Proceedings of the 12th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS 2026)
August 16 – 18, 2026 – London, United Kingdom
Paper No. ICBES 120
DOI: 10.11159/icbes26.120

When Place Matters: Age-Dependent Effects of Neighborhood Deprivation on Type 2 Diabetes Burden in the *All of Us* Program

Christopher Lee[1], Veronica Gomez-Godinez[2], Linda Shi[2]
[1]High school student participating in IEM OPALS program
chris.leon.lee2020@gmail.com
[2]Institute of Engineering in Medicine, University of California San Diego
9500 Gilman Drive, La Jolla, United States
gomezvg@gmail.com; zshi@ucsd.edu

Abstract - Type 2 diabetes (T2D) remains a major contributor to morbidity and healthcare costs, with persistent disparities linked to socioeconomic conditions. Many prior studies rely on self-reported surveys or modeled estimates, and fewer have examined how socioeconomic gradients in clinically observed T2D burden vary across adulthood. We evaluated associations between area-level socioeconomic status (SES) and ZIP3-level T2D prevalence using large-scale electronic health records from the *All of Us* Controlled Tier Dataset v8, linked to standardized ZIP3-level American Community Survey indicators, including a composite deprivation index and poverty rate. We first modeled ZIP3-level T2D case counts using binomial regression with decade-based age groups and gender, including SES-by-age interaction terms and ZIP3-clustered robust standard errors. We then used exploratory scatter plots and Pearson correlations to visualize age-specific gradients. Convergent evidence from these analyses identified a contrast emerging around age 60, which guided downstream summaries. We estimated age-standardized T2D prevalence at the 10th and 90th percentiles of each SES indicator within age-by-gender strata to quantify absolute and relative disparities. Socioeconomic gradients were modest at younger ages but substantially stronger at older ages, with composite deprivation showing the largest and most consistent differences. Among women aged ≥ 60 years, high poverty was associated with a 10.6 percentage point higher age-standardized T2D prevalence than low poverty (risk ratio 1.51), while high composite deprivation was associated with an 11.5 percentage point higher prevalence (risk ratio 1.57). These findings highlight pronounced, age-dependent socioeconomic disparities in clinically observed T2D burden in a national, EHR-based cohort and demonstrate a scalable approach for linking real-world clinical data with privacy-preserving area-level context to inform targeted prevention and care management later in life.

Keywords: type 2 diabetes; socioeconomic status; health disparities; electronic health records; life-course epidemiology

1. Introduction

Type 2 diabetes (T2D) is a leading cause of cardiovascular disease, kidney failure, and premature mortality in the United States [1]–[3]. The neighborhood environment, or "place", plays a fundamental part in shaping chronic disease and progression [4]. Neighborhood socioeconomic status (SES) influences type 2 diabetes (T2D) risk through access to healthy foods, physical activity opportunities, and chronic psychosocial stress [5]–[9]. However, most studies provide only a snapshot of this relationship, often failing to account for how these disparities shift as populations age [10].

Using the All of Us Research Program—a massive, diverse longitudinal cohort [11]–[12]—this study applies a life-course perspective to real-world evidence (RWE). We hypothesize that the socioeconomic gradient of T2D is not static but rather intensifies with age due to the cumulative impact of living in deprived environments. By linking Electronic Health Records (EHR) to ZIP3-level socioeconomic indices, we aim to quantify these age-dependent disparities and identify which populations are most vulnerable to the long-term effects of neighborhood deprivation.

2. Methods

2.1. Study population

Participants were selected from the *All of Us* Research Program Controlled Tier Dataset v8. The analytic cohort included all adult participants with EHR data available. Participants were uniquely identified by person_id.

2.2. Demographics

Demographic characteristics were obtained using the *All of Us* Demographics Prepackaged Concept Sets, including person_id, gender, race, ethnicity, and sex at birth.

2.3. Area-level socioeconomic status

Area-level socioeconomic indicators were obtained from the *All of Us* Controlled Tier Dataset v8 ZIP3 socioeconomic mapping resource (zip3_ses_map). Participant ZIP3 was derived from ZIP3-related records in the observation table (observation_source_concept_id = 1585250). The numeric ZIP3 prefix was parsed from observation.value_as_string and linked to zip3_ses_map, and the most recent ZIP3 observation was retained for each participant. To protect participant privacy and minimize re-identification risk, particularly in sparsely populated regions, the *All of Us* Controlled Tier Dataset restricts geographic identifiers to the ZIP3 level. Although this level of aggregation is coarser than ZIP5 codes or census tracts, it provides a standardized approach for linking large-scale clinical outcomes with area-level socioeconomic indicators while adhering to data security requirements.

2.4. Type 2 diabetes (T2D) phenotype

T2D was identified from the condition_occurrence table using either (a) standard condition concept identifiers (condition_concept_id IN (4193704, 201826)) or (b) source diagnosis codes mapped through the concept table, including ICD-10-CM E11% and ICD-9-CM 250.%0 or 250.%2, based on concept.vocabulary_id and concept.concept_code. T2D burden was summarized as the count of qualifying condition_occurrence records per participant, and T2D prevalence was defined as T2D_count ≥ 1.

2.5. Analytic dataset construction

A participant-level analytic dataset was created by left-joining demographics, the most recent ZIP3 socioeconomic record, and T2D counts onto the cohort participant list by person_id, producing one record per participant. The final analytic dataset included person_id, birth_datetime, gender, race, ethnicity, sex_at_birth, observation_datetime, zip3, and T2D_found flag.

2.6. ZIP3-level aggregation and stability thresholds

ZIP3-level summaries were generated by stratifying participants by decade-based age groups and gender. Within each stratum, T2D prevalence was computed as the proportion of participants with T2D diagnosis found in EHRs. To improve stability of estimates, strata with small denominators were excluded using a minimum sample-size threshold (default N = 50 per ZIP3 by age group by gender stratum). The final aggregated dataset contained zip3, age_bin, gender, n_patients, n_T2D, and ZIP3-level SES indicators.

2.7. Statistical analysis

We evaluated whether associations between area-level SES and ZIP3-level T2D prevalence varied across age using formal interaction models. ZIP3-level prevalence was modeled using binomial regression with counts of T2D cases and total patients as the outcome. As specified in Equation (1), each model included a standardized SES indicator (ses_z), decade-based age groups (age_bin), and gender as main effects, along with interaction terms between SES and age group to test for age-specific modification of SES–T2D associations.

$$T2D\ prevalence \sim ses_z + age_bin + gender + ses_z \times age_bin \quad (1)$$

To account for correlation among observations within the same ZIP3, standard errors were clustered at the ZIP3 level using robust sandwich estimators. Joint Wald tests were used to assess the statistical significance of SES-by-age interaction blocks.

We also conducted exploratory visualization using scatter plots of ZIP3-level T2D prevalence versus SES indicators, stratified by age group and gender. Pearson correlation coefficients were used descriptively to summarize effect size, with emphasis placed on associations with $|r| \geq 0.2$ rather than statistical significance alone. These exploratory analyses informed consolidation of age groups for downstream analyses.

To quantify the magnitude of SES disparities, we estimated age-standardized T2D prevalence at the 10th (P10) and 90th percentiles (P90) of each SES indicator within age–gender strata. Absolute risk differences and risk ratios between age-standardized prevalence at P10 and P90 were computed to compare high versus low SES environments.

All analyses were conducted using Python (version 3.14) and R (version 4.5).

3. Results

A total of 393,601 adult participants had available EHR data in the *All of Us* Research Program Controlled Tier Dataset v8. Among these, 376,904 participants (95.8%) were successfully linked to ZIP3-level socioeconomic indicators derived from the American Community Survey. A total of 78,310 participants (19.9%) met criteria for type 2 diabetes based on clinically recorded diagnoses. The final participant-level analytic dataset, defined by having both a diagnosis-based T2D phenotype and ZIP3-linked socioeconomic data, included 68,719 participants.

Formal interaction models demonstrate significant age-related differences in baseline T2D prevalence across successive decade-based age groups. Joint tests of socioeconomic status (SES)-by-age interactions confirm that the relationship between neighborhood context and T2D prevalence is modified by age, particularly for the composite deprivation index ($p = 0.0039$) and poverty measures ($p = 0.0023$). Exploratory scatter plots provided complementary visual evidence. Across decade-based age groups, associations between SES indicators and T2D prevalence were generally modest in younger adults but became steeper and more consistent with increasing age. Visual inspection suggested a contrast beginning around age 60, with stronger socioeconomic gradients observed in adults aged 60 years or older across multiple SES measures and in both genders.

Based on convergence between interaction models and exploratory patterns, we defined two age categories for subsequent analyses: adults younger than 60 years and adults aged 60 years or older. All subsequent correlation summaries and age-standardized estimates were conducted using these final age groups.

Fig.s 1A–1B show ZIP3-level associations between T2D prevalence and two key socioeconomic indicators: poverty rate, and a composite deprivation index, stratified by age group (< 60 vs ≥ 60 years) and gender. Across both measures, associations with T2D prevalence were weak or inconsistent among adults younger than 60 years.

In contrast, among adults aged ≥ 60 years, clear socioeconomic gradients were observed across both indicators. Higher poverty and higher composite deprivation,, were consistently associated with higher T2D prevalence in this older age group. These patterns were evident in both men and women but were more pronounced among women, as indicated by larger absolute Pearson correlation coefficients and smaller p-values across SES measures (Fig.s 1A-1B).

Across the two figures, gradients differed markedly by age, with substantially stronger and more consistent associations observed among older adults. To quantify the magnitude of these disparities, we estimated age-standardized T2D prevalence at the 10th and 90th percentiles of each SES indicator within age–gender strata (Table 1). Among adults younger than 60 years, absolute differences in standardized prevalence were generally modest, typically below 5 percentage points. In contrast, among adults aged ≥ 60 years, pronounced differences were observed. Among women aged ≥ 60 years, residence in high-poverty ZIP3s was associated with a 10.6 percentage point higher age-standardized T2D prevalence compared with low-poverty areas (31.4% vs 20.8%), corresponding to a risk ratio of 1.51 (Table 1). High composite deprivation index was associated with an 11.5 percentage point higher prevalence (31.8% vs 20.3%), corresponding to a risk ratio of 1.57. Patterns among men aged ≥ 60 years were directionally similar but attenuated.

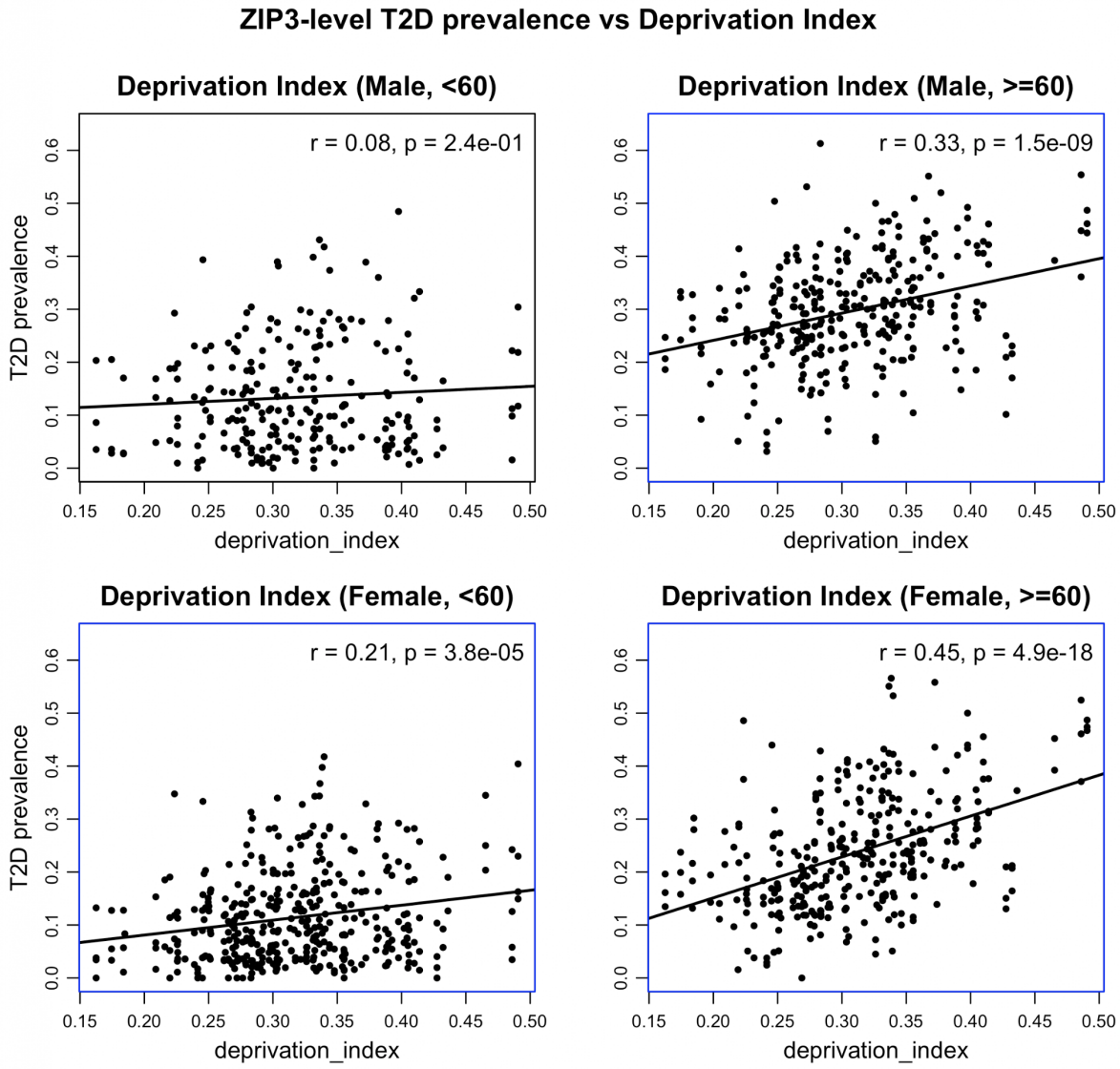

Fig. 1A. ZIP3-level type 2 diabetes (T2D) prevalence versus *composite deprivation index*, stratified by age group and gender. Scatter plots show ZIP3-level T2D prevalence in relation to a composite deprivation index, stratified by age (< 60 vs ≥ 60 years) and gender (men and women). Each point represents a ZIP3 meeting minimum data thresholds. Solid lines indicate linear fits, and Pearson correlation coefficients (r) with corresponding p-values are shown in each panel. Associations were modest among adults younger than 60 years and stronger among adults aged 60 years or older, particularly among women

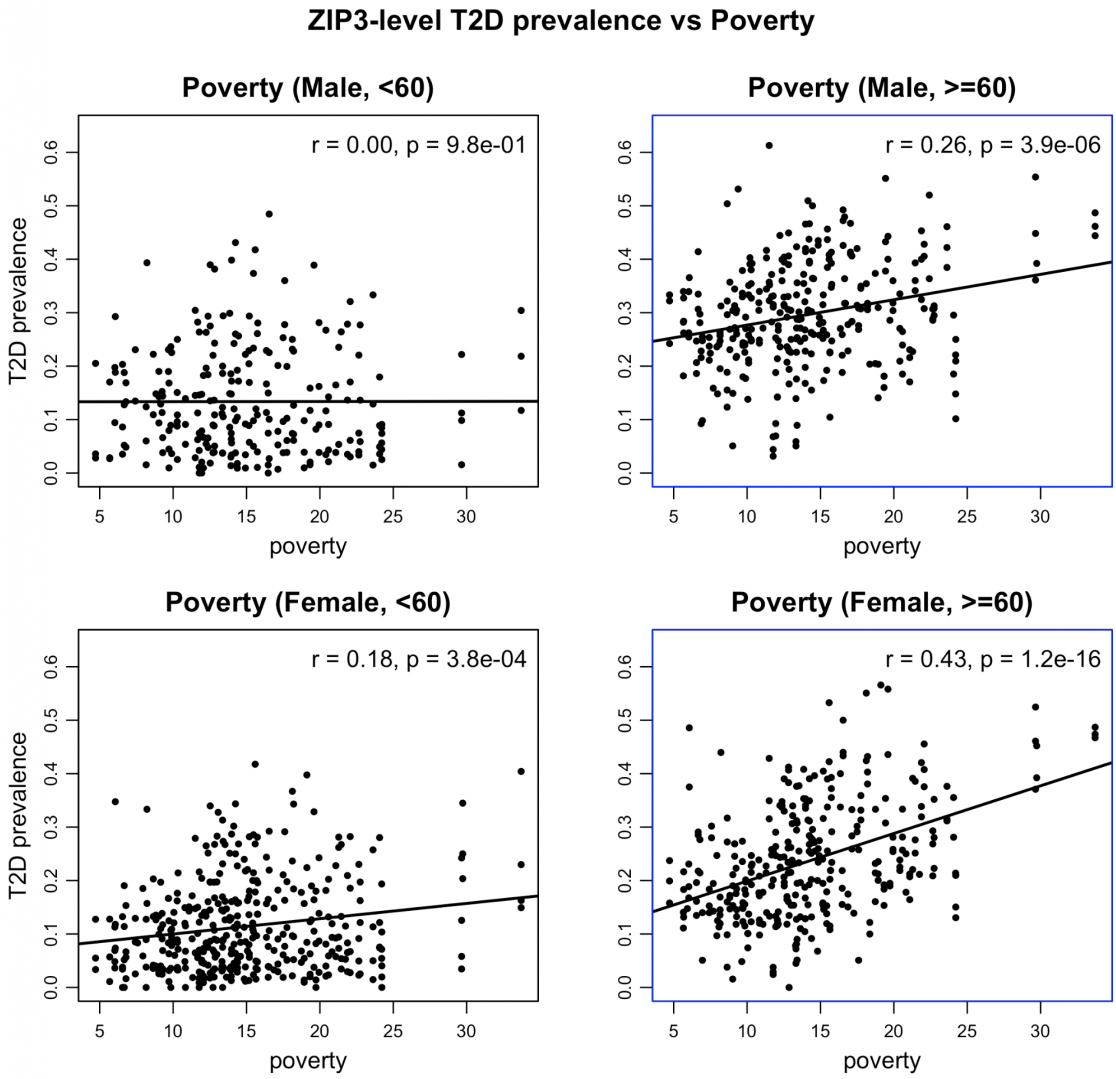

Fig. 1B. ZIP3-level type 2 diabetes (T2D) prevalence versus *poverty rate*, stratified by age group and gender. Scatter plots show ZIP3-level T2D prevalence in relation to poverty rate, stratified by age (< 60 vs ≥ 60 years) and gender (men and women). Each point represents a ZIP3 meeting minimum data thresholds. Solid lines indicate linear fits, and Pearson correlation coefficients (r) with corresponding p-values are shown in each panel. Associations were modest among adults younger than 60 years and stronger among adults aged 60 years or older, particularly among women.

Across age and gender strata, the composite deprivation index showed the largest and most consistent absolute and relative differences.

Socioeconomic Gradients in Type 2 Diabetes Prevalence by Age and Sex

strata	ses_var	ses_p10	ses_p90	std_prev_p10	std_prev_p90	risk_diff	risk_ratio	n_zip3	total_patients	total_t2d
< 60 Male	poverty	8.21	22.65	0.14	0.13	0.00	0.97	85	48,878	6,587
< 60 Male	deprivation_index	0.24	0.40	0.13	0.14	0.01	1.09	85	48,878	6,587
>=60 Male	poverty	7.79	21.45	0.28	0.31	0.03	1.12	135	66,790	19,764
>=60 Male	deprivation_index	0.23	0.39	0.27	0.32	0.05	1.18	135	66,790	19,764
< 60 Female	poverty	8.65	21.88	0.11	0.14	0.04	1.34	124	105,115	13,218
< 60 Female	deprivation_index	0.24	0.40	0.10	0.15	0.04	1.40	124	105,115	13,218
>=60 Female	poverty	7.99	21.28	0.21	0.31	0.11	1.51	146	90,443	23,954
>=60 Female	deprivation_index	0.23	0.39	0.20	0.32	0.11	1.57	146	90,443	23,954

4. Discussion

The results demonstrate that the association between neighborhood socioeconomic context and T2D burden is highly age-dependent. The pronounced widening of the socioeconomic gradient in the ≥ 60 cohort suggests that chronic exposure to adverse environmental determinants—including limited nutritional access and psychosocial stressors—results in cumulative physiological degradation over the lifespan. In younger cohorts, these disparities are present but likely latent, manifesting fully only after decades of residential exposure.

A critical factor in the observed intensification of the gradient is residential stability. Older adults typically exhibit lower rates of geographic mobility compared to younger populations. Consequently, the ZIP3-level socioeconomic indicators in the ≥ 60 group likely serve as a more accurate proxy for long-term environmental exposure. The more modest associations in the < 60 cohort may be partially attributed to higher residential turnover, which introduces "noise" into the relationship between current neighborhood characteristics and clinical outcomes.

ZIP3-level data, while necessary for participant privacy, represents a relatively coarse geographic scale. This aggregation likely leads to conservative estimates of the true socioeconomic disparity, as it masks finer-grained intra-community variations in deprivation [13]–[14]. Furthermore, while the EHR-based phenotyping is robust, it relies on clinical touchpoints, potentially underestimating prevalence in the most deprived quintiles where healthcare access is limited [15].

5. Conclusions

This study leverages a massive RWE dataset to demonstrate that the socioeconomic gradient of type 2 diabetes is not a static feature of the population but an escalating burden that intensifies with age. The finding that women in high-deprivation areas aged 60 and older face a disproportionately high T2D prevalence highlights a critical target for public health intervention. As the All of Us Research Program continues to expand, integrating more granular longitudinal data will be essential to further untangle the causal pathways between neighborhood deprivation and metabolic health. Ultimately, these results suggest that clinical and policy interventions must adopt a life-course perspective, recognizing that the health disparities observed in older age are the culmination of lifelong environmental exposures.

Acknowledgements

We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available

the participant data examined in this study. The *All of Us* Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

This project was supported by a generous gift from the Beckman Laser Institute Inc. to the UCSD Biophotonics Technology Center. We also extend our sincere gratitude to the donors of the UCSD OPALS Program—including the Leighton Foundation Inc., Ming-Wei Wu, Inc., Dr. Shu Chien (UCSD Bioengineering), Dr. Lizhu Chen (CorDx Inc.), Dr. Xinhua Zheng, David & Leslie Lee, and Mingwei Hu & Wen Shi—for their invaluable support and generosity.

References

[1] GBD 2021 Diabetes Collaborators, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," *Lancet*, vol. 402, no. 10397, pp. 203–234, Jul. 2023.

[2] W. B. Kannel and D. L. McGee, "Diabetes and cardiovascular disease. The Framingham study," *JAMA*, vol. 241, no. 19, pp. 2035–2038, May 1979.

[3] K. M. V. Narayan, J. P. Boyle, T. J. Thompson, S. W. Sorensen, and D. F. Williamson, "Lifetime risk for diabetes mellitus in the United States," *JAMA*, vol. 290, no. 14, pp. 1884–1890, Oct. 2003.

[4] A. V. Diez Roux and C. Mair, "Neighborhoods and health: Neighborhoods and health," *Ann. N. Y. Acad. Sci.*, vol. 1186, no. 1, pp. 125–145, Feb. 2010.

[5] A. H. Auchincloss, A. V. Diez Roux, D. G. Brown, C. A. Erdmann, and A. G. Bertoni, "Neighborhood resources for physical activity and healthy foods and their association with insulin resistance," *Epidemiology*, vol. 19, no. 1, pp. 146–157, Jan. 2008.

[6] R. Saelee, I. A. Hora, M. E. Pavkov, G. Imperatore, Y. Chen, S. R. Benoit, C. S. Holliday, and K. M. Bullard, "Diabetes prevalence and incidence inequality trends among U.s. adults, 2008-2021," *Am. J. Prev. Med.*, vol. 65, no. 6, pp. 973–982, Dec. 2023.

[7] A. V. Diez Roux, S. S. Merkin, D. Arnett, L. Chambless, M. Massing, F. J. Nieto, P. Sorlie, M. Szklo, H. A. Tyroler, and R. L. Watson, "Neighborhood of residence and incidence of coronary heart disease," *N. Engl. J. Med.*, vol. 345, no. 2, pp. 99–106, Jul. 2001.

[8] D. Kuh, Y. Ben-Shlomo, J. Lynch, J. Hallqvist, and C. Power, "Life course epidemiology," *J. Epidemiol. Community Health*, vol. 57, no. 10, pp. 778–783, Oct. 2003.

[9] P. Clarke, J. Morenoff, M. Debbink, E. Golberstein, M. R. Elliott, and P. M. Lantz, "Cumulative exposure to neighborhood context: consequences for health transitions over the adult life course: Consequences for health transitions over the adult life course," *Res. Aging*, vol. 36, no. 1, pp. 115–142, Jan. 2014.

[10] F. Hill-Briggs, N. E. Adler, S. A. Berkowitz, M. H. Chin, T. L. Gary-Webb, A. Navas-Acien, P. L. Thornton, and D. Haire-Joshu, "Social determinants of health and diabetes: A scientific review," *Diabetes Care*, vol. 44, no. 1, pp. 258–279, Nov. 2020.

[11] A. H. Ramirez, L. Sulieman, D. J. Schlueter, A. Halvorson, J. Qian, F. Ratsimbazafy, R. Loperena, K. Mayo, M. Basford, N. Deflaux, K. N. Muthuraman, K. Natarajan, A. Kho, H. Xu, C. Wilkins, H. Anton-Culver, E. Boerwinkle, M. Cicek, C. R. Clark, E. Cohn, L. Ohno-Machado, S. D. Schully, B. K. Ahmedani, M. Argos, R. M. Cronin, C. O'Donnell, M. Fouad, D. B. Goldstein, P. Greenland, S. J. Hebbring, E. W. Karlson, P. Khatri, B. Korf, J. W. Smoller, S. Sodeke, J. Wilbanks, J. Hentges, S. Mockrin, C. Lunt, S. A. Devaney, K. Gebo, J. C. Denny, R. J. Carroll, D. Glazer, P. A. Harris, G. Hripcsak, A. Philippakis, D. M. Roden, and All of Us Research Program, "The All of Us Research Program: Data quality, utility, and diversity," *Patterns (N. Y.)*, vol. 3, no. 8, p. 100570, Aug. 2022.

[12] All of Us Research Program Investigators, J. C. Denny, J. L. Rutter, D. B. Goldstein, A. Philippakis, J. W. Smoller, G. Jenkins, and E. Dishman, "The 'All of Us' Research Program," *N. Engl. J. Med.*, vol. 381, no. 7, pp. 668–676, Aug. 2019.

[13] G. K. Singh, "Area deprivation and widening inequalities in US mortality, 1969-1998," *Am. J. Public Health*, vol.

93, no. 7, pp. 1137–1143, Jul. 2003.
[14] R. Pampalon, D. Hamel, P. Gamache, and G. Raymond, “A deprivation index for health planning in Canada,” *Chronic Dis. Can.*, vol. 29, no. 4, pp. 178–191, 2009.
[15] K. J. O’Malley, K. F. Cook, M. D. Price, K. R. Wildes, J. F. Hurdle, and C. M. Ashton, “Measuring diagnoses: ICD code accuracy,” *Health Serv. Res.*, vol. 40, no. 5 Pt 2, pp. 1620–1639, Oct. 2005.