Addressing Health Disparities by Integrating Electronic Health Records with Census Data: First Steps

Aubrey Limburg, Joanna Motro, Lucinda Dalzell, and Victoria Udalova; Enhancing Health Data (EHealth) Program, U.S. Census Bureau
Barbara Entwisle[1], Jordan Young[1], Tim Carey[2], and Paul Chelminski[3]; University of North Carolina at Chapel Hill (UNC)

Interdisciplinary Association for Population Health Science (IAPHS)
Minneapolis, MN
September 20-23, 2022

OBJECTIVE

To investigate the nature and size of various sociodemographic differences in the success of linking medical records to Protected Identification Keys (PIKs). PIKs are anonymous person identifiers developed by the Census Bureau to facilitate linking across files while protecting privacy[1]. PIK assignments are the first step in the creation of an integrated dataset.

SIGNIFICANCE

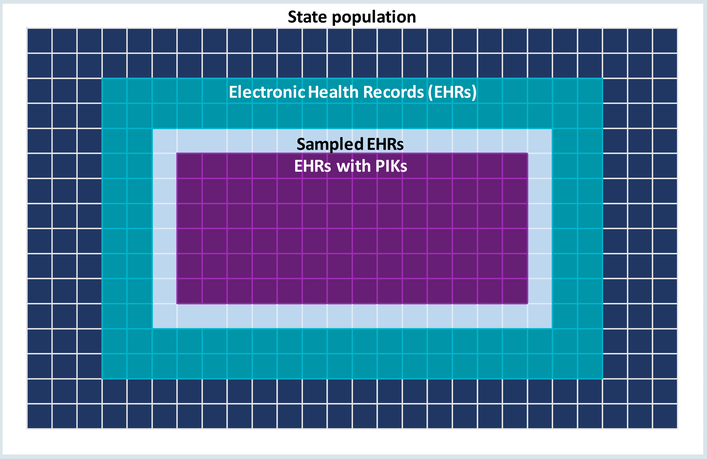
Electronic health records (EHRs) have unique potential to expand the study of population health since they contain detailed information on patient visits, diagnoses, and medications and are based on large sample sizes. However, there are challenges including:

- Missing race/ethnicity
- Limited social characteristics
- Questionable population representation

Previous research has demonstrated strong feasibility for linking EHR data and Census microdata via PIKs[2], but potential differences related to race, ethnicity, and other characteristics have not yet been addressed.

DATA & METHODS

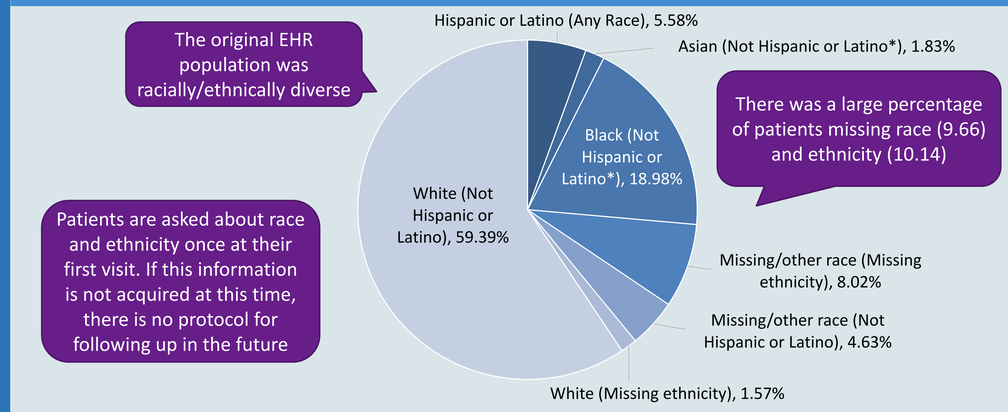
- We linked 1) EHRs from a healthcare system in the southeast to 2) Census records
- Disproportionate stratified random sample of 200,000 patients
  - 25-74 years of age
  - At least 2 visits between 01/2016 and 12/2019
- We estimated logistic models regressing whether an observation was assigned a PIK (1=yes; 0=no) based on 1) race, 2) ethnicity, 3) sex, 4) language, 5) birth cohort, 6) health insurance, and 7) residence (in state or not)



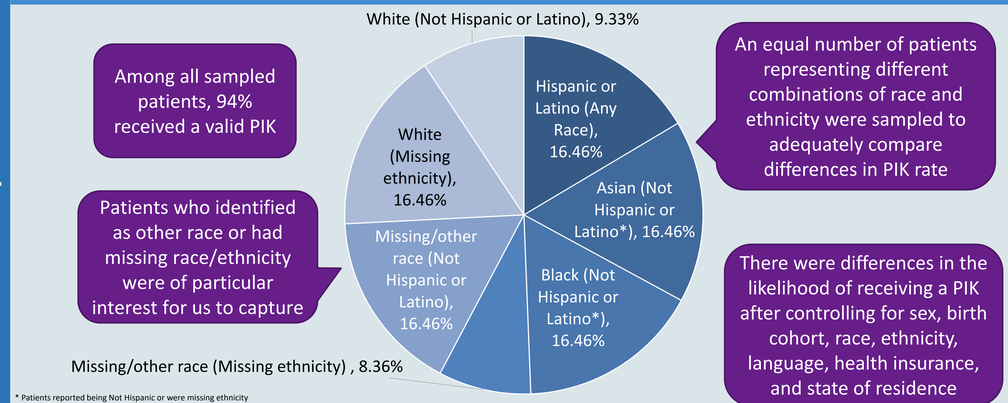
Note: Proportions presented above are only suggestive and do not represent the true proportions

RACE/ETHNICITY DISTRUTION IN EHR DATA

EHR POPULATION



EHR SAMPLE



* Patients reported being Not Hispanic or were missing ethnicity
Note: Patients who were "missing" on race or ethnicity either were unknown (did not answer) or refused to answer; all race and ethnicity information was derived from EHR data

LOGISTIC REGRESSION RESULTS

Who is more likely to receive a PIK?

- White patients compared to Asian (OR=0.32), other race (OR=0.42), and missing race (OR=0.49) patients
- Patients who were not Hispanic or Latino compared to patients who were Hispanic or Latino (OR=0.37)
- Patients who spoke English compared to patients who only spoke Spanish (OR=0.18) and other languages (OR=0.33)
- Patients with private health insurance (OR=6.65), Medicare (OR=10.19), and Medicaid (OR=1.65) compared to those without

Table 1 provides the predicted probability of receiving a PIK for 5 illustrative patient cases

Table 1. Predicted probability of receiving a PIK*

Race	Ethnicity	Language**	Health insurance	Probability
White	Not Hispanic/Latino	English	Private	0.996
White	**Hispanic/Latino**	English	Private	0.989
Other race	Hispanic/Latino	English	Private	0.973
Other race	Hispanic/Latino	**Spanish**	Private	0.866
Other race	Hispanic/Latino	Spanish	**Uninsured/self-pay**	0.353

Highest probability of receiving a PIK

Lowest probability of receiving a PIK

Note: All variables presented come from EHR data
* For the purposes of this table, all probabilities are for patients who reported an address in the state, were born in 1985 or later, and identified as male
** Language was indicative of a patient needing a translator in (e.g., language proficiency)

DISCUSSION

What accounts for differences in PIK assignment?
PIK assignments are primarily based on Social Security Numbers (SSNs). If SSNs are not available, PIKs are assigned probabilistically based on names, addresses, gender, and date of birth:

- **SSNs**: foreign-born less likely to have SSN (e.g., Asian, Hispanic)
- **Names**: some patients may have multiple last names (Hispanics); name changes (women)
- **Address**: some patients may live in group quarters, less attached to household; change address more frequently; do not want to be found

SUMMARY / IMPLICATIONS

Despite success assigning PIKs to EHRs, this success is not uniform. PIK rates were lower for patients who identified as other race, Hispanic/Latino, spoke only Spanish (needed a translator), and were uninsured. Disparities in PIK assignment suggest:

- Limitations of information available in EHRs
- Differential coverage of patients in Census reference files (EHRs may be a better source for some people)

Researchers should consider these selectivities when relying on these sources to study population health since these can result in statistical bias due to exclusion.

FUTURE RESEARCH

Strategic repurposing of already available data (EHRs and Census data) provides unique and innovative ways in which to study population health. Given this, future research interested in expanding the field of population health via record linkages should explore the following:

- Look to patients who did not receive a PIK to better understand PIK process
- Look to patients who did receive a PIK to understand coverage/representativeness of EHRs
- Identify whether characteristics reported in EHRs parallel those in Census data sources

ACKNOWLEDGEMENTS

We are grateful for support from the UNC Translational and Clinical Sciences Institute (Clinical and Translational Science Awards UL1TR002489) and the Carolina Population Center (National Institute of Child Health and Human Development P2C HD050924).

United States® Census Bureau

UNC SCHOOL OF MEDICINE North Carolina Translational and Clinical Sciences Institute

UNC CAROLINA POPULATION CENTER

Affiliations
1. Department of Sociology and Carolina Population Center
2. Department of Medicine, School of Medicine
3. Departments of Allied Health Science and Medicine, School of Medicine

References
1. Bond, Brittany, J. David Brown, Adela Luque, and Amy O'Hara. 2014. The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey. Working Paper. #2014-08. U.S. Census Bureau.
2. Udalova, Victoria, Timothy S. Carey, Paul Roman Chelminski, Lucinda Dalzell, Patricia Knoepp, Joanna Motro, and Barbara Entwisle. 2022. "Linking Electronic Health Records to the American Community Survey: Feasibility and Process." American Journal of Public Health 112(6):923–30. doi: 10.2105/AJPH.2022.306783.