# Answering calls for rigorous health equity research: a cross-sectional study leveraging electronic health records for data disaggregation in Latinos

John Heintzman,[1] Dang Dinh,[1] Jennifer A Lucas [ORCID],[1] Elena Byhoff,[2] Danielle M Crookes,[3] Ayana April-Sanders,[4] Jorge Kaufmann,[1] Dave Boston,[5] Audree Hsu,[6] Sophia Giebultowicz,[5] Miguel Marino[1]

[1]Family Medicine, Oregon Health & Science University, Portland, Oregon, USA
[2]Department of Medicine, University of Massachusetts, Boston, Massachusetts, USA
[3]College of Social Sciences and Humanities, Northeastern University, Boston, Massachusetts, USA
[4]Rutgers School of Public Health, Piscataway, New Jersey, USA
[5]OCHIN, Portland, Oregon, USA
[6]California University of Science and Medicine, Colton, California, USA

**Correspondence to**
Dr John Heintzman;
heintzma@ohsu.edu

## ABSTRACT

**Introduction** Country of birth/nativity information may be crucial to understanding health equity in Latino populations and is routinely called for in health services literature assessing cardiovascular disease and risk, but is not thought to co-occur with longitudinal, objective health information such as that found in electronic health records (EHRs).

**Methods** We used a multistate network of community health centres to describe the extent to which country of birth is recorded in EHRs in Latinos, and to describe demographic features and cardiovascular risk profiles by country of birth. We compared geographical/demographic/clinical characteristics, from 2012 to 2020 (9 years of data), of 914 495 Latinos recorded as US-born, non-US-born and without a country of birth recorded. We also described the state in which these data were collected.

**Results** Country of birth was collected for 127 138 Latinos in 782 clinics in 22 states. Compared with those with a country of birth recorded, Latinos without this record were more often uninsured and less often preferred Spanish. While covariate adjusted prevalence of heart disease and risk factors were similar between the three groups, when results were disaggregated to five specific Latin countries (Mexico, Guatemala, Dominican Republic, Cuba, El Salvador), significant variation was observed, especially in diabetes, hypertension and hyperlipidaemia.

**Conclusions** In a multistate network, thousands of non-US-born, US-born and patients without a country of birth recorded had differing demographic characteristics, but clinical variation was not observed until data was disaggregated into specific country of origin. State policies that enhance the safety of immigrant populations may enhance the collection of health equity related data. Rigorous and effective health equity research using Latino country of birth information paired with longitudinal healthcare information found in EHRs might have significant potential for aiding clinical and public health practice, but it depends on increased, widespread and accurate availability of this information, co-occurring with other robust demographic and clinical data nativity.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Survey-based research and national recommendations suggest that the birth country of Latino patients (the largest ethnic minority) in the USA should be collected and is associated with health outcomes.
⇒ Few if any real-world healthcare datasets contain this place of birth information. Therefore, real-world research, especially research over time verifying differential healthcare outcomes and/or utilisation, is lacking.

## WHAT THIS STUDY ADDS

⇒ Thousands of Latino patients had place of birth collected in a large network of hundreds of federally funded clinics.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY.

⇒ Clinics and health systems should prioritise the further understanding of disaggregated data in Latinos to better the health outcomes of the largest ethnic minority in the USA.

## INTRODUCTION

The Latino population is the largest ethnic minority[1] in the USA and is heterogeneous, especially with respect to birth country.[2] Country of birth is sometimes studied as an important health and healthcare factor in Latino patients in the USA,[3 4] and may be associated with differences in disease prevalence, social disadvantage, immigration status and/or acculturation.[5 6] Along these lines, there have been numerous calls in the health services literature for 'data disaggregation' when studying minority populations,[7–9] including Latinos,[10] in order to better understand health and healthcare inequity. Disaggregating country of birth among Latino patients is complex, however. Rigorous and effective health services research using Latino country of birth information might

have significant potential for aiding clinical and public health practice, but it depends on widespread and accurate availability of this information, co-occurring with other robust demographic and clinical data.[7 11] Many commonly used sources of demographic information used in population-based research do not have robust race/ethnicity data,[12–14] let alone country of birth information in Latinos,[15 16] making evaluation of country of birth and its impact on healthcare impossible in many circumstances. Long-term cohort studies on Hispanic/Latino health (such as the Hispanic Community Health Study/Study of Latinos)[17 18] may have country of origin data, but not detailed health system data.

Electronic health records (EHRs) are a burgeoning data source in public health and health services research in Latino populations, including in numerous Latino subpopulations[19 20] and evaluations of quality of care in Latinos.[20 21] To date, there are few examples of Latino country of birth reported in EHR. Community health centres (CHCs) disproportionately care for low-income Latino patients,[22] and therefore, are an important setting in which to study the availability and utility of country of birth information in a clinical dataset. The OCHIN network of CHCs is one of the largest hosted linked EHR-based networks in the USA.[23] Country of birth is reported in a proportion of Latino patients in the OCHIN network. Our objective was to describe the extent to which place of birth is recorded in EHRs for Latinos across a multistate network of CHCs. Furthermore, to better understand the clinical and public health utility of place-of birth information, we aimed to describe the characteristics of Latino patients with and without a recorded place of birth as well as the clinics that collect place of birth information by comparing USA)-born, non-US-born and place-of-birth not recorded Latino patients in this network.

Cardiovascular disease is the biggest cause of mortality in the USA and continues to be the main driver of mortality in Latinos. Previous studies suggest that Latino individuals experience more frequent and more poorly controlled risk factors for cardiac events,[24] and evidence suggests that cardiovascular risks and care may not be uniform across Latinos of varying national origins.[25] Therefore, we examined cardiovascular disease risk factors as a test case for the comparison of demographic, disease risk and care measures in US-born and non-US-born Latinos, and among Latinos where country of birth was not recorded.

## METHODS
### Data sources
We used EHRs from the OCHIN network, a hosted, linked multistate EHR network of CHCs across the USA. All study data, including country of birth, are collected in the routine course of clinical care in structured, pre-existing fields in OCHIN Epic (EHR software), and was not collected specifically for this study.

### Population
Our study population consisted of 914495 Latino patients (ages 9–79 years) who had at least one ambulatory or telemedicine visit between 1 January 2012 and 31 January 2020, within an OCHIN clinic. All Latinos were included regardless of having country of birth information recorded in the EHR.

### Main independent variables
Our primary independent variable was a mutually exclusive three-category grouping of the combination of ethnicity and nativity: Latinos with a recorded birthplace of the USA (ie, US-born), Latinos with a recorded birthplace outside of the USA (ie, non-US-born) and Latinos without a place of birth recorded in the EHR. This birthplace information was self-reported by patients and was extracted from a discrete variable field used/built for the purpose of recording country of birth. While we use 'Latino' in the majority of our discussion because it is often preferred in our study population, the actual ethnicity information collected by clinics is Hispanic and non-Hispanic, 'Hispanic' and "Latino" differ slightly in their definitions.[26]

### Cardiovascular disease prevention outcomes
We described our three ethnicity/nativity groups by basic clinical features relevant to cardiovascular disease prevention (visits per year, smoking status, obesity, hyperlipidaemia diagnosis and ever receipt of a lipid test, diabetes diagnosis and ever a haemoglobin a1c test, hypertension diagnosis, heart disease diagnosis). All diagnoses data come from the 'problem list' in the EHR; labs were resulted labs to the clinic. A patient's most frequently visited clinic was used as the index clinic in this portion of the analysis.

### Potential confounders
In descriptive and modelling analyses, we considered common patient-level demographics including sex, age at first visit, insurance status, income, race, preferred language, annual ambulatory visit rates and smoking status. While we account for clustering by a patient clinic (see "Statistical analysis"), we describe these potential confounders by clinic characteristics as well.

### Statistical analysis
We described our patient characteristics by ethnicity/nativity groups using means, SD, frequencies and percentages. We also described patient characteristics by specific country of birth among non-US-born Latino patients. Next, we visually described through a histogram the number and per cent of Latino patients with a country of birth recorded by state. In addition, we described select characteristics of clinics (eg, total number of patients, total number of clinics, patient-panel characteristics) that collect this information and those that do not collect this information.

We then reported unadjusted and covariate-adjusted prevalences of International Classification of Disease

(ICD)-9/10 coded heart disease, obesity, diabetes, hypertension and hyperlipidaemia by our three ethnicity/nativity groups. We also described the EHR reported demographic factors and ICD-9/10 coded disease prevalence of patients with specific countries of birth among the non-US-born Latino patients in our sample. To estimate covariate-adjusted prevalences for heart disease, obesity, diabetes, hypertension and hyperlipidaemia by three ethnicity/nativity groups, we used generalised estimating equations (GEE) logistic regression that included indicators for ethnicity/nativity groups and the potential confounders listed above. We accounted for clustering on the patient's most frequented clinic through the use of a robust sandwich variance estimator with exchangeable correlation structure.

Lastly, among a subset of non-US-born patients within five Latin countries with large sample sizes (Mexico, El Salvador, Guatemala, Dominican Republic, Cuba), we estimated unadjusted and adjusted prevalences of heart disease, obesity, diabetes, hypertension and hyperlipidaemia diagnoses using a similar GEE logistic model replacing the indicators for ethnicity/nativity groups with indicators for country of birth. We performed this analysis on the entire sample and in those age 18 and over. Analyses were performed in Stata V.15 and R V.1.1 and two-sided testing with set 5% type I error.

### Patient and public involvement
Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

## RESULTS
### Frequency of country of birth reported in the EHR
Descriptive statistics of our patient sample (N=914495) are in table 1. Of the 914495 Latino patients who met study criteria, 127138 (13.9%) had a reported place of birth in the EHR. Of the 127138 Latinos with a country of birth, 81427 (64.0%) were non-US-born and 45711 (36%) were US-born.

### Comparison of Latinos with and without a place of birth reported
Latino patients with record of their place of birth were less often uninsured, and more likely to prefer Spanish than patients without a country of birth reported (table 1). Fewer identified as black compared with those with a place of birth reported. Latino patients with a place of birth documentation had similar diagnoses prevalence (of diabetes, heart disease, hyperlipidaemia and hypertension) to Latinos without a country of birth recorded, but had a lower prevalence of having a lipid test or haemoglobin a1c in the study period.

### Patient characteristics of US-born and non-US-born Latinos
Of Latinos with a country of birth recorded, non-US-born Latinos were more often always under 138%

of the federal poverty level, more often uninsured, and more preferred Spanish than US-born Latinos (table 1). Non-US-born Latinos had a higher rate of clinic visits per year as well. They more frequently had documented diagnoses of diabetes, hyperlipidaemia and hypertension, and more often had at least one Hba1c and lipid test in the study period. Of note, the prevalence of these measures (eg, insurance, language, diabetes, hypertension, hyperlipidaemia, Hba1c, lipid screen) in the US-born and non-US-born Latino groups differed from the more general groupings (country of birth recorded or not). More non-US born Latinos spoke Spanish and had diabetes, hypertension, and hyperlipidaemia diagnoses, as well as ever had a haemoglobin A1c or lipid test. Latinos without a place of birth recorded were more often uninsured.

The specific countries of birth of the non-US-born Latino patient group, with demographic information and disease prevalence, are described in online supplemental appendix tables 1–6. Nineteen countries of birth outside of the USA were represented among non-US-born Latino patients with a country of birth recorded. The most frequently represented countries in this data were Mexico (44% of non-US-born Latinos), Guatemala (17%), El Salvador (13%) and the Dominican Republic (12%). Across several patient demographics (eg, sex, income), we observe varying distribution of those demographics by country of birth.

### Clinic characteristics
While Latino patients in the OCHIN network spanned 1811 clinics in 26 states, only 782 (43%) clinics across 22 states recorded at least 1 patient's country of birth. Despite country of birth being collected in clinics in 22 states, greater than 98% of Latinos with a country of birth reported were seen at clinics in California, Massachusetts, Texas, Minnesota, Oregon, New Jersey, Wisconsin or Washington (figure 1).

### Characteristics of clinics collecting or not collecting country of birth information
In comparing clinics in the OCHIN network who did or did not collect place of birth information, there were numerous demographic and basic utilisation similarities in Latino patients in each category of clinic (table 2). A few differences were that more Latino patients in clinics who did not collect place of birth information were uninsured, fewer preferred Spanish, more had >5 visits per year and fewer had one visit per year. Fewer Latino patients overall visited these clinics.

Covariate-adjusted prevalences of heart disease and cardiovascular disease risk factor diagnoses among the three major ethnicity/nativity groups are displayed in figure 2 and online supplemental appendix table 7. Of note, foreign-born Latinos had a lower adjusted prevalence of obesity than the other groups, but diagnosis prevalences of all conditions were similar.

**Table 1** Characteristics of Latino patients with at least one ambulatory visit at an OCHIN clinic in 2012–2020, by reported country of birth, N (column %)

| | Latino with country of birth reported | | | Latino without country of birth reported |
| | US-born latino | Non-US-born latino | Overall | Overall |
|---|---|---|---|---|
| Total | 45711 | 81427 | 127138 | 787357 |
| Sex | | | | |
| Female | 26560 (58.1) | 50967 (62.6) | 77527 (61.0) | 463365 (58.9) |
| Male | 19151 (41.9) | 30460 (37.4) | 49611 (39.0) | 323992 (41.1) |
| Age at first study visit | | | | |
| 9–12 | 6305 (13.8) | 2012 (2.5) | 8317 (6.5) | 67202 (8.5) |
| 13–17 | 6946 (15.2) | 4510 (5.5) | 11456 (9.0) | 86854 (11.0) |
| 18–29 | 15205 (33.3) | 16185 (19.9) | 31390 (24.7) | 196306 (24.9) |
| 30–49 | 11241 (24.6) | 36814 (45.2) | 48055 (37.8) | 281896 (35.8) |
| 50–64 | 5157 (11.3) | 17580 (21.6) | 22737 (17.9) | 119657 (15.2) |
| 65 and over | 857 (1.9) | 4326 (5.3) | 5183 (4.1) | 35442 (4.5) |
| Federal poverty level* | | | | |
| Always ≥138% | 2397 (5.2) | 4176 (5.1) | 6573 (5.2) | 68681 (8.7) |
| Always <138% | 23721 (51.9) | 62592 (76.9) | 86313 (67.9) | 497596 (63.2) |
| Above and below 138% | 2648 (5.8) | 8233 (10.1) | 10881 (8.6) | 80782 (10.3) |
| Never reported | 16945 (37.1) | 6426 (7.9) | 23371 (18.4) | 140298 (17.8) |
| Insurance§ | | | | |
| Never insured | 3772 (8.3) | 11839 (14.5) | 15611 (12.3) | 197210 (25.0) |
| Some private insurance | 3496 (7.6) | 7073 (8.7) | 10569 (8.3) | 95828 (12.2) |
| Combined private and public | 4825 (10.6) | 13955 (17.1) | 18780 (14.8) | 54034 (6.9) |
| Some public insurance | 33618 (73.5) | 48560 (59.6) | 82178 (64.6) | 440285 (55.9) |
| Preferred language | | | | |
| English | 32083 (70.2) | 9232 (11.3) | 41315 (32.5) | 320939 (40.8) |
| Spanish | 13628 (29.8) | 72195 (88.7) | 85823 (67.5) | 466418 (59.2) |
| Visits per year | | | | |
| <1 | 14511 (31.7) | 18607 (22.9) | 33118 (26.0) | 237254 (30.1) |
| (1, 3) | 16662 (36.5) | 26067 (32.0) | 42729 (33.6) | 255670 (32.5) |
| (3, 5) | 6229 (13.6) | 15399 (18.9) | 21628 (17.0) | 113767 (14.4) |
| 5+ | 8309 (18.2) | 21354 (26.2) | 29663 (23.3) | 180666 (22.9) |
| Black race | 2823 (6.2) | 5116 (6.3) | 7939 (6.2) | 20885 (2.7) |
| Smoking status | | | | |
| Never smoked | 30454 (66.6) | 62768 (77.1) | 93222 (73.3) | 532750 (67.7) |
| Formerly smoked | 3128 (6.8) | 6308 (7.7) | 9436 (7.4) | 54655 (6.9) |
| Currently smokes | 6787 (14.8) | 6312 (7.8) | 13099 (10.3) | 78929 (10.0) |
| Secondhand smoke exposure | 348 (0.8) | 500 (0.6) | 848 (0.7) | 8089 (1.0) |
| Unknown | 4994 (10.9) | 5539 (6.8) | 10533 (8.3) | 112934 (14.3) |
| Obese ever† | | | | |
| No | 24859 (54.4) | 42475 (52.2) | 67334 (53.0) | 429910 (54.6) |
| Yes | 20852 (45.6) | 38952 (47.8) | 59804 (47.0) | 357447 (45.4) |
| Heart disease ever‡ | | | | |
| No | 45284 (99.1) | 80514 (98.9) | 125798 (98.9) | 779537 (99.0) |
| Yes | 427 (0.9) | 913 (1.1) | 1340 (1.1) | 7820 (1.0) |
| Diabetes ever ‡ | | | | |
| No | 42294 (92.5) | 69384 (85.2) | 111678 (87.8) | 706636 (89.7) |
| Yes | 3417 (7.5) | 12043 (14.8) | 15460 (12.2) | 80721 (10.3) |

Continued

**Table 1** Continued

| | Latino with country of birth reported | | | Latino without country of birth reported |
|---|---|---|---|---|
| | US-born latino | Non-US-born latino | Overall | Overall |
| Hyperlipidaemia ever‡ | | | | |
| No | 39 577 (86.6) | 61 117 (75.1) | 100 694 (79.2) | 671 808 (85.3) |
| Yes | 6134 (13.4) | 20 310 (24.9) | 26 444 (20.8) | 115 549 (14.7) |
| Hypertension ever‡ | | | | |
| No | 40 453 (88.5) | 65 920 (81.0) | 106 373 (83.7) | 681 650 (86.6) |
| Yes | 5258 (11.5) | 15 507 (19.0) | 20 765 (16.3) | 105 707 (13.4) |
| HbA1c ever‡ | | | | |
| No | 26 996 (59.1) | 36 016 (44.2) | 63 012 (49.6) | 483 645 (61.4) |
| Yes | 18 715 (40.9) | 45 411 (55.8) | 64 126 (50.4) | 303 712 (38.6) |
| Lipid screening ever‡ | | | | |
| No | 24 166 (52.9) | 35 796 (44.0) | 59 962 (47.2) | 463 855 (58.9) |
| Yes | 21 545 (47.1) | 45 631 (56.0) | 67 176 (52.8) | 323 502 (41.1) |

*Income information is collected annually; therefore, patients level may change over time.
†Obese results are extracted from adult and children BMI. An adult is considered obese if their BMIs have ever been greater than 30. A child is considered obese if their BMIs have ever been higher than 95th percentile BMI calculated using the 'childsds' package in R based on age, sex, weight and height.
‡A diagnosis is flagged whenever its result is non-missing and greater than zero.
§Insurance status is measured at each visit; 'some' indicates they had this insurance type at some proportion of their total visits and no other insurance types at other visits.
BMI, body mass index; HbA1c, glycosylated haemoglobin.

Covariate-adjusted disaggregated prevalences of heart disease and its risk factors are shown in figure 3 and online supplemental appendix table 8 for the five Latin countries with the highest sample sizes. Of note, when this data are disaggregated into specific country of origin, there is increased variability in findings between countries, especially in hyperlipidaemia, hypertension and diabetes. In sensitivity analyses for the above outcomes including only adults (age 18 and over), the results did not appreciably differ between the entire sample and the adult sample.
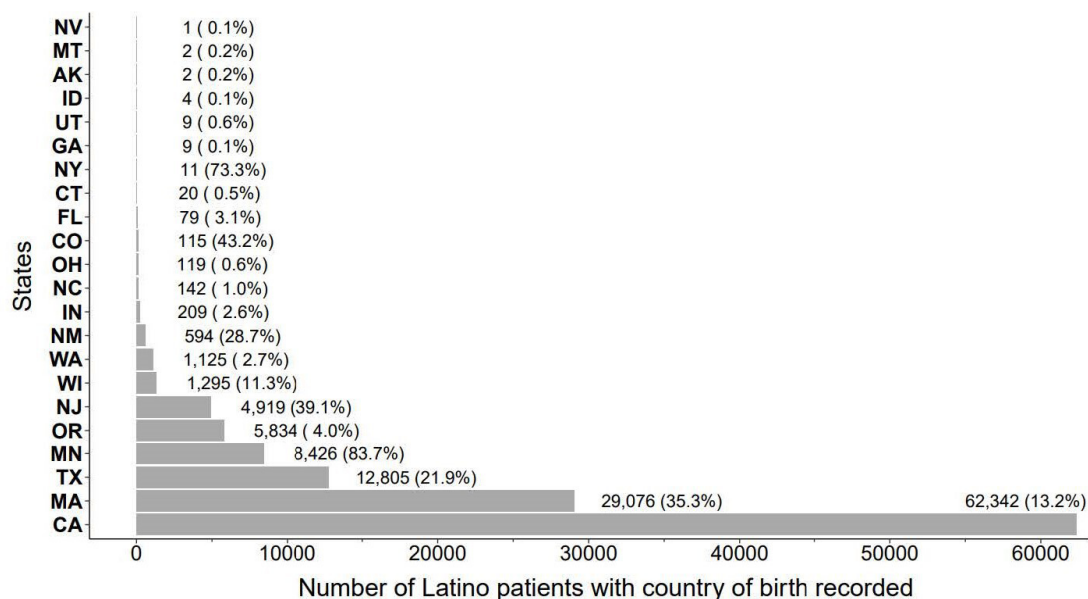


**Figure 1** Number and per cent of Latino patients with country of birth recorded by state (2012–2020). The denominator for each state included as Latinos who visited an OCHIN clinic during the study period in that state regardless of whether country of birth was recorded or not. AK, Alaska; CA, California; CO, Colorado; CT, Connecticut; FL; Florida; GA, Georgia; ID, Idaho; IN, Indiana; MA, Massachusetts; MN, Minnesota; MT, Montana; NC, North Carolina; NJ, New Jersey; NM, New Mexico; NV, Nevada; NY, New York; OH; Ohio; OR, Oregon; TX, Texas; UT, Utah; WA, Washington; WI, Wisconsin.

**Table 2** Clinic characteristics of clinics who did and did not record a country of birth in the OCHIN electronic health record

| Characteristic | Clinics who collect place of birth information | Clinics who do not collect place of birth information |
|---|---|---|
| Total no patients | 769 196 | 145 299 |
| Total no health systems | 116* | 121* |
| Mean patients per health system (SD) | 6631 (10 784) | 1201 (3920) |
| Total no clinics | 782 | 1029 |
| Mean patients per clinic (SD) | 984 (1751) | 141 (408) |
| Sex, N (column %) | | |
| Female | 457 067 (59.4) | 83 825 (57.7) |
| Male | 312 129 (40.6) | 61 474 (42.3) |
| Age at first study visit, N (column %) | | |
| 9–12 | 59 354 (7.7) | 16 165 (11.1) |
| 13–17 | 77 348 (10.1) | 20 962 (14.4) |
| 18–29 | 196 184 (25.5) | 31 512 (21.7) |
| 30–49 | 282 280 (36.7) | 47 671 (32.8) |
| 50–64 | 120 486 (15.7) | 21 908 (15.1) |
| 65 and over | 33 544 (4.4) | 7081 (4.9) |
| Federal poverty level, N (column %) | | |
| Always ≥138% | 60 250 (7.8) | 15 004 (10.3) |
| Always <138% | 486 544 (63.3) | 97 365 (67.0) |
| Above and below 138% | 84 743 (11.0) | 6920 (4.8) |
| Never reported | 137 659 (17.9) | 26 010 (17.9) |
| Insurance, N (column %) | | |
| Never insured | 174 263 (22.7) | 38 558 (26.5) |
| Some private insurance | 84 243 (11.0) | 22 154 (15.2) |
| Combined private and public | 67 073 (8.7) | 5741 (4.0) |
| Some public insurance | 443 617 (57.7) | 78 846 (54.3) |
| Preferred language, N (column %) | | |
| English | 297 286 (38.6) | 64 968 (44.7) |
| Spanish | 471 910 (61.4) | 80 331 (55.3) |
| Visits per year, N (column %) | | |
| <1 | 241 342 (31.4) | 29 030 (20.0) |
| (1, 3) | 253 677 (33.0) | 44 722 (30.8) |
| (3, 5) | 112 226 (14.6) | 23 169 (15.9) |
| 5+ | 161 951 (21.1) | 48 378 (33.3) |
| Black race, N (column %) | 26 677 (3.5) | 2147 (1.5) |

*There are 101 overlapping health systems between two columns.

## DISCUSSION

We endeavoured to describe, in one of the largest multistate data networks serving low-income patients in the USA, the frequency and characteristics of Latino patients with country of birth information recorded, as compared with those without such information recorded, using cardiovascular disease risk factors/care as a clinical test

case. Our goal was to more fully understand the potential clinical, policy and research implications of using EHRs to address calls for data disaggregation in health equity research.[7–10]

There were several evident themes in this analysis. First, place of birth was reported across hundreds of health systems and clinics in a network of CHCs. This reinforces the widely distributed capacity to collect this information across CHCs caring for thousands of Latino patients. CHCs are an environment uniquely suited to collect this information; previous work has shown that immigrant Latino populations engage care at CHCs and are comfortable disclosing sensitive information in this environment.[27–29] This environment, coupled with a robust, linked EHR may allow the pairing of country of birth data with longitudinal healthcare information that can be missing from national surveys to further improve health equity research. This expands on prior long-term cohort studies of risk factor prevalence[30 31] by instead examining the real-world documentation of these risk factors in a multistate network.

While country of birth information is collected across numerous and varied health systems and clinics, the collection of this information is concentrated in several states. This increased collection may be associated with states with a high non-US-born population,[32] although not in all cases. The collection of nativity data in Latino patients may not always be perceived to be safe, for fear that disclosure of nativity among undocumented immigrants may lead to discovery by federal authorities and eventual deportation. Specific state policies may make this disclosure of nativity (and therefore, possibly immigration status) more or less risky.[33 34] Indeed, the 8 states where most country of birth data were collected rank in the top 15 of immigrant-friendly policy states in some evaluations.[35] Our analysis was not designed to formally evaluate this association, and unadjusted confounders (eg, number of clinics) between states make comparisons difficult. Further research should investigate this relationship. Because of safety concerns, one could hypothesise that clinics who never collect place of birth information do so because their Latino patients are more likely to be immigrants, and disclosure of this information may involve risk of discovery by authorities and deportation for those who may be undocumented. While clinics not collecting place of birth information did see more Latino patients who were never insured (which might be a proxy for immigration status),[19] a substantial percentage of Latino patients who were never insured were seen at clinics who collected birth country information more frequently. Clinics never collecting place of birth information were less likely to see Spanish-preferring patients, which also has been an oft-used proxy for immigration-status or acculturation level. Therefore, these findings do not suggest a clear association between birthplace collection (or lack of it) and proxies for recent immigration.

Our description and analysis of individual patient characteristics by US-born, non-US-born and not-recorded
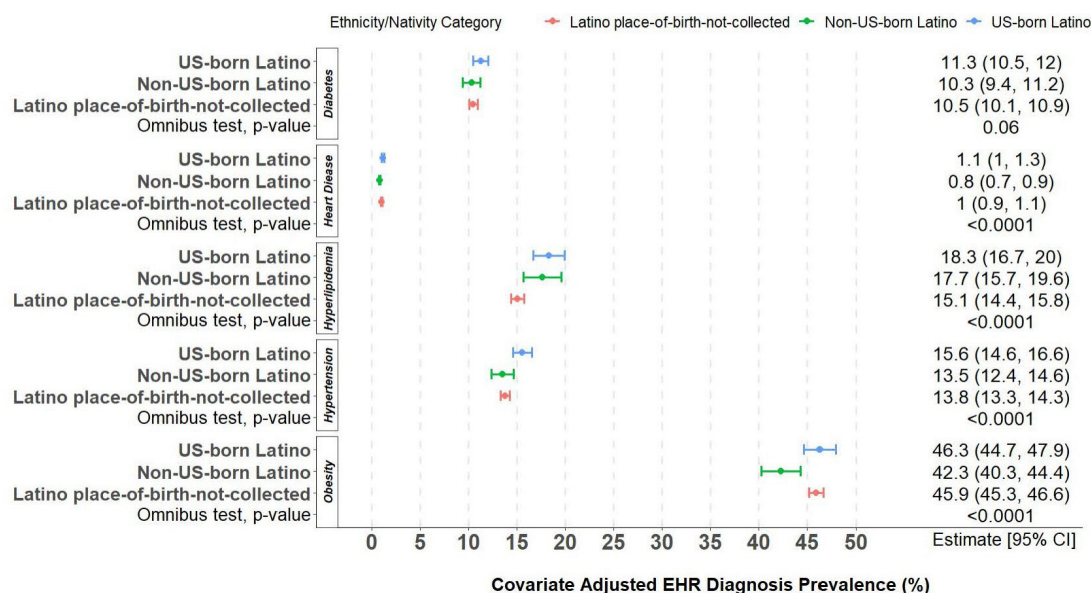
**Figure 2** Covariate-adjusted prevalence of heart disease and cardiovascular disease risk factor diagnosis, by ethnicity/nativity category. EHR, electronic health record.

groups, compared with our disaggregated results, demonstrates why it should be a health equity research priority to understand this data more fully. Our analysis of heart disease and cardiovascular risk factor diagnosis in our three broad categories reveal very similar prevalences of diagnoses. However, when disaggregated, we observed wide variation in some conditions. Hypertension prevalence ranged from 15% to 23%, hyperlipidaemia from 15% to 27% and diabetes varied from 5% to 18%. This suggests that Latino populations may vary in their basic cardiovascular risk and/or care by specific country of origin, and combining them into a single broad category (or even smaller broad categories, like 'non-US-born') inappropriately combines a heterogeneous population that may have very different needs for prevention and treatment of heart disease and cardiovascular disease risk factors. Despite the challenges, health systems and researchers should continue to pursue safe ways to collect nativity information with co-occurring longitudinal, multilevel clinical data in order to better prevent and/or mitigate cardiovascular disease inequities.
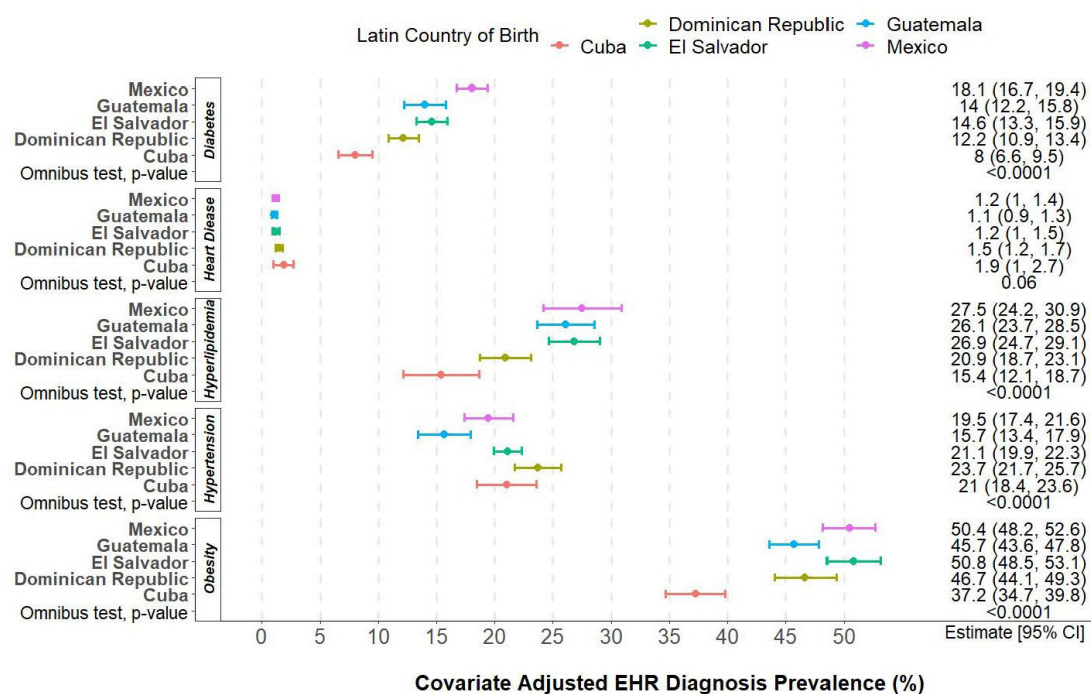


**Figure 3** Covariate-adjusted prevalence of heart disease and cardiovascular disease risk factor diagnosis prevalences, by Latin country of birth. EHR, electronic health record.

## Limitations

Country of birth information was present for a small subset of our total Latino population, which could produce greater variation in results. These findings and their implications could change with a larger sample. We were not able to consider time in the USA, generation after immigration or subgroups of US-born patients (eg, US-born Mexican Americans or US-born Cuban Americans) which may play a significant role in the healthcare utilisation of this population. We did not analyse local or state policies or environments that may affect clinic workflow and practice in the collection of country of birth. CHCs are likely a unique environment for collecting sensitive information,[28] and may not represent other clinical environments.

## Conclusion

In a multistate network of CHCs with a linked EHR, country of birth was collected in numerous Latino patients in many clinics and health systems, but not equally across all states in our network. Non-US-born, US-born and Latino patients without a place of birth reported had differing demographic and clinical characteristics. State policies that enhance the safety of immigrant populations may enhance the collection of health equity related data. CHCs should strongly consider the collection of country of birth information, as the utilisation and risk factors of Latinos from specific countries of origin may differ, and population differences may not be observable when broader categories than specific country of origin are used. Rigorous and effective health equity research using Latino country of birth information paired with longitudinal healthcare information such as those found in EHRs might have significant potential for aiding clinical and public health practice. However, this depends on increased, widespread and accurate availability of this information, co-occurring with other robust demographic and clinical data nativity.

**ORCID iD**
Jennifer A Lucas http://orcid.org/0000-0002-2895-9809

## REFERENCES

1 United States Department of Health and Human Services, Office of Minority Health. Profile: Hispanic Americans; 2022.
2 Pew Research Center. Key facts about U.S. Latinos for Hispanic heritage month. Available: https://www.pewresearch.org/fact-tank/2021/09/09/key-facts-about-u-s-latinos-for-national-hispanic-heritage-month/ [Accessed 1 Dec 2021].
3 Camacho-Rivera M, Kawachi I, Bennett GG, et al. Revisiting the Hispanic health paradox: the relative contributions of nativity, country of origin, and race/ethnicity to childhood asthma. *J Immigr Minor Health* 2015;17:826–33.
4 Costas-Muñiz R, Jandorf L, Philip E, et al. Examining the impact of Latino nativity, migration, and acculturation factors on colonoscopy screening. *J Community Health* 2016;41:903–9.
5 Afable-Munsuz A, Liang S-Y, Ponce NA, et al. Acculturation and colorectal cancer screening among older Latino adults: differential associations by National origin. *J Gen Intern Med* 2009;24:963–70.
6 Do DP, Frank R. The diverging impacts of segregation on obesity risk by nativity and neighborhood poverty among Hispanic Americans. *J Racial Ethn Health Disparities* 2020;7:1214–24.
7 Heintzman J, Marino M. Race and ethnicity data in research. *JAMA* 2019;321:1217–8.
8 Kauh TJ, Read JG, Scheitler AJ. The critical role of racial/ethnic data disaggregation for health equity. *Popul Res Policy Rev* 2021;40:1–7.
9 Rubin. Counting a diverse nation: disaggregating data on race and ethnicity to advance a culture of health. 2018. Available: https://www.policylink.org/sites/default/files/Counting_a_Diverse_Nation_08_15_18.pdf
10 Alcántara C, Suglia SF, Ibarra IP, et al. Disaggregation of latina/o child and adult health data: a systematic review of public health surveillance surveys in the United States. *Popul Res Policy Rev* 2021;40:61–79.
11 Duran DG, Pérez-Stable EJ. Novel approaches to advance minority health and health disparities research. *Am J Public Health* 2019;109:S8–10.
12 Howland RE, Tsao TY. Evaluating race and ethnicity reported in hospital discharge data and its impact on the assessment of health disparities. *Med Care* 2020;58:280–4.
13 Jarrín OF, Nyandege AN, Grafova IB, et al. Validity of race and ethnicity codes in Medicare administrative data compared with gold-standard self-reported race collected during routine home health care visits. *Med Care* 2020;58:e1–8.
14 Ng JH, Ye F, Ward LM, et al. Data on race, ethnicity, and language largely incomplete for managed care plan members. *Health Aff (Millwood)* 2017;36:548–52.
15 Rodriguez-Lainz A, McDonald M, Fonseca-Ford M, et al. Collection of data on race, ethnicity, language, and nativity by US public health surveillance and monitoring systems: gaps and opportunities. *Public Health Rep* 2018;133:45–54.

16 Thiel de Bocanegra H, Gany F, Fruchter R. Epidemiologic data on new York's Latino population: a critical review of the literature. *Ethn Dis* 1993;3:413–26.

17 Lavange LM, Kalsbeek WD, Sorlie PD, *et al*. Sample design and cohort selection in the Hispanic community health Study/Study of Latinos. *Ann Epidemiol* 2010;20:642–9.

18 Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, *et al*. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* 2010;20:629–41.

19 Heintzman J, Marino M, Clark K, *et al*. Using electronic health record data to study Latino immigrant populations in health services research. *J Immigr Minor Health* 2020;22:754–61.

20 Heintzman J, Marino M, Hoopes M, *et al*. Using electronic health record data to evaluate preventive service utilization among uninsured safety net patients. *Prev Med* 2014;67:306–10.

21 Heintzman JD, Bailey SR, Muench J, *et al*. Lack of lipid screening disparities in obese Latino adults at health centers. *American Journal of Preventive Medicine* 2017;52:805–9.

22 National Association of Community Health Centers. Community health center chartbook 2020. 2020. Available: https://www.nachc. org/wp-content/uploads/2020/01/Chartbook-2020-Final.pdf [Accessed 19 May 2020].

23 OCHIN Inc. OCHIN 2020 annual report. Available: https://ochin.org/ 2020-annual-report [Accessed 17 Jan 2022].

24 Campos CL, Rodriguez CJ. High blood pressure in Hispanics in the United States: a review. *Curr Opin Cardiol* 2019;34:350–8.

25 Qureshi WT, Kaplan RC, Swett K, *et al*. American College of cardiology/american heart association (ACC/AHA) class I guidelines for the treatment of cholesterol to reduce atherosclerotic cardiovascular risk: implications for us hispanics/latinos based on findings from the Hispanic community health Study/Study of Latinos (HCHS/SOL). *JAHA* 2017;6.

26 Center PR. Latinx used by just 3 % of U.S. Hispanics. About One-in-Four have heard of it. 2020. Available: https://www.pewresearch.org/ hispanic/2020/08/11/about-one-in-four-u-s-hispanics-have-heard-of-latinx-but-just-3-use-it/

27 DeVoe J, Angier H, Likumahuwa S, *et al*. Use of qualitative methods and user-centered design to develop customized health information technology tools within federally qualified health centers to keep children insured. *J Ambul Care Manage* 2014;37:148–54.

28 López-Cevallos DF, Lee J, Donlan W. Fear of deportation is not associated with medical or dental care use among mexican-origin farmworkers served by a federally-qualified health center -- faith-based partnership: an exploratory study. *J Immigr Minor Health* 2014;16:706–11.

29 Byhoff E, De Marchis EH, Gottlieb L, *et al*. Screening for immigration-related health concerns in a federally qualified health center serving a diverse latinx community: a mixed methods study. *J Immigr Minor Health* 2020;22:988–95.

30 Kaplan RC, Avilés-Santa ML, Parrinello CM, *et al*. Body mass index, sex, and cardiovascular disease risk factors among Hispanic/Latino adults: Hispanic community health Study/Study of Latinos. *JAHA* 2014;3.

31 Guadamuz JS, Durazo-Arvizu RA, Daviglus ML, *et al*. Immigration status and disparities in the treatment of cardiovascular disease risk factors in the Hispanic community health Study/ Study of Latinos (visit 2, 2014-2017). *Am J Public Health* 2020;110:1397–404.

32 Pew Research Center. Nativity by state 2016. Available: https:// assets.pewresearch.org/wp-content/uploads/sites/7/2018/09/ 12103811/PH_2016-Foreign-Born-Statistical-Portraits_Current-Data_ 45_Nativity-by-state.png [Accessed 8 Jan 2021].

33 House bill 56. In: Alabama so, editor. 2011.

34 Senate bill 1070. In: Arizona so, editor. 2010.

35 Samari G, Nagle A, Coleman-Minahan K. Measuring structural xenophobia: US state immigration policy climates over ten years. *SSM - Population Health* 2021;16:100938.