



## Neighborhood built environment, obesity, and diabetes: A Utah siblings study

Quynh C. Nguyen<sup>a</sup>, Tolga Tasdizen<sup>b</sup>, Mitra Alirezaei<sup>b</sup>, Heran Mane<sup>a</sup>, Xiaohe Yue<sup>a</sup>,  
Junaid S. Merchant<sup>a</sup>, Weijun Yu<sup>a</sup>, Laura Drew<sup>a</sup>, Dapeng Li<sup>c</sup>, Thu T. Nguyen<sup>a,\*</sup>

<sup>a</sup> Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, MD, United States

<sup>b</sup> Department of Electrical and Computer Engineering, Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT, United States

<sup>c</sup> Department of Geography and the Environment, University of Alabama, Tuscaloosa, AL, United States

### ARTICLE INFO

#### Keywords:

Neighborhood built environment  
Chronic diseases  
Google street view  
Convolutional neural networks  
Siblings

### ABSTRACT

**Background:** This study utilizes innovative computer vision methods alongside Google Street View images to characterize neighborhood built environments across Utah.

**Methods:** Convolutional Neural Networks were used to create indicators of street greenness, crosswalks, and building type on 1.4 million Google Street View images. The demographic and medical profiles of Utah residents came from the Utah Population Database (UPDB). We implemented hierarchical linear models with individuals nested within zip codes to estimate associations between neighborhood built environment features and individual-level obesity and diabetes, controlling for individual- and zip code-level characteristics (n = 1,899,175 adults living in Utah in 2015). Sibling random effects models were implemented to account for shared family attributes among siblings (n = 972,150) and twins (n = 14,122).

**Results:** Consistent with prior neighborhood research, the variance partition coefficients (VPC) of our unadjusted models nesting individuals within zip codes were relatively small (0.5%–5.3%), except for HbA1c (VPC = 23%), suggesting a small percentage of the outcome variance is at the zip code-level. However, proportional change in variance (PCV) attributable to zip codes after the inclusion of neighborhood built environment variables and covariates ranged between 11% and 67%, suggesting that these characteristics account for a substantial portion of the zip code-level effects. Non-single-family homes (indicator of mixed land use), sidewalks (indicator of walkability), and green streets (indicator of neighborhood aesthetics) were associated with reduced diabetes and obesity. Zip codes in the third tertile for non-single-family homes were associated with a 15% reduction (PR: 0.85; 95% CI: 0.79, 0.91) in obesity and a 20% reduction (PR: 0.80; 95% CI: 0.70, 0.91) in diabetes. This tertile was also associated with a BMI reduction of  $-0.68$  kg/m<sup>2</sup> (95% CI:  $-0.95$ ,  $-0.40$ ).

**Conclusion:** We observe associations between neighborhood characteristics and chronic diseases, accounting for biological, social, and cultural factors shared among siblings in this large population-based study.

### 1. Introduction

Neighborhood characteristics play a powerful role in shaping people's health and a growing body of research has demonstrated the particular importance of built environment factors associated with human-made spaces that individuals engage with on a regular basis (Cummins & Jackson, 2001; Gary et al., 2007; Renalds et al., 2010; Tuckett et al., 2018). This includes physical characteristics like walkability and green spaces that influence health behaviors, as well as factors like public transportation and zoning regulations that constrain

individuals' food environments, access to healthcare, and social support (Eibich et al., 2016; Frank et al., 2022; Gomez et al., 2015; Zick et al., 2013). However, much of the extant research on built environment and health has relied on self-report measures or existing demographic datasets (Liu et al., 2018). Research investigating objective measures to examine associations between built environment and health has been emerging. A group of researchers built an objective measuring tool for built environment characteristics. They evaluated 200 older people living in diverse housing and neighborhoods in England. Their results showed that neighborhood characteristics can significantly affect the

\* Corresponding author.

E-mail address: [txn@umd.edu](mailto:txn@umd.edu) (T.T. Nguyen).

<https://doi.org/10.1016/j.ssmph.2024.101670>

Received 6 October 2023; Received in revised form 2 April 2024; Accepted 5 April 2024

Available online 19 April 2024

2352-8273/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

health of older residents and highlighted their tool as a potential evidence-based design guide for places for older individuals (Burton et al., 2011). Another study conducted in New Zealand surveyed 2033 residents in 48 neighborhoods and found a positive correlation between five built environment characteristics (street connectivity, residential density, destination accessibility, street environment quality, and land use diversity) and physical activity (Witten et al., 2012).

The current study employs a data-driven approach that leverages Google Street View images to derive neighborhood characteristics and examines the associations with health using a dataset that includes familial information about siblings and twins. This research builds upon emerging research that leverages innovative, data-rich technology such as Google Street View (Hystad et al., 2022; Keralis et al., 2020; Larkin et al., 2022) and contributes to the existing body of knowledge regarding the links between built environment characteristics and shared genetic and social factors (Cohen-Cline et al., 2015; Duncan et al., 2022, 2023; Figaroa et al., 2023).

Neighborhood characteristics create circumstances that are more or less conducive to healthy nutritional choices and increased physical activity (Christensen et al., 2010; Cohen, 2019; Coleman et al., 2008; Franco et al., 2008; Morland et al., 2002), and research has linked these factors to chronic conditions such as obesity and diabetes that rank high among pervasive health challenges in the U.S. (Baskin et al., 2005; Engelgau et al., 2004; Mokdad et al., 2001). For instance, the presence of neighborhood establishments offering healthy foods is associated with reduced incidence of diabetes (Auchincloss et al., 2009; Christine et al., 2015; Kanchi et al., 2021). Similarly, neighborhood walkability is associated with a lower risk for diabetes, obesity, and hypertension (Chandrabose et al., 2019; Creatore et al., 2016; Liu et al., 2018; Sundquist et al., 2015; Zick et al., 2013), and exposure to environments that encourage physical activity is associated with decreased diabetes and obesity incidence (Christine et al., 2015; Ellaway et al., 2005; Kanchi et al., 2021). Research also suggests that aspects of built environments promote health through indirect mechanisms. For instance, green spaces can increase positive perceptions about the quality of outdoor spaces, which incentivizes physical activity and outdoor social activities (Richardson et al., 2013; Sullivan et al., 2004), and research has found associations between green spaces and stress reduction, physical activity, and social cohesion (Beyer et al., 2014; Kweon et al., 1998; Sugiyama et al., 2008). On the other hand, emerging international work suggests that the visibility of environmental factors indicating disorder and danger, like visible wires, can disincentivize walking and other health behaviors out of safety concerns (Remigio et al., 2019). Although compelling, the aforementioned findings are observational, and unmeasured confounding of individual-level risk factors remains a viable threat to causal inference.

Novel research approaches leveraging familial relationships have been developed to disentangle genetic factors from neighborhood and environmental effects. A few studies have used data collected from siblings and twins to partition variance associated with shared environments and genetic factors. One study involving 9918 women in Utah found that residing in a walkable neighborhood was associated with decreases in obesity risk for women. This protective effect was observed for women with and without a familial risk of obesity based on siblings' BMI (Kowaleski-Jones et al., 2017). Another study also found consistent associations between neighborhood socioeconomic status and heart disease incidence among siblings, especially in women (Forsberg et al., 2018)—providing more support for the impact of environment on health.

However, a study in Sweden involving 415,540 middle-aged full-brothers that evaluated the relationship between increased risk of ischemic heart disease (IHD) and socioeconomically disadvantaged neighborhoods found that neighborhood boundaries were limited in predicting IHD risk. This highlights the potential caveat of unaccounted confounders including unshared environmental and genetic factors (Merlo et al., 2013). Similarly, another Swedish co-sibling analysis that

adjusted for shared genetic and familial environmental factors did not show significant differences in health outcomes among siblings residing in neighborhoods with varying walkability (Sundquist et al., 2015).

These findings provide a framework for investigating the health impacts of neighborhood characteristics, but the limited and mixed findings indicate further investigations are needed. Some of the challenges in identifying ecological effects on population health include determining the individual and neighborhood units of analysis, determining whether to focus on specific neighborhood attributes such as presence of sidewalks versus defining overall neighborhood causal estimates (Subramanian et al., 2007).

### 1.1. Study aims

In this study, we aim to observe the relationships between built environment characteristics and health outcomes. We utilized a database of 1.4 million Google Street View images and individual-level health outcome data from the Utah Population Database to analyze the interconnections between health and built environment characteristics. We implemented models accounting for shared genetic and familial factors between siblings (shared biological mother; shared biological father; shared biological mother and father; and identical and fraternal twins) to address potential confounding bias by these background characteristics. We hypothesized that individuals residing in neighborhoods with higher rates of walkability (sidewalks), improved neighborhood aesthetics (green streets), urban development (non-single family-homes, multiple lane roads), and lower physical disorder (visible utility wires overhead) would exhibit a reduced prevalence of chronic disease (Remigio et al., 2019).

## 2. Methods

### 2.1. Google Street View image data collection and processing

Using the Google Street View (GSV) Image API, we collected street view image data from all primary and secondary roads in Utah, specifically sampling street intersections and locations along road segments at 50-m intervals. This process, conducted in November 2019, yielded 1,394,442 images, with each coordinate set offering images facing west, east, north, and south. To provide training data for our computer vision models, between December 2016 and February 2017, we manually annotated 18,700 images from various locations, including Chicago, Salt Lake City, and Charleston. This effort, which involved the principal investigator and three graduate research assistants labeling each image for the presence (or absence) of the neighborhood characteristics (e.g., single lane road, sidewalks). The raters achieved an inter-rater agreement exceeding 85% for neighborhood indicators. The data was split into training (80%) and testing sets (20%). Using Visual Geometry Group VGG-19 (Simonyan & Zisserman, 2015) and TensorFlow (Abadi et al., 2016) with sigmoid cross entropy with logits as the loss function, we trained a deep convolutional neural network, achieving high accuracy rates for various recognition tasks: including street greenness at 89%, single lane roads at 88%, and overhead utility wires at 83%. eTable 1 displays the data dictionary for these indicators. To build our sidewalk indicator, we used 24,316 manually labeled Google Street View images. We then trained a standard deep convolutional neural network architecture- ResNet-18, which produced an accuracy of 84.5% and a F1 score of 81.0. Further methodological details were elaborated in previous studies (Nguyen et al., 2017, 2022).

### 2.2. Health outcomes

*Individual level health data.* We merged our neighborhood data with individual health outcome data from the Utah Population Database (UPDB). The data cover the entire state of Utah, including rural and frontier counties. It is the only database of its kind in the United States

**Table 1**  
Descriptive characteristics for neighborhood and individual characteristics.

	N	Mean (SD)
<i>Zip code level built environment characteristics</i>		
Non-single-family home	30,556	25.62 (21.10)
Visible utility wires	30,556	44.14 (16.81)
Sidewalk	30,556	19.50 (24.31)
Single lane road	30,556	65.47 (14.31)
Green street	30,556	87.08 (15.70)
<i>Individual-level characteristics</i>		
Age (years)	1,968,451	46.23 (17.37)
% Female	1,964,485	49.55 (50.00)
% Married	1,786,137	61.82 (48.58)
% Nonwhite race	1,832,286	4.73 (21.23)
% white race	1,832,286	95.27 (21.23)
% Hispanic ethnicity	1,703,412	11.05 (31.35)
% Less than high school	1,208,101	9.11 (28.77)
% High school	1,208,101	30.00 (45.83)
% Some college	1,208,101	31.90 (46.61)
% College or greater	1,208,101	28.99 (45.37)
% Obese	1,952,993	28.75 (45.26)
% Diabetes	1,968,451	5.22 (22.24)
Body mass index (kg/m <sup>2</sup> )	1,952,993	27.67 (6.52)
Fasting glucose (mg/dL)	137,589	94.91 (32.15)
Hemoglobin A1c (%)	382,100	6.06 (2.17)

Data sources: Google Street View-derived built environment characteristics from Utah; Utah Population Database, University of Utah Health Science Center Data Warehouse; Intermountain Healthcare Data Warehouse.

because a central component of UPDB is the extensive set of Utah family histories, interlinked with demographic and medical data. The UPDB contains records covering over 11 million individuals tracing back to the late 18th century and is a dynamic database updated annually by data contributors. Data for health diagnoses and lab values were obtained from Intermountain Healthcare, the University of Utah Health Sciences Center Enterprise Data Warehouse (UUHSC EDW) and the Utah Department of Health and Human Services (DHHS) records. Other UPDB data contributors for administrative data include the Utah DHHS Office of Vital Records and Statistics (for birth certificates) and the Utah Driver License Division.

Each individual's multiple records are linked via a unique person ID. The UPDB can only be utilized for biomedical and health-related research with confidentiality and privacy of individuals strictly protected. The UPDB is administered by the Pedigree and Population Resource (PPR) of Huntsman Cancer Institute at the University of Utah. More information on the UPDB can be found at <https://uofuhealth.utah.edu/huntsman/utah-population-database>.

**Cohort of adults:** Our analytic sample consisted of adults in Utah who were aged 20 years or older in 2015. **Obesity:** Obesity was defined dichotomously based on the body mass index (BMI) (kg/m<sup>2</sup>)  $\geq 30$  from clinical records. If BMI was missing from clinical records (42%) because individuals did not seek clinical care from one of the data contributors, self-reported height and weight were extracted from Utah's Driver License Division. Utah law requires individuals to renew their driver license every five years and to renew it in person every ten years. BMI data were drawn from the most recent driver license data spanning from 2011 to 2017.

**Diabetes** was assessed via diagnoses by a health care provider using ICD-9 ('250.00', '250.01', '250.02', '250.03', '250.70', '250.71', '250.72', '250.73', '250.40', '250.41', '250.42', '250.43', '250.50', '250.51', '250.52', '250.53', '250.60', '250.61', '250.62', '250.63') and ICD-10 codes (E11.40', 'E10.9', 'E10.29', 'E11.51', 'E11.29', 'E11.36', 'E10.36', 'E10.39', 'E10.51', 'E10.40', 'E10.65', 'E11.21', 'E11.319', 'E10.319', 'E11.65', 'E11.311', 'E11.39', 'E10.21', 'E10.311', 'E11.9'). Lab values for fasting glucose (mg/dL) and glycated hemoglobin (HbA1c (%)) were also examined as health outcomes.

### 2.3. Covariates

Analyses controlled for the following individual-level characteristics: age (years), sex (male/female), marital status (married vs. divorced/single), nonwhite race, ethnicity, and education (less than high school, high school, some college, and college or greater). These data were obtained from administrative and medical records. We sourced an individual's education level from the Utah birth certificates of their children (if any) or medical records. Mother's and father's education levels were included in Utah birth certificates.

**Zip code characteristics.** We adopted zip code boundaries to define neighborhoods. We used the 2011–2015 American Community Survey's (ACS) five-year estimates (U.S. Census Bureau, 2016). The ACS, administered by the U.S. Census Bureau, is the premier source for detailed population and housing information about the United States. The following zip code level ACS variables were included in analyses to control for potential confounding of other area-level characteristics in estimating the association between built environment characteristics and health outcomes: population density, percentage of the population aged 65 years and older, percentage of the population that is Hispanic, percentage of the population that is Black, and median household income. We did not include property values as a covariate in the models due to its high correlation ( $r = 0.93$ ) with median household income. eTable 2 displays the correlation matrix for the remaining zip code ACS variables. Correlations ranged from  $-0.05$  to  $0.50$ .

In sensitivity analyses, we investigated census tracts as alternative neighborhood boundaries but found that would involve dropping about 500,000 individuals from analyses that were missing census tract identifiers, and thus zip codes were selected to conserve more of the sample. Fig. 1 displays our path diagram indicating the size of the original population and successive selection criteria that led to the final analytic sample and siblings subsamples. eTable 3 displays the rate of missingness for each variable included in analyses.

### 2.4. Analytic approach

To aggregate the image results into distinct geographic areas, we spatially joined the image's geocoordinates to its zip code location. Choropleth maps were created using ArcGIS Desktop software (ESRI, Inc.) and the 2016 U.S. Census TIGER/Line Shapefiles. We utilized quartile breaks in displaying spatial data patterns. We produced descriptive statistics of area-level built environment characteristics as well as sociodemographics and health outcomes of our analytic sample. To examine the relationship between built environment characteristics and individual-level health outcomes, we merged administrative and medical records from the Utah Population Database to zip code summaries of built environment characteristics.

Using the complete analytic sample of approximately 1.9 million individuals, we implemented hierarchical linear models for each outcome, with individuals nested within zip codes and robust standard error estimation using Stata MP16 (StataCorp, 2015). Following recent developments in modeling geographic inequalities, we first calculated a null model that only included a random intercept term for zip code to estimate general zip code-level effects on outcomes. These effects were quantified as the variance partition coefficient (VPC), which is equivalent in this context to the intra-class correlation. The VPC represents the percentage of the total variance (at both the individual- and zip code-level) that is associated with zip code (Merlo, 2003; Merlo et al., 2018, 2019). We then calculated the fully adjusted models and compared the zip code-level variance to the estimate obtained from the null model to estimate proportional change in variance (PCV), which percent of the total zip code-level variation that is explained by inclusion of the built environment variables and covariates (Evans & Erickson, 2019; Merlo, 2018; Merlo et al., 2016). For dichotomous outcomes (i.e., obesity and diabetes status), we implemented log Poisson models with zip code as random intercept to estimate prevalence ratios (95% CIs)

**Table 2**  
Google Street View-derived predictors of adult obesity and diabetes.<sup>a</sup>

Built environment characteristics	Log Poisson Regression for dichotomous outcomes		Linear regression for continuous outcomes		
	Obese	Diabetes	Body Mass Index (kg/m <sup>2</sup> )	Fasting glucose (mg/dL)	HbA1c (%)
	Prevalence Ratio (95% CI) <sup>b</sup>	Prevalence Ratio (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>
Non-single family home					
3rd tertile (highest)	0.85 (0.79, 0.91)	0.80 (0.70, 0.91)	-0.68 (-0.95, -0.40)	-1.51 (-3.24, 0.22)	-0.11 (-0.66, 0.43)
2nd tertile	0.96 (0.91, 1.02)	0.91 (0.81, 1.01)	-0.16 (-0.39, 0.07)	-1.20 (-2.44, 0.04)	-0.11 (-0.57, 0.36)
Visible utility wires					
3rd tertile (highest)	1.10 (1.05, 1.17)	1.15 (1.03, 1.28)	0.46 (0.27, 0.65)	2.13 (0.64, 3.63)	0.04 (-0.08, 0.16)
2nd tertile	1.06 (1.00, 1.12)	1.15 (1.04, 1.28)	0.26 (0.08, 0.44)	1.51 (0.28, 2.75)	0.10 (-0.06, 0.26)
Sidewalk					
3rd tertile (highest)	0.89 (0.85, 0.94)	0.98 (0.89, 1.09)	-0.53 (-0.76, -0.29)	-1.75 (-4.05, 0.54)	-0.18 (-0.49, 0.13)
2nd tertile	0.93 (0.88, 0.97)	0.97 (0.87, 1.08)	-0.35 (-0.58, -0.11)	-2.14 (-4.42, 0.14)	-0.15 (-0.32, 0.03)
Single lane road					
3rd tertile (highest)	1.01 (0.96, 1.06)	1.11 (0.98, 1.24)	0.07 (-0.13, 0.26)	1.24 (-0.14, 2.62)	0.02 (-0.22, 0.26)
2nd tertile	1.02 (0.98, 1.06)	1.15 (1.04, 1.28)	0.07 (-0.09, 0.23)	1.51 (0.61, 2.42)	0.05 (-0.26, 0.37)
Green street					
3rd tertile (highest)	0.88 (0.80, 0.97)	0.83 (0.71, 0.97)	-0.39 (-0.71, -0.07)	-0.60 (-2.05, 0.86)	0.01 (-0.36, 0.38)
2nd tertile	0.92 (0.88, 0.96)	0.96 (0.88, 1.05)	-0.36 (-0.53, -0.19)	-0.06 (-1.00, 0.87)	0.03 (-0.13, 0.19)
N	1,899,861	1,910,875	1,899,861	369,953	132,835

<sup>a</sup> Data source for health outcome: Utah Population Database and Intermountain Healthcare Enterprise Data Warehouse on Utah adults 20 years and older.

<sup>b</sup> Adjusted hierarchical linear regression models were run for each outcome separately, with zip code included as a random intercept. For dichotomous outcomes such as obesity and diabetes (0 = no; 1 = yes), log Poisson regression hierarchical models were utilized. For continuous variables like body mass index, linear regression hierarchical models were used. Models controlled for age, sex, race (white/nonwhite), Hispanic ethnicity, education, marital status as well as the following zip code area characteristics: population density, percent of the population 65 years and older, percent Hispanic, percent black, household income. Indicator variables were created for missing data on covariates. Built environment characteristics were categorized into tertiles, with the lowest tertile serving as the referent group.

representing associations between tertiles of zip code level built environment characteristics and the individual-level chronic diseases. For continuous outcomes, random intercept linear regression models were used. For siblings subsamples of approximately 1.0 million siblings and 14,000 (maternal and fraternal) twins, we implemented similar hierarchical models, but with siblings as random intercepts to account for family background characteristics shared between siblings and estimated robust standard errors. The study was approved by Institutional Review Boards at the University of Utah and University of Maryland.

### 3. Results

Table 1 displays the study’s descriptive statistics for the analytic sample. Built environment characteristics were summarized at the zip code level. Across zip codes, on average, about 25% of the images had a non-single-family home, while 20% of images had sidewalks. Utility wires suspended overhead were captured in 44% of the images. Single-lane roads and green streets were also prevalent and present in 65% and 87% of the images, respectively. The mean age of participants was 46 years old, with women making up half of the participants. Most participants were married (62%), identified as white (95%), and 11% classified themselves as of Hispanic ethnicity. Regarding educational background, 29% had achieved a college degree or more. For health outcomes, the prevalence rates for obesity and diabetes were 29% and 5%, respectively. eTable 4 provides data on exposure-discrepant siblings across the different siblings samples. Across these siblings samples, about 71%–75% of siblings live in different zip codes.

Figs. 2 and 3 display the geographical distribution of GSV-derived built environment and health outcomes across zip codes in Utah. The exurban areas along the Wasatch Range in the Salt Lake City–Provo–Ogden metropolitan area have large greenness values (Fig. 2 (a)). The urban areas in the Salt Lake City–Provo–Ogden metropolitan area and the eastern part of the Logan metropolitan area have more sidewalks (Fig. 2(b)). Non-single family homes are more concentrated in the rural areas in the western and eastern parts of the state (Fig. 2 (c)). Single-lane roads dominate the exurban areas in the Salt Lake City–Provo–Ogden metropolitan area and southwestern Utah (Fig. 2(d)). The rural areas across Utah have more visible wires (Fig. 2(e)). Diabetes

seems to have higher prevalence in the southwestern Utah (Fig. 3(b)). The western part of Utah has higher obesity and BMI prevalence (Fig. 3 (a) and (c)).

#### 3.1. Built environment and health outcomes

Using a null hierarchical logistic model that only includes zip code as a random intercept as the only predictor, we found that a small percent (3.5% VPC) of variance in obesity can be accounted for by zip code level differences. After adjusting for built environment exposures and covariates, zip code level variance is 1.4%, indicating that 61.47% proportional change in variance (PCV) at the zip code level can be accounted for by our built environment exposures and covariates. A similar pattern of findings is observed for the other outcome variables, such that a small percentage of variance of the outcome is at the zip code-level but that much of this variance can be attributed to our built environment exposures and covariates (eTable 5).

We further explored the associations between the built environment and specific health outcomes. Table 2 presents regression results for the full analytic sample, examining the relationship between the built environment and health outcomes. Notably, the presence of non-single-family homes, sidewalks, and green streets were associated with reduced chronic health conditions. For example, zip codes in the third tertile for non-single-family homes, an indicator of mixed land use, had a 15% reduction (PR: 0.85; 95% CI: 0.79, 0.91) in obesity and a 20% reduction (PR: 0.80; 95% CI: 0.70, 0.91) in diabetes. Non single-family homes, sidewalks, and green streets were associated with reductions in BMI. For example, zip codes in the third tertile of non-single-family homes were associated with a BMI reduction of -0.68 kg/m<sup>2</sup> (95% CI: -0.95, -0.40) (Table 2). Table 2 only displays main associations. The full table of results that display associations between zip code level- and individual level covariates and health outcomes are displayed in eTable 6. Higher zip code level median income was associated with lower individual-level obesity and diabetes. Non-white individual-level race was associated with higher diabetes (eTable 6).

In sensitivity analyses, we implemented multilevel models for obesity and compared across data sources: 1) driver license only, 2) driver licenses and clinical records, 3) clinical records only (eTable 7).

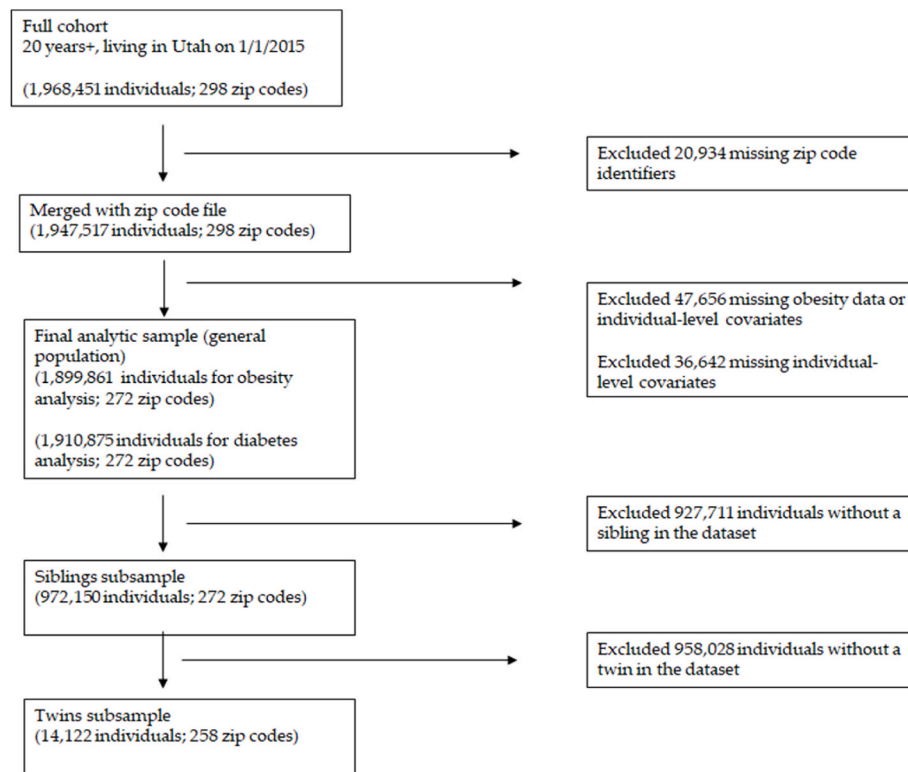


Fig. 1. Path diagram describing final analytic sample and exclusions.

While clinically derived BMI (mean (sd) = 28.7 (7.0)) were higher than those reported on driver licenses (mean (sd) = 26.4 (5.6)), multilevel regression results across the data sources show very similar associations with GSV-derived neighborhood variables, with changes in the second decimal place; for example PR = 0.86 (95% CI: (0.79, 0.94)) vs. PR = 0.85 (95% CI: 0.80, 0.91) for the association between the third tertile of non-single family home for driver license data only and clinical data only, respectively (eTable 7). In additional sensitivity analyses, we controlled for the length of time spent in the most recent neighborhood (eTable 8). Mean time spent in the current neighborhood was 2.8 years (Q1-Q3 range was 1.7 years–3.8 years). Results were almost identical controlling and not controlling for time spent in most recent residence.

In addition, analyses examined fasting glucose and lower glycated hemoglobin (HbA1c) as these outcomes were only available on a subset of participants. More sidewalks were associated with lower fasting glucose (PR for 3rd tertile vs 1st:  $-1.75$  mg/dL; 95% CI:  $-4.05, 0.54$ ) and lower glycated hemoglobin (HbA1c) (PR for 3rd tertile vs 1st  $-0.18$ ; 95% CI:  $-0.49, 0.13$ ), although these bordered statistical significance (Table 2).

### 3.2. Siblings sharing biological mother

To account for the influence of shared genetic and familial factors, we conducted analyses restricted to siblings sharing a biological mother. Across health outcomes, the pattern of estimated effects remains consistent with non-single family homes, sidewalks, and green streets being associated with lower individual-level obesity and diabetes and visible utility wires and single lane roads associated with increased obesity and diabetes, although the magnitude of associations was slightly attenuated when controlling for shared family characteristics of siblings (Table 3). For instance, among the subsample of siblings sharing a biological mother, zip codes in the highest tertile of non-single family homes had 14% (PR: 0.86, 95% CI: 0.84, 0.88) and 23% (PR: 0.77; 95% CI: 0.73, 0.82) lower obesity and diabetes, respectively compared to 15% and 20% in the full analytic sample. In the siblings' sample, zip

codes abundant with green streets had reductions of 18% (PR: 0.82; 95% CI: 0.80, 0.84) and 21% (PR: 0.79; 95% CI: 0.74, 0.83) in obesity and diabetes, respectively. In contrast, zip codes with a higher presence of single-lane roads showed a 7% (PR: 1.07; 95% CI: 1.05, 1.09) and 8% (PR: 1.08; 95% CI: 1.04, 1.13) increase in obesity and diabetes, respectively.

Residents living in zip codes with more non-single-family homes, sidewalks, and green streets had lower BMI and glycated hemoglobin (HbA1c). Utility wires were associated with increased BMI and fasting glucose. Single-lane roads were associated with increased BMI, fasting glucose, and HbA1c. Similar associations were seen for individuals sharing biological mothers and fathers (eTable 9) and those sharing biological fathers (eTable 10).

### 3.3. Twins in the analytic sample

To further refine the analysis and explore genetically similar groups, we restricted the analytic sample to twins (identical and fraternal) and further explored these associations (Table 4). Despite the reduced sample size of 14,122 twins, estimated protective effects for non-single-family homes and green streets on obesity and diabetes were similar in magnitudes to the full analytic sample. Single-lane roads were associated with increased diabetes (PR: 1.46; 95% CI: 0.98, 2.17). Visible utility wires and sidewalks were no longer associated with obesity and diabetes. Green streets were associated with reductions in BMI (PR:  $-0.71$  kg/m<sup>2</sup>; 95% CI:  $-1.33, -0.08$ ), and single-lane roads were associated with increases in HbA1c (PR: 0.32, 0.04, 0.61).

## 4. Discussion

The current study conducted a one-of-a-kind, state-wide investigation into the relationship between built environments and health that utilized family background information to account for genetic and shared environment effects. We examined records from 1.9 million individuals, which included 1.0 million siblings and 14,000 twins in Utah,

**Table 3**  
Sibling (shared biological mother<sup>a</sup>) random effects model: Google Street View-derived predictors of individual health outcomes.

Built environment characteristics	Log Poisson Regression for dichotomous outcomes		Linear regression for continuous outcomes		
	Obese	Diabetes	Body Mass Index (kg/m <sup>2</sup> )	Fasting glucose (mg/dL)	HbA1c (%)
	Prevalence Ratio (95% CI) <sup>b</sup>	Prevalence Ratio (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>
Non-single family home					
3rd tertile (highest)	0.86 (0.84, 0.88)	0.77 (0.73, 0.82)	-0.63 (-0.70, -0.56)	0.38 (-0.83, 1.60)	0.06 (0.01, 0.11)
2nd tertile	0.92 (0.90, 0.93)	0.83 (0.80, 0.87)	-0.33 (-0.39, -0.27)	-0.09 (-1.14, 0.96)	0.04 (0.00, 0.08)
Visible utility wires					
3rd tertile (highest)	1.04 (1.03, 1.06)	1.02 (0.98, 1.07)	0.16 (0.10, 0.21)	1.02 (0.01, 2.03)	0.00 (-0.05, 0.06)
2nd tertile	1.01 (1.00, 1.03)	1.01 (0.97, 1.05)	0.01 (-0.04, 0.06)	0.54 (-0.33, 1.40)	-0.01 (-0.05, 0.04)
Sidewalk					
3rd tertile (highest)	0.91 (0.88, 0.93)	0.96 (0.89, 1.03)	-0.41 (-0.52, -0.30)	-1.69 (-4.40, 1.03)	-0.18 (-0.29, -0.07)
2nd tertile	0.93 (0.90, 0.96)	0.95 (0.88, 1.03)	-0.35 (-0.47, -0.23)	-1.55 (-4.36, 1.27)	0.04 (-0.10, 0.18)
Single lane road					
3rd tertile (highest)	1.07 (1.05, 1.09)	1.08 (1.04, 1.13)	0.26 (0.20, 0.32)	1.20 (0.15, 2.25)	0.09 (0.05, 0.13)
2nd tertile	1.05 (1.04, 1.07)	1.12 (1.08, 1.16)	0.21 (0.16, 0.26)	1.23 (0.30, 2.15)	0.07 (0.04, 0.11)
Green street					
3rd tertile (highest)	0.82 (0.80, 0.84)	0.79 (0.74, 0.83)	-0.83 (-0.90, -0.75)	-0.16 (-1.45, 1.14)	-0.07 (-0.12, -0.01)
2nd tertile	0.91 (0.90, 0.92)	0.94 (0.91, 0.96)	-0.36 (-0.40, -0.32)	0.67 (-0.01, 1.35)	-0.04 (-0.07, -0.01)
N	972,150	972,150	972,150	65,502	181,180

<sup>a</sup> Data source for health outcome: Utah Population Database and Intermountain Healthcare Enterprise Data Warehouse on Utah adults 20 years and older. Analyses restricted to siblings with a shared biological mother.

<sup>b</sup> Siblings random effects models were run to account for shared family background characteristics. Adjusted regression models were run for each outcome separately. For dichotomous outcomes such as obesity and diabetes (0 = no; 1 = yes), log Poisson models were utilized. For continuous variables like body mass index, linear regression was used. Models controlled for age, sex, race (white/nonwhite), Hispanic ethnicity, education, and marital status as well as the following zip code area characteristics: population density, percent of the population 65 years and older, percent Hispanic, percent black, median household income. Built environment characteristics were categorized into tertiles, with the lowest tertile serving as the referent group.

to evaluate associations between neighborhood characteristics and diabetes and obesity prevalence. Across all three samples, our results consistently found associations between chronic disease burden and neighborhood environment’s characteristics including indicators of walkability, physical disorder, and urban development. Our findings suggest that built environment characteristics play an important role in chronic disease burden even when accounting for familial risk factors.

Across the study samples, the presence of non-single-family homes and green streets were consistently associated with reduced obesity and

diabetes rates. We used non-single-family homes as an indicator of mixed land use, representing areas where residential and commercial entities coexist. Mixed land use is related to more non-vehicular commuting (Cervero, 1996) and walking (Cervero & Kockelman, 1997; Colom et al., 2021), which often relates to healthier BMIs (Brown et al., 2009; Colom et al., 2021). Additionally, these results suggest that urban development could potentially reduce the chronic disease burden through increasing walkability and access to foods (Giles-Corti et al., 2016; Luo & Wang, 2022).

**Table 4**  
Twins<sup>a</sup> random effects model: Google Street View-derived predictors of individual health outcomes.

Built environment characteristics	Log Poisson Regression for dichotomous outcomes		Linear regression for continuous outcomes		
	Obese	Diabetes	Body Mass Index (kg/m <sup>2</sup> )	Fasting glucose (mg/dL)	HbA1c (%)
	Prevalence Ratio (95% CI) <sup>b</sup>	Prevalence Ratio (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>	Beta (95% CI) <sup>b</sup>
Non-single family home					
3rd tertile (highest)	0.75 (0.62, 0.91)	0.81 (0.49, 1.32)	-0.38 (-0.94, 0.18)	-4.14 (-14.47, 6.19)	0.16 (-0.20, 0.52)
2nd tertile	0.77 (0.66, 0.91)	0.95 (0.62, 1.44)	-0.27 (-0.76, 0.21)	-1.04 (-10.67, 8.59)	0.16 (-0.13, 0.44)
Visible utility wires					
3rd tertile (highest)	0.97 (0.83, 1.13)	1.14 (0.75, 1.72)	-0.05 (-0.51, 0.41)	-2.07 (-9.31, 5.16)	-0.16 (-0.45, 0.14)
2nd tertile	0.94 (0.82, 1.08)	1.05 (0.73, 1.51)	-0.23 (-0.62, 0.17)	-1.14 (-7.38, 5.09)	-0.17 (-0.42, 0.08)
Sidewalk					
3rd tertile (highest)	0.97 (0.71, 1.31)	1.14 (0.53, 2.43)	-0.56 (-1.49, 0.37)	-4.21 (-13.00, 4.57)	-0.01 (-0.65, 0.63)
2nd tertile	1.00 (0.72, 1.38)	0.99 (0.44, 2.20)	-0.54 (-1.55, 0.47)	-4.92 (-15.46, 5.62)	-0.03 (-0.79, 0.73)
Single lane road					
3rd tertile (highest)	1.01 (0.87, 1.18)	1.46 (0.98, 2.17)	0.36 (-0.09, 0.81)	2.14 (-3.49, 7.77)	0.32 (0.04, 0.61)
2nd tertile	1.08 (0.95, 1.23)	1.22 (0.86, 1.73)	0.52 (0.14, 0.91)	0.22 (-4.40, 4.84)	0.26 (-0.04, 0.55)
Green street					
3rd tertile (highest)	0.78 (0.63, 0.96)	0.82 (0.48, 1.39)	-0.71 (-1.33, -0.08)	-3.69 (-16.17, 8.78)	-0.30 (-0.63, 0.03)
2nd tertile	0.88 (0.80, 0.98)	0.97 (0.75, 1.25)	-0.40 (-0.68, -0.11)	0.83 (-3.34, 5.00)	-0.06 (-0.24, 0.13)
N	14,122	14,151	14,122	835	2307

<sup>a</sup> Data source for health outcome: Utah Population Database and Intermountain Healthcare Enterprise Data Warehouse on Utah adults 20 years and older. Analyses restricted to fraternal and maternal twins.

<sup>b</sup> Siblings random effects models account for shared family background characteristics. Adjusted regression models were run for each outcome separately. For dichotomous outcomes such as obesity and diabetes (0 = no; 1 = yes), log Poisson models were utilized. For continuous variables like body mass index, linear regression was used. Models controlled for age, sex, race (white/nonwhite), Hispanic ethnicity, education, and marital status as well as the following zip code area characteristics: population density, percent of the population 65 years and older, percent Hispanic, percent black, median household income. Built environment characteristics were categorized into tertiles, with the lowest tertile serving as the referent group.

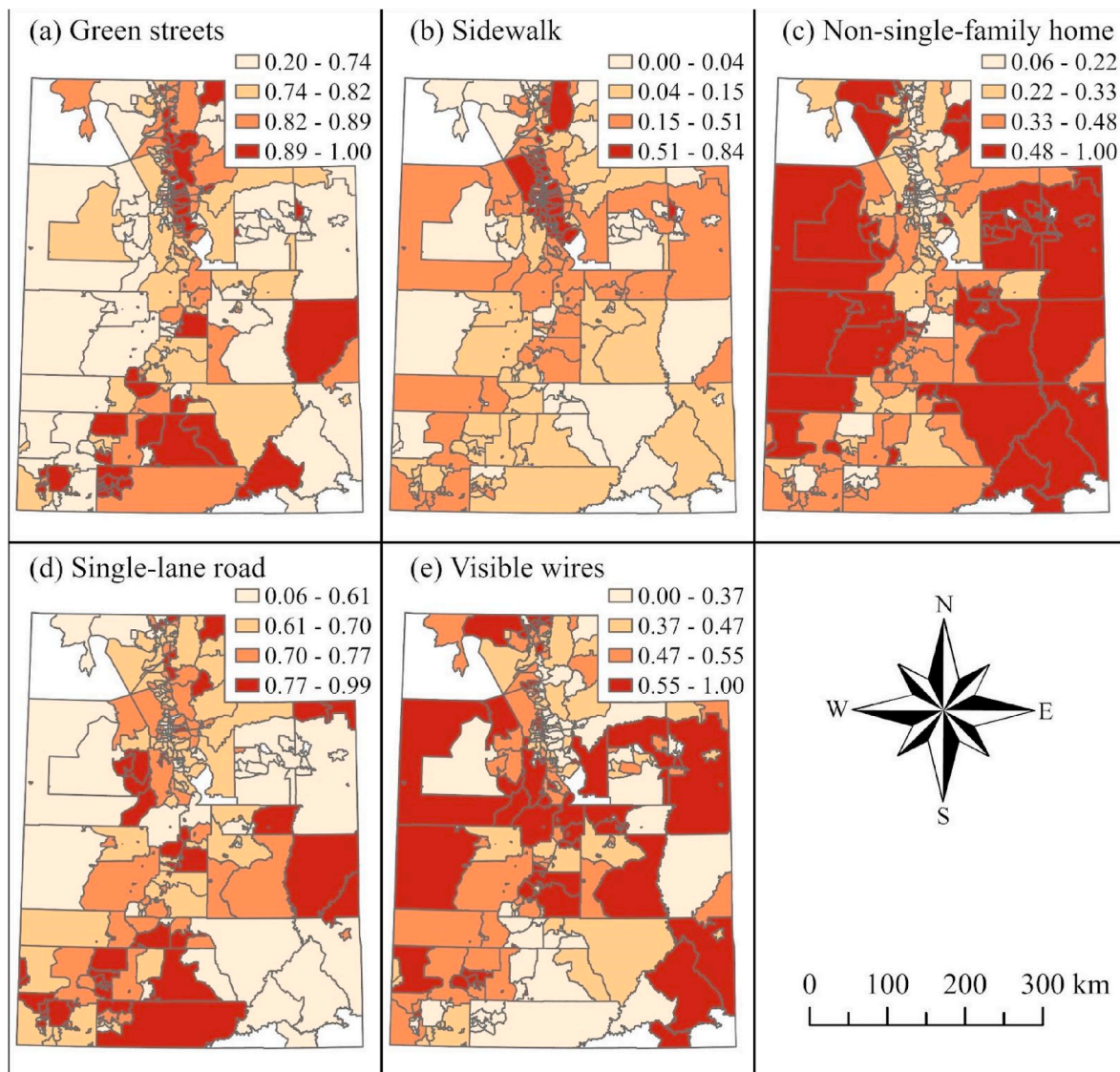


Fig. 2. Zip code distribution of Google Street View-derived neighborhood built environments.

Employing (identical and fraternal) twin data allows the examination of associations between built environment characteristics and chronic health conditions, controlling for shared genetic factors and familial backgrounds. In this data, both non-single-family homes and green streets were inversely associated with obesity and diabetes. Specifically, neighborhoods with more mixed-use properties had 15% and 20% lower obesity and diabetes, respectively, while green streets were associated with a 12% and a 17% decrease in these health outcomes. These trends were also observed in the sibling dataset. In contrast, single-lane roads (an indicator of lower urban development) and overhead utility wires (an indicator of physical disorder) were associated with increased chronic disease. By demonstrating associations between health and novel neighborhood characteristics like mixed-use properties and green streets, our results extend the findings of other US-based studies reporting associations between built-environment factors and obesity, even when adjusting for familial risk (Kowaleski-Jones et al., 2017), as well as international work showing associations with coronary heart disease (Forsberg et al., 2018).

Moreover, environmental factors can potentially influence disease risk before reaching clinical thresholds (Babić Leko et al., 2021; Ladd-Acosta & Fallin, 2016). We examined lab values for BMI, fasting glucose, and HbA1c. The presence of sidewalks was associated with

improved BMI, fasting glucose, and HbA1c levels suggests positive health impacts of walkable neighborhoods. This demonstrates the sensitivity of our built environment measures for capturing neighborhood characteristics that relate to health, even at sub-clinical levels, and highlights the potential for using computer vision approaches for public health research examining the role of built environment on health.

#### 4.1. Study strengths and limitations

The incorporation of both twin and sibling data helps mitigate the effects of confounding factors, allowing for stronger causal inference. Our comprehensive dataset, encompassing approximately 2 million individuals which included 1 million siblings, is another strength. UPDB has partnerships with a multitude of data providers to link vital records, health facilities and claims records, driver license data, and census data, enabling a comprehensive examination of characteristics of interest. Furthermore, the inclusion of disease biomarkers such as fasting glucose, HbA1c, and BMI, allows for more examination of pre-clinical conditions.

However, this study has some limitations. Given that our dataset is exclusively drawn from Utah, the findings may not generalize to other regions whose populations may present different health behaviors and

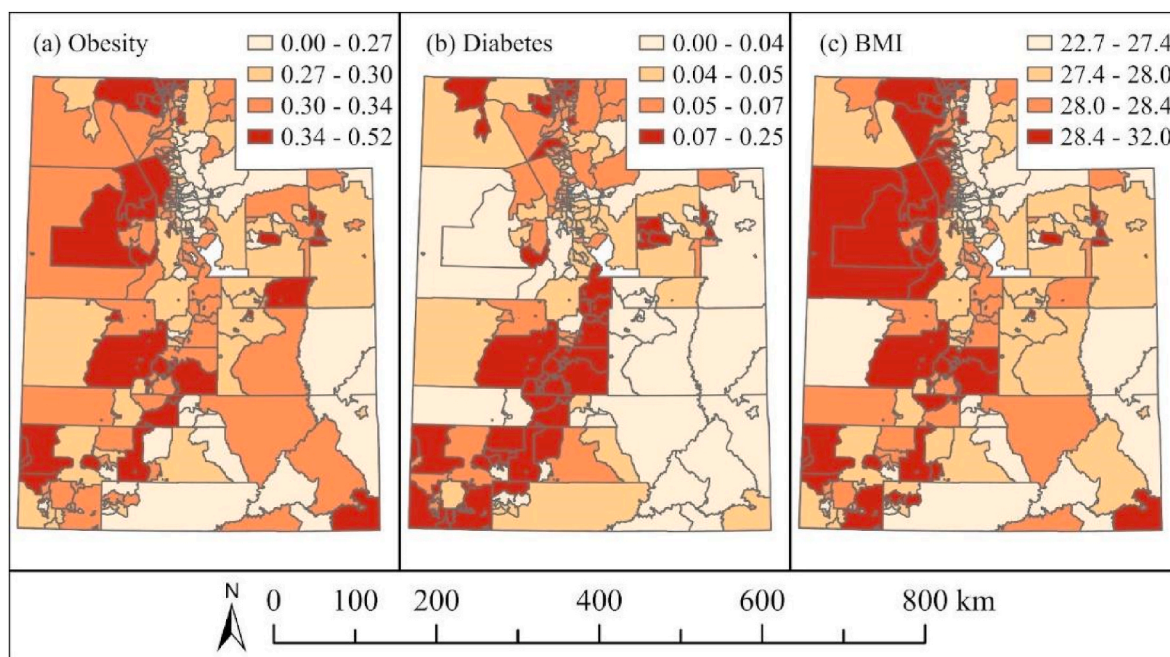


Fig. 3. Zip code distribution of health outcomes.

population characteristics. We do not have information on the extent of their engagement with their neighborhood or community patterns. The observational nature of neighborhood studies leaves room for unmeasured confounding biases including self-selection into neighborhoods (Oakes, 2004; Robbins et al., 2020).

Zip codes as neighborhood boundaries pose a myriad of issues including their coarseness and variation in size (Krieger et al., 2002). Prior research has demonstrated paradoxical effects when using larger geographic areas for investigating health impacts, such that when the general geographic context has little relevance to individual-level outcomes, specific effects of the context (i.e., our built environment variables) can be estimated more precisely, which increases statistical significance (Merlo et al., 2018, 2019). Indeed, except for HbA1c (VPC = 23%), the VPCs of each of our hierarchical models nesting individuals within zip code were relatively small (0.5%–5.3%), suggesting that general zip code effects are small and may contribute to a misleading picture of the role of built environment characteristics (Merlo et al., 2012, 2013). Although the variance associated within zip code similarity is relatively small, this does not necessarily invalidate our findings when considering how much the contributions of the zip code- and individual-level factors account for the effects we report. That is, even though other social factors (e.g., strata at the intersections of SES, multiple race/ethnicity categories, and other social identity dimensions) play an important role in determining health, our analyses found small but meaningful environmental effects that explained the health outcomes. Moreover, although we did not have more granular intersectional strata measures for some of these social factors, we did include relevant individual level covariates (i.e., gender, education, Black/White race, and Hispanic ethnicity) as fixed effects that were nested within zip codes so that we can better account for their main effects in our examination of the built environment characteristics.

Research has found zip code characteristics to be associated with health (Orminski, 2021; Thomas et al., 2006). An analysis of the NYC Community Health Survey found valuable heterogeneity in health outcomes at the zip code level to inform local interventions, although pooling across 5-years of data was necessary to produce reliable small-area estimates (Bi et al., 2020). A study comparing zip code- and census-based income measures of long-term mortality found both were similar and significant predictors across various causes of death

(Thomas et al., 2006). Zip codes as boundaries are easily understood and data contributors such as hospitals and administrative data using residential addresses have zip code data readily available. Converting residential addresses to block groups or census tracts, alternative neighborhood boundaries, requires geocoding expertise, which is not always available among research teams and hospitals. Additionally, health surveys that have interest in neighborhood locations may opt to ask for residential zip codes rather than exact addresses to protect participants' identities.

Nonetheless, researchers have identified issues with zip codes for research. Originally established by the United States Postal Service for efficient mail delivery, zip codes do not align with census tracts and blocks groups, which are designed to be relatively homogeneous with respect to population characteristics (Krieger et al., 2002). Additionally, there can be important spatiotemporal mismatches between zip codes and census-derived ZIP code tabulation areas (ZCTAs) (Grubestic & Matisziw, 2006). Thus, future work could build off the current findings to examine how built environment factors at other geographic levels contribute to health outcomes.

Our data was available prior to the COVID-19 pandemic, and it remains unclear how built environments will impact health outcomes after the arrival of COVID-19. There was also a temporal mismatch between the Google Street View data collection (2019) and assembly of the Utah cohort (2015). For some neighborhoods, built environment characteristics may have remained relatively unchanged and for others, new developments may cause a mismatch between time of exposure assessment (built environment) and health outcome assessment. Additionally, a substantial proportion of BMI data (about 58%) came from self-reported information found on drivers' licenses, which may not be as accurate or updated as clinical measurements. Clinical data were only available if the healthcare facility was a UPDB data contributor, which included major healthcare providers in the state such as the University of Utah and Intermountain Healthcare. Nonetheless, sensitivity analyses restricting to only clinical data, yielded similar results—which are in line with a previous study that found that the two sources of BMI (drivers' licenses and clinical records) when used as predictors of type 2 diabetes yielded similar risk. This is because, while individuals tend to overestimate their height and underestimate their weight, accurate categorization of obesity is still possible (Chernenko et al., 2019).



Additionally, covariate information was sourced from administrative records and medical records from UPDB. As such these records might not contain the latest or most complete information as the data update frequency depends on how frequently the individuals interact with the healthcare system or administrative agencies, and the willingness of data contributors to share data and provide updated information (DuVall et al., 2012). Nonetheless, the breadth and depth of data available from data partners participating in the UPDB provides an invaluable and cost-efficient resource for studying population health and family histories (Smith et al., 2022). Moreover, while we adjusted for individual- and neighborhood-level sociodemographic characteristics in analyses, we did not adjust for residential segregation, zoning laws, and other potentially relevant contextual characteristics.

In the United States, residential segregation is prevalent (Massey et al., 1994; Quillian, 2012). Notably, residential segregation by household income has grown substantially over the past decades, with Black and Hispanic families in particular living in increasingly income-segregated communities (Bischoff & Reardon, 2014; Watson, 2009). This has potential implications for inequities in the spatial distribution of resources. Examining neighborhoods in Seattle, San Diego and Baltimore, Thornton and colleagues found neighborhoods with lower socioeconomic status and higher proportions of racial/ethnic minority groups had more unmaintained buildings, graffiti, broken windows, and litter (Thornton et al., 2016). Residential segregation can impact access to quality education, employment opportunities, and increase exposures to physical and chemical hazards (Williams et al., 2019). It can also have neighborhood composition effects that can influence distribution of role models, peers, and social networks (Watson, 2009) and create areas of concentrated poverty and low quality housing stock (Williams et al., 2019). Racial segregation is a widely recognized institutional mechanism by which racism impacts health and health disparities (Williams et al., 2019). The challenge lies in separating the effects of person-level characteristics from neighborhood characteristics, because the sorting of people into neighborhoods is not random and can be dictated by people's ability to afford housing in a particular area, conditional on preferences and other social and structural facilitators and barriers (Swope et al., 2022).

## 5. Conclusions

Through our study using the Utah Population Database, we explored the associations between neighborhood characteristics and health outcomes (obesity, diabetes, BMI, fasting glucose, and HbA1c), accounting for the influence of shared familial backgrounds among siblings and twins. Although the variance in health outcomes attributable to the zip code level was modest, enhancements in neighborhoods that foster physical activity, recreation, and easier access to resources may be one of many strategies to improve population health. Future studies incorporating longitudinal analyses to examine changes in neighborhood environments and corresponding changes in health outcomes can further strengthen causal inferences about neighborhood effects. The UPDB is a unique dataset that allows for studying individual, family, and neighborhood influences on health. Continued support of population-based data sources that enable multidimensional understanding of interlocking factors affecting health could better inform health interventions and health policy.

## Funding

Research reported in this publication was supported by the National Library of Medicine under Award Number R01LM012849 (Q.C.N.) and National Institute on Minority Health and Health Disparities under Award Numbers R01MD016037 (Q.C.N.), R00MD012615 (T.T.N.) and R01MD015716 (T.T.N.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data

collection and analysis, decision to publish, or preparation of the manuscript.

Partial support for all datasets within the Utah Population Database was provided by the University of Utah Huntsman Cancer Institute and the Huntsman Cancer Institute Cancer Center Support grant, P30 CA2014 from the National Cancer Institute.

## Ethics statement

The University of Utah and University of Maryland's Institutional Review Board approved the study.

## Declaration of competing interest

The authors declare no conflicts of interest.

## CRediT authorship contribution statement

**Quynh C. Nguyen:** Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Tolga Tasdizen:** Writing – review & editing, Supervision, Resources, Conceptualization. **Mitra Alirezaei:** Writing – review & editing, Visualization, Formal analysis. **Heran Mane:** Writing – review & editing, Formal analysis. **Xiaohe Yue:** Writing – review & editing, Formal analysis. **Junaid S. Merchant:** Writing – review & editing. **Weijun Yu:** Writing – review & editing, Writing – original draft. **Laura Drew:** Writing – review & editing. **Dapeng Li:** Writing – review & editing, Formal analysis. **Thu T. Nguyen:** Writing – review & editing, Supervision, Funding acquisition.

## Data availability

The authors do not have permission to share data.

## Acknowledgements

We would like to thank Nataly Delcid, Isabelle Yang, Katrina Makres, and Melanie Kim for their help with the manuscript. We would like to acknowledge Dr. Clare Evans for her help in guiding the variance estimation. Partial support for all datasets within the Utah Population Database was provided by the University of Utah Huntsman Cancer Institute and the Huntsman Cancer Institute Cancer Center Support grant, P30 CA2014 from the National Cancer Institute. We thank the University of Utah Clinical and Translational Science Institute (CTSI) (funded by NIH Clinical and Translational Science Awards), the Pedigree and Population Resource, University of Utah Information Technology Services and Biomedical Informatics Core for establishing the Master Subject Index between the Utah Population Database, the University of Utah Health Sciences Center and Intermountain Healthcare.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ssmph.2024.101670>.

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). *TensorFlow: A system for large-scale machine learning*.
- Babić Leko, M., Gunjača, I., Pleić, N., & Zemunik, T. (2021). Environmental factors affecting thyroid-stimulating hormone and thyroid hormone levels. *International Journal of Molecular Sciences*, 22(12). <https://doi.org/10.3390/ijms22126521>. Article 12.
- Baskin, M. L., Ard, J., Franklin, F., & Allison, D. B. (2005). Prevalence of obesity in the United States. *Obesity Reviews*, 6(1), 5–7. <https://doi.org/10.1111/j.1467-789X.2005.00165.x>

- Beyer, F., Munte, T. F., Erdmann, C., & Krämer, U. M. (2014). Emotional reactivity to threat modulates activity in mentalizing network during aggression. *Social Cognitive and Affective Neuroscience*, 9(10), 1552–1560. <https://doi.org/10.1093/scan/nst146>
- Bi, Q., He, F., Konty, K., Gould, L. H., Immerwahr, S., & Levanon Seligson, A. (2020). ZIP code-level estimates from a local health survey: Added value and limitations. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 97(4), 561–567. <https://doi.org/10.1007/s11524-020-00423-z>
- Bischoff, K., & Reardon, S. F. (2014). *Residential segregation by income, 1970–2009: Vol. Diversity and disparities: America enters a new century* (43rd ed.).
- Brown, B. B., Yamada, I., Smith, K. R., Zick, C. D., Kowaleski-Jones, L., & Fan, J. X. (2009). Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. *Health & Place*, 15(4), 1130–1141. <https://doi.org/10.1016/j.healthplace.2009.06.008>
- Burton, E. J., Mitchell, L., & Stride, C. B. (2011). Good places for ageing in place: Development of objective built environment measures for investigating links with older people's wellbeing. *BMC Public Health*, 11(1), 839. <https://doi.org/10.1186/1471-2458-11-839>
- Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American housing survey. *Transportation Research Part A: Policy and Practice*, 30(5), 361–377. [https://doi.org/10.1016/0965-8564\(95\)00033-X](https://doi.org/10.1016/0965-8564(95)00033-X)
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3), 199–219. [https://doi.org/10.1016/S1361-9209\(97\)00009-6](https://doi.org/10.1016/S1361-9209(97)00009-6)
- Chandrabose, M., Rachele, J. N., Gunn, L., Kavanagh, A., Owen, N., Turrell, G., Giles-Corti, B., & Sugiyama, T. (2019). Built environment and cardio-metabolic health: Systematic review and meta-analysis of longitudinal studies. *Obesity Reviews*, 20(1), 41–54. <https://doi.org/10.1111/obr.12759>
- Chernenko, A., Meeks, H., & Smith, K. R. (2019). Examining validity of body mass index calculated using height and weight data from the US driver license. *BMC Public Health*, 19(1), 100. <https://doi.org/10.1186/s12889-019-6391-3>
- Christensen, K., Holt, J. M., & Wilson, J. F. (2010). Effects of perceived neighborhood characteristics and use of community facilities on physical activity of adults with and without disabilities. *Preventing Chronic Disease*, 7(5), A105. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938399/>
- Christine, P. J., Auchincloss, A. H., Bertoni, A. G., Carnethon, M. R., Sánchez, B. N., Moore, K., Adar, S. D., Horwich, T. B., Watson, K. E., & Diez Roux, A. V. (2015). Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: The multi-ethnic study of atherosclerosis (MESA). *JAMA Internal Medicine*, 175(8), 1311–1320. <https://doi.org/10.1001/jamainternmed.2015.2691>
- Cohen, N. (2019). SNAP at the community scale: How neighborhood characteristics affect participation and food access. *American Journal of Public Health*, 109(12), 1646–1651. <https://doi.org/10.2105/AJPH.2019.305363>
- Cohen-Cline, H., Turkheimer, E., & Duncan, G. E. (2015). Access to green space, physical activity and mental health: A twin study. *Journal of Epidemiology & Community Health*. <https://doi.org/10.1136/jech-2014-204667>
- Coleman, K. J., Rosenberg, D. E., Conway, T. L., Sallis, J. F., Saelens, B. E., Frank, L. D., & Cain, K. (2008). Physical activity, weight status, and neighborhood characteristics of dog walkers. *Preventive Medicine*, 47(3), 309–312. <https://doi.org/10.1016/j.ypmed.2008.05.007>
- Colom, A., Mavoa, S., Ruiz, M., Wärnberg, J., Muncunill, J., Konieczna, J., Vich, G., Barón-López, F. J., Fitó, M., Salas-Salvado, J., & Romaguera, D. (2021). Neighbourhood walkability and physical activity: Moderating role of a physical activity intervention in overweight and obese older adults with metabolic syndrome. *Age and Ageing*, 50(3), 963–968. <https://doi.org/10.1093/ageing/afaa246>
- Creator, M. I., Glazier, R. H., Moineddin, R., Fazli, G. S., Johns, A., Gozdyra, P., Matheson, F. I., Kaufman-Shriqui, V., Rosella, L. C., Manuel, D. G., & Booth, G. L. (2016). Association of neighborhood walkability with change in overweight, obesity, and diabetes. *JAMA*, 315(20), 2211–2220. <https://doi.org/10.1001/jama.2016.5898>
- Cummins, S. K., & Jackson, R. J. (2001). The built environment and CHILDREN'S health. *Pediatric Clinics of North America*, 48(5), 1241–1252. [https://doi.org/10.1016/S0031-3955\(05\)70372-2](https://doi.org/10.1016/S0031-3955(05)70372-2)
- Duncan, G. E., Avery, A. A., Hurvitz, P., Vernez-Moudon, A., & Tsang, S. (2022). Cross-sectional associations between neighbourhood walkability and objective physical activity levels in identical twins. *BMJ Open*, 12(11), Article e064808. <https://doi.org/10.1136/bmjopen-2022-064808>
- Duncan, G. E., Sun, F., Avery, A. R., Hurvitz, P. M., Moudon, A. V., Tsang, S., & Williams, B. D. (2023). Cross-Sectional study of location-based built environments, physical activity, dietary intake, and body mass index in adult twins. *International Journal of Environmental Research and Public Health*, 20(6). <https://doi.org/10.3390/ijerph20064885>. Article 6.
- DuVall, S. L., Fraser, A. M., Rowe, K., Thomas, A., & Mineau, G. P. (2012). Evaluation of record linkage between a large healthcare provider and the Utah Population Database. *Journal of the American Medical Informatics Association: JAMIA*, 19(e1), e54–e59. <https://doi.org/10.1136/amiajnl-2011-000335>
- Eibich, P., Krekel, C., Demuth, I., & Wagner, G. G. (2016). Associations between neighborhood characteristics, well-being and health vary over the life course. *Gerontology*, 62(3), 362–370. <https://doi.org/10.1159/000438700>
- Ellaway, A., Macintyre, S., & Bonnefoy, X. (2005). Graffiti, greenery, and obesity in adults: Secondary analysis of European cross sectional survey. *BMJ*, 331(7517), 611–612. <https://doi.org/10.1136/bmj.38575.664549.F7>
- Engelgau, M. M., Geiss, L. S., Saaddine, J. B., Boyle, J. P., Benjamin, S. M., Gregg, E. W., Tierney, E. F., Rios-Burrows, N., Mokdad, A. H., Ford, E. S., Imperatore, G., & Venkat Narayan, K. M. (2004). The evolving diabetes burden in the United States. *Annals of Internal Medicine*, 140(11), 945–950. <https://doi.org/10.7326/0003-4819-140-11-200406010-00035>
- Evans, C. R., & Erickson, N. (2019). Intersectionality and depression in adolescence and early adulthood: A mailda analysis of the national longitudinal study of adolescent to adult health, 1995–2008. *Social Science & Medicine*, 220, 1–11. <https://doi.org/10.1016/j.socscimed.2018.10.019>
- Figarola, M. N. S., Gielen, M., Casas, L., Loos, R. J. F., Derom, C., Weyers, S., Nawrot, T. S., Zeegers, M. P., & Bijmens, E. M. (2023). Early-life residential green spaces and traffic exposure in association with young adult body composition: A longitudinal birth cohort study of twins. *Environmental Health*, 22(1), 18. <https://doi.org/10.1186/s12940-023-00964-1>
- Forsberg, P. O., Ohlsson, H., & Sundquist, K. (2018). Causal nature of neighborhood deprivation on individual risk of coronary heart disease or ischemic stroke: A prospective national Swedish co-relative control study in men and women. *Health & Place*, 50, 1–5. <https://doi.org/10.1016/j.healthplace.2017.12.006>
- Franco, M., Roux, A. V. D., Glass, T. A., Caballero, B., & Brancati, F. L. (2008). Neighborhood characteristics and availability of healthy foods in Baltimore. *American Journal of Preventive Medicine*, 35(6), 561–567. <https://doi.org/10.1016/j.amepre.2008.07.003>
- Frank, L. D., Adhikari, B., White, K. R., Dummer, T., Sandhu, J., Demlow, E., Hu, Y., Hong, A., & Van den Bosch, M. (2022). Chronic disease and where you live: Built and natural environment relationships with physical activity, obesity, and diabetes. *Environment International*, 158, Article 106959. <https://doi.org/10.1016/j.envint.2021.106959>
- Gary, T. L., Stark, S. A., & LaVeist, T. A. (2007). Neighborhood characteristics and mental health among African Americans and whites living in a racially integrated urban community. *Health & Place*, 13(2), 569–575. <https://doi.org/10.1016/j.healthplace.2006.06.001>
- Giles-Corti, B., Vernez-Moudon, A., Reis, R., Turrell, G., Dannenberg, A. L., Badland, H., Foster, S., Lowe, M., Sallis, J. F., Stevenson, M., & Owen, N. (2016). City planning and population health: A global challenge. *The Lancet*, 388(10062), 2912–2924. [https://doi.org/10.1016/S0140-6736\(16\)30066-6](https://doi.org/10.1016/S0140-6736(16)30066-6)
- Gomez, S. L., Shariff-Marco, S., DeRouen, M., Keegan, T. H. M., Yen, I. H., Mujahid, M., Satariano, W. A., & Glaser, S. L. (2015). The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. *Cancer*, 121(14), 2314–2330. <https://doi.org/10.1002/ncr.29345>
- Grubestic, T. H., & Matisziw, T. C. (2006). On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International Journal of Health Geographics*, 5(1), 58. <https://doi.org/10.1186/1476-072X-5-58>
- Hystad, P., Amram, O., Oje, F., Larkin, A., Boakye, K., Avery, A., Gebremedhin, A., & Duncan, G. (2022). Bring your own location data: Use of Google smartphone location history data for environmental health research. *Environmental Health Perspectives*, 130(11), Article 117005. <https://doi.org/10.1289/EHP10829>
- Kanchi, R., Lopez, P., Rummo, P. E., Lee, D. C., Adhikari, S., Schwartz, M. D., Avramovic, S., Siegel, K. R., Rolka, D. B., Imperatore, G., Elbel, B., & Thorpe, L. E. (2021). Longitudinal analysis of neighborhood food environment and diabetes risk in the veterans administration diabetes risk cohort. *JAMA Network Open*, 4(10), Article e2130789. <https://doi.org/10.1001/jamanetworkopen.2021.30789>
- Keralis, J. M., Javanmardi, M., Khanna, S., Dwivedi, P., Huang, D., Tasdizen, T., & Nguyen, Q. C. (2020). Health and the built environment in United States cities: Measuring associations using Google Street View-derived indicators of the built environment. *BMC Public Health*, 20(1), 215. <https://doi.org/10.1186/s12889-020-8300-1>
- Kowaleski-Jones, L., Brown, B. B., Fan, J. X., Hanson, H. A., Smith, K. R., & Zick, C. D. (2017). The joint effects of family risk of obesity and neighborhood environment on obesity among women. *Social Science & Medicine*, 195, 17–24. <https://doi.org/10.1016/j.socscimed.2017.10.018>
- Krieger, N., Waterman, P., Chen, J. T., Soobader, M.-J., Subramanian, S. V., & Carson, R. (2002). Zip code caveat: Bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas—the public health disparities geocoding project. *American Journal of Public Health*, 92(7), 1100–1102. <https://doi.org/10.2105/AJPH.92.7.1100>
- Kweon, B.-S., Sullivan, W. C., & Wiley, A. R. (1998). Green common spaces and the social integration of inner-city older adults. *Environment and Behavior*, 30(6), 832–858. <https://doi.org/10.1177/001391659803000605>
- Ladd-Acosta, C., & Fallin, M. D. (2016). The role of epigenetics in genetic and environmental epidemiology. *Epigenomics*, 8(2), 271–283. <https://doi.org/10.2217/epi.15.102>
- Larkin, A., Krishna, A., Chen, L., Amram, O., Avery, A. R., Duncan, G. E., & Hystad, P. (2022). Measuring and modelling perceptions of the built environment for epidemiological research using crowd-sourcing and image-based deep learning models. *Journal of Exposure Science and Environmental Epidemiology*, 32(6). <https://doi.org/10.1038/s41370-022-00489-8>. Article 6.
- Liu, S. R., Kia-Keating, M., Santacrose, D. E., & Modir, S. (2018). Linking profiles of neighborhood elements to health and related outcomes among children across the United States. *Health & Place*, 53, 203–209. <https://doi.org/10.1016/j.healthplace.2018.08.013>
- Luo, Y., & Wang, S. (2022). Urban living and chronic diseases in the presence of economic growth: Evidence from a long-term study in southeastern China. *Frontiers in Public Health*, 10. <https://www.frontiersin.org/articles/10.3389/fpubh.2022.1042413>
- Massey, D. S., Gross, A. B., & Shibuya, K. (1994). Migration, segregation, and the geographic concentration of poverty. *American Sociological Review*, 59(3), 425–445. <https://doi.org/10.2307/2095942>

- Merlo, J. (2003). Multilevel analytical approaches in social epidemiology: Measures of health variation compared with traditional measures of association. *Journal of Epidemiology & Community Health*, 57(8), 550–552. <https://doi.org/10.1136/jech.57.8.550>
- Merlo, J. (2018). Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) within an intersectional framework. *Social Science & Medicine*, 203, 74–80. <https://doi.org/10.1016/j.socscimed.2017.12.026>
- Merlo, J., Ohlsson, H., Chaix, B., Lichtenstein, P., Kawachi, I., & Subramanian, S. V. (2013). Revisiting causal neighborhood effects on individual ischemic heart disease risk: A quasi-experimental multilevel analysis among Swedish siblings. *Social Science & Medicine*, 76, 39–46. <https://doi.org/10.1016/j.socscimed.2012.08.034>
- Merlo, J., Viciana-Fernández, F. J., & Ramiro-Fariñas, D. (2012). Bringing the individual back to small-area variation studies: A multilevel analysis of all-cause mortality in andalusia, Spain. *Social Science & Medicine*, 75(8), 1477–1487. <https://doi.org/10.1016/j.socscimed.2012.06.004>
- Merlo, J., Wagner, P., Austin, P. C., Subramanian, S., & Leckie, G. (2018). General and specific contextual effects in multilevel regression analyses and their paradoxical relationship: A conceptual tutorial. *SSM - Population Health*, 5, 33–37. <https://doi.org/10.1016/j.ssmph.2018.05.006>
- Merlo, J., Wagner, P., Ghith, N., & Leckie, G. (2016). An original stepwise multilevel logistic regression analysis of discriminatory accuracy: The case of neighbourhoods and health. *PLoS One*, 11(4), Article e0153778. <https://doi.org/10.1371/journal.pone.0153778>
- Merlo, J., Wagner, P., & Leckie, G. (2019). A simple multilevel approach for analysing geographical inequalities in public health reports: The case of municipality differences in obesity. *Health & Place*, 58, Article 102145. <https://doi.org/10.1016/j.healthplace.2019.102145>
- Mokdad, A. H., Bowman, B. A., Ford, E. S., Vinicor, F., Marks, J. S., & Koplan, J. P. (2001). The continuing epidemics of obesity and diabetes in the United States. *JAMA*, 286(10), 1195–1200. <https://doi.org/10.1001/jama.286.10.1195>
- Morland, K., Wing, S., Roux, A. D., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, 22(1), 23–29. [https://doi.org/10.1016/S0749-3797\(01\)00403-2](https://doi.org/10.1016/S0749-3797(01)00403-2)
- Nguyen, Q. C., Belnap, T., Dwivedi, P., Deligani, A. H. N., Kumar, A., Li, D., Whitaker, R., Keralis, J., Mane, H., Yue, X., Nguyen, T. T., Tasdizen, T., & Brunisholz, K. D. (2022). Google street view images as predictors of patient health outcomes, 2017–2019. *Big Data and Cognitive Computing*, 6(1), 15. <https://doi.org/10.3390/bdcc6010015>
- Nguyen, Q. C., Brunisholz, K. D., Yu, W., McCullough, M., Hanson, H. A., Litchman, M. L., Li, F., Wan, Y., VanDerslice, J. A., Wen, M., & Smith, K. R. (2017). Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Scientific Reports*, 7, Article 16425. <https://doi.org/10.1038/s41598-017-16573-1>
- Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science & Medicine*, 58(10), 1929–1952. <https://doi.org/10.1016/j.socscimed.2003.08.004>
- Orminski, E. (2021). Your zip code is more important than your genetic code » NCRC. <https://ncrc.org/your-zip-code-is-more-important-than-your-genetic-code/>
- Quillian, L. (2012). Segregation and poverty concentration: The role of three segregations. *American Sociological Review*, 77(3), 354–379. <https://doi.org/10.1177/0003122412447793>
- Remigio, R. V., Zulaika, G., Rabello, R. S., Bryan, J., Sheehan, D. M., Galea, S., Carvalho, M. S., Rundle, A., & Lovasi, G. S. (2019). A local view of informal urban environments: A mobile phone-based neighborhood audit of street-level factors in a Brazilian informal community. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 96(4), 537–548. <https://doi.org/10.1007/s11524-019-00351-7>
- Renalds, A., Smith, T. H., & Hale, P. J. (2010). A systematic review of built environment and health. *Family & Community Health*, 33(1), 68. <https://doi.org/10.1097/FCH.0b013e3181c4e2e5>
- Richardson, E. A., Pearce, J., Mitchell, R., & Kingham, S. (2013). Role of physical activity in the relationship between urban green space and health. *Public Health*, 127(4), 318–324. <https://doi.org/10.1016/j.puhe.2013.01.004>
- Robbins, M. W., Griffin, B. A., Shih, R. A., & Slaughter, M. E. (2020). Robust estimation of the causal effect of time-varying neighborhood factors on health outcomes. *Statistics in Medicine*, 39(5), 544–561. <https://doi.org/10.1002/sim.8423>
- Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional Networks for large-scale image recognition* (arXiv:1409.1556). arXiv. <https://doi.org/10.48550/arXiv.1409.1556>
- Smith, K. R., Fraser, A., Reed, D. L., Barlow, J., Hanson, H. A., West, J., Knight, S., Forsythe, N., & Mineau, G. P. (2022). The Utah population database. A model for linking medical and genealogical records for population health research. *Historical Life Course Studies*, 12, 58–77. <https://doi.org/10.51964/hlcs11681>
- StataCorp. (2015). *Stata 13 base reference manual*. StataCorp LLC [Computer software] <https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/>
- Subramanian, S. V., Glymour, M. M., & Kawachi, I. (2007). Identifying causal ecologic effects on health: A methodological assessment. In *Macrosocial determinants of population health* (pp. 301–331). New York: Springer. [https://doi.org/10.1007/978-0-387-70812-6\\_15](https://doi.org/10.1007/978-0-387-70812-6_15)
- Sugiyama, T., Leslie, E., Giles-Corti, B., & Owen, N. (2008). Associations of neighbourhood greenness with physical and mental health: Do walking, social coherence and local social interaction explain the relationships? *Journal of Epidemiology & Community Health*, 62(5). <https://doi.org/10.1136/jech.2007.064287>. ARTN e9.
- Sullivan, W. C., Kuo, F. E., & DePooter, S. F. (2004). The fruit of urban nature: Vital neighborhood spaces. *Environment and Behavior*, 36(5), 678–700. <https://doi.org/10.1177/0193841X04264945>
- Sundquist, K., Eriksson, U., Mezuk, B., & Ohlsson, H. (2015). Neighborhood walkability, deprivation and incidence of type 2 diabetes: A population-based study on 512,061 Swedish adults. *Health & Place*, 31, 24–30. <https://doi.org/10.1016/j.healthplace.2014.10.011>
- Swope, C. B., Hernández, D., & Cushing, L. J. (2022). The relationship of historical redlining with present-day neighborhood environmental and health outcomes: A scoping review and conceptual model. *Journal of Urban Health*. <https://doi.org/10.1007/s11524-022-00665-z>
- Thomas, A. J., Eberly, L. E., Davey Smith, G., & Neaton, J. D., for the Multiple Risk Factor Intervention Trial (MRFIT) Research Group. (2006). ZIP-Code-based versus tract-based income measures as long-term risk-adjusted mortality predictors. *American Journal of Epidemiology*, 164(6), 586–590. <https://doi.org/10.1093/aje/kwj234>
- Thornton, C. M., Conway, T. L., Cain, K. L., Gavand, K. A., Saelens, B. E., Frank, L. D., Geremia, C. M., Glanz, K., King, A. C., & Sallis, J. F. (2016). Disparities in pedestrian streetscape environments by income and race/ethnicity. *SSM - Population Health*, 2, 206–216. <https://doi.org/10.1016/j.ssmph.2016.03.004>
- Tuckett, A. G., Banchoff, A. W., Winter, S. J., & King, A. C. (2018). The built environment and older adults: A literature review and an applied approach to engaging older adults in built environment improvements for health. *International Journal of Older People Nursing*, 13(1), Article e12171. <https://doi.org/10.1111/ijn.12171>
- U.S. Census Bureau. (2016). *2011–2015 American community survey 5-year public use microdata samples*. Census.Gov. <https://www.census.gov/programs-surveys/acs/microdata/access.html>
- Watson, T. (2009). *Inequality and the Measurement of residential Segregation by income in American neighborhoods (working paper 14908)*. National Bureau of Economic Research. <https://doi.org/10.3386/w14908>
- Williams, D. R., Lawrence, J. A., & Davis, B. A. (2019). Racism and health: Evidence and needed research. *Annual Review of Public Health*, 40(1), 105–125. <https://doi.org/10.1146/annurev-publhealth-040218-043750>
- Witten, K., Blakely, T., Bagheri, N., Badland, H., Ivory, V., Pearce, J., Mavoja, S., Hinckson, E., & Schofield, G. (2012). Neighborhood built environment and transport and leisure physical activity: Findings using objective exposure and outcome measures in New Zealand. *Environmental Health Perspectives*, 120(7), 971–977. <https://doi.org/10.1289/ehp.1104584>
- Zick, C. D., Hanson, H., Fan, J. X., Smith, K. R., Kowaleski-Jones, L., Brown, B. B., & Yamada, I. (2013). Re-visiting the relationship between neighbourhood environment and BMI: An instrumental variables approach to correcting for residential selection bias. *International Journal of Behavioral Nutrition and Physical Activity*, 10(1), 27. <https://doi.org/10.1186/1479-5868-10-27>