decreased by 13 percentage

[11] The ACS shows a slight increase in the share of all immigrants who are naturalized citizens. Thus, trends regarding citizenship's contribution to BC accuracy are likely to differ between the ACS and the NAWS.

CORRECTED PROOFS

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

points for undocumented accuracy. In the online appendix, we show that the recent rise in Medicaid enrollments, which accelerated in 2014 after the ACA implementation, is in great part responsible for these changes in Medicaid legal and undocumented accuracy rates. Combined, these Medicaid effects on legal and undocumented accuracy cancel out, in part because the undocumented share is greater than the legal share, so that the effect on overall accuracy changes very little over the sample period.[12] Despite this negligible effect on overall accuracy, as we show below, these large impacts of Medicaid on BC legal and undocumented accuracy can have strong implications for parameter estimates.

Medicaid and Performance of the BC Algorithm in the ACS and CPS

We expand on our analysis in the NAWS to show some of the implications of including Medicaid in the BC algorithm when applied in the ACS and CPS. If our Medicaid findings in the NAWS are comparable to those in the ACS or CPS, Medicaid inclusion could in theory have strong effects on estimates of some parameters of the overall immigrant population, and these effects could be magnified by Medicaid expansion. While we are unable to replicate our accuracy analysis in the ACS and CPS because they lack a measure of legal status, we can nonetheless use these datasets to test for the effects of Medicaid inclusion on such parameter estimates. We first show how Medicaid inclusion affects the size of the imputed legal share. Next, we show that Medicaid expansion can dramatically influence results from econometric models using the BC legal status variable, though the severity of this issue can vary significantly depending on the application. For brevity, we refer to the group of Medicaid recipients without any other legal markers and their spouses as the "Medicaid group" or the "Medicaid-only group."

Medicaid Expansion and Proportion Imputed as Legal

We now show that for some low-income immigrant subgroups, many of which are the focus of research relying on logical imputation (e.g., immigrants from Mexico), the effects of Medicaid inclusion on the BC legal share are large and increased substantially in response to ACA Medicaid expansion. Because Medicaid expansion is unlikely to have any meaningful effect on the true legal share, we argue that, in general, it is highly problematic to interpret changes in the BC legal shares for these groups as reflective of the evolution of the true legal share.[13]

How might we expect Medicaid inclusion and ACA Medicaid expansion to affect the BC legal share? For any immigrant subpopulation, we can define the "the Medicaid effect" on the BC legal share, P_L^M, as the difference between the imputed legal share

[12] Note that ACA implementation could improve the BC algorithm's overall accuracy in the ACS because the legal share is much higher in the ACS than in the NAWS. See the online appendix for more details.

[13] Proponents of the "welfare magnet effect" argue that immigrants move to states that provide them with relatively generous public benefits. However, Yasenov et al. (2020)'s ACS study found no migration effects stemming from Medicaid expansion. In Figure A7, we show that Medicaid expansion seems to have had no impact on the true legal share in the NAWS.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

under the full algorithm, P_L^F, and the legal share under the base algorithm, P_L^B. Note that P_L^M is simply the share of the total immigrant population in the Medicaid-only group:

$$P_L^M = P_L^F - P_L^B = \frac{n_L}{N}\frac{n_L^M}{n_L} + \frac{n_U}{N}\frac{n_U^M}{n_U} = \frac{n^M}{N} \geq 0, \tag{3}$$

where $n_L^M (n_U^M)$ is the number of legal (undocumented) immigrants classified as legal by Medicaid alone, $n_L (n_U)$ is the total number of legal and undocumented immigrants, and $N = n_L + n_U$. Because some Medicaid recipients will not have other legal markers, P_L^M should be positive and increasing with Medicaid expansion. Intuitively, while Medicaid expansion should have little influence on the true share of legal immigrants, it should move some people previously imputed as undocumented, whether correctly or not, to the imputed legal category, and thus the share deemed as legal should increase. For brevity, we refer to P_L^M as the "Medicaid-only share."

In Figure 3 we show the evolution of the Medicaid-only share for select immigrant groups from 2008 through 2018. Panel a shows the Medicaid-only share of agricultural workers in the ACS and the equivalent figure in the NAWS. The NAWS and ACS trends track each other closely over the period. The increase after 2013 is of a similar magnitude in both surveys, which might suggest that factors driving changes in reported Medicaid participation by both legal and undocumented workers in the NAWS are also prevailing in the ACS.

Panel b shows the Medicaid-only share of immigrants aged 21 to 64, for all, "low-skill" (no college), and "high-skill" (some college) groups. These shares also increase between 2013 and 2016, likely as many states took advantage of the ACA's incentives, though the increase for the all-immigrant group is only moderate (about two percentage points). While the increase is much higher for low-skill workers, we also see an increase for high-skill immigrants. We argue that this is in part because the high-skill category is defined as all workers with at least some college education, and some immigrants in this category, perhaps because of underemployment, have family incomes below the Medicaid eligibility threshold.

As a more formal analysis of the effect of Medicaid expansion on the BC legal share, we test whether this share increased by a larger amount in expansion states than in non-expansion states after ACA implementation. We implement a difference-in-differences strategy using regression models of the form

$$P_{L,st}^M = \alpha_s + \mu_t + \sum_{\tau=2008}^{2012} \delta_\tau M_s I(t=\tau) + \sum_{\tau=2014}^{2018} \beta_\tau M_s I(t=\tau) + \epsilon_{st}, \tag{4}$$

where $P_{L,st}^M$ is the Medicaid-only share in state s and year t, M_s is an indicator variable equal to 1 if state s is an expansion state and 0 otherwise, and α_s and μ_t are state and year fixed effects. To evaluate pretrends (δ_τ's) and post-treatment effects (β_τ's), we include an indicator variable for the sample year $I(t=\tau)$ and exclude the reference year, which we set to 2013. We use the Medicaid-only share as the dependent variable, $P_{L,st}^M$, instead of the BC legal share, $P_{L,st}^F$, to account for differential pretrends in other legal markers, that is, those in $P_{L,st}^B$, across states.[14]

[14] In the online appendix, we show very similar results when using $P_{L,st}^F$ as the dependent variable.

CORRECTED PROOFS

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

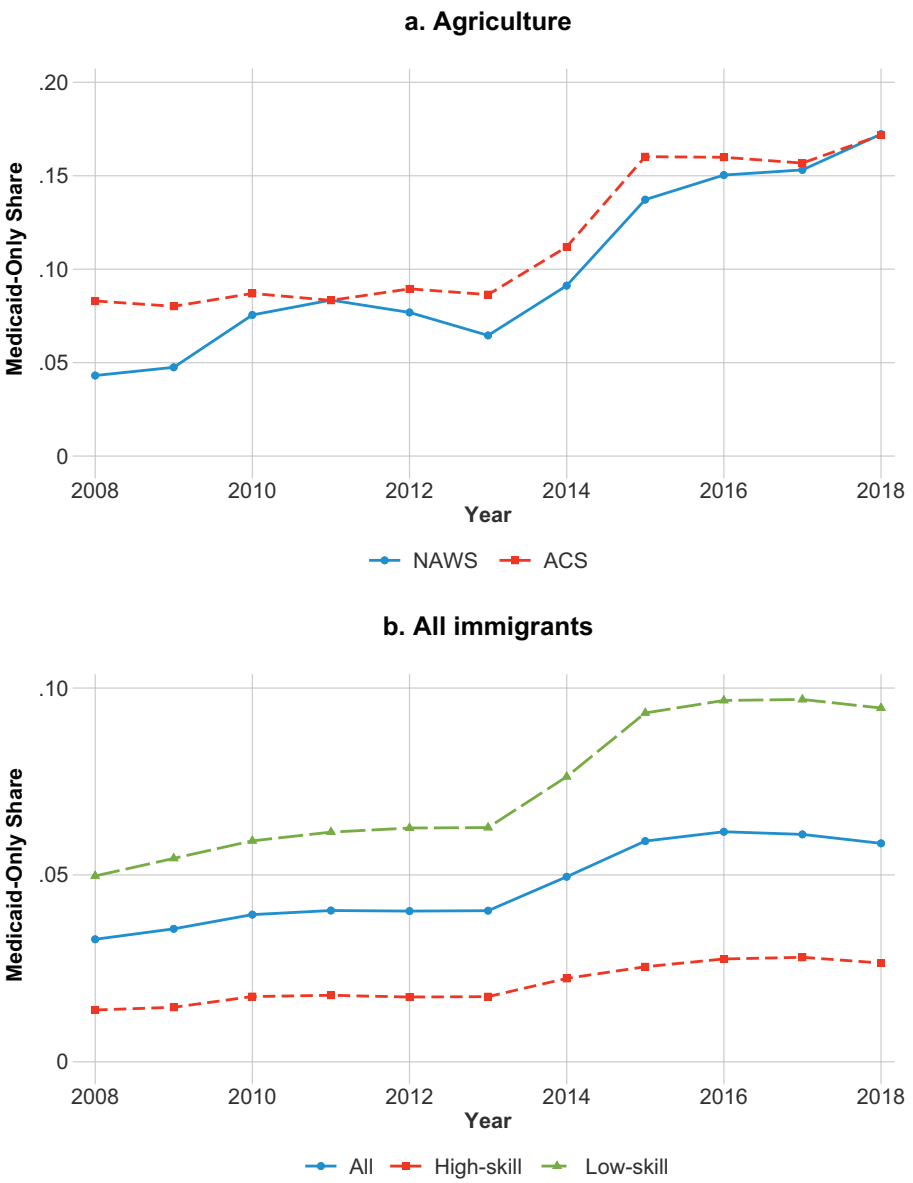


Fig. 3 Share of immigrants in the Medicaid-only group. Panel a shows the share of agricultural workers classified as legal using the BC algorithm owing to Medicaid alone. Panel b shows the corresponding summary for all, high-skill, and low-skill immigrants. "Low-skill" are workers who have a high school education or less, and "high-skill" are those with more than a high school education.

Figure 4 shows coefficients of interest, δ_τ and β_τ, and associated 95% confidence intervals from implementing Eq. (4) for select immigrant groups. The panels show that Medicaid expansion increased the proportion of immigrants classified as legal owing to Medicaid alone. Panel a shows that these effects are quite large for agricultural workers: from 2013 to 2016, there was a 10-percentage-point increase in

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

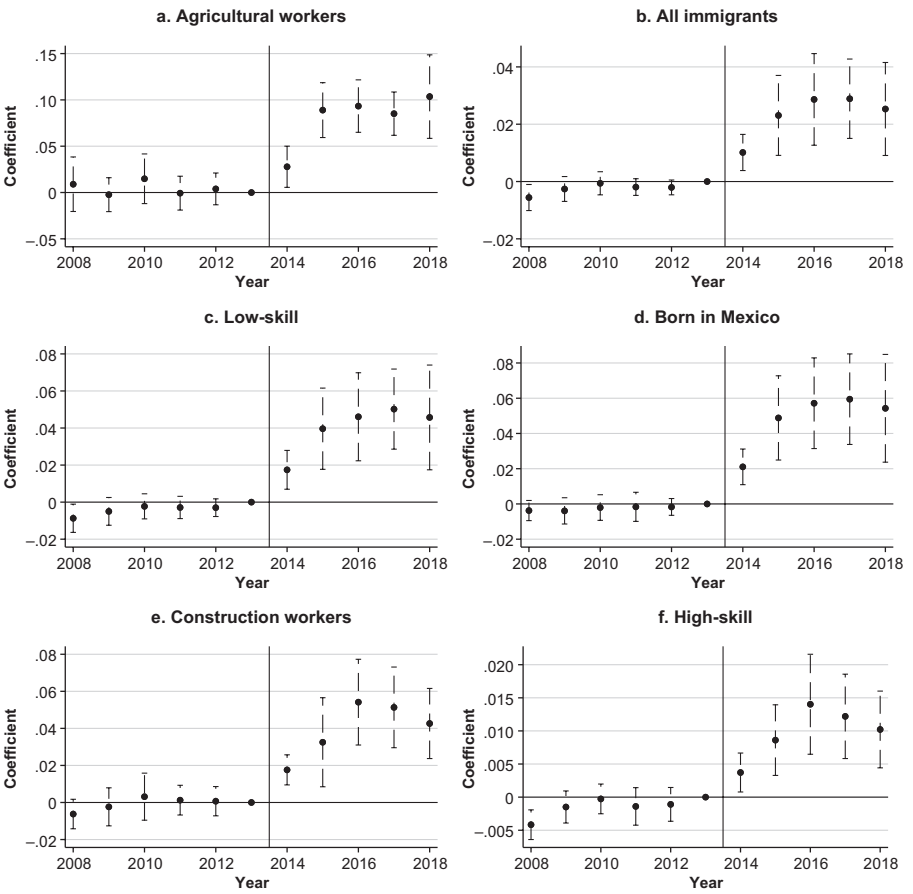


Fig. 4 ACA expansion effect on the share of immigrants in the Medicaid-only group. All panels show coefficients and 95% confidence intervals from state-clustered standard errors from implementing Eq. (4). We weigh the regressions by the state size of the relevant group. Coefficients capture differences between expansion and nonexpansion states in the share of immigrants in the indicated group who are categorized as legal by the Medicaid criteria alone, relative to differences in 2013. Panel a includes all immigrant agricultural workers in the ACS, and panel b includes all immigrants in the ACS. We follow the delineation of expansion and nonexpansion states in Kandilov and Kandilov (2022). Expansion states consist of all states with no or limited expansion prior to 2014 who expanded Medicaid eligibility in 2014, and nonexpansion states consist of states that did not expand Medicaid eligibility or that expanded Medicaid eligibility prior to 2014. Medicaid expansion states are: AR, AZ, CA, CO, CT, HI, IA, IL, KY, MD, MI, MN, ND, NH, NJ, NM, NV, OH, OR, RI, WA, WV.

the imputed legal share in expansion states relative to nonexpansion states. Panel b shows that these effects are smaller for the broader immigrant population but still statistically significant: by 2016, Medicaid expansion had resulted in a three-percentage-point increase in the imputed legal share in expansion states relative to nonexpansion states. These results are not surprising given the relatively high prevalence of low-income workers in the agricultural sector; low-income workers are more likely to have increased access to Medicaid because of the ACA compared with workers in

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

higher earning sectors. The other panels show results that support this claim. For low-skilled immigrant workers, immigrants from Mexico, and construction workers, Medicaid expansion increased the proportion imputed as legal in expansion states by five to six percentage points, while for high-skill immigrants the increase was only about one percentage point.

Medicaid Expansion and Wage Gaps

Next, we replicate results in Borjas and Cassidy (2019), who found that the wage premium to having legal status, estimated using ACS data, declined between 2013 and 2016. Using a simple decomposition, we explain why including Medicaid in the algorithm can be challenging for this analysis. In particular, covariate-adjusted average wages are significantly lower for immigrants in the Medicaid-only group than for other immigrants. Thus, even if rising Medicaid enrollments cause only moderate increases in the imputed share of legal immigrants, they can lead to large declines in wage gaps estimated using the full algorithm. We show evidence that ACA Medicaid expansion can explain most of the decline in the BC wage gap.

Following Borjas and Cassidy (2019), we define the true wage gap as the cofficient (β) on the true legal status variable (L) in a population regression of the log wage (W) on legal status and a vector of covariates ($\mathbf{X}$). This population regression can be written as

$$W = \beta L + \mathbf{X}'\gamma + \epsilon, \tag{5}$$

where L and $\mathbf{X}$ are uncorrelated with ϵ by construction. Borjas and Cassidy (2019) estimated Eq. (5) using L^F as a proxy for L, separately by year and worker group. Let β^j denote the coefficient on L^j in a population-level regression of W on L^j and $\mathbf{X}$, and $\hat{\beta}^j$ denote the ordinary least-squares (OLS) estimator of β^j. We refer to β^F as the "BC wage gap" and β^B as the "base wage gap." Our goal is to study how Medicaid inclusion in the algorithm affects the BC wage gap (i.e., why and by how much β^F and β^B differ), and how this effect changes with Medicaid expansion.[15]

Figure 5 replicates Figure 2A in Borjas and Cassidy (2019), which documents the trends in the OLS estimate of the BC wage gap, $\hat{\beta}^F$, between (imputed) legal and undocumented immigrant men for the entire workforce, and for low-skill and high-skill workers (panel a). For all three groups, the figure shows a small and somewhat constant wage gap from 2008 to 2013, which declines monotonically starting in 2013. The decline for high-skill workers is particularly sharp. The wage gap for this group declined from 6.1% in 2013 to 2.7% in 2016, for a 58% decrease.[16]

In panel b of Figure 5, we show the OLS estimate of the base wage gap, $\hat{\beta}^B$. Panels a and b show notable differences between $\hat{\beta}^B$ and $\hat{\beta}^F$, both in levels and in

[15] Note that, given population differences, using NAWS to correct for measurement error bias in ACS-based estimates is not recommended. Moreover, signing the bias of the BC wage gap is generally not possible.

[16] Borjas and Cassidy (2019) acknowledged the difficulty in pinpointing exactly why the wage gap decreased and that their findings rely on the imputation algorithm's validity.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

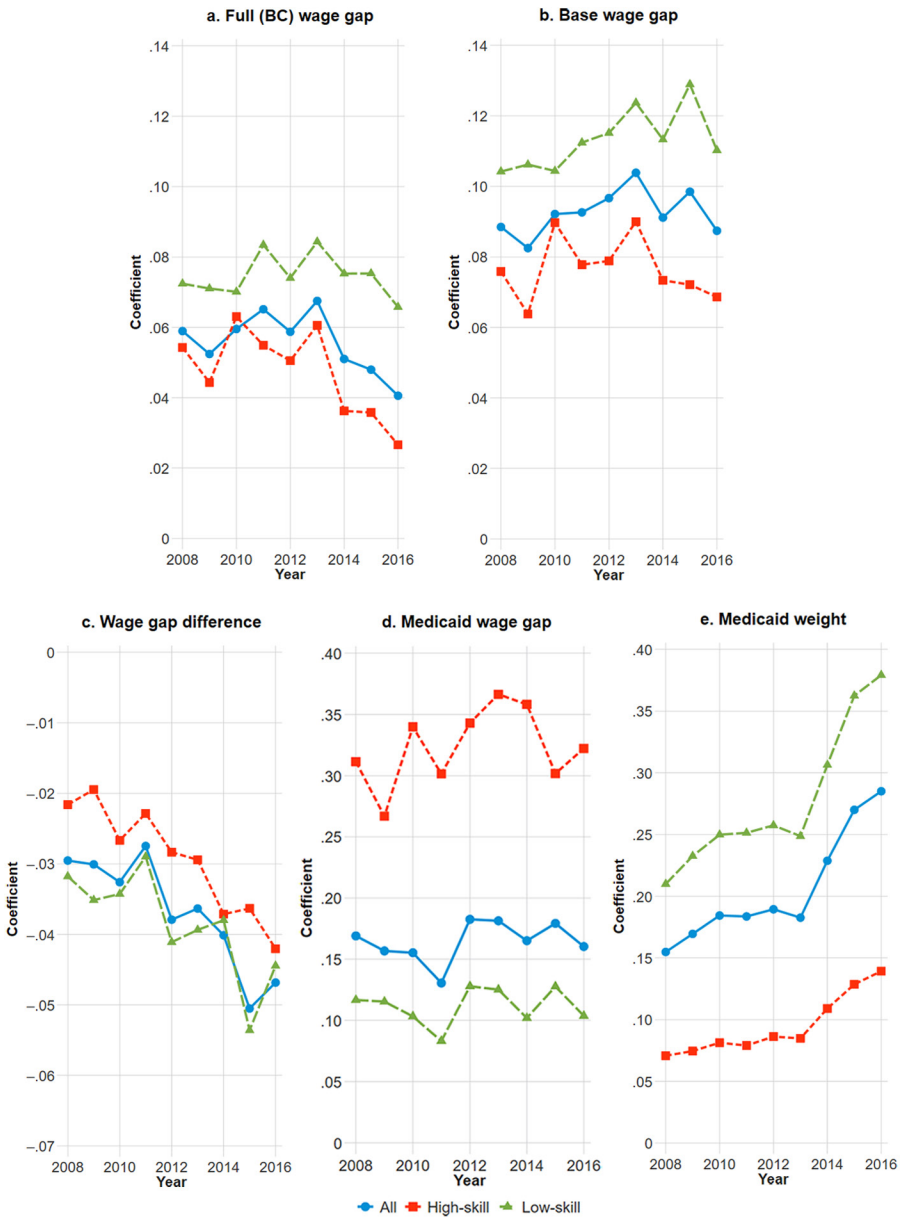


Fig. 5 Full and base algorithm wage gaps, wage gap difference, Medicaid group wage gap, and Medicaid group weight. This figure shows estimates of the wage penalty between legal and undocumented immigrants as in Borjas and Cassidy (2019). We regress (separately in each year) the log hourly wage on a dummy variable for legal status while controlling for age, educational attainment, country of birth, state of residence, years since migration, birthplace, and English language proficiency. In panel a, the legal status variable is the BC legal status indicator; in panel b, the legal status variable excludes Medicaid in its construction. Panel c shows the difference between panels a and b. In any given year, standard errors are very small (about 0.006) and are omitted from the graphs for clarity. Panel d shows estimates of the adjusted wage gap ($|\hat{\beta}^{M}|$) between the Medicaid-only group and all other immigrants. The Medicaid-only group consists of all immigrants classified as legal solely because of Medicaid usage. Panel e shows estimates of the Medicaid weight, $\hat{w}_{M}$.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

trends. While the Medicaid group constitutes a relatively small fraction of those imputed as legal by the full algorithm, Medicaid inclusion drastically alters wage gap estimates—$\hat{\beta}^B$ ranges from 1.5 to 2 times the size of $\hat{\beta}^F$. In contrast to the BC wage gaps, the base wage gaps are increasing moderately from 2008 to 2013 and do not monotonically decline after 2013. In fact, for the average male immigrant, the estimated base wage gaps are roughly equal in 2016 and 2008, and for the low-skill group, the base wage gaps are higher in 2016 than in 2008. To highlight the differences in trends between $\hat{\beta}^F$ and $\hat{\beta}^B$, panel c of Figure 5 shows the difference between the two, $\hat{\beta}^F - \hat{\beta}^B$. The figure makes it clear that the overall decline in $\hat{\beta}^F$ is explained in great part by including Medicaid in the algorithm.

To highlight the role of Medicaid expansion in the decline of $\hat{\beta}^F - \hat{\beta}^B$, we can use the fact that we decompose $\hat{\beta}^F$ into a weighted sum of $\hat{\beta}^B$ and the OLS estimator of the wage gap between the Medicaid-only group and all other immigrants, $\hat{\beta}^M$, namely:[17]

$$\hat{\beta}^F = \hat{\beta}^B \hat{w}_B + \hat{\beta}^M \hat{w}_M, \tag{6}$$

where the weights, $\hat{w}_B$, $\hat{w}_M > 0$, depend on the (residual) sizes of the group consisting of immigrants with legal markers other than Medicaid and the Medicaid group, respectively.

Perhaps unsurprisingly, considering that Medicaid's eligibility hinges on having a lower income, Medicaid-only immigrants have lower adjusted wages than other immigrants in our sample, so that $\hat{\beta}^M < 0$. Moreover, for all years in our sample, $\hat{w}_B \approx 1$ and thus $\hat{\beta}^M \hat{w}_M \approx \hat{\beta}^F - \hat{\beta}^B$ for all three groups. Hence, $\hat{\beta}^F - \hat{\beta}^B$ is declining because either (1) the gap between the wages of the Medicaid-only group and everyone else is widening (i.e., $|\hat{\beta}^M| \uparrow$), (2) the Medicaid group weight is getting larger (i.e., $\hat{w}_M \uparrow$), or (3) both. In panels d and e of Figure 5, we plot $|\hat{\beta}^M|$ and $\hat{w}_M$. Trends in both terms for all three groups follow a similar trajectory. Panel d shows no obvious trend in $|\hat{\beta}^M|$ over time and no evidence of a change following Medicaid expansion—the estimates are of similar magnitude in 2008 and 2016. Panel b shows that the Medicaid group weights, $\hat{w}_M$, are trending slightly upward from 2008 to 2013 (increasing by 27% for all immigrants) and then increase considerably from 2013 to 2016 (increasing by more than 50% for all immigrants). This latter increase coincides with the timing of Medicaid expansion (see panel b of Figure 3).

Panels d and e show that the long-term downward trends in $\hat{\beta}^F - \hat{\beta}^B$ shown in panel c are driven almost entirely by increases in $\hat{w}_M$. This is the term that we would expect to respond the most to changes in Medicaid enrollments, as it depends on the relative size of the Medicaid group.[18] Consequently, since β^B should be unaffected by increased Medicaid enrollments, and $\hat{\beta}^B$ is not declining monotonically, the sharp increase in $\hat{w}_M$ shown in Figure 5 suggests that the steady post-2013 decline in $\tilde{\beta}^F$ highlighted in Borjas and Cassidy (2019) is likely caused by Medicaid expansion. In

[17] In the online appendix, we show the full derivation of this decomposition of $\hat{\beta}^F$ as well as a more general decomposition.

[18] Note that, unlike $\hat{\beta}^M$, $\hat{w}_M$ is very unlikely to reflect changes in legal and undocumented wages and thus should be invariant to changes in the true wage gap β.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

Table 3 Labor supply elasticities (1994–2014) for the full and base algorithms

	Labor Force Participation Rate (1)	Positive Hours Worked Last Year (2)	Average Log Hours Worked (3)	Log Average Hours Worked (4)
A. Legal				
Full	0.0610*	0.1008***	–0.0701	0.1066**
	(2.11)	(4.01)	(–1.97)	(2.92)
Base	0.0754*	0.1078***	–0.0585	0.1185**
	(2.63)	(4.22)	(–1.63)	(3.03)
B. Undocumented				
Full	0.0509	0.0272	–0.0621	–0.0243
	(1.71)	(0.63)	(–0.94)	(–0.27)
Base	0.0376	0.0266	–0.075	–0.027
	(1.22)	(0.67)	(–1.25)	(–0.33)

Notes: All regressions include age–year, nativity–year, education–year, and age–education–nativity fixed effects. Standard errors are clustered at the age–education level (*t* statistics are reported in parentheses). The first two rows replicate columns 1–4 of rows 2–3 in Table 4 of Borjas (2017). The last two rows use the base algorithm indicator to construct the legal and undocumented indicators. The unit of analysis is the age–education–nativity–year cell. $N = 2{,}268$.

*$p < .05$; **$p < .01$; ***$p < .001$

Figure A12, we provide additional evidence to support this claim by showing that $\tilde{\beta}^F$ declines after 2013 only for the group of expansion states.

Medicaid Expansion and Labor Supply Elasticities

We replicate results in Borjas (2017), who found that CPS estimates of labor supply elasticities are highest for native-born workers, lower for legal immigrants, and near zero for the undocumented. Unlike our analysis of wage gaps, here we find that the results are robust to including the Medicaid criteria in the classification algorithm.

Our main specification follows that in Borjas (2017) and is given by

$$h = \phi_{at}^F + \phi_{st}^F + \phi_{nt}^F + \phi_{asn}^F + \sigma_L^F WL^F + \sigma_U^F WU^F + \sigma_N^F WN + \epsilon^F, \quad (7)$$

where h is a measure of labor supply for all workers in an age–education–year–legal status group, and N, L^F, and $U^F = 1 - L^F$ are indicators for native-born, foreign-born legal, and foreign-born undocumented workers, respectively. W is the mean log wage for workers in a given age–education–year group. The interaction of wages with indicators for worker legal status allows us to estimate separate elasticities for each legal status group, denoted by σ_N^F, σ_L^F, and σ_U^F. ϕ_{at}^F, ϕ_{st}^F, and ϕ_{nt}^F are year by age, education, and legal status fixed effects, and ϕ_{asn} are age–education–legal status fixed effects.

Table 3 shows our replication of the first four columns of Table 4 in Borjas (2017) for men only. In panel A, we show the estimates of the specification above using either the full algorithm, L^F (first row), or the base algorithm, L^B (second row), to assign legal status. Panel B shows the corresponding estimates for undocumented workers. For legal workers, the elasticities for the various labor supply measures are

CORRECTED PROOFS

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

smaller when using the full algorithm compared with the base one, while the opposite is true for undocumented workers. In both cases, these differences are not statistically significantly different from zero. Importantly, whether we use the full or base algorithm to assign legal documentation status, the labor supply elasticities are higher for legal workers than for undocumented workers.

In the online appendix, we express the full algorithm legal and undocumented elasticities in terms of the elasticities of the underlying groups (the group with other legal markers, the group without any legal markers, and the Medicaid-only group). Our decomposition shows that the results are similar between the two algorithms for three reasons. First, the labor supply elasticities between the Medicaid group and the other two groups are similar in magnitude. Second, the size of the Medicaid group is small. Third, the residual variance in labor supply within the Medicaid group is small relative to the other groups.

Conclusion

In this article, we evaluate the BC imputation algorithm, demonstrating that Medicaid inclusion significantly affects its performance. Using the NAWS, we show how Medicaid criteria contributed to recent fluctuations in the algorithm's accuracy, particularly as a result of ACA Medicaid expansion, which improved the classification of legal immigrants but worsened the classification of undocumented immigrants. Using the ACS, we show that despite its unlikely influence on the true legal share, ACA Medicaid expansion considerably increased the share of immigrants imputed as legal, particularly among low-income groups. We further examine the implications of Medicaid inclusion through two replications. Our results show significant effects in one case and minimal impact in the other, underscoring the need for researchers to carefully evaluate and justify the decision to include or exclude Medicaid from their imputation algorithm. More broadly, our findings suggest that the decision to include certain criteria in a logical imputation algorithm likely depends on the specific context and, thus, should be tailored to the specific research question and methodological approach.

Among studies that are closest to our work, a small literature has shown that estimates derived from logical imputation methods may be biased (Altman et al. 2023; Altman et al. 2021; Ro and Van Hook 2022; Spence et al. 2020; Van Hook et al. 2015). These studies evaluate bias by comparing model estimates using logically imputed legal status against those obtained from self-reported legal statuses in the SIPP. They mainly use linear probability or logistic regression models on cross-sectional national data. Van Hook et al. (2015) found that logical imputation can skew estimates of the link between legal status and health coverage, a finding echoed in subsequent research examining its effects on poverty (Spence et al. 2020), public health insurance (Altman et al. 2021), self-rated health (Ro and Van Hook 2022), and sick days (Altman et al. 2023). Our study adds to this literature in multiple ways. It is the first to track changes in imputation accuracy over time and across space and to demonstrate how specific markers of legal status drive accuracy. Second, to the best of our knowledge, it is the first to use the NAWS data in a validation exercise, despite its robust measure of legal status. Finally, we apply a measurement error framework

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

that has not been used in this setting. With this framework, we show that parameter estimates derived from the full and base algorithms can sometimes differ dramatically; in contrast, prior studies primarily focused on comparing estimates from the full algorithm to those from the self-reported measures in the SIPP.

Our findings provide insights for researchers applying logical imputation methods and for future studies assessing the performance of the logical imputation method. We have shown how researchers can decompose OLS estimates into components that reflect the importance of each legal marker in their algorithm. This decomposition reveals how the outcome of interest differs among the groups associated with each legal marker and the weight each marker receives in the final estimate. Researchers could apply these methods to demonstrate how each legal marker contributes to their final estimates and, if applicable, how these contributions evolve over time. When using these methods, researchers may want to consider whether these contributions are reasonable within their study's context.

As noted by Ro and Van Hook (2022) and Altman et al. (2023), BC legal markers may be correlated with unobserved variables influencing the outcome of interest. Failing to address these correlations could lead to biased estimates when using the BC imputed legal variable as a proxy for legal status. Altman et al. (2023) suggested that researchers evaluate the potential magnitude and direction of the biases associated with including certain markers in their algorithm, recognizing that these biases will vary with the research question and the corresponding outcome of interest. For this purpose, researchers could use measurement error models, like the one presented here. Such models could help researchers determine which markers to include or exclude from their algorithm and devise strategies to minimize potential biases (e.g., instrumental-variables-type strategies).

Improvements can also be made to logical imputation algorithms. One approach is to link survey data to administrative records, such as social security numbers, to impute legal status. However, U.S. federal statistical agencies are usually unwilling to allow researchers to form these linkages or even provide access when such linkages already exist (Van Hook et al. 2015). An alternate, more feasible approach is for future work to perform validation exercises aimed at improving logical imputation algorithms by adjusting the variable set or adding filters to certain variables. For example, our validation exercise shows that including spousal Medicaid participation in the algorithm reduces overall accuracy. This may be due to the fact that Medicaid eligibility is determined by household income. If an individual is not receiving Medicaid but their spouse is, the individual might be undocumented. This is because the nonrecipient individual is also eligible for Medicaid based on their household income. We recommend removing or filtering this criterion in future work.

Careful validation exercises can also help researchers clarify scenarios for which logical imputation is most appropriate. For instance, our replication of the BC wage gap study suggests that including Medicaid criteria in the algorithm may complicate analyses of trends when conducted over periods with substantive changes in Medicaid enrollment, such as during the ACA expansion period. However, these criteria may still be useful in cross-sectional analyses. In our case, the accuracy of the BC algorithm as applied to low-income subpopulations in the ACS could have improved markedly post-ACA in expansion states. Hence, cross-sectional analyses of these subpopulations could prove insightful, provided they

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685 5a4 by guest on 29 February 2024

exclude states where undocumented individuals are likely to misreport Medicaid use (e.g., California).

Lastly, researchers might explore the use of combined-sample multiple imputation (CSMI) methods, a statistical imputation method, when feasible. For CSMI to work effectively, two key assumptions need to be met: the donor and target datasets should sample the same underlying population, and both datasets should contain all variables of interest. There is evidence that this method can outperform logical imputation when these assumptions hold true (Hamilton et al. 2022; Van Hook et al. 2015). Notably, CSMI, like other statistical imputation methods (e.g., single imputation), can be combined with logical imputation. Building on Van Hook et al. (2015), future research could explore using statistical imputation methods to improve the logical method in scenarios where CSMI is not feasible. ■

Acknowledgments We would like to thank Tim Beatty, Phil Martin, Daniel Carroll, Brandon Restrepo, William Johnson, Wendy Zhang, Sophia Tanner, and seminar participants at the 2018 AAEA annual meeting for helpful comments. Any remaining errors are our own. This work was initiated while Tom Hertz was employed by the Economic Research Service, U.S. Department of Agriculture. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. government determination policy.

References

Aigner, D. J. (1973). Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics, 1,* 49–59.

Albert, C. (2021). The labor market impact of immigration: Job creation versus job competition. *American Economic Journal: Macroeconomics, 13*(1), 35–78.

Altman, C. E., Bachmeier, J. D., Spence, C., & Hamilton, C. (2023). Sick days: Logical versus survey identification of the foreign-born population in the United States. *International Migration Review, 57,* 395–420.

Altman, C. E., Spence, C., Hamilton, C., & Bachmeier, J. D. (2021). Health insurance coverage: Logical versus survey identification of the foreign-born. *Journal of Immigrant and Minority Health, 23,* 606–614.

Amuedo-Dorantes, C., & Antman, F. (2017). Schooling and labor market effects of temporary authorization: Evidence from DACA. *Journal of Population Economics, 30,* 339–373.

Amuedo-Dorantes, C., Arenas-Arroyo, E., & Sevilla, A. (2018). Immigration enforcement and economic resources of children with likely unauthorized parents. *Journal of Public Economics, 158,* 63–78.

Amuedo-Dorantes, C., Arenas-Arroyo, E., & Wang, C. (2020). Is immigration enforcement shaping immigrant marriage patterns? *Journal of Public Economics, 190,* 104242. https://doi.org/10.1016/j.jpubeco.2020.104242

Bachmeier, J. D., Van Hook, J., & Bean, F. D. (2014). Can we measure immigrants' legal status? Lessons from two U.S. surveys. *International Migration Review, 48,* 538–566.

Bhandari, R. A., Feigenberg, B., Lubotski, D., & Medina-Cortina, E. M. (2021). *Projecting trends in undocumented and legal immigrant populations in the United States* (NBER Center Paper, NB21-16). Cambridge, MA: National Bureau of Economic Research. Retrieved from https://www.nber.org/programs-projects/projects-and-centers/retirement-and-disability-research-center/center-papers/nb21-16

Bohn, S., Lofstrom, M., & Raphael, S. (2014). Did the 2007 Legal Arizona Workers Act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics, 96,* 258–269.

Borjas, G. J. (2017). The labor supply of undocumented immigrants. *Labour Economics, 46,* 1–13.

Borjas, G. J., & Cassidy, H. (2019). The wage penalty to undocumented immigration. *Labour Economics, 61,* 101757. https://doi.org/10.1016/j.labeco.2019.101757

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

Borjas, G. J., & Slusky, D. J. G. (2022). Health, employment, and disability. *American Journal of Health Economics, 8,* 1–29.

Bozick, R., & Miller, T. (2014). In-state college tuition policies for undocumented immigrants: Implications for high school enrollment among non-citizen Mexican youth. *Population Research and Policy Review, 33,* 13–30.

California Department of Health Care Services. (2017). *Medi-Cal's optional adult ACA expansion population—October 2016* (Medi-Cal Statistical Brief, No. 2017-002). California Department of Health Care Services, Research and Analytic Studies Division. Retrieved from https://www.dhcs.ca.gov/dataandstats/statistics/Documents/Expansion_Adults_201610.pdf

Capps, R., Bachmeier, J. D., Fix, M., & Van Hook, J. (2013). *A demographic, socioeconomic, and health coverage profile of unauthorized immigrants in the United States* (MPI Issue Brief, No. 5). Washington, DC: Migration Policy Institute.

Capps, R., Bachmeier, J. D., & Van Hook, J. (2018). Estimating the characteristics of unauthorized immigrants using U.S. Census data: Combined sample multiple imputation. *Annals of the American Academy of Political and Social Science, 677,* 165–179.

Capps, R., Gelatt, J., Ruiz Soto, A. G., & Van Hook, J. (2020). *Unauthorized immigrants in the United States* (Fact sheet). Migration Policy Institute. Retrieved from https://www.migrationpolicy.org/research/unauthorized-immigrants-united-states-stable-numbers-changing-origins

Chung, Y.-K., & Leigh, J. P. (2015). Medicaid use by documented and undocumented farm workers. *Journal of Occupational and Environmental Medicine, 57,* 329–333.

Clark, R. L., & King, R. B. (2008). Social and economic aspects of immigration. *Annals of the New York Academy of Sciences, 1136,* 289–297.

Docquier, F., Turati, R., Valette, J., & Vasilakis, C. (2019). Birthplace diversity and economic growth: Evidence from the U.S. states in the post–World War II period. *Journal of Economic Geography, 20,* 321–354.

Flores, S. M. (2010). The first state Dream Act: In-state resident tuition and immigration in Texas. *Educational Evaluation and Policy Analysis, 32,* 435–455.

Grzywacz, J. G., Quandt, S. A., Chen, H., Isom, S., Kiang, L., Vallejos, Q., & Arcury, T. A. (2010). Depressive symptoms among Latino farmworkers across the agricultural season: Structural and situational influences. *Cultural Diversity and Ethnic Minority Psychology, 16,* 335–343.

Hall, M., Greenman, E., & Farkas, G. (2010). Legal status and wage disparities for Mexican immigrants. *Social Forces, 89,* 491–513.

Hamilton, C., Altman, C. E., Bachmeier, J. D., & Spence, C. (2022). Legal status gradients in health insurance coverage: Estimates from a combine-survey multiple imputation approach. *Demographic Research, 47,* 453–488. https://doi.org/10.4054/DemRes.2022.47.16

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, 56,* 1–24.

Heinzel, J., Heller, R., & Tawil, N. (2021). *Estimating the legal status of foreign-born people* (Working Paper, No. 2021-02). Washington, DC: Congressional Budget Office. Retrieved from https://www.cbo.gov/publication/57022

Kandilov, A. M. G., & Kandilov, I. T. (2022). The impact of the Affordable Care Act Medicaid expansions on agricultural workers' health insurance coverage, medical care utilization, and labor supply. *American Journal of Agricultural Economics, 104,* 1026–1049.

Kaushal, N. (2006). Amnesty programs and the labor market outcomes of undocumented workers. *Journal of Human Resources, XLI,* 631–647.

KFF. (2023, September 17). *Key facts on health coverage of immigrants*. Kaiser Family Foundation. Retrieved from https://www.kff.org/racial-equity-and-health-policy/fact-sheet/health-coverage-of-immigrants/

Lessem, R., & Nakajima, K. (2019). Immigrant wages and recessions: Evidence from undocumented Mexicans. *European Economic Review, 114,* 92–115.

Lopez, M. H., Passel, J. S., & Cohn, D. (2021, April 13). *Key facts about the changing us unauthorized immigrant population*. Pew Research Center. Retrieved from https://www.pewresearch.org/fact-tank/2021/04/13/key-facts-about-the-changing-u-s-unauthorized-immigrant-population/

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024

Orrenius, P., & Zavodny, M. (2016). *Unauthorized Mexican workers in the United States: Recent inflows and possible future scenarios* (Working Paper, No. 436). Washington DC: Center for Global Development. Retrieved from https://www.cgdev.org/publication/unauthorized-mexican-workers-united-states-inflows

Passel, J. S., & Clark, R. L. (1998). *Immigrants in New York: Their legal status, incomes, and taxes* (Report). Washington, DC: Urban Institute. Retrieved from http://webarchive.urban.org/publications/407432.html

Passel, J. S., & Cohn, D. (2019, June 12). *Mexicans decline to less than half the U.S. unauthorized immigrant population for the first time*. Pew Research Center. Retrieved from https://pewrsr.ch/31s3twf

Passel, J. S., Van Hook, J., & Bean, F. D. (2004). *Estimates of legal and unauthorized foreign-born population for the United States and selected states, based on Census 2000* (Report). Washington, DC: Urban Institute.

Pena, A. A. (2010). Poverty, legal status, and pay basis: The case of U.S. agriculture. *Industrial Relations: A Journal of Economy and Society, 49,* 429–456.

Potochnick, S. (2014). How states can reduce the dropout rate for undocumented immigrant youth: The effects of in-state resident tuition policies. *Social Science Research, 45,* 18–32.

Richards, T. J. (2018). Immigration reform and farm labor markets. *American Journal of Agricultural Economics, 100,* 1050–1071.

Ro, A., & Van Hook, J. (2022). Comparing the effectiveness of assignment strategies for estimating likely undocumented status in secondary data sources for Latino and Asian immigrants. *Population Research and Policy Review, 41,* 449–464.

Sohn, H., & Pebley, A. (2020). *New approaches to estimating immigrant documentation status in survey data* (CCPR Population working papers). Los Angeles: California Center for Population Research, University of California, Los Angeles.

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,* 2327–2335.

Tamborini, C. R., & Villarreal, A. (2021). Immigrants' employment stability over the great recession and its aftermath. *Demography, 58,* 1867–1895. https://doi.org/10.1215/00703370-9429459

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, 52,* 329–354.

Van Hook, J., Morse, A., Capps, R., & Gelatt, J. (2021). Uncertainty about the size of the unauthorized foreign-born population in the United States. *Demography, 58,* 2315–2336. https://doi.org/10.1215/00703370-9491801

Warren, R. (2014). Democratizing data about unauthorized residents in the United States: Estimates and public-use data, 2010 to 2013. *Journal on Migration and Human Security, 2,* 305–328.

Warren, R., & Passel, J. S. (1987). A count of the uncountable: Estimates of undocumented aliens counted in the 1980 United States census. *Demography, 24,* 375–393.

Yasenov, V. I., Lawrence, D., Mendoza, F. S., & Hainmueller, J. (2020). Public health insurance expansion for immigrant children and interstate migration of low-income immigrants. *JAMA Pediatrics, 174,* 22–28.

Marcelo Castillo (corresponding author)
marcelo.castillo@usda.gov

Castillo • Economic Research Service, U.S. Department of Agriculture, Washington, DC, USA; https://orcid.org/0000-0002-8462-1896

Hill • Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA, USA; https://orcid.org/0000-0002-3996-1541

Hertz • Unaffiliated

CORRECTED PROOFS

Downloaded from http://read.dukeupress.edu/demography/article-pdf/doi/10.1215/00703370-11189687/2064639/11189687.pdf?guestAccessKey=2d55290b-8897-435b-a5aa-2d3662685a4 by guest on 29 February 2024