

Linking Electronic Health Records to the American Community Survey: Feasibility and Process

Victoria Udalova, PhD, Timothy S. Carey, MD, MPH, Paul Roman Chelminski, MD, MPH, Lucinda Dalzell, MA, Patricia Knoepp, MPH, Joanna Motro, PhD, and Barbara Entwisle, PhD

 See also Cantor, p. 821.

Objectives. To assess linkages of patient data from a health care system in the southeastern United States to microdata from the American Community Survey (ACS) with the goal of better understanding health disparities and social determinants of health in the population.

Methods. Once a data use agreement was in place, a stratified random sample of approximately 200 000 was drawn of patients aged 25 to 74 years with at least 2 visits between January 1, 2016, and December 31, 2019. Information from the sampled electronic health records (EHRs) was transferred securely to the Census Bureau, put through the Census Person Identification Validation System to assign Protected Identification Keys (PIKs) as unique identifiers wherever possible. EHRs with PIKs assigned were then linked to 2001–2017 ACS records with a PIK.

Results. PIKs were assigned to 94% of the sampled patients. Of patients with PIKs, 15.5% matched to persons sampled in the ACS.

Conclusions. Linking data from EHRs to ACS records is feasible and, with adjustments for differential coverage, will advance understanding of social determinants and enhance the ability of integrated delivery systems to reflect and affect the health of the populations served. (*Am J Public Health.* 2022; 112(6):923–930. <https://doi.org/10.2105/AJPH.2022.306783>)

Patterns of population health and the persistence of health disparities demonstrate the importance of going beyond clinical settings to understand the conditions in which people live, opportunities to which they have access, environments to which they are exposed, and how they are treated by others, all of which are powerfully shaped by race, ethnicity, and income.^{1–3} Nonprofit health care systems have a statutory responsibility to benefit the communities they serve,⁴ but their ability to leverage their clinical data to develop such an understanding is limited.^{5,6} Information needed to describe the sociodemographic characteristics of the

patient population is generally restricted to age, sex, race, and ethnicity, often with substantial missing data on race and ethnicity. Patients may or may not be representative of the communities in which they live.⁷ A potential solution to these problems is to link clinical records to social, economic, and demographic information collected for individuals in the American Community Survey (ACS).

The ACS is an ongoing sample survey conducted by the US Census Bureau that collects information on the US population.⁸ The ACS asks about a full range of personal characteristics such as education, employment status, occupation, income, marital status, detailed

race and ethnicity, language, health insurance, and disabilities; housing characteristics such as type and age of dwelling, number of rooms, monthly rent or mortgage payment, and Internet access; and derivable household characteristics such as size, presence of children, family composition, and poverty status.⁹ Each year, the ACS collects data for more than 2 million persons living in households and 150 000 persons living in group quarters.^{10,11} Properly weighted, ACS data are representative of national, state, and local populations.

Ours is the first study to our knowledge to link electronic health records

(EHRs) with ACS microdata. Others have appended ACS-based publicly available data for small geographic areas to EHRs to improve screening for chronic disease,¹²⁻¹⁴ to identify and characterize neighborhood contexts that contribute to or potentially exacerbate medical conditions,^{15,16} and to develop population-adjusted prevalence rates for various medical conditions.¹⁷ However, there are limitations to this area-based approach. The lowest level of geography for which ACS data are publicly available is the block group, and, because of disclosure concerns, not all information of potential interest is available at this level. Furthermore, areal ACS data do not always add meaningfully to the prediction of health outcomes.¹⁸ With individual-level linkages between EHR and ACS data, it would be possible to systematically evaluate use of areal estimates as proxies for social determinants missing from clinical databases, correct for potential bias in EHR-based studies, and assess the extent to which integrated delivery systems care for representative populations.

To realize the value of EHR-ACS linkage requires a collaboration between a health care system and the Census Bureau that can successfully meet a series of challenges, from developing a data-use agreement that meets the stringent data protection requirements of both entities to determining whether linkages meet necessary quality standards. The health care system that is the focus of our study is composed of a large academic health center, 11 community hospitals, and hundreds of community practices across the state. Care is offered to all residents of the state regardless of ability to pay, with a generous charity care program and a vigorous population health outreach

program, resulting in a diverse patient population, including citizens and non-citizens. Within the broad mix of racial, ethnic, and cultural diversity, both urban and rural populations are represented. Through expansion, the total patient population, operationalized as number of unique patients served over a 2-year period, increased from 1.9 million patients in 2016 to 3.4 million patients in 2021. This article describes the process, reports on lessons learned, and discusses future promise of an attempt to match records for a sample of these patients.

METHODS

To link EHR-derived data to data from the ACS required the following steps:

1. develop a data-use agreement between the health care system and the US Census Bureau;
2. design, select, and transmit a sample of patient records to the Census Bureau;
3. pass the records through the Census Bureau's Person Identification Validation System (PVS) to assign Protected Identification Keys (PIKs);
4. use the PIKs to match to ACS records with PIKs; and
5. assess the quality of linkage.

Each of these steps is briefly described.

Health Care System-Census Bureau Agreements

Linking data required collaboration between 2 entities with very different missions, processes, and cultures. The health care system is a not-for-profit academic integrated care delivery system and a safety-net provider for the state. Its medical records are protected

by the Health Insurance Portability and Accountability Act as well as confidentiality agreements signed by all employees, medical staff, students, volunteers, vendors, and others who access these records in the process of conducting their business. Policy issues related to data use are governed by an oversight committee, with representation from the clinical enterprise, the school of medicine, other health science schools, and 2 patient representatives.

The US Census Bureau is the largest federal statistical agency with a mission to "serve as the nation's leading provider of quality data about its people and economy."¹⁹ This project was done under the Census Bureau's legal authority, Title 13 USC, and all individuals' information was safeguarded under the confidentiality and use restrictions in 13 USC § 8,9 and in accordance with the Census Bureau's Data Stewardship Program.²⁰ There are severe penalties for violating the oath to protect data at the Census Bureau. Though the primacy of data confidentiality is shared by both organizations, differences in procedures resulted in a complex negotiation to establish a governance process to link ACS and EHR data. The critical first step was, thus, to develop the necessary agreements.

A data use agreement was negotiated and approved by the health care system, including institutional review board approval and authorization of the necessary disclosures of protected health information to the Census Bureau to conduct the study. Team members made presentations regarding the long-term goals of the collaboration with the Census Bureau, the value of a joint project to assess population health, and proposed measures to ensure confidentiality of patient data. In turn, health care personnel were educated regarding

Census Bureau data governance and extensive nondisclosure and privacy policies. The resulting data use agreement, signed in late 2019, specifies the legal authorization to participate in this joint statistical project as well as its purpose, mutual interests and responsibilities of the parties, data confidentiality, system security, disclosure avoidance, and research plan. The agreement requires that the confidential data be used only for statistical purposes as described in the research plan and not disclosed or published in any way that permits identification of a particular individual or entity. For this project, only authorized Census Bureau staff can analyze the data sets and only within the Census Bureau's information technology environment. All results must pass a formal disclosure review conducted by the Census Bureau's Disclosure Review Board. This process ensures that there is no information that can identify an individual, either alone or when combined with other publicly available information. All results reported in this article have passed this review.

Electronic Health Records

The health care system uses a single, enterprise-level system (Epic) to manage and store EHR data from hospitals and outpatient practices. Data are transferred daily to a Clinical Data Warehouse and are used for both operations and research. Patient identifiers and sociodemographic information such as race and ethnicity were provided by patients and recorded at the time of registration either online or by staff interview. As part of clinical activity, some variables such as patient address are updated regularly, but, other than de-duplication to correct for multiple records in a few cases

because of multiple surnames, name changes, and John Doe admissions to the emergency department, no formal data "cleaning" was performed as part of this linkage project, making these results applicable to other large integrated delivery systems.

For the purposes of this study, we drew a disproportionate stratified random sample of 200 000 patients aged 25 to 74 years with at least 2 visits between January 1, 2016, and December 31, 2019, from the Clinical Data Warehouse. The goal was to achieve approximately equal numbers of patients representing different combinations of race and ethnicity, although numbers were not always sufficient in some groups to meet this goal. The sample selected for study was as follows:

- White/not Hispanic or Latino: 32 922;
- Black/not Hispanic or Latino, or missing: 32 922;
- Any race/Hispanic or Latino: 32 922;
- Asian/not Hispanic or Latino, or missing: 16 721;
- Missing or other race/missing ethnicity: 32 922;
- Missing or other race/not Hispanic or Latino: 32 922; and
- White/missing ethnicity: 18 670.

We drew the identifying information needed to link their records with ACS data: name, address, date of birth, sex, and Social Security number (SSN). We also drew a limited number of sociodemographic variables: race, Hispanic ethnicity, language, and health insurance. These were analyzed separately.

Protected Identification Key Assignment

The record linkage identifiers used at the Census Bureau do not contain any

direct identifiers. Instead, PIKs replace identifying information and are used to anonymously link to ACS microdata. PIKs are based on exact and probabilistic matching by comparing information in a given input file against a reference file generated from Social Security Administration data and other administrative records.²¹ To assign PIKs, each record is passed through successive modules comparing it with the reference file based on SSN, address, name, gender, and date of birth. When a linkage can be made between the incoming record and the reference file, the PIK is appended to the record in the file transferred from the health care system. Only records that were not found move on to the next module. First, the data go through the SSN module. Those records that were not found in the SSN module move to the GEO search module. In the GEO search module, the program is looking for a match using name plus date of birth plus sex within a certain geographic radius (no more than the surrounding neighborhoods of first 3 digits of zip code from the address provided). If there is still no match, the record moves to the name module, which also uses date of birth. The match to SSN is exact; matches involving other identifiers are probabilistic.²²

After secure transfer, Census personnel attempted to assign PIKs to approximately 199 000 patient records. The difference between this number and the 200 000 records drawn is attributable to removal of a few duplicates and the decision to drop patient records indicating nonbinary or "other" for sex because of sample-size considerations to avoid potential disclosure. All numbers were rounded to meet disclosure requirements. Once PIKs were assigned, personally identifying information was removed. None of the analytic research

files for this project contain personally identifying information such as name or Social Security number, as these fields are used only in the initial PIK-generation phase with authorized access for only a few employees at the Census Bureau.

Linking to American Community Survey Data

The health care system serves mainly residents of the state (2% of patients reside elsewhere), so we first considered limiting the assessment of linkages between EHRs and the ACS to in-state addresses sampled in the ACS. After further consideration, we opted for a more expansive approach because of significant migration into and out of the state over the 17-year span of ACS data utilized, which could undermine future analyses and conclusions regarding social determinants given the intrinsic mobility of residents.²³

Linking EHRs to the ACS means linking a subset of the population with a

sample of the population. Figure 1 illustrates this process. The EHRs refer to a selected subpopulation from the state consisting of 2.1 million patients during 2016 to 2019; the sample selected for the study is a subset (~200 000), and those assigned PIKs a further subset (~187 000; Figure 1a). ACS data are a representative sample of the state population (Figure 1c). The intersection of these 2 sets (i.e., matched EHRs and ACS records) is shown in Figure 1b. Not shown in the figure is coverage error in the ACS, which increased over time.²⁴ Furthermore, and relatedly, not all ACS data were assigned PIKs.

RESULTS

Overall, PIKs were assigned to 187 000 of the 199 000 records, a PIK rate of 94.0%. This compares favorably to the PIK rate for the ACS (90.8%–94.4%). As shown in Table 1, 77.5% of patients in the data from the health care system were assigned PIKs based on SSN, an

exact match. An additional 12.3% were found using name, date of birth, and sex within a geographic radius determined from their address, and a further 4.1% were found using name, sex, and date of birth. The latter PIK assignments are probabilistic. Of those records not successfully assigned PIKs, the large majority either lacked an SSN or provided an SSN that did not match the Census Bureau's reference file. Clearly, SSNs played an outsized role in the PIK assignment process. SSN matches provide reassurance on the quality of the linkages because they are exact. Yet, relying on exact SSN matches may not be realistic in the long run because health systems are increasingly moving away from collecting SSNs as mandatory fields. To investigate the possible consequences on non-SSN matches, we resubmitted the EHRs for PIK assignments without SSNs and using only name, address, sex, and date of birth, and succeeded in assigning PIKs for 90.0% of the sampled patients.

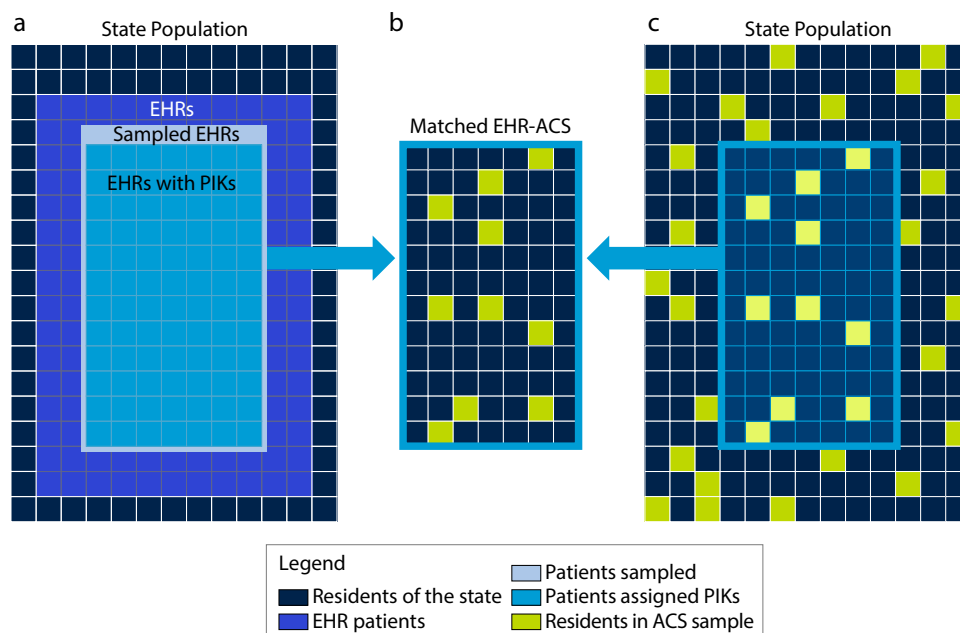


FIGURE 1— Matching Electronic Health Records (EHRs) to American Community Survey (ACS) Sample: United States

TABLE 1— Electronic Health Records (EHRs) Protected Identification Key Quality: United States

	No. (%)
Found, Social Security number	154 000 (77.54)
Found in geographic search	24 500 (12.34)
Found in name search	8 100 (4.08)
Found in date-of-birth search	100 (0.05)
Not found	11 900 (5.99)
Total	198 600 (100.00)

Note. The US Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release, CBDRB-FY21-POP001-0087. All numbers are rounded according to US Census Bureau disclosure protocols. The discrepancy between the total of 198 600 shown in the table and the references to 199 000 in the text reflects rounding error.

Source. American Community Survey data (2001–2017) and EHRs obtained from a health care system in the southeastern United States.

Most patients in the EHR subset will not be sampled in the ACS, and many respondents in the ACS will not be patients of the health care system. For these groups, there is no possibility of a match. Without knowing the population of eligible potential matches, it is not possible to say exactly what match rate would indicate that all eligible potential matches have been identified and linked, but, because the ACS is a true probability sample, we can put rough bounds around it. From the perspective of the EHRs, the maximal match rate will depend on the individuals sampled for the ACS each year (1%–1.5%) cumulated over years of observation (2001–2017), taking account of coverage (91.8%–94.1%), PIK rates (90.8%–94.4%), and smaller sampling fractions in 2001 to 2004 (0.2%–0.3%).¹¹ The maximal match rate likely lies between 12% and 18%, although we cannot know for sure. The match rate for our sample of EHRs with PIKs is within that range: 15.5%.

Table 2 shows the distribution of matched observations by year of the ACS record to which it matched. The distribution added up to 15.5%, the overall match rate. Match rates

were higher for ACS data collected in 2012 or later (1.23% or higher) than for ACS data collected in earlier years, (0.21%–0.27% in the 2001–2004 period). These patterns point to a tradeoff: the greater the number of years of ACS data included in the potential match, the higher the match rate will be, but the greater time difference in the reference year for the 2 sets of data may undermine the utility. It is questionable whether social, economic, and housing characteristics in, say, 2001 add much value in an analysis of health records collected in 2016 to 2019. If we restrict our attention to more recent ACS data, say 2013 to 2017, the match rate will be lower—in this instance, 6.47%.

Table 3 enables us to look at the time trends in a different way. For each ACS year, we can examine the percentage of individuals in the survey who match to a patient with at least 2 visits in 2016 to 2019. Match rates varied between 0.043% and 0.054%. The percentages were very small because we considered for a potential match all individuals included in the ACS for a particular year, not only in the state but also

across the country. Had we limited our attention to ACS records from the state, match rates would have been higher and the trend more pronounced.

DISCUSSION

Census Bureau data, such as the ACS, are collected to describe populations, follow trends in populations, and make demographic inferences useful to policymakers. In contrast, health data are collected for the express purpose of being applied to the individuals who provide it—though integrated delivery systems also routinely use aggregated data to track health care use and outcomes over time to improve care, and EHR data are increasingly used for research and to inform public health. Navigating a series of steps required for data integration, the study demonstrated a successful collaboration between a health care system and the US Census Bureau to advance a project germane to their disparate missions.

Data integration required internal review at both institutions and a data use agreement acceptable to both parties. Selected information from a sample of EHRs selected from the Clinical Data Warehouse was transmitted by secure means. Data integration proceeded smoothly: the PIK rate was high (94%), and the match rate to EHR data for observations with PIKs (15.5%) was within expectations given sampling fractions in the ACS and the 17-year span of ACS data used. Match rates are less with narrower windows—for example, 3.8% if ACS data were limited to the 2015–2017 period and 1.3% if limited to 2017. The narrower the window, the fewer the number of matched cases.

TABLE 2— Matches of Patients Having Electronic Health Records (EHRs) With Protected Identification Keys (PIKs) to Respondents in the American Community Survey (ACS) : United States

	No. of Matches	Total No. of EHRs With PIKs (Constant Over Time)	Match Rates of EHRs With PIKs
Overall, all ACS years	29 000	187 000	15.508
2001	450	187 000	0.241
2002	400	187 000	0.214
2003	500	187 000	0.267
2004	450	187 000	0.241
2005	1 700	187 000	0.909
2006	1 900	187 000	1.016
2007	1 700	187 000	0.909
2008	1 800	187 000	0.963
2009	1 800	187 000	0.963
2010	1 900	187 000	1.016
2011	1 900	187 000	1.016
2012	2 300	187 000	1.230
2013	2 400	187 000	1.283
2014	2 500	187 000	1.337
2015	2 500	187 000	1.337
2016	2 400	187 000	1.283
2017	2 300	187 000	1.230

Note. The US Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release, CBDRB-FY21-POP001-0087. All numbers are rounded according to US Census Bureau disclosure protocols.

Source. ACS data (2001–2017) and EHRs from a health care system in the southeastern United States (2016–2019).

Importantly, the expected number of matches is sufficient to support subsequent analyses. For patients seen in a particular year for a relatively common condition such as type 2 diabetes matched to 3 years of ACS data, we would expect more than 3000 matches. Even if we narrowed to 1 year of ACS data, there would be more than 1000 cases. For less common conditions, information from more than 1 medical center might be used, assuming that appropriate agreements could be negotiated. We could also boost the number of matches by expanding the Census Bureau data sources used—for

example, the Decennial Census or additional sources of administrative records such as program participation from states have much greater coverage than the ACS.

Future uses could include transfer of a greater number of EHR cases to Census, addition of structured clinical fields including *International Classification of Diseases, 10th Revision* (Geneva, Switzerland: World Health Organization; 1992) diagnostic codes, medication type and refills, procedures performed, details of inpatient hospital stays, and laboratory result data. Scaling up to include other medical centers is certainly possible

and would be facilitated through use of a common data model. From the Census perspective, the clinical detail of EHR-derived data will be helpful to provide clinical correlates of demographic and symptom report information. Uses can also include examination of the detailed data on household size, socioeconomic status, and structure, information that clinical systems generally lack. Clinical and health services researchers can examine the extent to which populations cared for within large integrated delivery systems are or are not representative of populations in geographic areas. In addition, the relationship of the Census “gold-standard” measures of social determinants of health and population demographic characteristics to clinical outcomes can be assessed.⁷

An issue that any research along these lines will face is timing. As noted, the wider the window of ACS data used for matching, the larger the number of matches. The advantages of having more cases must be balanced against the increasing possibility that the social, demographic, economic, and housing characteristics measured in the ACS have changed since they were collected. This timing difference is less of a problem for variables such as sex, race, ethnicity, age (which changes in predictable ways), or educational attainment (which changes little for most adults²⁵) than it is for more dynamic social determinants of health such as income,²⁶ household composition, or place of residence. For example, according to data from the Survey of Income and Program Participation, 12.2% of the population changed residence in 2013, and 14.9% experienced a change in household composition that year.²⁷ Exploring these tradeoffs in the future will be instructive

TABLE 3— Matches of Respondents in the American Community Survey (ACS) With Protected Identification Keys (PIKs) to Patients Having Electronic Health Records (EHRs) With PIKs: United States

	No. of Matches	Total No. of ACS Observations With PIKs (Varies by ACS Year)	Match Rates of All ACS Observations With PIKs
Overall, all ACS years	29 000	57 160 000	0.05074
2001	450	1 038 000	0.04335
2002	400	890 000	0.04494
2003	500	1 015 000	0.04926
2004	450	1 019 000	0.04416
2005	1 700	3 323 000	0.05116
2006	1 900	3 492 000	0.05441
2007	1 700	3 440 000	0.04942
2008	1 800	3 589 000	0.05015
2009	1 800	3 567 000	0.05046
2010	1 900	3 729 000	0.05095
2011	1 900	4 045 000	0.04697
2012	2 300	4 528 000	0.05080
2013	2 400	4 612 000	0.05204
2014	2 500	4 870 000	0.05133
2015	2 500	4 856 000	0.05148
2016	2 400	4 670 000	0.05139
2017	2 300	4 502 000	0.05109

Note. The US Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release, CBDRB-FY21-POP001-0087. All numbers are rounded according to US Census Bureau disclosure protocols.

Source. ACS data (2001–2017); EHRs from a health care system in the southeastern United States (2016–2019).

in determining appropriate cutpoints for different research questions.

Conclusions

Matched EHR–ACS data offer a unique window into population health, health disparities, and social determinants. ACS data tell us about life circumstances. EHR data tell us about health circumstances. While the ACS and other demographic databases lack the medical specificity that would make them useful population health tools, EHR data provide only a limited view of social determinants as well as potentially

skewed and biased depictions of the population, although the extent of the bias is uncertain. Linking Census Bureau and EHR data, though, is a promising avenue to reconcile demographic data with clinical data in a manner that improves our understanding of how disease operates in populations and the patterns of care that emerge and the role of the delivery system in assessing and improving population health.

Public Health Implications

Integrating population-representative data on social determinants of health in

the ACS and data on health outcomes in EHRs makes possible a better description of population health and a deeper examination of the social determinants of health and health disparities than is possible with either source alone. *AJPH*

ABOUT THE AUTHORS

Victoria Udalova, Lucinda Dalzell, and Joanna Motro are with the Enhancing Health Data Program, Demographic Directorate, US Census Bureau, Suitland, MD. Timothy S. Carey is with the Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill (UNC). Paul Roman Chelminski is with the Departments of Allied Health Science and Medicine, School of Medicine, UNC. Patricia Knoepp is with the Sheps Center for Health Services Research, UNC. Barbara Entwisle is with the Department of Sociology and Carolina Population Center, UNC.

CORRESPONDENCE

Correspondence should be sent to Barbara Entwisle, Carolina Population Center, 123 W Franklin St, Suite 210, Chapel Hill, NC 27516 (e-mail: entwisle@unc.edu). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints” link.

PUBLICATION INFORMATION

Full Citation: Udalova V, Carey TS, Chelminski PR, et al. Linking electronic health records to the American Community Survey: feasibility and process. *Am J Public Health*. 2022;112(6):923–930. Acceptance Date: February 7, 2022.

DOI: <https://doi.org/10.2105/AJPH.2022.306783>

CONTRIBUTORS

V. Udalova, T.S. Carey, P.R. Chelminski, L. Dalzell, P. Knoepp, J. Motro, and B. Entwisle contributed to the conceptualization and writing of the article. V. Udalova and B. Entwisle designed the analysis. J. Motro conducted the analysis. B. Entwisle developed the first draft of the article.

ACKNOWLEDGMENTS

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).

This article would not have been possible without help and support from both institutions. At UNC, we thank Denise Ammons, Allison Aiello, Elizabeth Frankenberg, Abigail Haydon, Emily Pfaff, and Jordan Young. We also want to acknowledge the very helpful comments and suggestions of the 5 anonymous reviewers. At the US Census Bureau, we thank Eloise Parker, who championed this project from its early days, and the Enhancing Health Data program for supporting this partnership.

Note. The Census Bureau reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization numbers CBDRB-FY21-POP001-0087 and CBDRB-FY21-POP001-0182. This article is intended to inform interested parties of ongoing research and to encourage discussion. Any views expressed are those of the authors and not necessarily those of the US Census Bureau.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

HUMAN PARTICIPANT PROTECTION

The study was approved by the institutional review board at UNC.

REFERENCES

- Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav.* 1995;35:80–94. <https://doi.org/10.2307/2626958>
- Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav.* 2010;51(1 suppl):S28–S40. <https://doi.org/10.1177/0022146510383498>
- Roundtable on Population Health Improvement, Board on Population Health and Public Health Practice, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. *Population Health Science in the United States: Trends, Evidence, and Implications for Policy: Proceedings of a Joint Symposium.* Washington, DC: National Academies Press; 2021:25631. <https://doi.org/10.17226/25631>
- James J. Nonprofit hospitals' community benefit requirements. *Health Affairs Health Policy Brief.* Washington, DC: Health Affairs; February 25, 2016. <https://doi.org/10.1377/hpb20160225.954803>
- Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health.* 1992;82(5):703–710. <https://doi.org/10.2105/AJPH.82.5.703>
- Bazemore AW, Cottrell EK, Gold R, et al. "Community vital signs": incorporating geocoded social determinants into electronic records to promote patient and population health. *J Am Med Inform Assoc.* 2016;23(2):407–412. <https://doi.org/10.1093/jamia/ocv088>
- Schinasi LH, Auchincloss AH, Forrest CB, Diez Roux AV. Using electronic health record data for environmental and place based population health research: a systematic review. *Ann Epidemiol.* 2018;28(7):493–502. <https://doi.org/10.1016/j.annepidem.2018.03.008>
- US Census Bureau. About the American Community Survey. Available at: <https://www.census.gov/programs-surveys/acs/about.html>. Accessed December 10, 2021.
- US Census Bureau, Economics and Statistics Administration. The American Community Survey 2019. August 2018. Available at: <https://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2019/quest19.pdf>. Accessed December 6, 2021.
- American Community Survey Office, US Census Bureau. Sample size. Available at: <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size>. Accessed December 10, 2021.
- US Census Bureau. ACS design and methodology (January 2014) - Chapter 2: Program history. January 2014. Available at: https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch02_2014.pdf. Accessed March 15, 2022.
- Buehler JW, Castro JC, Cohen S, Zhao Y, Melly S, Moore K. Personal and neighborhood attributes associated with cervical and colorectal cancer screening in an urban African American population. *Prev Chronic Dis.* 2019;16:E118. <https://doi.org/10.5888/pcd16.190030>
- Kurani SS, McCoy RG, Lampman MA, et al. Association of neighborhood measures of social determinants of health with breast, cervical, and colorectal cancer screening rates in the US midwest. *JAMA Netw Open.* 2020;3(3):e200618. <https://doi.org/10.1001/jamanetworkopen.2020.0618>
- Ghazi L, Oakes JM, MacLehose RF, Luepker RV, Osypuk TL, Drawz PE. Neighborhood socioeconomic status and identification of patients with CKD using electronic health records. *Am J Kidney Dis.* 2021;78(1):57–65.e1. <https://doi.org/10.1053/j.ajkd.2020.10.019>
- Tomayko EJ, Flood TL, Tandias A, Hanrahan LP. Linking electronic health records with community-level data to understand childhood obesity risk: electronic health records and childhood obesity. *Pediatr Obes.* 2015;10(6):436–441. <https://doi.org/10.1111/jjpo.12003>
- Lê-Scherban F, Ballester L, Castro JC, et al. Identifying neighborhood characteristics associated with diabetes and hypertension control in an urban African-American population using geo-linked electronic health records. *Prev Med Rep.* 2019;15:100953. <https://doi.org/10.1016/j.pmedr.2019.100953>
- Flood TL, Zhao YQ, Tomayko EJ, Tandias A, Carrel AL, Hanrahan LP. Electronic health records and community health surveillance of childhood obesity. *Am J Prev Med.* 2015;48(2):234–240. <https://doi.org/10.1016/j.amepre.2014.10.020>
- Bhavsar NA, Gao A, Phelan M, Pagidipati NJ, Goldstein BA. Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data. *JAMA Netw Open.* 2018;1(5):e182716. <https://doi.org/10.1001/jamanetworkopen.2018.2716>
- US Census Bureau. What we do. Available at: <https://www.census.gov/about/what.html>. Accessed July 22, 2021.
- US Census Bureau. Data Stewardship Executive Policy Committee. Available at: https://www.census.gov/about/policies/privacy/data_stewardship/dsep_committee.html. Accessed July 22, 2021.
- Finlay K, Genadek KR. Measuring all-cause mortality with the Census Numident file. *Am J Public Health.* 2021;111(suppl 2):S141–S148. <https://doi.org/10.2105/AJPH.2021.306217>
- Wagner D, Layne M. The Person Identification Validation System (PVS): applying the Center for Administrative Records Research and Applications' (CARRA) record linkage software. Washington, DC: US Census Bureau Center for Administrative Records Research and Applications; 2014.
- Spring A, Tolnay SE, Crowder K. Moving for opportunities? Changing patterns of migration in North America. In: White MJ, ed. *International Handbook of Migration and Population Distribution.* Vol 6. International Handbooks of Population. Dordrecht, the Netherlands: Springer Netherlands; 2016:421–448. https://doi.org/10.1007/978-94-017-7282-2_19
- American Community Survey Office. Coverage rates. Available at: <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/coverage-rates>. Accessed July 22, 2021.
- US Census Bureau. School enrollment in the United States: October 2019—detailed tables. Available at: <https://www.census.gov/data/tables/2019/demo/school-enrollment/2019-cps.html>. Accessed July 22, 2021.
- Hirschl TA, Rank MR. The life course dynamics of affluence. *PLoS ONE.* 2015;10(1):e0116370. <https://doi.org/10.1371/journal.pone.0116370>
- Monte L. Family complexity and changing household dynamics as measured in the 2014 Survey of Income and Program Participation. In: The Survey of Income and Program Participation. Fertility and Family Statistics Branch of US Census Bureau. 2017. Available at: <https://www.census.gov/content/dam/Census/library/working-papers/2017/demo/SEHSD-WP2017-119.pdf>. Accessed March 15, 2022.