Title

Comparing imputation approaches for immigration status in ED visits: Implications for using electronic medical records

Journal

Health Services Research, 60(1)

ISSN

0017-9124

Authors

Axeen, Sarah

Gorman, Anna

Schneberk, Todd

et al.

Publication Date

2025-02-01

DOI

10.1111/1475-6773.14397

Peer reviewed

DOI: 10.1111/1475-6773.14397

METHODS ARTICLE

HSR Health Services Research

Comparing imputation approaches for immigration status in ED visits: Implications for using electronic medical records

Sarah Axeen PhD[1,2] | Anna Gorman MPH[3] | Todd Schneberk MD, MSHPM[1,4] | Annie Ro PhD, MPH[5]

[1]Department of Emergency Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[2]Leonard D. Schaeffer Center for Health Policy and Economics, University of Southern California, Los Angeles, California, USA

[3]Los Angeles County Department of Health Services, Alhambra, California, USA

[4]Gehr Family Center for Health Systems Science and Innovation, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[5]Department of Health, Society, & Behavior, Wen School of Public Health, University of California, Irvine, Irvine, California, USA

Correspondence
Annie Ro, Department of Health, Society, & Behavior, Wen School of Public Health, University of California, Irvine, Irvine, CA, USA.
Email: roa@hs.uci.edu

Funding information
National Institute on Minority Health and Health Disparities, Grant/Award Number: R21MD019086

Abstract

Objective: This study aimed to compare imputation approaches to identify the likely undocumented patient population in electronic health record (EHRs). EHR are a promising source of information on undocumented immigrants' medical needs and care utilization, but there is no verified way to identify immigration status in the data. Different approaches to approximating immigration status in EHR introduce unique biases, which in turn has major implications on our understanding of undocumented immigrant patients.

Study setting and design: We used a dataset of all emergency department (ED) visits from 2016 to 2019 in the Los Angeles Department of Health Services (LADHS) merged across patient medical records, demographic data, and claims data. We included all ED visits from our patient groups of interest and limited to patients at or over the age of 18 years at the time of their ED visit and excluded empty encounter records ($n = 1{,}106{,}086$ ED encounters).

Data sources and analytic sample: We created three patient groups: (1) US-born, (2) foreign-born documented, and (3) undocumented using two different imputation approaches: a logical approach versus statistical assignment. We compared predicted probabilities for two outcomes: an ED visit related to a behavioral health (BH) disorder and inpatient admission/transfer to another facility.

Principal findings: Both approaches provide comparable estimates among the three patient groups for ED encounters for a BH disorder and inpatient admission/transfer to another facility. Undocumented immigrants are less likely to have a BH diagnosis in the ED and are less likely to be admitted or transferred compared to the US-born.

Conclusions: Researchers should consider expanding EHR with administrative data when studying the undocumented patient population and may prefer a logical

approach to estimate immigration status. Researchers who rely on payer status alone (i.e., restricted Medicaid) as a proxy for undocumented immigrants in EHR should consider how this may bias their results. As Medicaid expands for undocumented immigrants, statistical assignment may become the preferred method.

KEYWORDS
electronic medical records, health disparities, immigrants, immigration status, imputation

What is known on this topic

- Electronic health records (EHRs) are a promising source of information on undocumented immigrants' medical needs and care utilization, but there is no verified way to identify immigration status.
- Researchers use proxy measures to define the likely undocumented population but different approximation approaches introduce unique biases.
- In survey data, statistically assigning immigration status is preferred over a logical algorithm, but it is unknown whether the same holds true in EHRs.

What this study adds

- This paper uses a unique EHR dataset from the Los Angeles County health system to compare different imputation approaches to identify the likely undocumented patient population in ED encounters.
- A statistical approach and a logical algorithm provide comparable estimates among patient groups with varying nativity and immigration status for behavioral health disorders and inpatient admission/transfer to another facility.
- When demographic data is merged with EHR, a logical approach to estimate immigration status is preferred given the ease of implementation. Statistical approaches may become important with Medicaid expansion.

1 | INTRODUCTION

Understanding the health status and medical needs of undocumented immigrants has become a major field of inquiry. This population has many social and economic risk factors due to their legal status: they face structural barriers to health care because of federal policy exclusions, they are exposed to unhealthy workplace conditions and potential exploitation in the informal economy, and they often live in economic precarity.[1] One growing source of information about undocumented immigrants' health needs and utilization comes from their medical visits, which are recorded in electronic health records (EHRs). While undocumented immigrants are explicitly ineligible for federal insurance (i.e., Medicaid and Medicare), undocumented immigrants can use medical services in the safety net system through limited federal coverage for emergency health needs or state and local programs that offer health coverage for their undocumented residents. For instance, California, the setting of this study, has gradually expanded Medi-Cal (the state-run Medicaid program) for undocumented immigrants and fully expanded the program to all residents, regardless of immigration status, in January 2024. This increased access to care will make EHR an even more important source of data as more undocumented immigrants utilize formal health care services.

Examinations of medical records have found undocumented patients to have some better hospitalization outcomes, but poorer outcomes for specific conditions such as End Stage Renal Disease when compared to their documented counterparts.[2–5] While EHR can provide important insight into undocumented immigrants' health status at the point of needing medical care and their utilization patterns in the formal health care system, there is no way to definitively determine immigration status in medical records. As a result, researchers must use proxy measures to define the likely undocumented population. Different approaches to approximating immigration status in EHR introduce unique biases, which in turn has major implications on our understanding of health care utilization and outcomes for undocumented immigrants.

One of the most common ways to approximate immigration status in EHR has been to use payer status. A number of studies have used restricted/emergency Medicaid to define likely undocumented patients in EHR,[6–8] as undocumented immigrants have been explicitly ineligible for full-scope Medicaid under federal law. Conversely, patients in full-scope Medicaid programs are considered documented. Yet this proxy can overestimate health care utilization, as undocumented patients enrolled in restricted Medicaid are likely those who are more familiar with the formal health care system and are less vulnerable and fearful of interacting with bureaucratic institutions.[9] Focusing on the restricted Medicaid population alone also leaves out other undocumented immigrants who are uninsured or are above the federal income requirement for Medicaid.

Given the drawbacks of using payer status alone, it is imperative to find other ways to approximate immigration status in the EHR.

Researchers in the social sciences often use demographic or socioeconomic features, such as place of birth or enrollment in federal safety net programs (i.e., Social Security, food stamps) to define undocumented status. This is called a "logical" approach and is used in survey data when immigration status is not explicitly collected.[10] While the EHR does not routinely have detailed demographic information, it can be supplemented with administrative health system data to get more patient information. For instance, health system registration data often collects information on presence/absence of social security number, place of birth, preferred language, and detailed address. However, even having demographic data in the EHR may not fully solve the proxy problem. In survey data, it appears that there are larger biases in the logical approach when compared to other imputation approaches.[11–13] In particular, methods that statistically assign people to different immigration status using probabilistic models have less bias than logical methods, as the former uses missing data patterns to impute immigration status. To our knowledge, there has yet to be comparable work comparing imputation approaches in EHR data.

This paper uses a unique dataset from the Los Angeles Department of Health Services (LADHS) merged across patient medical records, demographic data, and claims data to compare different imputation approaches to identify the likely undocumented patient population. We move beyond previous approaches that have heavily relied on payer status and supplement standard EHR with hospital administrative data to get more patient demographic information. We aim to understand how different definitions of immigration and/or documentation status impact estimates of patient-related outcomes. We compare two different imputation approaches in our unique dataset of EHR—one based on a logical approach and another based on statistical assignment. We also compare our estimates for likely undocumented and documented patients across two very different types of outcomes, whether a patient had a mental health emergency department (ED) visit and whether a patient was transferred to another facility, to evaluate consistency in our definitions.

Our data is especially well-suited to conduct this analysis, as we have harvested more demographic data from the LADHS EHR than is typically used by researchers. This detailed patient information enables us to build a robust set of predictors of immigration status. We also leverage the unique local health care landscape of Los Angeles County to identify a subgroup of patients who are enrolled in a local health coverage program (MyHealth LA, or MHLA) for low-income residents who were not eligible for Medicaid at the time because of their immigration status (i.e., undocumented). We use this known population of undocumented patients to check and refine our imputation approaches. Our findings have major implications on future work that uses patient medical records to understand undocumented immigrants' health utilization and outcomes.

2 | METHODS

2.1 | Data and sample

We obtained a novel dataset from the LADHS of all ED encounters from 2016 to 2019 from all three major public hospitals with an ED in Los Angeles County. The dataset contains a patient identifier, location of service, disposition for the visit, discharge diagnoses for the ED encounter (or associated inpatient stay), primary payer for the visit, and a detailed collection of patient demographics. Included in the demographic detail are self-reported measures of race, ethnicity, country of birth, language spoken, zip code of residence, presence or absence of a social security number, gender, and age. We augmented the existing dataset with information on patients enrolled in a local health coverage program only available to undocumented immigrants. From that program's data, we obtained matched patient identifiers and dates of enrollment in the program to support the analytic approach. We limited our analysis to patients at or over the age of 18 years at the time of their ED visit and excluded those visits where the treatment facility was missing or where the registration was in error as there is a lack of documentation of diagnoses available for these patients. Our final sample was 1,106,086 ED encounters across all patient types.

2.2 | Defining immigration status

We examined three categories of immigration status: (1) United States (US)-born, (2) foreign-born documented, and (3) undocumented. We classified patients into these categories using two different imputation approaches. The first, hereafter our logical definition (LD), classifies patients in the following hierarchical manner: If the country of birth is reported as the United States, the patient is classified as US-born and all other responses (including missing) are categorized as non-US born. Among the non-US born, those individuals who have any of the following: a non-missing Social Security number, Medicare coverage, or full-scope Medicaid coverage, are classified as documented immigrants. Finally, the remaining group is classified as undocumented immigrants. These individuals lack a reported Social Security number, are enrolled in the local health coverage program referenced above, or have coverage through other public programs earmarked for undocumented immigrants, such as restricted Medicaid.

The second imputation approach uses statistical assignment. A notable feature of our LD is that it treats missing information as if it is systematically missing, and therefore meaningful, and not just omitted at random. As a result, we construct a competing definition, hereafter our multiple imputation definition (MID). The MID makes the same assumption as the LD, that individuals reporting the United States as their country of birth are US-born and those reporting another country are non-US born, but patients without country of birth information are not classified and treated as missing. Among those born outside the United States, those with full-scope Medicaid or Medicare coverage are classified as documented, and those with payers explicitly available only to the undocumented are classified as undocumented, including restricted Medi-Cal and the county program for undocumented immigrants. This approach allows us to categorize approximately 80% of our sample into US-born, documented immigrants, and undocumented immigrants. We consider the remaining 20% of our sample as missing immigration status information and categorized them using multiple imputation.

Specifically, we employ multiple imputation with chained equations and predictive mean matching relying on the 10 nearest neighbor matches to estimate our outcomes of interest. We compared our known undocumented population (i.e., those enrolled in the local coverage program) with the non-US born population to determine variables that provided meaningful information on characteristics associated with undocumented status. We identified language, facility of treatment, health insurance status, region of birth (country of birth recategorized as US, Mexico, Other Central America, Asia, and Other), and percent of their home Public Use Microdata Area (PUMA) that is undocumented as differing significantly between these populations. Therefore, the variables included in our imputation equation are gender, region of birth, ethnicity, race, share of the patient's PUMA of residence that is undocumented, presence of a Social Security number, age, insurance status, year of encounter, quarter of encounter, treatment facility, language spoken, and our outcomes of interest: a binary flag for visits with a behavioral health (BH) visit and discharge status from the ED encounter. As gender, region of birth, ethnicity, race, and PUMA documentation share also included missing values, those measures were also multiply imputed from the same set of donor variables. See Table 1 for a complete account of rates of missing information.

2.3 | Analytic approach

To understand how varying estimates of immigration status impact patient or encounter-level findings, we perform two sets of multivariate logistic regressions. One examines the likelihood that patients are diagnosed with a BH disorder (CCSR codes MBD001–MBD014 and MDB017–MBD026) and the other the likelihood that a patient is admitted to the same facility or transferred to another health care facility, based on the recorded discharge disposition from the patient's ED encounter. The BH measure captures variation in patterns of care by diagnosis while inpatient admissions measure variations of patterns of care by severity of the patient's ED encounter. For each outcome, we perform multivariate logistic regressions using either the LD or MID coding of immigration status as our key independent variable. Regressions also control for age, race, ethnicity, language spoken at home, treatment facility, insurance status, gender, year, and quarter of the encounter. We report predicted probabilities.

3 | RESULTS

From 2016 to 2019, there were 1,106,086 ED encounters across three hospitals in the LADHS healthcare system. Within those encounters, 221,737 (20% of the cases) in our MID of immigration status had missing values for immigration status. In addition, 2.7% observations were missing ethnicity, 3.3% were missing race, and 2.3% were missing region of birth. Only 55 were missing gender. After multiple imputation, all variables were successfully imputed. See Table 1 for a complete accounting of missing observations.

Among the more than 1 million ED visits examined in the study, 174,391 (15.8%) patients had a BH diagnosis and 217,196 (19.6%) had visits of high enough severity to warrant either inpatient admission to the same facility or transfer to another facility. As shown in Table 1, the vast majority of the population utilizing the LADHS safety net health care system were Hispanic (63.1%), born in the United States (45.7%) or Mexico (28.6%), and had some form of insurance payer (whether public or private). We use the term "Hispanic" to be consistent with the ethnicity coding in the EHR. A slight majority of patients were male (54.5%), and the average age of patients was about 47 years old.

When comparing all patients to those who had any BH-related visit, there is a marked difference in the demographics of the patient population. Most notably, patients with a BH diagnosis are disproportionately likely to be US-born (70.0%), Black (20.2%), non-Hispanic (50.9%), male (67.3%), uninsured (11.3%), and English speaking (80.0%). However, patients who are admitted or transferred tend to more closely resemble the whole LADHS population reviewed than does the BH population, though are slightly more likely to be non-Hispanic (41.1%), male (60.7%), and insured (95.0%) than the broader population.

We compared the demographic characteristics of the undocumented population derived by the MID and LD to two known undocumented populations in the data: patients enrolled in restricted Medicaid and patients enrolled in the local health coverage program for undocumented immigrants, MyHealth LA (MHLA) (see Table S1). We did this to confirm that our empirically derived undocumented patient populations had comparable characteristics to known undocumented populations, as well as to highlight potential biases in the undocumented sample when relying on payer status alone. In general, we found that LD and MID undocumented populations look comparable to the known undocumented groups in their demographic characteristics and BH disorders and inpatient admission/transfer to another facility. However, both the restricted Medicaid and MHLA populations were more likely to be Hispanic, female, insured, and Spanish-speaking that the MID or LD populations. This suggests that the restricted Medicaid and MHLA programs may be capturing a select population of undocumented immigrants. For instance, the relatively higher precent of Spanish language use among the restricted Medicaid (87%) and MHLA (90%) patients compared to the LD (80%) and MID (82%) groups may point to unequal outreach to and enrollment of particular subsets of the undocumented population. Further, both the LD and MID provided a much larger sample of undocumented immigrants than using the restricted Medicaid proxy measure or MHLA patient population alone.

Next, we compared overall ED visits across immigration statuses by our two imputation approaches. Figure 1 shows that the unadjusted proportion of ED visits by immigration status varies slightly between the LD and the multiple imputation definition. The LD slightly understates the share of patients who are US-born and overstates the share who are immigrants relative to the MID imputation by about 3.6 percentage points. In particular, the MID estimates that 49.3% of visits are by the US-born, 19.0% by documented immigrants, and 31.7% by undocumented immigrants compared to 45.7%, 21.2%,

TABLE 1 Description of population utilizing LA County DHS Emergency Departments (ED).

	All ED encounters		Behavioral health		Admission or transfer	
	N	%	*N*	%	*N*	%
Observations	1,106,086	100	174,391	15.8	217,196	19.6
Region of birth						
United States	505,433	45.7	122,133	70.0	106,097	48.8
Mexico	316,232	28.6	21,298	12.2	55,602	25.6
Asia	53,310	4.8	4092	2.3	13,020	6.0
Central America	153,185	13.8	9491	5.4	24,163	11.1
Other	52,145	4.7	6913	4.0	9827	4.5
Missing	25,781	2.3	10,464	6.0	8487	3.9
Social security number						
Any SSN	719,017	65.0	126,775	72.7	143,920	66.3
No SSN	387,069	35.0	47,616	27.3	73,276	33.7
Race						
White	111,021	10.0	23,471	13.5	23,637	10.9
Black	155,245	14.0	35,204	20.2	33,395	15.4
Asian	47,290	4.3	4574	2.6	12,112	5.6
Other	756,283	68.4	103,181	59.2	140,832	64.8
Missing	36,247	3.3	7961	4.6	7220	3.3
Ethnicity						
Hispanic	697,889	63.1	77,506	44.4	121,193	55.8
Not Hispanic	377,789	34.2	88,706	50.9	89,159	41.1
Missing	30,408	2.7	8179	4.7	6844	3.2
Gender						
Male	603,288	54.5	117,293	67.3	131,845	60.7
Female	502,743	45.5	57,091	32.7	85,334	39.3
Missing	55	0.0	7	0.0	17	0.0
LA DHS facility						
Facility 1	545,335	49.3	90,189	51.7	111,283	51.2
Facility 2	291,099	26.3	43,700	25.1	56,364	26.0
Facility 3	269,652	24.4	40,502	23.2	49,549	22.8
Insurance status						
Insured	1,005,140	90.9	154,764	88.7	206,313	95.0
Uninsured	100,946	9.1	19,627	11.3	10,883	5.0
Any Medicare	106,620	9.6	19,312	11.1	38,083	17.5
Any full scope Medicaid managed care	272,398	24.6	52,360	30.0	53,042	24.4
Any undocumented program	106,778	9.7	5144	2.9	15,014	6.9
Any restricted Medicaid	217,416	19.7	12,874	7.4	36,286	16.7
Preferred language						
English	613,590	55.5	139,465	80.0	129,688	59.7
Spanish	444,350	40.2	29,320	16.8	76,931	35.4
Other	48,146	4.4	5606	3.2	10,577	4.9

Note: All estimates are at the encounter level. Any behavioral health visit indicates ED encounters with any diagnosis related to behavioral health. Admission or transfer indicates any ED visit that resulted in the patient being admitted to the same facility's inpatient floor or to another facility for additional care.

Abbreviation: ED, emergency department.

Source: Authors' analysis of electronic medical record data from the LA County DHS system from 2016 to 2019. Underlying data are a combination of patient enrollment data and codes resulting from physician treatment decisions.

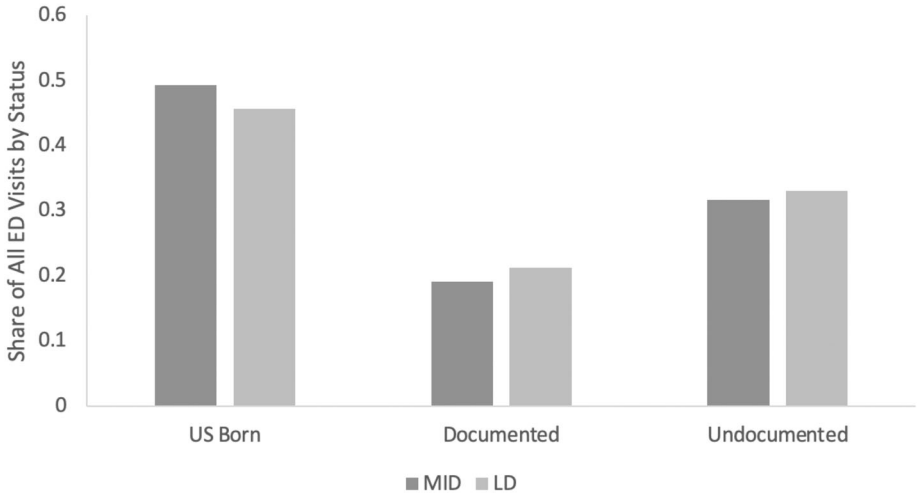

FIGURE 1 Total emergency department utilization by immigration status definition. All estimates are at the encounter level. MID indicates the multiple imputation definition of immigration status, and LD indicates the logical definition of immigration status. *Source*: Authors' analysis of electronic medical record data from the LA County Department of Health Services system from 2016 to 2019. Underlying data are a combination of patient enrollment data and codes resulting from physician treatment decisions.

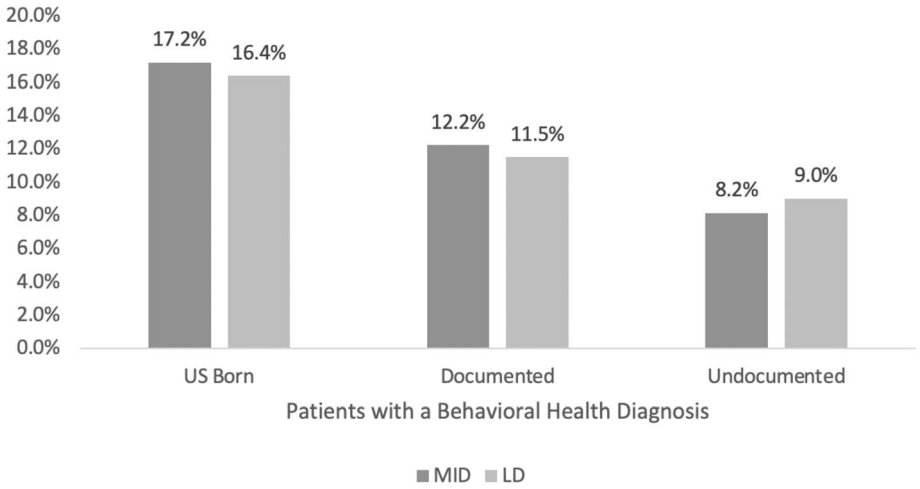

FIGURE 2 Regression-adjusted probability of emergency department visits for behavioral health diagnoses by immigration status definition. All estimates are at the encounter level. MID indicates the multiple imputation definition of immigration status, and LD indicates the logical definition of immigration status. *Source*: Authors' analysis of electronic medical record data from the LA County Department of Health Services system from 2016 to 2019. Underlying data are a combination of patient enrollment data and codes resulting from physician treatment decisions.

and 33.1%, respectively, for the LD. Notably, when compared to a commonly used measure of documentation status—use of restricted Medicaid services—both estimates are markedly larger than the share of visits covered by restricted Medicaid (19.7%). For full results, see Table S2.

Key to our comparison of how these estimates of immigration status compare is to test their predictions across a variety of patient-related outcomes. As shown in Figure 2, the predicted probability of an ED encounter with a BH diagnosis is similar across the two estimates. The predicted probability using the MID is slightly higher for the US-born and documented and lower for the undocumented when compared with the LD, but there is no re-ordering of the categories and differences are smaller than 1 percentage point. When compared to patients with restricted Medicaid, both the MID and LD are larger (6.2% and 7.4% compared to 5.9%, see Table S1).

When examining the predicted probability that patients are admitted or transferred, the two definitions are also quite similar. There is a 1.3 percentage point difference in the predicted probability for documented patients comparing the MID to the LD, but again there is no re-ordering of categories. However, in the MID, the documented appear to behave more similarly to the US-born, while in the LD, they appear to behave more similarly to the undocumented. When compared to patients with restricted Medicaid, both estimates are lower, with a 0.9 percentage point difference between the MID and restricted Medicaid. For full regression results, see Table S2, Figure 3.

We also conducted the MID again without insurance status in the prediction model and re-ran the logistic regressions and calculated the predicted probabilities of our outcomes (Table S4). While the point estimates differed, the patterns remained the same: undocumented patients were the least likely to have a mental health ED visit and an ED visit that resulted in a hospital admission or transfer.

4 | DISCUSSION

Researchers use proxy measures to examine the health utilization and outcomes of undocumented immigrants, given the absence of a definitive record of immigration status. In this paper, we compared two approaches to defining immigration status in EHR. Medical records are a growing source of information on undocumented immigrants' utilization and health status, but researchers have not examined the

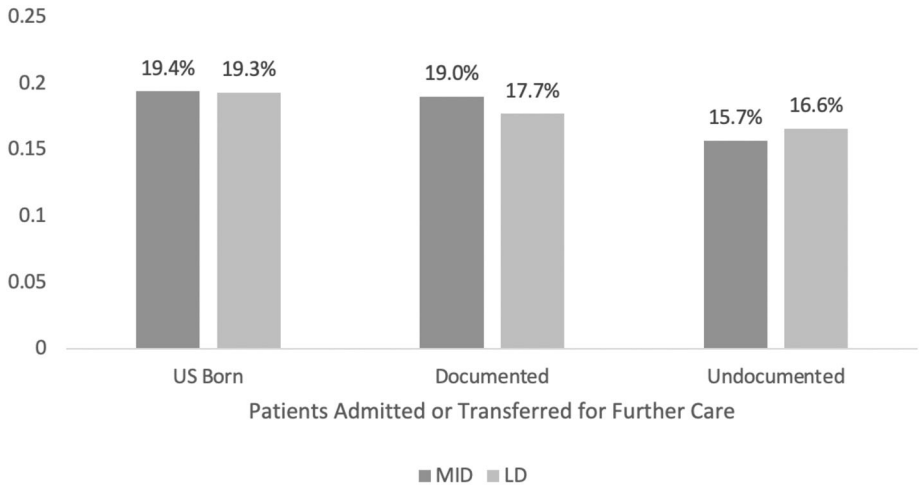

FIGURE 3 Regression-adjusted probability of emergency department visits resulting in admission or transfer by immigration status definition. All estimates are at the encounter level. MID indicates the multiple imputation definition of immigration status, and LD indicates the logical definition of immigration status. *Source*: Authors' analysis of electronic medical record data from the LA County Department of Health Services system from 2016 to 2019. Underlying data are a combination of patient enrollment data and codes resulting from physician treatment decisions.

validity of different imputation approaches. We found that both logical and statistical definitions provide comparable estimates among three immigration status categories for two different outcomes; there is no re-ordering of the categories, and nearly all findings are significant. These findings hold true for outcomes where the patient population is quite similar to the full sample (admit/transfer) and where the patient population diverges quite a bit from the full sample (BH visit).

Our finding that the logical approach is comparable to the statistical assignment approach diverges from other research that has found weaknesses in the logical method.[11–13] We suspect that our detailed demographic information, particularly presence/absence of Social Security number, is the reason for the similarity between the logical and statistical approach. Even though Social Security numbers can be false or inaccurate, they can be useful to triangulate with other demographic characteristics or payment sources. We acknowledge that our EHR may be uniquely detailed and we are unsure whether we would have found similar results between the two approaches with less demographic information. We therefore recommend that researchers interested in using EHR to study undocumented immigrants seek out robust administrative data to ensure the best approximation of immigration status. We found that the restricted Medicaid patient group in our sample was more Spanish-speaking, female, and Hispanic than the undocumented population empirically derived from our imputation approaches. Researchers who are using restricted Medicaid as a proxy for undocumented immigrants in EHR consider how this may bias their results.

We believe that the methods we propose in this paper will still be pertinent after Medicaid expansion in states like California. First, it will be critical that researchers who are interested in undocumented immigrants' medical needs and healthcare utilization augment standard EHR with more demographic information. Payer status alone will not determine patient immigrant status, as documented and undocumented immigrants alike will have full-scope Medicaid; it will be up to researchers to seek out creative data solutions to improve health care for this marginalized population. In sensitivity checks with insurance status removed from the MID imputation, we found the predicted probabilities of our outcomes for the undocumented population to be comparable to the original approach, underscoring the value of additional demographic variables in predicted immigration status. While combining administrative variables with medical records may be uncommon in the academic literature, these data are available in health care centers and systems. We acknowledge that this kind of merging should undergo the highest security review and rigorous IRB approval to ensure patient privacy protection.

However, we believe that pre-expansion payer information can still be useful to impute post-expansion immigration status in states with expanded Medicaid if data is maintained longitudinally. Researchers can either apply the LD or MID approaches to pre-expansion patient data that can then be extended to post-expansion data. Alternatively, they can treat post-expansion encounters as "missing" immigration status and then use MID to impute the missing data. For other states that have not yet expanded Medicaid, the logical approach may be preferred if researchers have access to detailed demographic data merged with EHR and the undocumented immigrants in their sample still utilize restricted Medicaid, given the ease of implementation over a statistical approach.

The main purpose of our paper was to compare imputation approaches, but we did find significant disparities in ED visits related to BH and whether the patient was admitted or transferred from the ED among the undocumented and documented immigrant and US-born citizen populations. There are many theories to why this utilization is significantly different among the groups, including the immigrant paradox, structural vulnerability, racial oppression, and poor access to care. There is currently no gold standard that provides the "true" estimates for our ED outcomes for undocumented patients, and we caution against using our estimates to establish prevalence of mental health ED visits and admission/transfers. Instead, we highlight the relative comparison of the three patient populations and how the imputation methods differed. Nonetheless, we found that the two empirically derived undocumented patients groups (the LID and MID undocumented) look comparable to the known undocumented groups in their demographic characteristics and BH disorders and inpatient admission/transfer to another facility. However, we note that while our outcomes are broadly consistent, the patient characteristics of the

restricted Medicaid population differs from our undocumented populations highlighting the need for more nuanced definitions of documentation status. Further studies should be performed to understand the mechanisms for these disparities.

We also acknowledge potential underreporting of our outcomes. LADHS does not perform physician billing, and the recorded diagnoses may only represent those related to the presenting complaint for the current visit. This may especially affect patients with minimal connection to the healthcare system, contributing to lower counts of mental health diagnoses, and past medical history in general of the undocumented population. Additionally, as with all ED data, certain encounters are more prone to be missing, such as high acuity visits, patient dead on arrival, and those with incomplete visits, such as left without being seen or left before treatment was completed.[14] Lastly, demographic data are collected by trained multilingual registration staff, but human error still exists.

5 | CONCLUSIONS

Undocumented patient populations experience significant disparities in outcomes, related to structural barriers to their health. Accurate determination of immigration status in EHRs is critical for understanding health care utilization and population health outcomes. We demonstrate that a simplified LD can be used to approximate this population using demographic information abstracted from health system claims data. As more instead of relying solely on restricted Medicaid status to estimate undocumented immigrant populations in healthcare, health systems and hospitals can use administrative, demographic, financial, and insurance data to better identify their undocumented populations to understand population utilization and outcomes. By documenting health outcomes among this structurally vulnerable population, researchers can identify crucial areas of interventions, especially in health systems that contain similar population health coverage programs. With more robust data, and better ways to approximate the undocumented population, health systems will be better equipped to identify root causes health disparities and deploy interventions to improve the health of undocumented communities.

ACKNOWLEDGMENTS

The authors would like to acknowledge Chun Nok Lam for his preparation of the data, the staff at the USC Leonard D. Schaeffer Center for data support, and and the anonymous peer-reviewers of the article.

FUNDING INFORMATION

The research reported in this paper was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under award number R21MD019086.

ORCID

Sarah Axeen https://orcid.org/0000-0002-1538-0013
Annie Ro https://orcid.org/0000-0001-9684-5566

REFERENCES

1. Samra S, Taira BR, Pinheiro E, Trotzky-Sirr R, Schneberk T. Undocumented patients in the emergency department: challenges and opportunities. *West J Emerg Med*. 2019;20(5):791-798. doi:10.5811/westjem.2019.7.41489
2. Cervantes L, Tuot D, Raghavan R, et al. Association of emergency-only vs standard hemodialysis with mortality and health care use among undocumented immigrants with end-stage renal disease. *JAMA Intern Med*. 2018;178:188-195. doi:10.1001/jamainternmed.2017.7039
3. Ro A, Yang HW, Du S, Hanlon CL, Young AS. Severity of inpatient hospitalizations among undocumented immigrants and Medi-Cal patients in a Los Angeles, California, hospital: 2019. *Am J Public Health*. 2021;111(11):2019-2026. doi:10.2105/AJPH.2021.306485
4. Nguyen OK, Vazquez MA, Charles L, et al. Association of scheduled vs emergency-only dialysis with health outcomes and costs in undocumented immigrants with end-stage renal disease. *JAMA Intern Med*. 2019;179:175. doi:10.1001/jamainternmed.2018.5866
5. Siddharthan K, Ahern M. Inpatient utilization by undocumented immigrants without insurance. *J Health Care Poor Underserved*. 1996;7:355-363. doi:10.1353/hpu.2010.0068
6. Ro A, Bruckner TA, Huynh MP, Du S, Young A. Emergency department utilization among undocumented Latino patients during the COVID-19 pandemic. *J Racial Ethn Health Disparities*. 2022;10:2020-2027. doi:10.1007/s40615-022-01382-8
7. DuBard CA, Massing MW. Trends in emergency Medicaid expenditures for recent and undocumented immigrants. *JAMA*. 2007;297(10):1085-1092. doi:10.1001/jama.297.10.1085
8. Swartz JJ, Hainmueller J, Lawrence D, Rodriguez MI. Expanding prenatal care to unauthorized immigrant women and the effects on infant health. *Obstet Gynecol*. 2017;130(5):938-945. doi:10.1097/AOG.0000000000002275
9. Saluja S, McCormick D, Cousineau MR, et al. Barriers to primary care after the affordable care act: a qualitative study of Los Angeles safety-net patients' experiences. *Health Equity*. 2019;3(1):423-430. doi:10.1089/heq.2019.0056
10. Borjas GJ. The labor supply of undocumented immigrants. *Labour Econ*. 2017;46:1-13. doi:10.1016/j.labeco.2017.02.004
11. Ro A, Van Hook J. Comparing Assignment Strategies for Estimating Likely Undocumented Immigrants in Secondary Data Sources. 2020.
12. Spence C, Bachmeier JD, Altman CE, Hamilton C. The association between legal status and poverty among immigrants: a methodological caution. *Demography*. 2020;57(6):2327-2335. doi:10.1007/s13524-020-00933-0
13. Altman CE, Spence C, Hamilton C, Bachmeier JD. Health insurance coverage: logical versus survey identification of the foreign-born. *J Immigr Minor Health*. 2020;23:606-614. doi:10.1007/s10903-020-01045-y
14. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. *J Electron Health Data Methods*. 2013;1(3):1035.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Axeen S, Gorman A, Schneberk T, Ro A. Comparing imputation approaches for immigration status in ED visits: Implications for using electronic medical records. *Health Serv Res*. 2025;60(1):e14397. doi:10.1111/1475-6773.14397